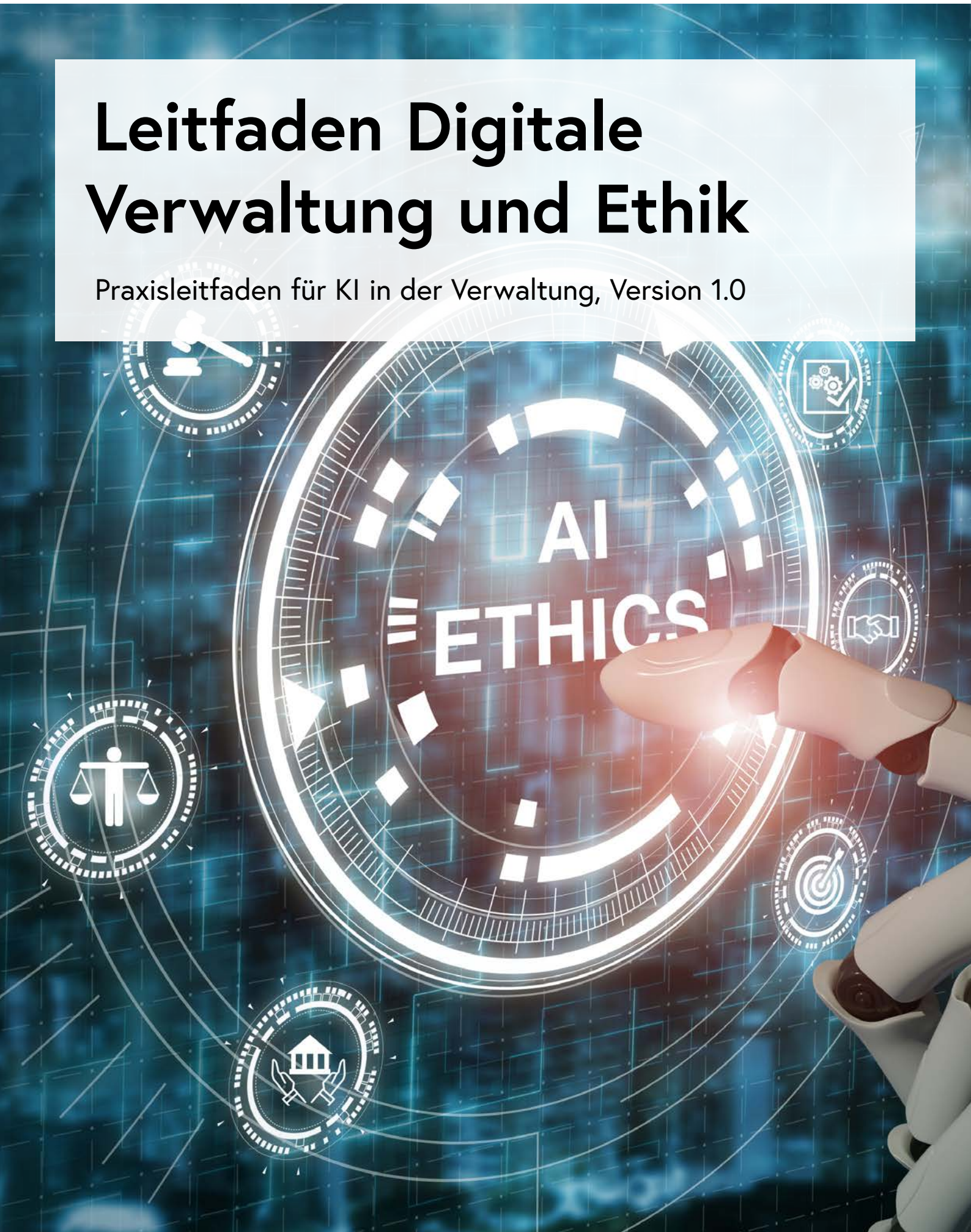


Leitfaden Digitale Verwaltung und Ethik

Praxisleitfaden für KI in der Verwaltung, Version 1.0



Leitfaden Digitale Verwaltung und Ethik

Praxisleitfaden für KI in der Verwaltung, Version 1.0

Wien, Oktober 2023

Impressum

Medieninhaber, Verleger und Herausgeber:

Bundesministerium für Kunst, Kultur, öffentlichen Dienst und Sport (BMKÖS) Sektion III –
Öffentlicher Dienst und Verwaltungsinnovation Hohenstaufengasse 3, 1010 Wien bmkoes.gv.at
Autor:innen: PD Dr. Peter Biegelbauer (Projektleitung), Caroline Lackinger, BA, Dr. Sven Schlarb,
Edgar Subak, BA, Pia Weinlinger, MA

Projektleitung: Mag.a Ursula Rosenbichler, Michael Huber, LL.B, MSc., Ralf M. Tatto, BA MA MA
Fotonachweis: BMKÖS

Layout: BKA Design & Grafik

Druck: Druckerei des BMF

Copyright und Haftung:

Jede kommerzielle Verwertung (auch auszugsweise) ist ohne schriftliche Zustimmung des Medieninhabers unzulässig. Dies gilt insbesondere für jede Art der Vervielfältigung, der Übersetzung, der Mikroverfilmung, der Wiedergabe in Fernsehen und Hörfunk, sowie für die Verbreitung und Einspeicherung in elektronische Medien wie z. B. Internet oder CDROM. Im Falle von Zitierungen im Zuge von wissenschaftlichen Arbeiten sind als Quellenangabe „BMKÖS“ sowie der Titel der Publikation und das Erscheinungsjahr anzugeben. Es wird darauf verwiesen, dass alle Angaben in dieser Publikation trotz sorgfältiger Bearbeitung ohne Gewähr erfolgen und eine Haftung des BMKÖS und der Autorin / des Autors ausgeschlossen ist. Rechtausführungen stellen die unverbindliche Meinung der Autorin / des Autors dar und können der Rechtsprechung der unabhängigen Gerichte keinesfalls vorgreifen.

Kontakt und Rückmeldungen:

Bundesministerium für Kunst, Kultur, öffentlichen Dienst und Sport Abteilung III/C/9 –
Strategisches Performancemanagement und Verwaltungsinnovation
Ralf Tatto, MA; ralf.tatto@bmkoes.gv.at

Diese Publikation ist abrufbar und unter oeffentlicherdienst.gv.at/publikationen zum Download verfügbar.

ISBN 978-3-903097-55-1

Wien, 2023

Vorwort des Herrn Vizekanzlers

Liebe Mitarbeiter:innen,
liebe Leser:innen,

Die Bundesregierung hat es sich zur Aufgabe gemacht die Vorteile der Digitalisierung in allen Lebensbereichen und flächendeckend nutzbar zu machen. Digitalisierung richtig verstanden setzt die Kontrolle über die eigenen Daten, Transparenz von Politik sowie öffentlicher Verwaltung unter Einhaltung aktueller ethischer Standards und vor allem einen menschenzentrierten Ansatz voraus.

Der vorliegende Leitfaden ist das Ergebnis intensiver Arbeit und widerspiegelt das Engagement unserer Verwaltung, sich den Herausforderungen und Chancen der Digitalisierung zu stellen.

Die digitale Transformation verändert unsere Gesellschaft, unsere Wirtschaft und auch unsere Verwaltung in einer Weise, die wir noch vor wenigen Jahren nicht für möglich gehalten hätten. Sie eröffnet uns neue Wege und Möglichkeiten, um Dienstleistungen effizienter und bürgernäher zu gestalten, stellt uns aber auch vor ethische und rechtliche Herausforderungen.

Im Zuge der politischen Dimensionen dieses Themas sehen wir es als unsere Pflicht, die Rahmenbedingungen für einen effizienten und ethisch korrekten Einsatz digitaler Werkzeuge im öffentlichen Dienstes zu gewährleisten.

Der vorliegende Leitfaden ist ein wertvolles Instrument, um uns auf diesem Weg zu begleiten. Er hilft uns, die komplexen Aspekte der Digitalisierung zu verstehen und den richtigen Kurs für eine moderne, digitale und ethisch verantwortungsvolle Verwaltung zu steuern.

Ich danke allen, die zu diesem Leitfaden beigetragen haben, für ihre harte Arbeit und ihr Engagement. Gemeinsam werden wir einen öffentlichen Dienst gewährleisten, der auch für die zukünftigen Herausforderungen und Chancen des digitalen Zeitalters gewappnet ist.

Mag. Werner Kogler

Vizekanzler und Bundesminister für Kunst, Kultur, öffentlichen Dienst und Sport



Vizekanzler
Mag. Werner Kogler

Vorwort des Herrn Sektionsleiters



Sektionsleiter
Mag. Christian Kemperle

Liebe Mitarbeiter:innen,
liebe Leser:innen,

Inmitten der rapiden Fortschritte der digitalen Technologien erleben wir eine Transformation, die unser tägliches Leben und die Art und Weise wie wir arbeiten tiefgreifend verändert. Diese Fortschritte bieten große Chancen, bergen aber auch Risiken und stellen uns vor neue Herausforderungen. Insbesondere Künstliche Intelligenz (KI) ist in den Fokus gerückt und hat entscheidende Auswirkungen auf viele Bereiche unseres Lebens und unserer Arbeit, einschließlich der öffentlichen Verwaltung.

Als Mitarbeiter:innen der öffentlichen Verwaltung haben wir die Pflicht, sicherzustellen, dass unsere Institutionen mit diesen Fortschritten Schritt halten und die Vorteile nutzen, die sie bieten, um den Bürger:innen den bestmöglichen Service zu bieten. Dabei ist es unerlässlich, dass wir uns bewusst sind, dass diese technologischen Werkzeuge ethisch einwandfrei eingesetzt werden müssen, um die Rechte Aller zu schützen und Vertrauen zu fördern.

Dieser erste Leitfaden „Digitale Verwaltung und Ethik“ wurde erstellt, um Ihnen bei dieser wichtigen Aufgabe zu helfen. Er bietet eine eingehende Untersuchung der ethischen Fragen im Zusammenhang mit der Verwendung von KI in der Verwaltung, der technischen Grundlagen und den gesetzlichen Rahmenbedingungen. Dabei werden Chancen, Herausforderungen und Handlungsbedarf, die KI in der Verwaltung mit sich bringt, dargestellt. Durch die Betrachtung von Datenschutz, Bias und Nachhaltigkeit zielt der Leitfaden darauf ab, Ihnen einen umfassenden Überblick über die Anwendung von KI in der Verwaltung zu geben, Orientierung zu bieten und Ihnen dabei zu helfen, fundierte Entscheidungen zu treffen.

Unser Anliegen ist es, den öffentlichen Sektor dabei zu unterstützen, die Vorteile der Digitalisierung zu nutzen, während wir gleichzeitig ethische Standards gewährleisten. Es ist unsere Überzeugung, dass wir durch einen verantwortungsvollen und ethisch korrekten Einsatz digitaler Werkzeuge unsere Arbeit effizienter und effektiver gestalten und zugleich den Schutz der Grundrechte und den Dienst an den Bürger:innen verbessern können.

Wir hoffen, dass Sie diesen Leitfaden nützlich finden und dass er Ihnen dabei hilft, Ihre Aufgaben in dieser neuen digitalen Landschaft zu erfüllen. So rapide wie sich eingangs erwähnt die Technologie weiterentwickelt, so rapide entwickeln sich auch die in diesem ersten Leitfaden verwendeten Inhalte weiter. Es ist daher geplant diesen ersten Leitfaden stetig weiterzuentwickeln und in regelmäßigen Abständen, zumindest in digitaler Form, aktualisiert zur Verfügung zu stellen. Scheuen sich daher bitte nicht der Einladung der Autor:innen Folge zu leisten und Kritik, Anregungen und Vorschläge zu übermitteln.

Mag. Christian Kemperle
Leiter der Sektion III – Öffentlicher Dienst und Verwaltungsinnovation

Inhalt

1 Einleitung	7
1.1 Praxisleitfaden Ziele und Zielgruppen.....	8
2 Das Projekt DVuE und seine Ziele	11
2.1 Methode und Vorgehen.....	11
2.2 Workshops.....	12
3 Technik: Was ist KI?	14
4 Handlungsbedarf	19
5 Chancen und Herausforderungen beim Einsatz von KI in der Verwaltung	23
5.1 KI und Auswirkungen auf die Arbeitswelt in der öffentlichen Verwaltung.....	23
5.2 KI in der Verwaltung und Auswirkungen auf die Bevölkerung	26
5.3 KI und Ökologie.....	30
5.4 Digitale Souveränität in der Verwaltung.....	33
6 Abwägung: Datengetriebene KI in der Verwaltung, oder nicht?	36
7 Rechtlicher Rahmen	39
7.1 Die Grundlage des Verwaltungshandelns.....	39
7.2 Datenschutzgrundverordnung.....	41
7.3 Die EU regelt KI: der AI Act	43
7.4 Weitere EU-Normierungen.....	45
7.5 Artificial Intelligence Mission Austria 2030.....	46
8 Ethische KI: Prinzipien und Leitlinien	48
8.1 Ethische Leitlinien: Governance durch „Soft Law“	48
8.2 Ethische Leitlinien für die österreichische Verwaltung.....	50
9 KI-Folgenabschätzung	52
9.1 Kriterien- und Maßnahmenkatalog für ethische KI in der Verwaltung EKIV.....	52
9.2 EU-Bewertungsliste für vertrauenswürdige künstliche Intelligenz (ALTAI).....	56
9.3 Ethics Self-Assessment für EU-Forschungsförderung und EU „Ethics by Design“ KI-Forschungsleitfaden.....	57

9.4 VCIO-Modell	58
9.5 Folgenabschätzung für Grundrechte und Algorithmen (FRAIA).....	58
9.6 Data Ethics Decision Aid (DEDA).....	59
9.7 Weitere KI-Folgenabschätzungsinstrumente.....	59
10 Empfehlungen für mögliche weitere Schritte:	
Ziel menschenzentrierte KI-Governance.....	61
Quellenverzeichnis.....	68
Anhang.....	75
Glossar für Fachbegriffe.....	81
Abbildungsverzeichnis.....	82
Tabellenverzeichnis.....	82

1 Einleitung

Ziel dieses Leitfadens ist es, Verwaltungsbedienstete in Planung, Design, Erstellung bzw. Vergabe, Einsatz und Evaluierung digitaler, insbesondere aber KI-basierter Anwendungen zu unterstützen. Für die ethischen Überlegungen hierbei ist die Komplexität des eingesetzten Systems, also z. B. ob es als Künstliche Intelligenz (KI) definiert werden kann oder nicht, weniger zentral, als die Beantwortung der Frage welche Auswirkungen digital unterstützte Prozesse und Entscheidungen auf Individuen und die Gesellschaft haben. Vertrauen in öffentliche Institutionen kann bei negativen Auswirkungen von Verwaltungshandlungen rasch beschädigt werden, weshalb die Verwaltung als zentraler Kontaktpunkt zwischen Bürger:innen und Staat hier eine besondere Verantwortung trägt.

Im Fokus des Interesses soll trotzdem KI stehen, erstens weil diese Technologie ein großes Potenzial für die öffentliche Verwaltung birgt, zweitens weil Chancen und Herausforderungen beim Einsatz von KI stellvertretend für andere Daten verarbeitende Anwendungen stehen und drittens, weil sich diese Technologie rasant ausbreitet.

Das vorliegende Dokument versteht sich als praxisbezogener Leitfaden, der Verwaltungsbediensteten zur Orientierung dienen soll. Hier finden sich Einführungen in verschiedene Themenbereiche, Empfehlungen für Vorgehensweisen, ein Kriterien- und Maßnahmenset zur ethischen und rechtlichen Auseinandersetzung sowie eine Checkliste für KI-Anwendungen.

Der Leitfaden richtet sich gleichermaßen an Anwender:innen, Entwickler:innen und Management in der öffentlichen Verwaltung, ebenso wie an die vom Verwaltungshandeln betroffene Öffentlichkeit.

Der Leitfaden versteht sich als Version 1.0. Spätestens mit der Verabschiedung des AI Acts und der Überarbeitung der österreichischen KI-Strategie wird eine Überarbeitung notwendig sein. In diesem Sinn soll es sich also um ein lebendiges Dokument handeln. Vor diesem Hintergrund nehmen wir auch gerne Anregungen entgegen, bitte unter den folgenden E-Mail-Adressen:

iii9@bmkoes.gv.at, office.isp@ait.ac.at

1.1 Praxisleitfaden Ziele und Zielgruppen

Digitalisierung verändert die Gesellschaft rasant. Mehr als drei Viertel der Österreicher:innen nutzt ein Smartphone zum Surfen im Internet und der Großteil von ihnen benützt dabei Anwendungen, die auf Künstlicher Intelligenz (KI) beruhen, wie Google Search, Google Maps oder soziale Medien (Statistik Austria 2021). Ein großer Teil der Bewerber:innen bei internationalen Firmen wird mittlerweile durch KI-Anwendungen beurteilt. Weniger als 10 % des Börsenhandels wird durch menschliche Akteure durchgeführt – Maschinen erstellen Risikoanalysen und treffen autonom Entscheidungen in Bruchteilen von Sekunden (Stahl 2021). Seit der Einführung von ChatGPT als erster generativer KI, die der breiten Öffentlichkeit weitgehend ohne Zutrittschürden zugänglich ist, entstehen täglich tausende Anwendungen, die einer weltweiten Nutzung offenstehen.

Der Einsatz von KI eröffnet dabei umfangreiche Möglichkeiten, birgt aber ebenso Gefahren. Die in vielen beruflichen wie privaten Kontexten eingesetzten Anwendungen, beispielsweise zur Spracherkennung, Übersetzung, Optimierung von Mobilitätsbedürfnissen, aber auch zur automatisierten Gesichtserkennung oder zur Analyse der Produktivität von Mitarbeiter:innen (wie etwa durch Microsoft „Workplace Analytics“, Christl 2021) beruhen auf der (gegenwartsbezogenen oder prädiktiven) Analyse personenbezogener Daten. Viele Anwendungen versprechen Effizienzsteigerung und Arbeitserleichterung. Es sind jedoch auch verschiedene Problembereiche zu bedenken.

Ein Beispiel ist die in manchen interaktiven KI-Anwendungen verwendete Analyse natürlicher Sprache, welche im Zuge des aktuellen Fortschritts neuronaler Netze zunehmend auf Sprachmodellen basieren, die Zusammenhänge und Muster aus sehr umfangreichen Datensätzen, wie z. B. Wikipedia bzw. Social Networks, wie z. B. Reddit, gelernt haben. Die historischen Sprachdaten dienen dabei als statistische Basis für die Analyse aktueller Trends. Auf diese Weise können sich jedoch in den für das Training einer KI notwendigen Daten enthaltene Vorurteile sowie explizit rassistische und diskriminierende Aussagen auf die Vorhersagen und Schlussfolgerungen in KI-Anwendungen auswirken.

Sowohl die Existenz, Vorhaltung und Verarbeitung personenbezogener Daten als auch ethische Vorgangsweisen von KI-Anwendungen werfen vielfältige Fragen zu Privatsphäre, Datenschutz, Verzerrungen, Nachhaltigkeit und Ethik auf, die in diesem Leitfaden berücksichtigt werden.

Diesen Fragen muss sich auch die österreichische Verwaltung täglich stellen. Während die Anforderungen an die öffentliche Verwaltung stetig steigen, wird mit der Pensionierung der Babyboomer-Generation ein erheblicher Anteil der Verwaltungsbediensteten aus dem aktiven Dienst ausscheiden. Gleichzeitig spielt der Ruf nach Einsparungen im öffentlichen Diskurs nach wie vor eine wichtige Rolle. Dementsprechend findet sich hier ein Betätigungsfeld für die Anwendung von digitalen Lösungen und insbesondere KI mit dem Ziel der Steigerung von Effizienz und der Entlastung von Mitarbeiter:innen bei möglichst gleichzeitigen Verbesserungen im Servicebereich.

Die öffentliche Verwaltung hat allerdings eine besondere Verantwortung im Hinblick auf die Vielzahl staatlicher Aufgaben, die sie wahrnimmt. Besonders hervorzuheben sind hier handlungsleitende Grundsätze wie Legalitätsprinzip, Rechenschaftspflicht, Transparenzgebot sowie Autoritäts- und Verantwortungsketten, welche Politik und Verwaltung verbinden. Sie bilden den Rahmen, der die Aktivitäten der öffentlichen Verwaltung ermöglicht und, im Sinne von Rechtsstaat und Demokratie, auch einschränkt (Holzinger et al. 2013).

Gleichzeitig ist Gesetzeskonformität im täglichen Verwaltungshandeln nicht ausreichend. Gesetzliche Regelungen können mit der rasanten Entwicklung neuer Technologien wie KI nicht Schritt halten. Deshalb ist es hier besonders wichtig, die Verankerung der öffentlichen Verwaltung in gesellschaftlichen Werten, die einer modernen, offenen, demokratischen Grundhaltung entsprechen, zu betonen.

Im Rahmen der öffentlichen Verwaltung sind dabei die zuvor angesprochenen Themenbereiche wie Datenschutz, Privatsphäre, Bias¹ und (KI) Ethik besonders brisant, geht es hier doch um Kernbereiche staatlicher Aufgaben, die unter den zuvor angesprochenen besonderen rechtsstaatlichen und Demokratie unterstützenden Bedingungen erfüllt werden müssen.

In diesem Zusammenhang ist der Schutz von Grundrechten von besonderer Bedeutung, denn dieser zählt zu den Verwaltungsaufgaben. Hier sind insbesondere Freiheit, Selbstbestimmung, Privatsphäre, Gleichbehandlungsgebot, Menschenwürde, Gerechtigkeit, Verfahrensgarantien, Sicherheit, aber – in einem modernen Verständnis von Staatsaufgaben – auch Demokratie und Nachhaltigkeit zu nennen (vgl. Latonero 2018, WWW Foundation 2018, Initiative D21 2019).

Im Kontext der „Ethics Guidelines for Trustworthy AI“ der „EU High-Level Expert Group on Artificial Intelligence“ (European Commission 2019), der Vorschläge der europäischen Institutionen für eine Regulierung von Digitalisierung und KI als Teil des „AI Act“ (im deutschen Sprachraum auch KI-Verordnung; European Commission 2021c), entsprechender Aktivitäten von OECD (2019) und UNESCO (2022), der KI-Strategie der österreichischen Bundesregierung „AIM AT 2030“ (BMK und BMDW 2021a), aber auch paralleler Diskussionen in Österreich (TÜV 2021; BRZ 2020; Nentwich et al. 2021), Deutschland (Deutscher Bundestag 2019, Deutscher Ethikrat 2023), Großbritannien (UK Parliament 2021) und den USA (The White House 2022) zur Regulierung von KI werden immer wieder ähnliche Prinzipien diskutiert: menschliche Letztentscheidungsverantwortung, (Daten-)Sicherheit, Einhaltung von Privatsphäre, Transparenz, Diversität,

1 Bias in der KI stellt eine Unverhältnismäßigkeit bzw. Verzerrtheit beim Output von maschinellen lernenden Algorithmen dar, die beispielsweise aufgrund von systematischen Fehlannahmen und Vorurteilen bei der Entwicklung des KI-Algorithmus oder aufgrund unvollständiger, unausgewogener bzw. verzerrter Eingabedaten erfolgen kann. KI-Systeme, deren Entscheidungen oder Vorhersagen einen solchen Bias enthalten bzw. davon beeinflusst sind, bilden die Realität nicht wirklichkeitsgetreu ab und können daher in weiterer Folge zur Diskriminierung von Menschen führen (vgl. Dilmegani 2022).

Fairness und Diskriminierungsverbot, soziale und ökologische Nachhaltigkeit sowie Rechenschaftspflicht (vgl. auch Jobin et al. 2019).

Diese Prinzipien haben vor dem Hintergrund der Aufgabenstellungen der öffentlichen Verwaltung für deren Arbeit eine besondere Bedeutung, wobei hier eine Analyse der gegenwärtigen Bedingungen sowie Vorschläge für ein weiteres Vorgehen gemacht werden sollen.

Vor diesem Hintergrund zeigt dieser Leitfaden die Voraussetzungen und Hürden auf, die bei der Integration von KI in die öffentliche Verwaltung zu beachten sind. Mit dem Ziel, einen robusten Rahmen für die Implementierung von KI in der öffentlichen Verwaltung zu schaffen, wird ein spezielles Augenmerk auf ethische Prinzipien und regulatorischen Diskussionen gelegt. Die Struktur dieses Leitfadens ist dabei wie folgt: zuerst wird der Hintergrund des Projektes, das zu diesem Leitfaden geführt hat, umrissen. Dann wird der Begriff KI erklärt, der Handlungsbedarf in Bezug auf deren Einsatz in der öffentlichen Verwaltung geklärt, sowie Chancen und Herausforderungen beim Einsatz von KI in der Verwaltung im Hinblick auf Gesellschaft, Ökologie und digitale Souveränität diskutiert. In einem weiteren Schritt geht es um die Abwägung, ob KI in einem gegebenen Fall in der Verwaltung eingesetzt werden soll oder nicht. Als Nächstes werden einerseits der Rechtsrahmen und andererseits die ethischen Prinzipien und Standards beim Einsatz von KI, sowie Möglichkeiten der Folgenabschätzung besprochen. Im letzten Abschnitt werden schließlich Empfehlungen für nächste Schritte auf einem Weg hin zu einer menschenzentrierten KI gemacht.

2 Das Projekt DVuE und seine Ziele

Dieser Abschnitt führt allgemein in das Projekt und seine Ziele ein. Die Methoden, insbesondere auch die Formate der Workshops, deren Inhalte sowie Organisation und dahinterstehenden Personen werden kurz vorgestellt.

Das Projekt „Digitale Verwaltung und Ethik“ beschäftigte sich von Juni 2022 bis Mai 2023 mit dem ethischen Einsatz digitaler Lösungen, insbesondere von KI, in der Verwaltung. Die Projektziele waren:

- einen Reflexionsrahmen für den Umgang mit Digitalisierung und KI aufzuspannen,
- ethische Standards zu diesen Themen zu entwickeln,
- Rahmenbedingungen für eine Folgenabschätzung zu schaffen, sowie
- Standards für die Aus- und Weiterbildung von Verwaltungsbediensteten zu setzen.

2.1 Methode und Vorgehen

Der vorliegende Praxisleitfaden basiert auf verschiedenen Quellen. Neben Recherchen in Bereichen wie KI-Ethik, ethische Softwareentwicklung, Verwaltung und Digitalisierung, Debatten im Projektteam und mit den Auftraggeber:innen im BMKÖS, waren die Diskussionen in fünf Workshops von besonderer Bedeutung. Die Verwaltungsbediensteten, die an den Workshops teilgenommen haben, haben dabei aktiv Ideen formuliert und Vorschläge sowie Anliegen eingebracht, die für die Erarbeitung dieses Leitfadens bedeutsam waren.

In den Workshops wurden einerseits neue Ideen durch Vorträge, die unterschiedliche technische, sozialwissenschaftliche, Ethik-, verwaltungsbezogene und zivilgesellschaftliche Perspektiven eröffneten, getestet und diskutiert. Andererseits wurden Lernprozesse auf Seiten des Projektteams wie auch der Workshop-Teilnehmer:innen durch zahlreiche Diskussions- und Feedbackelemente ermöglicht.

2.2 Workshops

Im Rahmen des Projekts wurden von September 2022 bis März 2023 fünf Workshops durchgeführt.

2022							2023				
J	J	A	S	O	N	D	J	F	M	A	M
			Kick-Off 19.09.22		W1 04.11.22						
					W2 17.11.22						
							W3 25.01.23				
									W4 08.03.23		

Tabelle 1: Workshop-Zeitverlauf

Die Workshops wurden von der Abteilung für Strategisches Performance Management und Verwaltungsinnovation (III/C/9) des BMKÖS und dem AIT Austrian Institute of Technology gemeinsam organisiert und umfassten TeilnehmerInnen aus dem Bundesministerium für Arbeit und Wirtschaft (BMAW), dem Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF), dem Bundesministerium für Finanzen (BMF) und dem Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK). Teilnehmer:innen aus einer Reihe österreichischer und europäischer Institutionen und Organisationen bereicherten die Veranstaltung, darunter Vertreter:innen des Bundesrechenzentrums (BRZ), der österreichischen Bundesagentur Austria Tech, des Research Institutes, der finnischen Forschungseinrichtung VTT, der Universität Utrecht in den Niederlanden, der Technischen Universität Darmstadt in Deutschland, sowie der NGOs Chaos Computer Club und AlgorithmWatch sowie des Unternehmens leiwand.ai.

Arbeitsschwerpunkte, welche im Rahmen des Workshops bearbeitet wurden, beinhalteten folgende zentrale Fragen:

- Warum will die Verwaltung KI einführen?
- Welche Bedingungen braucht es dafür? Welche Fähigkeiten brauchen die Mitarbeiter:innen?
- Was kommt auf die österreichische Bundesverwaltung im Hinblick auf KI-Prinzipien und AI-Act zu?
- Wie reagiert die österreichische Bundesverwaltung auf KI-Prinzipien und AI-Act?

Die Diskussionen und Inhalte der Workshops reichten von Definitionen und Anwendungen von KI über Beispiele für maschinelles Lernen wie Chatbots und den Nutzen von KI im Personalwesen und bei der Prozessoptimierung bis hin zu ethischen Bewertungsinstrumenten für KI-Anwendungen und der Bedeutung der digitalen Souveränität. Die Workshops beschäftigten sich auch mit der Einführung von KI und den damit verbundenen Anforderungen in der Verwaltung, der Kompetenzentwicklung der Mitarbeiter:innen und der Reaktion der Verwaltung auf KI-Grundsätze und AI-Act.

In interaktiven Dialogen wurden Aspekte wie Vertrauen in Institutionen, öffentliche KI-Akzeptanz, Kompetenzentwicklung, Verwaltungskooperation und die ökologischen Auswirkungen von KI behandelt. Darüber hinaus wurden „Reallabore“ vorgeschlagen, um neue KI-Anwendungen und -Tools in einem geschützten Umfeld zu testen. Die Workshops entpuppten sich letztlich als kollektive Lernräume, in denen die Teilnehmer:innen aus verschiedenen Fachbereichen ihr Wissen und ihre Erfahrung zu KI-Themen zusammenführten und so eine symbiotische Lernerfahrung förderten.

Vortragende bei den Workshops waren die Mitglieder des AIT AI Ethics Labs, Peter Biegelbauer, Caroline Lackinger, Alexander Schindler, Sven Schlarb, Edgar Subak und Pia Weinlinger, mit zentralen technischen bzw. rechtlichen / sozialwissenschaftlichen Inputs. Wichtige Diskussionsbeiträge wurden außerdem von den Auftraggeber:innen aus dem BMKÖS, Ursula Rosenbichler, Ralf M. Tatto und Michael Huber (alle Abteilung III/C/9 – Strategisches Performancemanagement und Verwaltungsinnovation), geleistet. Ein Vortrag zum Stand der Verhandlungen des AI-Acts wurde von Julia Fuith (BMF) gehalten.

Beim Wissenschaftsworkshop bereicherten die externen Vortragenden Mirko Schäfer von der Universität Utrecht, Jaana Leikas und Mika Nieminen von der finnischen Forschungseinrichtung VTT sowie Wolfgang Kabelka vom Bundesrechenzentrum das Programm mit ihren Beiträgen.

Zu den externen Referent:innen des NGO-Workshops gehörten Rania Wazir von leiwand.ai, Benjamin Hättasch vom Computer Chaos Club und der TU Darmstadt sowie Matthias Spielkamp von der internationalen NGO AlgorithmWatch.

3 Technik: Was ist KI?

In diesem Abschnitt wird ein kurzer Überblick über die wesentlichen Begriffe und Anwendungsformen von KI gegeben. Dabei wird der aktuelle Stand der technologischen Entwicklung umrissen und im Anschluss ein Ausblick auf Aspekte, die für spätere Abschnitte dieses Leitfadens besonders relevant sind, gegeben.

Der Begriff „Künstliche Intelligenz“ wurde 1955 vom US-amerikanischen Informatiker John McCarthy in Form eines Vorschlags für einen Sommerworkshop am Dartmouth College geprägt und wird seither kontrovers diskutiert. Dabei ging es um Computersysteme, die dazu in der Lage sind, menschliche Intelligenz zu simulieren bzw. nachzuahmen. Die Vielschichtigkeit des Begriffs „Intelligenz“ ist allerdings auch der Grund dafür, dass sich der Begriff „Künstliche Intelligenz“ einer präzisen Definition entzieht.

Legg und Hutter erstellten eine Übersicht mit über 70 Definitionen (Legg und Hutter 2007) und Russel und Norvig reduzieren in ihrer Einführung zur KI die Definitionen auf acht Definitionsansätze, die sie in die vier Bereiche menschliches und rationales Denken, sowie menschliches und rationales Handeln gliedern (Russell und Norvig 2012, 23). Die Autoren skizzieren außerdem die konzeptionelle Sicht eines generischen KI-Systems, welches aus drei Hauptelementen besteht: Erstens *Sensoren*, welche Rohdaten aus der Umgebung sammeln, zweitens *Aktoren*, welche Maßnahmen ergreifen, und den Zustand der Umgebung ändern, und drittens die *Betriebslogik* (engl. *Operational Logic*), welche anhand bestimmter Zielvorgaben und basierend auf Eingabedaten von Sensoren eine Ausgabe für die Aktoren bereitstellt.

Die OECD zieht in ihrem Bericht zum Geltungsbereich der OECD KI-Prinzipien (OECD 2019) diese konzeptionelle Sicht eines generischen KI-Systems heran und sieht diese im Einklang mit einer sehr weit gefassten Definition „Künstlicher Intelligenz“ als das „Forschungsgebiet der computer-gestützten Berechnungen, die Wahrnehmung, Schlussfolgern und Handeln ermöglichen“ (engl. „the study of the computations that make it possible to perceive, reason, and act“) (Winston 1992, 14).

Für den Rahmen dieses Leitfadens lässt sich „Künstliche Intelligenz“ als die Fähigkeit einer Maschine charakterisieren, menschliche Fähigkeiten, wie logisches und kreatives Denken, kontinuierliches Dazulernen, strategische Planung sowie die Übertragung erworbenen Wissens auf neue Anwendungsgebiete, nachzuahmen.

Definition: Künstliche Intelligenz

Hinsichtlich der Problemlösungsfähigkeit wird außerdem häufig zwischen *schwacher* und *starker* KI unterschieden (Buxmann und Schmidt 2021, 6):

- Schwache KI bezeichnet Systeme, die spezifische Aufgabenstellungen in einem Anwendungsbereich anhand gegebener Eingabedaten lösen. Auf Basis der Eingabedaten wird eine Ausgabe als Lösung der Aufgabenstellung erreicht. Typische Aufgabenstellungen sind die Erkennung natürlicher gesprochener Sprache, die Transkription von Handschrift oder die Objekterkennung in Bildern.
- Starke KI bezeichnet Systeme, die selbständig handeln, lernen, sich weiterentwickeln, und neue, bislang unbekannte Aufgabenstellungen lösen können. Sie entwickeln Lernstrategien ohne jegliche menschliche Intervention, planen Aufgaben und setzen Ziele eigenständig. Sie weisen also Fähigkeiten auf, die üblicherweise ausschließlich menschlicher Intelligenz zugeschrieben werden.

Im vergangenen Jahrzehnt haben vor allem KI-Anwendungen auf der Basis neuronaler Netze, die durch sehr große Datenmengen trainiert wurden, in verschiedensten Aufgabebereichen, wie z. B. Bildklassifizierung, Gesichts-, Objekt- und Spracherkennung, etc., beindruckende Ergebnisse erzielt. Ein Beispiel, das in der jüngsten Zeit große mediale Aufmerksamkeit erregt hat, ist die aktuell auf dem „großen Sprachmodell“ (engl. „*Large Language Model*“, LLM) GPT4 basierende Chatbot-Schnittstelle ChatGPT, welche unter bestimmten Voraussetzungen den Eindruck menschlicher Dialogfähigkeit erwecken kann.

Dennoch gelten auch diese Arten der Anwendungen gemeinhin als Beispiele schwacher KI, da sie für die jeweils spezifische Aufgabe erstellt werden und darüber hinaus keine allgemeine Problemlösungsfähigkeit aufweisen (in einer Forschungsarbeit wurden Fähigkeit und Verlässlichkeit logischer Schlussfolgerung mit Hilfe von großen Sprachmodellen systematisch untersucht und grundsätzlich in Frage gestellt, siehe Valmeekam et al. 2023). Im Allgemeinen wird die Möglichkeit der Schaffung einer starken KI eher angezweifelt oder in Anbetracht des aktuellen Wissensstands sogar ganz ausgeschlossen, so wie es z. B. Humm et al. (2022, 14) zum Ausdruck bringen:

„Bis heute sind keine Ansätze für starke KI bekannt. Stand heute gibt es also keine Evidenz dafür, dass starke KI überhaupt jemals möglich sein wird. Die allerwenigsten KI-Forscher:innen und -Praktiker:innen rechnen auf absehbare Zeit mit einer solchen.“

Darüber hinaus können KI-Systeme anhand der Art der Wissensrepräsentation in *symbolische* und *sub-symbolische* Systeme unterteilt werden (Döbel et al. 2018, 11). Symbolische Systeme verwenden explizite und nachvollziehbare Regeln und Symbole, um Wissen darzustellen und logische Schlussfolgerungen abzuleiten. Im Gegensatz dazu nutzen sub-symbolische Systeme implizites, numerisch dargestelltes Wissen, das für Menschen nicht unmittelbar lesbar ist. Diese Systeme basieren auf numerischen Daten und statistischen Methoden und verwenden maschinelles Lernen und neuronale Netze, um Muster und Zusammenhänge in den Daten zu erkennen. Dies ist der Grund, warum Systeme dieser Art oft metaphorisch als „Schwarzer Kasten“ (engl. „*Black Box*“) bezeichnet werden, die kein Nachvollziehen der Entscheidungsfindung ermöglichen.

Allerdings gibt es Ansätze der Erklärbarkeit (engl. *explainability*) die versuchen, die entscheidungsrelevanten Informationen zumindest teilweise transparent zu machen. Für konkrete Anwendungen in der öffentlichen Verwaltung ist derzeit vor allem die Kategorie der Schwachen künstlichen Intelligenz, insbesondere datengetriebene Methoden des maschinellen Lernens, von Bedeutung. Dabei werden drei Formen von Lernen unterschieden (Russel und Norvig 2012, 811):

- **Nicht-überwachtes Lernen** (engl. *unsupervised learning*) ist das automatisierte Lernen aus Beispielen, bei welchem dem System Beispiele der gewünschten Ausgabe gezeigt werden. Die Clusteranalyse ist ein Beispiel zur Bildung von Gruppen aus einer Menge einzelner Elemente. Ein Beispiel dafür ist die Clusterung von Städten nach relevanten Merkmalen, wie zum Beispiel der Einwohnerzahl, Fläche, BIP pro Kopf, etc. Ähnliche Städte werden dann hinsichtlich dieser Merkmale zu Gruppen (Clustern) zusammengefasst, wobei sich Mitglieder einer Gruppe untereinander möglichst ähneln und zugleich ein möglichst großer Unterschied zu Mitgliedern anderer Gruppen besteht.
- **Überwachtes Lernen** (engl. *supervised learning*) ist ein Verfahren, bei dem die gewünschte Ausgabe anhand gekennzeichneter Elemente vorgegeben wird. Bei der binären Klassifizierung wird etwa für jedes Beispiel eine von zwei Kategorien zugeordnet. Werden beispielsweise einzelne Emails als „Spam“ und „Nicht-Spam“ gekennzeichnet, so lernt ein System aus diesen Beispielen den Unterschied zwischen unerwünschten und erwünschten Emails, und kann das Gelernte auf neue Emails anwenden.
- **Bestärkendes Lernen** (engl. *reinforcement learning*) umfasst verschiedene Methoden des maschinellen Lernens bei welchen ein System selbstständig eine Strategie entwickelt, um bei Belohnung für Erfüllung bzw. Bestrafung für Nicht-Erfüllung die maximale Belohnung hinsichtlich bestimmter Zielvorgaben zu erreichen. Ein einfaches Beispiel ist die auf Belohnung basierende Entscheidung für den nächsten Zug in einem Brettspiel, um mit jedem Zug entsprechend der maximalen Siegwahrscheinlichkeit auch die maximale Belohnung anzustreben.

Im vergangenen Jahrzehnt sind die Menge der verfügbaren Daten und die Rechenleistung exponentiell angestiegen. Neue Technologien wie Cloud Computing und Big Data ermöglichen außerdem eine immer effizientere Verarbeitung und Analyse dieser Datenmengen. Unter diesen Voraussetzungen konnten Methoden des Maschinellen Lernens in zahlreichen Anwendungsbereichen ihre Leistungsfähigkeit unter Beweis stellen. Buxmann und Schmidt beschreiben dazu Anwendungen aus unterschiedlichen Domänen und der Verwendung verschiedener Ansätze Künstlicher Intelligenz (Buxmann und Schmidt 2021, 41ff).

Der Begriff Big Data bezieht sich eigentlich auf große Mengen von Daten, die in hoher Geschwindigkeit und Vielfalt generiert werden. Der Begriff umfasst darüber hinaus aber auch die Methoden zur Sammlung, Speicherung, Verarbeitung und Analyse dieser Daten, um Erkenntnisse zu gewinnen und Regularitäten zu erkennen. Ein Beispiel ist die Analyse des Kaufverhaltens bestimmter Alters- und Interessensgruppen im Internet, für die auf großen Verkaufsplattformen sehr große Datenmengen in Bezug auf das Navigationsverhalten und die Kaufabschlüsse der Nutzer:innen verwendet werden. KI stellt Schlüsseltechnologien für die Big Data-Analyse bereit. Die Verarbeitung der Daten kann also, muss aber nicht, mit Methoden der KI durchgeführt werden.

Wissen: Big Data

Die Veranschaulichung von KI durch konkrete Anwendungsbeispiele unterliegt dem Wandel stetig fortschreitender wissenschaftlicher und industrieller Entwicklung. Während in den 80er Jahren noch regelbasierte Expertensysteme und Entscheidungsbäume als Beispiele dienten, werden heute eher künstliche Neuronale Netze angeführt (Ertel und Black 2016, 2). Die Frage inwiefern ein Computerprogramm, eine Anwendung oder Informationstechnologiesystem (IT-System) vollständig oder teilweise zum Bereich Künstlicher Intelligenz gezählt werden soll, ist daher vom jeweils aktuellen Technologiestand abhängig.

Neuronale Netze: Künstliche neuronale Netze (KNN) oder simulierte neuronale Netze (SNN) sind Teil des Maschinellen Lernens (ML) und bilden das Kernstück von Deep-Learning-Algorithmen. Sie sind benannt und strukturiert nach dem menschlichen Gehirn und ahmen die Kommunikation biologischer Neuronen durch Signale nach. Neuronale Netze dienen dazu, Informationen zu verarbeiten und komplexe Muster zu erkennen.

Wissen: neuronale Netze

Da es – wie zu Beginn des Abschnitts erläutert – keine einheitliche Definition von Künstlicher Intelligenz gibt, die allgemein akzeptiert wird, gibt es auch keine eindeutige Liste von Kriterien, die es ermöglicht, zu entscheiden, ob eine Anwendung oder ein System als Künstliche Intelligenz gilt oder nicht.

Entscheidend im Rahmen dieses Leitfadens ist jedoch nicht die Frage, ob eine Anwendung oder ein System spezifische technische Kriterien einer Künstlichen Intelligenz erfüllt, sondern ob die Anwendung oder das System Aufgaben von Entscheidungsrelevanz übernimmt, das heißt, entweder selbstständig Entscheidungen trifft oder Informationen für einen Menschen in Entscheidungsposition aufbereitet.

Wissen: KI im Leitfaden

Besonders wichtig ist hier die Frage, welche Auswirkungen durch KI unterstützte oder beeinflusste Entscheidungen möglicherweise auf Individuen und die Gesellschaft haben. Wird dieses Kriterium in den Vordergrund gerückt, ist es nicht mehr primär von Bedeutung, ob es sich um ein einfaches regelbasiertes System oder um ein auf großen Datenmengen trainiertes neuronales Netz als technologische Grundlage handelt. In Abschnitt 10 zur KI-Folgenabschätzung wird der diesbezüglich relevante Aspekt der Auswirkungen der KI-Technologien näher betrachtet.

4 Handlungsbedarf

Angesichts der in der Einleitung beschriebenen spezifischen Aufgaben und Herausforderungen der Verwaltung beim Einsatz von KI-Systemen und der im Abschnitt 4 angesprochenen technischen Komplexität der Technologie muss hier besonders auf einen ethischen Einsatz geachtet werden. Das Vertrauen in die Verwaltung ist zentral für die Demokratie und Skandale durch den unbedachten Einsatz von KI-Systemen in diesem Bereich können das Verhältnis der Bürger:innen zum Staat nachhaltig beeinträchtigen. In diesem Abschnitt werden zwei problematische Fälle des Einsatzes von KI ausgeführt, die Risiken in verschiedenen Anwendungsbereichen unterschieden und schließlich wird die zentrale Bedeutung von Wissen über KI für eine sichere Anwendung derselben ausgearbeitet.

In Australien wurde das Robodebt System eingesetzt, im Rahmen dessen die australische Steuerbehörde automatisiert jährliche Einkommensdaten mit dem von Sozialversicherungsempfänger:innen angegebenen Einkommen verglich. Wenn diese Daten nicht übereinstimmten und die Person auf eine Anfrage nicht reagierte, wurde eine Schuld dem Staat gegenüber festgestellt und in weiterer Folge eingetrieben. Tatsächlich war Robodebt fehlerhaft, was im Rahmen parlamentarischer Untersuchungen einer eigens eingesetzten Royal Commission und eines gerichtlichen Verfahrens ermittelt wurde. Die amtierende australische Regierung (und ebenso die dieser nach Neuwahlen nachfolgende Regierung) musste sich 2020 öffentlich entschuldigen und Zahlungen in Milliardenhöhe durchführen.

In den Niederlanden kam es zu Unregelmäßigkeiten rund um die Auszahlung der Sozialleistung „Kinderbetreuungsgeld“. Dabei wurden falsche Betrugsvorwürfe der Steuer- und Zollverwaltung erhoben, als diese versuchte die Zuteilung von Kinderbetreuungsgeldern zu automatisieren. Rund 26.000 Eltern wurden zu Unrecht betrügerische Leistungsanträge unterstellt und Zulagen mussten vollständig zurückgezahlt werden. Die Rückzahlungen umfassten zum Teil mehrere € 10.000, was zu Privatkonkursen, zum Entzug von Sorgerechten und schließlich zu mehreren Suiziden führte. Der Skandal führte 2021 schließlich zum Rücktritt der Regierung und zu Neuwahlen.

Anwendungsfall: Beispiele aus Australien und den Niederlanden

Beide Skandale blieben über mehrere Jahre hindurch unentdeckt, unter anderem aufgrund von „Automation Bias“, also der in experimentellen Studien gut dokumentierten Neigung von Menschen automatisierten Verfahren mehr Vertrauen zu schenken als menschlichen Entscheidungen (Goddard et al. 2012). Der beste Weg einem derartigen Bias entgegenzuwirken ist Sensibilisierung und Ausbildung, insbesondere zum besseren Verständnis von Potentialen, Arbeitsweisen und Einsatzformen von KI, hier KI-Literacy genannt (siehe Abschnitte 5 und 11).

KI Literacy heißt übersetzt KI „Alphabetismus“. Der Begriff bezieht sich auf die Fähigkeit, KI zu verstehen und zu verwenden. Eine sichere, selbstbestimmte und verantwortungsbewusste Nutzung von KI ist nur durch ein ausreichendes Verständnis der Arbeitsweise, Möglichkeiten und Herausforderungen dieser Technologie möglich.

Definition: KI Literacy

So konnte in einer Untersuchung mit mehr als 1300 niederländischen Verwaltungsbediensteten gezeigt werden, dass diese tatsächlich eine deutlich geringere Neigung zu einem derartigen Automation Bias hatten als eine ähnlich große Anzahl von niederländischen Bürger:innen. Die Autor:innen führten das vor allem auf die Sensibilisierung der Verwaltungsangehörigen durch den während der Untersuchung in der niederländischen Öffentlichkeit intensiv diskutierten Skandal um die automatisierte Auszahlung von Kinderbetreuungsgeld zurück (Alon-Barkat und Busuioc 2023).

Eine andere Möglichkeit, derartigen Problemen entgegenzuwirken, ist die Einbeziehung von Menschen in die unmittelbare Kontrolle von Ergebnissen einer KI-Anwendung. Hier spricht man auch häufig von „Human In the Loop“.

Human in the Loop: heißt übersetzt „Mensch in der Schleife“ und bedeutet im Kontext von KI, dass ein Mensch in bestimmte Prozesse eines KI-Systems eingreifen und somit die Ergebnisse und Auswirkungen eines KI-Systems beeinflussen kann. Ein Sonderfall ist „Human on the loop“, also übersetzt „Mensch auf der Schleife“, da hier ein Mensch eine Kontrollfunktion übernimmt. Ein Beispiel wäre die Anforderung, dass ein Mensch einen von einem KI-System erkannten Betrug prüfen und gegebenenfalls als strafrechtlich relevant bestätigen müsste.

Wissen: Human in the Loop

Im Hinblick auf die Digitalisierung und Einsatzgebiete wie die Ermittlung und Auszahlung sozialer Leistungen sind in Bezug auf die Verwaltung besonders große Effizienz-erwartungen vorhanden. Grundsätzlich sind diese Bereiche der öffentlichen Verwaltung vergleichsweise weniger problematisch als etwa Rechtsetzung und Rechtsprechung. Die angeführten internationalen Beispiele zeigen allerdings, dass auch im Bereich der Leistungsverwaltung bei wenig reflektiertem Einsatz von digitalen Technologien umfangreicher Schaden entstehen kann.

Tatsächlich sind die Chancen und Risiken nicht in jedem Bereich gleich gelagert. So ist die Einführung von Bürger:innen im Regelfall sehr gut angenommenen Chatbots („WienBot“, „FinanzOnline Fred“) als rund um die Uhr zur Verfügung stehende Ergänzung zu anderen Informationsmöglichkeiten oder von Mobilitäts-Apps („ÖBB Scotty“, „Wien-Mobil“) grundsätzlich weniger problematisch: im Regelfall halten sich potentielle negative Folgen einer beispielsweise gebiasteten / verzerrten Empfehlung in Grenzen.

Ein Chatbot ermöglicht textbasierte Dialoge, um mit einem technischen System zu kommunizieren. Benutzer können sich dabei üblicherweise in Alltagssprache mit dem System austauschen und Fragen stellen. Das System kann – muss aber nicht – Künstliche Intelligenz verwenden, um Fragen zu interpretieren und zu beantworten.

Die Entwicklung und Verwendung KI-basierter, das heißt, datengetriebener und gegebenenfalls kontinuierlich dazulernender, Chatbots bergen aus ethischer Perspektive große Risiken. Eine der zentralen Herausforderungen ist die Vermeidung von Verzerrung und Diskriminierung. Im Allgemeinen sollten Interaktionen und Entscheidungen der dem Chatbot zugrundeliegenden KI (in Form großer Sprachmodelle) frei von diskriminierenden Vorurteilen sein. Insbesondere jedoch darf der Chatbot keine menschenverachtenden, diskriminierenden oder anti-semitische Nachrichten verbreiten.

Der im März 2016 von der Microsoft Corporation 2016 veröffentlichte selbstlernende Chatbot „Tay“ reagierte innerhalb von weniger als 24 Stunden mit sexistischen und rassistischen Kommentaren, so dass der Bot abgeschaltet und zahlreiche Tweets wieder gelöscht werden mussten (Lobe 2022).

Die koreanische Firma Scatter Lab veröffentlichte im Dezember 2020 einen Chatbot mit dem Namen „Lee Luda“, der ebenfalls wieder offline genommen werden musste, da der Bot rassistische und homophobe Nachrichten aus-sendete (Wille 2021).

Anwendungsfall: Chatbot

Gänzlich anders gelagert sind die Bereiche der Rechtsetzung und der Rechtsprechung. Hier sind die Gebote von Transparenz, Nachvollziehbarkeit, Rechenschaftspflicht und Verhinderung von Bias von zentraler Bedeutung. So erscheint der Einsatz von algorithmischen Entscheidungsunterstützungssystemen vor dem Hintergrund der Eigenschaften gegenwärtiger Technologien hier nur in peripheren, beispielsweise vorbereitenden, Anwendungsbereichen sinnvoll.

Die öffentliche Verwaltung hat unter den staatlichen Institutionen einen besonderen Stellenwert, weil sie in der überwiegenden Mehrzahl der Fälle den Kontaktpunkt der Bürger:innen mit dem Staat darstellt und damit auch für den Staat steht. Somit spielt für das Vertrauen der Bürger:innen in den Staat das Vertrauen in die öffentliche Verwaltung eine zentrale Rolle. Außerdem wird die Arbeit der Verwaltung von der demokratischen Öffentlichkeit in besonderer Weise an Werten wie Rechtsstaatlichkeit, Rechenschaft, Transparenz, aber auch Gleichbehandlung, Menschenwürde und Sicherheit gemessen.

In Österreich belegen Meinungsumfragen ein anhaltend hohes Vertrauen in staatliche Institutionen wie Polizei, Bundesheer, Universitäten, Gerichtsbarkeit, Schulen und Finanzämter (OGM 2022). Wie die im Verlauf der letzten Jahre stark schwankenden Werte anderer öffentlicher Institutionen aus dem Bereich von Politik und Religion zeigen, sind derartige Vertrauenswerte allerdings nicht garantiert.

Somit trägt die Verwaltung bei der Anwendung neuer Technologien wie im Fall der Digitalisierung eine hohe Verantwortung: Hier ist bei der Veränderung der Entscheidungsprozesse besondere Achtsamkeit geboten. Vor diesem Hintergrund bietet dieser Leitfaden einen wichtigen Ansatzpunkt für die Beschäftigung mit Ethik bei der Einführung von KI.

5 Chancen und Herausforderungen beim Einsatz von KI in der Verwaltung

Dieser Abschnitt widmet sich den Chancen und Herausforderungen des Einsatzes von KI im Kontext der öffentlichen Verwaltung. Beginnend mit einer Beschreibung einiger KI-Anwendungen, die es in der Arbeitswelt bereits gibt, folgt eine Untersuchung der jeweiligen Chancen und Herausforderungen. Die Beispiele sind Personalauswahl und Überwachung am Arbeitsplatz.

Im Anschluss geht dieser Abschnitt auf mögliche Auswirkungen der Verwendung von KI in der öffentlichen Verwaltung sowie mögliche Folgen für Bürger:innen ein.

Ein Diskurs zu ökologischer bzw. nachhaltiger KI unter besonderer Berücksichtigung der Umweltauswirkungen bedingt durch den Einsatz von KI-Systemen zeigt mögliche Maßnahmen und Instrumente der öffentlichen Verwaltung im Sinne einer nachhaltigen Nutzung und Minimierung negativer Folgen auf.

Abschließend rückt noch das Thema der digitalen Souveränität und Datensouveränität in der öffentlichen Verwaltung in den Fokus. Nach einer einführenden Definition werden wesentliche Maßnahmen vorgeschlagen, die der öffentlichen Verwaltung eine Bewegung hin zu digitaler Souveränität erlauben.

5.1 KI und Auswirkungen auf die Arbeitswelt in der öffentlichen Verwaltung

Die öffentliche Verwaltung ist genauso wie andere Sektoren von den schnellen Fortschritten im Bereich der KI betroffen. Auch wenn einige der Anwendungen spartenspezifisch sind, gibt es viele Anwendungen, die auch in der öffentlichen Verwaltung prinzipiell verwendet werden können (Haslinger, 2022, 61f):

- Recruitingsoftware, die bei der Bewertung von Bewerbungen oder der Begründung eines Dienstverhältnisses unterstützt; dabei können sowohl Persönlichkeitstests zum Einsatz kommen oder Vorschläge gemacht werden, mit welchen Arbeitskolleg:innen die jeweilige Kandidat:in zusammenarbeiten soll,
- Vorgabe von Arbeitsschritten bei Standardabläufen,
- Kontrolle und Überwachung im Rahmen von Office-Software,
- HR (Personalmanagement, Lohnverrechnung, Personalentwicklung und -beurteilung), Dienst- und Arbeitszeiteinteilung,
- Verhaltenssteuerung durch Anreizsysteme und / oder Druck (Ampelsysteme, Gamification, Punktevergabe etc.),
- eigene Systeme wie z. B. Plattformarbeit.

Immer mehr Dienstleistungen werden über Online-Plattformen vermittelt und abgewickelt, darunter Fahrradbot:innen, Reinigungskräfte, Kreativschaffende, Übersetzer:innen, Clickworker:innen und Fahrer:innen. Diese Form der Arbeitsorganisation bietet für Arbeitssuchende aufgrund der geringen Einstiegsbarrieren (z. B. keine abgeschlossene Ausbildung bzw. Sprachkenntnisse erforderlich) Vorteile. Allerdings basieren Geschäftsmodelle der Plattformen häufig auf der (Schein-)Selbstständigkeit der Beschäftigten, wodurch die Unternehmen sich ihrer sozialen Verantwortung entziehen und für Arbeitnehmer:innen das Risiko der Ausbeutung besteht.

Wissen: Plattformarbeit

Ein zentrales Argument für die Einführung von KI ist die erwartete Effizienzsteigerung von Arbeits- und Verwaltungsprozessen. Im Allgemeinen birgt die Einführung von KI-Anwendungen das Potential, den Arbeitsaufwand für Verwaltungsbedienstete zu verringern. Dies gilt insbesondere für Routinetätigkeiten, wie zum Beispiel die Ausstellung von Dokumenten und Zertifikaten oder Beratung zu Themen, die häufig nachgefragt werden. KI-Anwendungen in der öffentlichen Verwaltung bringen jedoch auch Herausforderungen in Bezug auf die Akzeptanz der Mitarbeiter:innen mit sich. Ein besonderes Risiko ist die Auslagerung der Entscheidungsfindung von Menschen auf KI-Systeme, wodurch möglicherweise Risiken der Verletzung des Datenschutzes bzw. der Benachteiligung von Mitarbeiter:innen entstehen. Insbesondere die Erkennung von „abweichendem“ Verhalten im Sinne einer Optimierung der Arbeitseffizienz kann auf ungerechten Annahmen basieren und somit zur Diskriminierung von Arbeitnehmer:innen führen.

Bias in der KI stellt eine Unverhältnismäßigkeit bzw. Verzerrtheit beim Output von maschinell lernenden Algorithmen dar, die beispielsweise aufgrund von systematischen Fehlannahmen und Vorurteilen bei der Entwicklung des KI-Algorithmus oder aufgrund unvollständiger, unausgewogener bzw. verzerrter Eingabedaten erfolgen kann. KI-Systeme, deren Entscheidungen oder Vorhersagen einen solchen Bias enthalten bzw. davon beeinflusst sind, bilden die Realität nicht wirklichkeitsgetreu ab und können daher in weiterer Folge zur Diskriminierung von Menschen führen (vgl. Dilmegani 2022).

Geschlechterbias (engl. gender bias) bezeichnet beispielsweise in natürlicher Sprache enthaltene voreingenommene Vorstellungen oder Erwartungen, die über die Fähigkeiten, Eigenschaften und Rollen von Menschen aufgrund ihres Geschlechts urteilen. Ein prominentes Beispiel aus der Vergangenheit war ein KI-basiertes System, welches der Amazon-Konzern ab 2014 zur Vorauswahl von Bewerber:innen eingeführt hatte (Dastin 2022). Das System verwendete

zur Bewertung aktueller Bewerber:innen historische Daten. Da sich jedoch in der Vergangenheit überwiegend männliche Kandidaten beworben hatten, wurden männliche gegenüber weiblichen Kandidat:innen systematisch von der KI-Anwendung bevorzugt.

Wissen: Bias; Anwendung: Geschlechterbias

Beispielsweise bieten KI-gestützte Prozesse bei der Personalauswahl potentielle Effizienzgewinne und Kosteneinsparungen durch die automatisierte Verarbeitung von Bewerbungen und die Möglichkeit der Berücksichtigung eines größeren Pools von Kandidat:innen. Diese Prozesse beinhalten jedoch personenbezogene Daten, die dem Datenschutz unterliegen. Bewerber:innen haben gemäß Artikel 22 der DSGVO das Recht, nicht vollständig automatisierten Prozessen unterworfen zu werden. Eine transparente Kommunikation über den Auswahlprozess und insbesondere die Nennung von Gründen für Entscheidungen ist entscheidend, um die Fairness in solchen Auswahlprozessen zu gewährleisten.

Die KI-gestützte Überwachung am Arbeitsplatz wirft ebenfalls ethische Bedenken aufgrund möglicher problematischer Auswirkungen auf. Die Verletzung der Privatsphäre von Arbeitnehmer:innen, Diskriminierung aufgrund algorithmischer Entscheidungen und die Auswirkungen auf die psychische Gesundheit sind mögliche Probleme (vgl. Christl 2021).

Praktiken der Überwachung, die die Menschenwürde berühren und durch keine Betriebsvereinbarung geregelt sind, sind unzulässig und insofern abzulehnen.

Um eine erfolgreiche Einführung von KI-Systemen zu gewährleisten, ist es von entscheidender Bedeutung, die Verwaltungsbediensteten zu unterstützen und ihnen die notwendigen Schulungen und Kompetenzen zu vermitteln (siehe KI-Literacy: Definition in Abschnitt 5, Empfehlungen in Abschnitt 11). Dies versetzt sie in die Lage, effektiv mit KI-Systemen umzugehen und mit diesen zu interagieren. Durch die Förderung von KI-Kenntnissen innerhalb der Verwaltung kann außerdem die Abhängigkeit von externen Expert:innen oder Unternehmen verringert werden, was die Autonomie der Verwaltung fördert und die Verwaltungsbediensteten in die Lage versetzt, sich neue Methoden und Technologien anzueignen.

Wissen: KI-Kompetenzen in der Verwaltung

5.2 KI in der Verwaltung und Auswirkungen auf die Bevölkerung

Der Einsatz von KI-Anwendungen in der öffentlichen Verwaltung hat nicht nur Auswirkungen auf die Mitarbeiter:innen des öffentlichen Sektors, sondern auch auf die Bürger:innen. Ein Grund dafür ist, dass der Einsatz von KI in der öffentlichen Verwaltung zur Einführung neuartiger oder automatisierter Dienstleistungen bzw. Bearbeitungsformen für die Bürger:innen führt. Beispiele dafür sind in Österreich die App „Digitales Amt“, das Unternehmensserviceportal und FinanzOnline. In diesem Zusammenhang ist die Verantwortung der öffentlichen Verwaltung als Vermittler zwischen Bürger:innen und Politik entscheidend. Vertrauen, Akzeptanz und Legitimität sind in diesem Sinne wichtige Faktoren. Ethische Belange, einschließlich Transparenz, Fairness und Datenschutz, sind ebenfalls von Bedeutung. Die Beurteilung der Auswirkungen von KI auf die Bürger:innen, die Ermittlung des Mehrwerts für die Bürger:innen und die Klärung ethischer Fragestellungen sind daher vor dem Einsatz von KI unbedingt in den Vordergrund zu stellen.

Unter dem Aspekt des Mehrwerts für die Bürger:innen lassen sich im Zusammenhang mit dem Einsatz von KI in der öffentlichen Verwaltung vor allem die folgenden Vorteile hervorheben:

- Interaktion und Kommunikation mit den Bürger:innen: Die Einführung von KI in der öffentlichen Verwaltung kann die Art und Weise, in der Mitarbeiter des öffentlichen Sektors mit den Bürger:innen interagieren, verändern und möglicherweise verbessern. Insbesondere können virtuelle Assistenten den Bürger:innen helfen, effizienter auf Informationen und Dienstleistungen zuzugreifen und schneller Antworten auf ihre Fragen zu erhalten.

In Finnland hat die Einwanderungsbehörde einen KI-Chatbot namens Kamu eingeführt. Kamu beantwortet spezifische Fragen, um das Einwanderungsverfahren für ausländische Bürger:innen zu erleichtern.² Wenn Kamu eine bestimmte Frage nicht versteht, wird den Benutzer:innen angeboten, sie direkt an eine/n Mitarbeiter:in weiterzuleiten, welche/r dann ebenfalls über das Chat-Protokoll verfügt. Der Chatbot ist in Englisch und Finnisch verfügbar.

Anwendungsfall: Chatbot Kamu in Finnland

2 [Chatbot Kamu | Maahanmuuttovirasto \(migri.fi\)](#)

- **Effizienz und Personalisierung:** KI kann Routineaufgaben automatisieren und so die manuelle Arbeit von Verwaltungsbediensteten reduzieren, was zu einer schnelleren Bereitstellung von Dienstleistungen für die Bürger:innen führen kann. Außerdem kann KI dazu beitragen, die Bedürfnisse der einzelnen Bürger:innen schneller zu erkennen und so personalisierte Dienstleistungen anzubieten. KI hat beispielsweise das Potenzial, die Sozialfürsorge zu verbessern, indem sie gefährdete Personen mit besonderen Bedürfnissen frühzeitig erkennt und maßgeschneiderte Unterstützung bietet. Durch die Analyse verschiedener Datenquellen können KI-Algorithmen schnell und effizient diejenigen identifizieren, die Unterstützung benötigen. Durch die proaktive Erkennung von Personen, die möglicherweise Unterstützung benötigen, kann die öffentliche Verwaltung ein rechtzeitiges Eingreifen und die Bereitstellung personalisierter Dienstleistungen sicherstellen, die den individuellen Bedürfnissen der Betroffenen entsprechen.
- **Zugänglichkeit und Chancengleichheit:** KI kann dazu beitragen, öffentliche Dienstleistungen für Bürger:innen, die beispielsweise in abgelegenen Gebieten leben, durch die Erbringung von verwaltungsbezogenen Dienstleistungen aus der Ferne zugänglicher zu machen. Darüber hinaus können KI-Anwendungen die Bürger:innen unterstützen, z. B. beim Ausfüllen von Formularen oder bei der Bewältigung von Sprachbarrieren bei Fragen.

Neben ihren vielfältigen Vorteilen birgt die fortschreitende Entwicklung von KI auch das Potenzial, eine digitale Kluft („digital divide“) zu schaffen oder zu verstärken. Diese digitale Kluft äußert sich in einer erheblichen Diskrepanz zwischen Personen, die neue Technologien beherrschen, und solchen, die dies nicht tun. Bedauerlicherweise trägt eine derartige Kluft zur Verschärfung der bestehenden sozialen Ungleichheiten und zur weiteren Ausgrenzung bestimmter gesellschaftlicher Gruppen bei. Um die unerwünschten Folgen eines solchen Szenarios abzuwenden, ist es wichtig, die Bevölkerung mit den erforderlichen Kompetenzen auszustatten, um KI-Anwendungen optimal nutzen zu können. In diesem Zusammenhang kommt der KI-Kompetenz (KI-Literacy) eine zentrale Rolle zu, denn sie stellt sicher, dass der oder die Einzelne nicht unangemessen benachteiligt wird. KI-Kompetenz umfasst dabei nicht nur ein Verständnis grundlegender KI-Konzepte, sondern auch die Kultivierung von Fähigkeiten zum kritischen Denken, die für die effektive Nutzung von KI-basierten Technologien erforderlich sind.

Wissen: Bekämpfung der digitalen Kluft durch KI-Kompetenz in der Bevölkerung

Vertrauensprobleme / Akzeptanzprobleme

Ein wesentliches Vertrauensproblem der Bürger:innen in öffentliche KI-Anwendungen besteht darin, dass es der KI an demokratischer Rechenschaftspflicht mangelt (vgl. Starke und Lünich 2020). Transparenz und Nachvollziehbarkeit für KI-basierte Handlungen und Entscheidungen sind in diesem Zusammenhang wesentliche Faktoren. Denn Vertrauen in öffentliche Verwaltungsprozesse wird grundsätzlich dadurch geschaffen, dass Bürger:innen angemessene Begründungen für die Entscheidungen der öffentlichen Verwaltung erhalten. Darüber hinaus verringert die Einführung algorithmischer Entscheidungsfindung potenziell die menschliche Fähigkeit, sich an prekäre Situationen anzupassen und auf sie zu reagieren. KI fehlt das kontextbezogene Verständnis und die situative Anpassungsfähigkeit, die Menschen besitzen (siehe auch Aoki 2020). Diese Einschränkung kann insbesondere in sensiblen oder kritischen Szenarien, in denen menschliches Urteilsvermögen und Flexibilität bedeutsam sind, weitreichende Folgen haben.

Darüber hinaus stellt sich die Frage des Schutzes der Privatsphäre im Zusammenhang mit dem Einsatz von KI in der öffentlichen Verwaltung. Die öffentliche Verwaltung verfügt über umfangreiche Daten von Bürger:innen, sodass – auch unter Einhaltung datenschutzrechtlicher Bestimmungen – allein die Tatsache, dass personenbezogene Daten durch KI-Anwendungen verarbeitet werden, ein Unbehagen in der Bevölkerung hervorrufen kann.

Zwar kann die Nutzung von Daten zu einer Verbesserung der Dienstleistungen der öffentlichen Verwaltung beitragen, allerdings muss das Risiko der Verletzung der Privatsphäre berücksichtigt werden. Grundsätzlich wissen Bürger:innen oft nicht, welche ihrer Daten auf welche Weise in KI-Anwendungen verwendet werden, was ernsthafte Bedenken hinsichtlich der Einhaltung der Datenschutzbestimmungen aufwirft (Madan und Ashok 2022).

Je weniger die Bevölkerung über die eingesetzten KI-Anwendungen informiert ist und je weniger sie deren Funktionsweise verstehen kann, desto geringer ist ihr Vertrauen in solche KI-basierten Anwendungen. Da Vertrauensfragen eigentlich immer zu Akzeptanzproblemen bei der Einführung neuer Technologien wie KI führen, ist es wichtig, die Gründe für die Akzeptanz oder Ablehnung von KI im öffentlichen Dienst zu verstehen. Dies unterstützt letztlich die Gestaltung und Einführung von KI-Anwendungen im öffentlichen Dienst.

Maßnahmenoptionen

Im Folgenden geben wir eine Reihe von Handlungsempfehlungen, die der öffentlichen Verwaltung helfen, die Akzeptanz von KI-basierten Diensten zu fördern und damit den Einsatz von KI-Anwendungen für öffentliche Dienstleistungen zu erleichtern. Diese Maßnahmen sind als Ergänzung zur Beachtung allgemeiner Leitprinzipien (wie Transparenz, Sicherheit, Rechenschaftspflicht, Vermeidung von Verzerrungen usw.) für die Entwicklung und den Einsatz vertrauenswürdiger KI zu sehen. Die Maßnahmen, auf die im Folgenden kurz eingegangen wird, sind erstens *Co-Kreation und Partizipation* zur Förderung der Bürgerbeteiligung bei der Entwicklung von KI-Anwendungen und der Akzeptanz und

Einhaltung ethischer Standards. Zweitens die *Benutzerfreundlichkeit und KI-Literacy für die Öffentlichkeit* zur Schaffung von Transparenz und als Beitrag zur Aufklärung über KI-Anwendungen, um Chancengleichheit und digitale Kompetenzen in öffentlichen Dienstleistungen zu fördern. Und schließlich *Opt-out-Möglichkeiten*, welche Bürger:innen Kontrolle darüber geben, welche ihrer Daten von der öffentlichen Verwaltung gespeichert und verarbeitet werden, und zudem die Privatsphäre hinsichtlich der Verwendung der Daten durch KI-basierte Dienste der öffentlichen Verwaltung gewährleisten.

Co-Kreation und Partizipation

Die Zusammenarbeit mit den Bürger:innen wird oft als Schlüsseldeterminante für Akzeptanz von KI-Anwendungen in der öffentlichen Verwaltung betrachtet (Madan und Ashok 2022; Gesk und Leyer 2022). Durch die Einbeziehung der Bürger:innen kann die öffentliche Verwaltung sicherstellen, dass die KI-Systeme gesellschaftlichen Bedürfnissen entsprechen und mit den öffentlichen Werten und Zielen in Einklang stehen.

Co-Kreation bezieht sich auf den kollaborativen und partizipativen Prozess der Einbeziehung der Bürger:innen in die Entwicklung von KI-Anwendungen. Dieser Ansatz ermöglicht somit die Berücksichtigung unterschiedlicher Perspektiven.

Pilotprojekte oder Innovationslabore können ein nützlicher Rahmen für die Mitgestaltung und Beteiligung in der öffentlichen Verwaltung sein. Sie bieten eine sichere und kontrollierte Umgebung, um mit KI-Anwendungen zu experimentieren, Annahmen zu testen und die Bürger:innen in den Entwicklungsprozess einzubeziehen. Auf diese Weise können Fähigkeiten und Wissen für den Umgang mit KI-Anwendungen aufgebaut werden, um das Bewusstsein für potenzielle ethische und rechtliche Fragen zu fördern.

Benutzerfreundlichkeit und KI-Literacy für die Öffentlichkeit

Wenn KI-Anwendungen für öffentliche Dienstleistungen der Verwaltung eingesetzt werden, sollte zunächst sichergestellt werden, dass die KI-Anwendung und die mögliche Interaktion mit den Bürger:innen transparent, einfach und verständlich dargestellt wird. Dies kann durch einfache Erläuterungen von Vor- und Nachteilen des Einsatzes von KI, z. B. auf den jeweiligen Online-Portalen der öffentlichen Verwaltung, geschehen.

Um Chancengleichheit beim Einsatz von KI-Anwendungen für alle Bürger:innen zu garantieren, könnten Initiativen gesetzt werden, die darauf abzielen, öffentliches Wissen über KI-Technologien und Anwendungen zu erweitern. Ein Beispiel dafür ist die „Digitale Kompetenzoffensive“ von Bundesministerium für Finanzen (BMF), Bundesministerium für Kunst, Kultur, öffentlichen Dienst und Sport (BMKÖS), Bundesministerium für Wirtschaft und Arbeit (BMWA) und Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF), die das Ziel verfolgt, digitale Basiskompetenzen in der Bevölkerung sowie IT-Kompetenzen für die Wirtschaft zielgerichtet zu verbessern. Ressorts, Länder, Sozialpartner, Städte und Gemeinden, Unternehmen und Bildungsanbieter wollen unter wissenschaftlicher Begleitung beim Thema digitale Kompetenzen strategisch abgestimmt zusammenarbeiten. Ein Schlüsselprojekt der Digitalen Kompetenzoffensive ist das 2023 vorgestellte österreichische Kompetenzmodell für digitale Kompetenzen

„DigComp 2.3 AT“ im europäischen DigComp 2.1-Referenzrahmen. Letzteres ist ein von der Europäischen Kommission herausgegebenes Dokument, das sich mit dem Themenkomplex digitale Kompetenzen von Bürger:innen befasst und beschreibt, was digitale Kompetenzen konkret beinhalten.

Opt-out Möglichkeiten

Eine weitere wichtige Maßnahme ist die Bereitstellung von Opt-out-Optionen für Bürger:innen bei der Nutzung von KI-basierten Diensten. Opt-out-Optionen eröffnen die Möglichkeit, sich gegebenenfalls gegen automatisierte Dienste zu entscheiden und stattdessen die Hilfe von menschlichen Verwaltungsbediensteten in Anspruch zu nehmen. Diese Maßnahme stellt sicher, dass die Bürger:innen die Kontrolle über ihre Daten und ihre Privatsphäre behalten und bei Bedarf personalisierte Unterstützung erhalten können.

Die öffentliche Verwaltung sollte auch sicherstellen, dass die Bürger:innen über ihr Recht auf Ablehnung von KI-basierten Diensten informiert sind. Dies kann durch die Bereitstellung klarer Informationen über das Opt-out erreicht werden, beispielsweise durch benutzerfreundliche Schnittstellen.

5.3 KI und Ökologie

Im Diskurs um die Entwicklung nachhaltiger Systeme künstlicher Intelligenz (engl. *sustainable AI*) werden die Auswirkungen von KI-Systemen auf Mensch und Umwelt im Zusammenhang mit ethischer und verantwortungsvoller KI diskutiert (Coeckelbergh 2021). Dabei geht es unter anderem um die Frage, welcher Energieverbrauch und damit einhergehende Treibhausgasemissionen mit der Erstellung und Verwendung von KI-Systemen verbunden sind.

Im vergangenen Jahrzehnt haben insbesondere datengetriebene Methoden des Maschinellen Lernens in der Künstlichen Intelligenz weite Verbreitung erfahren. Aufgrund des erheblichen Ressourcenaufwands bei der Herstellung und Nutzung dieser Systeme haben ökologische Aspekte, wie der damit verbundene CO₂ Verbrauch, in den letzten Jahren zunehmend das öffentliche Interesse geweckt.

Ein Beispiel für die CO₂-Bilanz bei der Erstellung von statistischen Sprachmodellen ist die Verwendung vieler miteinander verbundener Computer (sogenannte Computer-Cluster) und Rechenzentren, um große Mengen an Textdaten zu verarbeiten und auszuwerten. Während des Lernprozesses werden beim maschinellen Lernen häufig viele verschiedene Modelle und Algorithmen wiederholt getestet, um ein optimales Ergebnis zu erzielen. Allerdings erfordert jeder einzelne diese Testdurchläufe erhebliche Rechenleistung, was insgesamt zu einem hohen Energieverbrauch und somit zu erhöhten Treibhausgasemissionen führt.

Laut einer Studie der Universität Massachusetts Amherst (Strubell et al. 2019) beträgt der ökologische Fußabdruck für das Training mehrerer großer KI-Modelle zur Spracherkennung ca. 284 Tonnen CO₂.

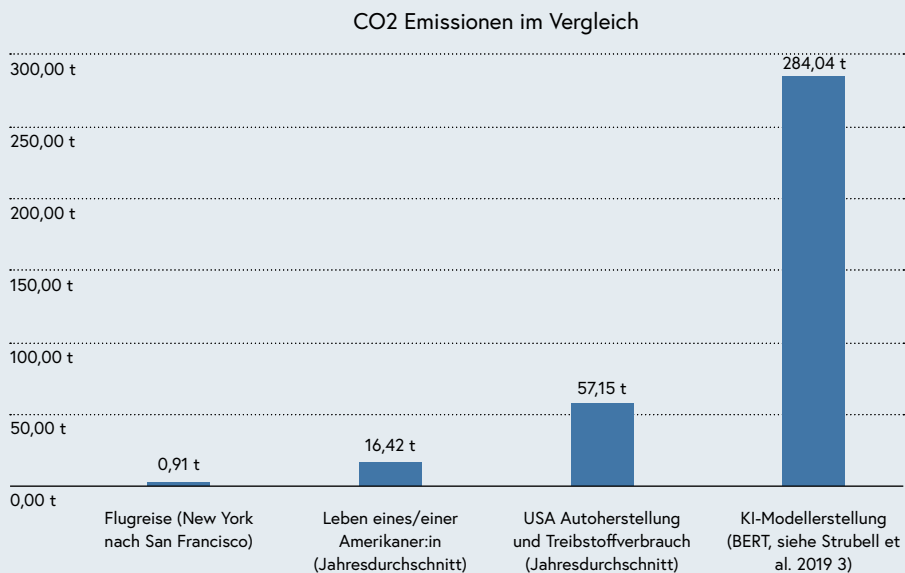


Abbildung 1: Vergleich der Treibhausgasemissionen einer Flugreise, des Lebens eines Menschen in den USA im Jahresdurchschnitt, eines Autos während der Lebensdauer und der Erstellung eines KI-Modells

Quelle: Studie der Universität von Massachusetts Amherst (Strubell et al. 2019).

Hierbei wurden insbesondere statistische Modelle zur Verarbeitung natürlicher Sprache betrachtet (Sprachmodell BERT, siehe Devlin et al. 2018).

Der CO₂ Ausstoß für die Entwicklung eines Sprachmodells entspricht fast dem Fünffachen der Emissionen eines durchschnittlichen amerikanischen Personenkraftwagens (einschließlich dessen Herstellung) heruntergerechnet auf ein Jahr.

Anwendungsfall: Die CO₂ Bilanz von KI am Beispiel von Sprachmodellen

Schließlich werden für den Betrieb eines KI-Systems bis zum Ende des Lebenszyklus Energie- und Ressourcen für Hardwareinfrastruktur, sowie für Wartung und Kühlung aufgebracht (Rohde et al. 2021, 43).

Eine Reihe von Maßnahmen und Entscheidungen, die auch seitens der öffentlichen Verwaltung beeinflusst bzw. gesteuert werden können, bergen das Potential den Ressourcenverbrauch zu reduzieren und so die negativen Auswirkungen auf die Umwelt zumindest abzumildern:

- Wird durch die finale Optimierung eines Algorithmus nur eine geringfügige Verbesserung erreicht und der Energie- und Ressourcenaufwand ist zugleich unverhältnismäßig hoch, sollte überlegt werden, welche Genauigkeit im vorliegenden Anwendungsszenario wirklich benötigt wird. Sowohl in einer Ausschreibung als auch in der Entwicklung können durch die Einschränkung der geforderten Genauigkeit Energie und Ressourcen gespart werden.
- Die Erstellung generalisierter Modelle, die freie Bereitstellung von Erkenntnissen und Ergebnissen (Open Access) und eine Kultur des Teilens vereinfacht die Übertragbarkeit auf neue Anwendungsszenarien und gewährleistet die Wiederverwendbarkeit durch andere Organisationen, was ein erhebliches Potential zur Energie- und Ressourceneinsparung birgt. Die Entwicklung einer Strategie und eigener Richtlinien in Bezug auf das Veröffentlichen und Teilen von Erkenntnissen und Ergebnissen kann Synergien zwischen Organisationen der Verwaltung fördern.
- Bereits bei der Planung der Einführung einer KI-Lösung sollte geprüft werden, welche Formen der Zusammenarbeit und Nutzung technischer Infrastruktur auf lokaler, nationaler, internationaler oder europäischer Ebene möglich sind. Dadurch können Ressourcen gemeinsam genutzt, Fachwissen gebündelt und Standards geschaffen werden.
- Relevant ist auch der Einsatz zertifizierter Rechenzentren, die eine ökologische Ausrichtung hinsichtlich Energieeffizienz vorweisen können und die beim Energiemix auf die Nutzung alternativer Energieformen (Solar, Wind, etc.) achten. Diese können in öffentlichen Ausschreibungen bei der Anschaffung von KI-Systemen als Bedingung definiert werden.

Es ist davon auszugehen, dass die Entwicklung und Anwendung von KI-Systemen in vielen Anwendungsbereichen ohne Berücksichtigung ökologischer Nachhaltigkeit erfolgt (Rohde et al. 2021, 20), und dass oft technologische Neuerung und Steigerung der Effizienz im Vordergrund stehen. Die öffentliche Verwaltung verfügt hier über die Möglichkeit, hohe ökologische Standards für die Entwicklung und den Einsatz von KI-Systemen zu setzen, beispielsweise indem verbindlicher Kriterien für die öffentliche Beschaffung solcher Systeme definiert und zugrunde gelegt werden.

Die UNESCO hat sich diesbezüglich in ihrer Empfehlung zur Ethik in der Künstlichen Intelligenz dafür ausgesprochen, dass ihre Mitgliedstaaten bei der Auswahl von KI-Methoden aufgrund des potenziellen daten- oder ressourcenintensiven Charakters der Technologien besonders auf die Auswirkungen auf die Umwelt achten sollten (UNESCO 2022, 31). Außerdem seien Technologien mit besserer Daten-, Energie- und Ressourcen-Effizienz zu bevorzugen, und KI-Technologien sollten nicht eingesetzt werden, wenn die Gefahr unverhältnismäßig negativer Auswirkungen auf die Umwelt besteht (UNESCO 2022, 31).

5.4 Digitale Souveränität in der Verwaltung

Im Allgemeinen bezeichnet der Begriff „Digitale Souveränität“ die Möglichkeit eines Staates oder einer Organisation, autonom Entscheidungen über die Technologieinfrastruktur bzw. deren Nutzung zu fällen, ohne von externen Akteuren in dieser Entscheidung eingeschränkt zu werden. In Bezug auf die öffentliche Verwaltung bedeutet dies die Fähigkeit, digitale Dienstleistungen und Prozesse unabhängig zu gestalten, die Kontrolle über ihre Daten, Systeme und Technologien zu behalten und dabei die Interessen der Bürger:innen und der öffentlichen Institutionen zu schützen. Sie ist die Grundlage für eine kompetentes und vertrauenswürdigen Handeln der öffentlichen Verwaltung. Ein wesentliches Kontrollinstrument zur Wahrung digitaler Souveränität ist das öffentliche Beschaffungswesen, wodurch übermäßige Abhängigkeiten von einzelnen externen Anbietern vermieden werden können.

Die Verwaltung sollte im Rahmen der politischen und rechtsstaatlichen Vorgaben entscheiden können, welche Dienstleistungen in welcher Form und über welchen Zeitraum für Bürger:innen angeboten werden. Externe Abhängigkeiten stellen dabei ein Risiko hinsichtlich der Kontinuität, Zuverlässigkeit und Qualität der Dienstleistung dar. Daher ist es wichtig, eigenständig Wissen und Erfahrung in Bezug auf die Entwicklung und den Einsatz von KI-Technologien aufzubauen. Generell steht die einseitige Abhängigkeit von einzelnen Ländern, Organisationen oder Firmen einer Technologiesouveränität entgegen (Rat für Forschung und Technologieentwicklung 2021). Dies bezieht sich insbesondere auf nicht-demokratische Staaten, die im Krisenfall unter Umständen wenig verlässliche Kooperationspartner darstellen.

In der EU wurde die entsprechende Diskussion durch das Dokument „2030 Digital Compass: the European way for the Digital Decade“ und dem entsprechenden Schwerpunkt im europäischen Forschungsrahmenprogramm „Horizon Europe“ befördert (Cluster 4: Digital, Industry and Space). Außerdem gibt es verschiedene Schritte in Richtung einer Industriestrategie, beispielsweise im Hinblick auf den „European Chips Act“, oder auch den „Data Governance Act“ (siehe Abschnitt 8.4).

Mögliche Instrumente der digitalen Souveränität umfassen die Förderung einer innovativen öffentlichen Beschaffung, die Beeinflussung von Standardisierungsprozessen, eine Bevorzugung und aktive Unterstützung von Open-Source Soft- und Hardware, sowie die Stärkung europäischer Prozesse bzw. auf internationaler Ebene des Multilateralismus, um Monopolstellungen zu vermeiden (Edler et al. 2020).

Sinnvoll scheint auch die Herstellung einer Datensouveränität, da Daten die Grundlage für die Digitalisierung und auch für KI-Anwendungen sind. Von besonderer Bedeutung ist dabei eine robuste Dateninfrastruktur der öffentlichen Verwaltung im Hinblick auf das Sammeln, Speichern und Verwalten großer Datenmengen (unter gleichzeitiger Wahrung von Privatsphäre und Sicherheit).

Im Hinblick auf das Wissen um Digitalisierung und KI ist die KI-Literacy auch hier von zentraler Bedeutung. Diese kann durch Schulungsmaßnahmen auf verschiedenen Ebenen hergestellt werden (siehe Empfehlungen in Abschnitt 11). Die Zertifizierung von Daten und KI-Modellen kann in Bezug auf digitale bzw. Datensouveränität unterstützend wirken.

Die größte Wirkung kann die öffentliche Verwaltung dabei durch die Vorgabe von Bedingungen und die Definition zu erfüllender Kriterien im Rahmen von Beschaffungsvorgängen erzielen (siehe ebenfalls Abschnitt 11). Die österreichische KI-Strategie „AIM AT 2030“ identifiziert die Beschaffung als „wichtiges strategisches Instrument [...], das zur Forcierung und Marktüberleitung von Innovationen eingesetzt werden kann. Der Staat kann z. B. als nachfragendes Organ für ethische und vertrauenswürdige KI agieren und dadurch Märkte definieren, Standards setzen und seine Effizienz steigern. Zugleich können innovative Lösungen von Start-ups, jungen Unternehmen und Kleinbetrieben davon profitieren“ (BMK und BMDW 2021, 56). Die öffentliche Verwaltung kann dabei durch ihre Vergabetätigkeit eine Vorbildwirkung entfalten. Bei der Beschaffung von KI-Anwendungen für KI-Projekte ist die Innovationsfördernde Öffentliche Beschaffung–Servicestelle (IÖB) ein Ansprechpartner für die Verwaltungsbediensteten. Hier ist neben dem Bundesvergabegesetz auch das White Paper der IÖB-Servicestelle eine erste Orientierung (IÖB 2021). Darüber hinaus sollten bei KI-Anwendungen wichtige ethische Grundsätze, wie sie im „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ (siehe Abschnitt 10.1) ausgeführt werden, schon zu Beginn von Entwicklungsprozessen, wie z. B. vom „Ethics by Design“-Ansatz vorgeschlagen (siehe Abschnitte 10.3 und 11), bedacht werden.

Ein Good-Practice-Beispiel hinsichtlich des Beschaffungswesens durch die öffentliche Verwaltung stammt aus Amsterdam. Die Gemeinde Amsterdam legt beim Zukauf von KI-Anwendungen vertraglich einen Rahmen für die Informationen fest, die die Anbieter bereitstellen müssen. Auf diese Weise kann die Gemeinde die Qualität und die Risiken von Anwendungen bewerten, ohne dass der Anbieter gezwungen ist, vertrauliche Unternehmensinformationen herauszugeben. Es gibt drei Arten von Informationen, die mit dem Anbieter in den Vertragsbedingungen vereinbart werden:

- **Technische Transparenz:** Sie gibt Einblick in die technische Funktionsweise der Anwendung, zum Beispiel in den Code, auf dessen Basis die Anwendung funktioniert. Technische Transparenz ist im Falle eines Audits oder wenn es notwendig ist, die „Erklärbarkeit“ (siehe unten) zu erläutern, von besonderer Bedeutung.
- **Verfahrenstransparenz:** Sie verdeutlicht den Zweck der Anwendung und die Schritte, in denen er erstellt wurde. Zum Beispiel eine Beschreibung der getroffenen Entscheidungen und Annahmen, welche Art von Daten verwendet wurden, und wie einer möglichen Verzerrung entgegengewirkt wurde. So kann überprüft werden, ob die richtigen Maßnahmen zur Qualitätssicherung und Risikominderung getroffen wurden.
- **Erklärbarkeit:** Wenn das Ergebnis einer Anwendung jemanden persönlich betrifft, gelten strengere Regeln. Der Anbieter muss mitwirken, um die Entscheidungsfindung der Anwendung auf individueller Ebene transparent zu machen. Dies wird als „Erklärbarkeit“ einer Anwendung bezeichnet. In weiterer Folge ermöglicht das den Bürger:innen, gegebenenfalls rechtlich gegen eine Entscheidung oder ein Ergebnis vorzugehen (City of Amsterdam 2021).

Diesem Beispiel folgend ist der Zugang zu Quellcodes (d.h. die einer KI-Anwendung zugrundeliegenden und nachvollziehbaren Anweisungen) zugekaufter KI-Anwendungen bedeutsam, um dem Kriterium der Transparenz gerecht zu werden. Ethische Prüfungen der in der Verwaltung verwendeten KI-Systeme sollen von verschiedenen Organisationen mit relevanter Expertise gemeinsam mit den Ministerien durchgeführt werden können, ohne dem Anbieter-Unternehmen Schaden durch eine etwaige Lüftung von Betriebsgeheimnissen zuzufügen.

Gegenüber Bürger:innen sollen KI-Anwendungen verständlich erklärt werden können, z. B. durch einen Chatbot, der Fragen bezüglich des Einsatzes von KI in der Verwaltung beantwortet („Mentoring-Bot“). Black-Box Anwendungen, also Applikationen, bei denen der Prozess zwischen Eingabedaten und Ergebnis nicht nachvollziehbar ist, und die keine Form der Erklärbarkeit anbieten, sollten nach Möglichkeit nicht eingesetzt werden, da sie Transparenz- und Vertrauensanforderungen nicht entsprechen. Anders formuliert: Black-Box Anwendungen sollen angewandt werden können, sofern Erklärbarkeit gewährleistet wird.

6 Abwägung: Datengetriebene KI in der Verwaltung, oder nicht?

Dieser Abschnitt beschäftigt sich systematisch mit der Frage, ob man eine KI-Anwendung in der öffentlichen Verwaltung einführen soll oder nicht.

Als Entscheidungshilfe bezüglich der Frage, ob und unter welchen Bedingungen datengetriebene KI-Technologie in der Verwaltung eingesetzt werden kann, dient der in Abbildung 2 gezeigte Entscheidungsbaum. Es sei darauf hingewiesen, dass sich das komplexe Gebiet der Einführung von KI-Technologien nicht vollumfassend auf die dargestellten Fragen reduzieren lässt. Vielmehr dient dieser Entscheidungsbaum dazu, zentrale Fragestellungen „auf einen Blick“ und nach Prioritäten geordnet darzustellen.

Der Wurzelknoten des Entscheidungsbaums beginnt mit der grundsätzlichen Frage, ob überhaupt geeignete Anwendungsfälle für den Einsatz dieser Technologien gegeben sind.

Bei datengetriebene KI-Technologien lernen Algorithmen die für spezifische Aufgabenstellungen benötigten Zusammenhänge und Muster aus den Daten. Daher macht es wenig Sinn, diese Technologien in Erwägung zu ziehen, wenn keine Daten für den Anwendungsfall verfügbar sind, beziehungsweise wenn diese auch nicht im Zuge der Einführung beschafft werden können. Darüber hinaus ist der Einsatz von KI-Technologie im Allgemeinen nur dann sinnvoll, wenn der Anwendungsfall bestimmte Kriterien, wie zum Beispiel mindestens eines der folgenden, erfüllt:

- Es gibt sich wiederholende Aufgaben, die sich automatisieren lassen, die also eine Maschine grundsätzlich übernehmen könnte.
- Die Aufgabenstellung stützt sich nicht überwiegend auf menschliche Urteilsfähigkeit. Handelt es sich um komplexe Aufgaben, deren Lösungsansatz erst im Zuge einer Problemanalyse ermittelt werden muss, so sind diese Aufgaben üblicherweise mit Methoden des maschinellen Lernens (vgl. schwache KI, Abschnitt 4) nicht zu lösen.
- Es gibt einen Bedarf der Entscheidungsunterstützung, das heißt, KI-Technologien können in unterstützender Form, zum Beispiel zur Aufbereitung und Visualisierung benötigter Informationen, einen Beitrag für bestimmte Aufgabenstellungen leisten.

Für den Fall, dass ein geeigneter Anwendungsfall vorliegt, stellt sich außerdem die grundlegende Frage, ob eine Rechtsgrundlage gegeben ist. Dabei handelt es sich um eine notwendige Voraussetzung, ohne die der Einsatz jeglicher Technologie (eingeschlossen der KI) für die öffentliche Verwaltung nicht ohne Verletzung rechtlicher Vorgaben möglich ist.

Ist das notwendige Kriterium der Rechtskonformität erfüllt, stellt sich die Frage, inwieweit personenbezogene Daten für die Erstellung von Modellen beziehungsweise Optimierung der KI-Technologie involviert sind. Werden hier keinerlei personenbezogene Daten benötigt, wie das beispielsweise bei der Verwendung von Umgebungstemperatur-Sensordaten der Fall wäre, so ist möglicherweise ein geeignetes Anwendungsfeld für datengetriebene KI-Technologie gegeben. Sind dagegen personenbezogene Daten involviert, so ist zu prüfen, ob datenschutzrechtliche Bedenken vorliegen, die nicht mit Verfahren der Anonymisierung, Pseudonymisierung beziehungsweise durch Abstrahierung ausgeräumt werden können. In jedem Fall ist hier streng auf die Einhaltung der DSGVO zu achten. Insbesondere, um zu verhindern, dass trotz des Einsatzes von Verfahren zur Gewährleistung der DSGVO-Konformität mittels einer Verschränkung mit externen Datenquellen letztendlich doch Rückschlüsse auf Einzelpersonen gezogen werden können. In jedem Fall muss bedacht werden, dass selbst dann, wenn alle Anstrengungen zur Einhaltung der DSGVO unternommen werden, bei Vorliegen personenbezogener Daten immer das Risiko der Verletzung datenschutzrechtlicher Vorgaben gegeben ist.

Schließlich ist zu erwähnen, dass datengetriebene Technologien oft sogenannte Basismodelle (engl. Foundation Models) verwenden, die statistische Zusammenhänge aus sehr vielen Eingabedaten, die häufig aus Social Media- und Internet-Datenquellen gesammelt wurden, erkennen. Algorithmen können dann mit dem Training an vergleichsweise wenigen Beispielen für einen spezifischen Anwendungsbereich optimiert werden. Dabei besteht jedoch das Risiko, Verzerrungen, die in den ursprünglichen Daten der Basismodelle enthalten sind, zu übernehmen und dadurch bestimmte Gruppen zu diskriminieren. Neben der Vermeidung von Vorurteilen ist darauf zu achten, dass Minderheiten adäquat repräsentiert sind und die KI-Technologie auch für Randgruppen erprobt wurde. Wurden diese Überprüfungen durchgeführt, ist möglicherweise ein geeigneter Anwendungsfall für datengetriebene KI-Technologie gegeben.



Abbildung 2: Entscheidungsbaum zur Verwendung daten-getriebener KI-Technologie

7 Rechtlicher Rahmen

Abschnitt 8 gibt einen Überblick über den rechtlichen Rahmen für die öffentliche Verwaltung und den Einsatz von KI. Zunächst wird die entscheidende Rolle des Rechts als Grundlage des Verwaltungshandelns hervorgehoben und das Legalitätsprinzip erörtert. Der Abschnitt befasst sich dann mit der Datenschutz-Grundverordnung (DSGVO), einer wichtigen EU-Verordnung für den Schutz personenbezogener Daten, und gibt ein anschauliches Anwendungsbeispiel für einen verantwortungsvollen Umgang mit Daten. Darüber hinaus wird in diesem Abschnitt der AI Act vorgestellt, der darauf abzielt, rechtliche Rahmenbedingungen spezifisch für die Nutzung von KI in der EU zu schaffen. Auf nationaler Ebene ist zusätzlich die KI Strategie Artificial Intelligence Mission Austria 2030 (AIM AT 2030) zentral, um die Ziele und Vorstellungen der österreichischen Bundesregierung zu verstehen. Ebenfalls kurz erwähnt werden einige zentrale EU-Normierungen, nämlich die Produkthaftungsrichtlinie, die KI-Haftungsrichtlinie und der Data Governance Act, die beschlossen wurden, um Verantwortlichkeit, Sicherheit und Schutz in den Mitgliedsstaaten im Bereich der Digitalisierung zu gewährleisten.

Die beschriebenen Governance-Instrumente, wie DSGVO, AI Act und KI Strategie, lassen sich in „Soft Law“ und „Hard Law“ Instrumente unterscheiden. Dabei steht „Hard Law“ für rechtlich bindende Normen wie Gesetze und Verordnungen (siehe Abschnitt 8), während unter „Soft Law“ rechtlich nicht bindende Leitlinien, Leitfäden, Strategien und Absichtserklärungen zu verstehen sind (siehe Abschnitt 9).

Definition: „Soft Law“ und „Hard Law“

7.1 Die Grundlage des Verwaltungshandelns

Verwaltungshandeln hat seine Basis in der Rechtsstaatlichkeit und baut immer auf einer Rechtgrundlage auf, dennoch ist nicht jede einzelne Verwaltungstätigkeit im Detail rechtlich vordefiniert. Es gibt unterschiedliche Verständnisse der Rolle der Verwaltung, entweder als ausschließlich vollziehendes oder als auch eigenmächtig tätiges Organ.

Die zentrale Bedeutung des Rechts als Handlungsgrundlage wird oft als Legalitätsprinzip bezeichnet (vgl. Art. 18 Abs. 1 B-VG). Aus diesem ergibt sich, dass Verwaltungshandeln sowohl inhaltlich als auch formell durch den Gesetzgeber determiniert sein muss.

Gleichzeitig wird der Großteil des Verwaltungshandelns nicht unmittelbar in Gesetzen beschrieben. Die Verwaltung vollzieht dann nicht unmittelbar gesetzliche Vorschriften, agiert

aber nicht im gesetzesfreien Raum, sondern hält sich an allgemeine rechtsstaatliche Grundsätze bzw. wird aufgrund allgemeiner Aufgabenbestimmungen tätig (Raschauer 2009, 217).

Hinzu kommen noch Freiräume im Sinne von Ermessensentscheidungen, nämlich dort, wo bindende Regelungen nicht vorliegen und Behörden freies Ermessen im Sinne eines Gesetzes haben (Raschauer 2009, 224).

Das Verwaltungshandeln lässt sich darüber hinaus in Hoheitsakte³ im engeren Sinn, das heißt typengebundenes Verwaltungshandeln, z.B. Bescheid, Weisung, Verordnung, sowie in schlicht-hoheitliches Verwaltungshandeln untergliedern. Mit schlicht-hoheitlichem Verwaltungshandeln sind Verwaltungsorganhandlungen gemeint, denen keine Hoheitsakte im engeren Sinn zugrunde liegen, die aber eine Verbindung zur staatlichen Hoheitsverwaltung aufweisen. Letzteres wird auch als „informelles Verwaltungshandeln“ verstanden.

Es ist umstritten, ob „schlicht-hoheitliches Verwaltungshandeln“ dem Art 18 Abs 1 B-VG unterliegt, da dies in der einschlägigen Literatur teilweise bejaht (Stöger 2014; Feik 2007; Karkulik 2014) bzw. verneint (Adamovich-Funk 1987, 308) wird. Eine Meinung (Berka 2021) dazu lautet, dass das „informelle Verwaltungshandeln“ abgeschwächten Bestimmtheitserfordernissen sowie verringerter Intensität unterliegt. Eine Ausnahme sollten allerdings Grundrechtseingriffe darstellen, die grundsätzlich einer gesonderten Untersuchung und Begründung bedürfen (Scheichenbauer und Seidl 2022). Genau an diesem Punkt setzt die Diskussion um den rechtlichen Rahmen der Digitalisierung der öffentlichen Verwaltung an: Grundfreiheiten wie das Recht auf Privatheit sind beim Einsatz neuer Technologien, insbesondere aufgrund der Möglichkeiten der automatisierten Verarbeitung von Daten, potenziell gefährdet und gerade für das Verwaltungshandeln von besonderer Wichtigkeit.

Bei einer geringen Regelungsdichte eines Bereiches, die beispielsweise durch die rasche Entwicklung von Technologien bewirkt wird, kann sich Verwaltungshandeln also auch auf „soft law“ stützen, z. B. auf Strategien und Leitfäden, wie die KI-Strategie oder auch den vorliegenden Praxisleitfaden.

Da es darüber hinaus nicht wünschenswert ist, grundsätzlich jedes Verwaltungshandeln durch Gesetze und Verordnungen zu regeln, können ergänzende Maßnahmen zu den detaillierteren und standardisierten Verwaltungsmaßnahmen durchaus auch durch „soft law“ Instrumente unterstützt werden. Entscheidend ist jedoch, dass der Kern des Verwaltungshandelns nach wie vor im Gesetz verankert ist.

Tatsächlich ist der Einsatz von KI vor dem Hintergrund spezifischer politischer Rahmenbedingungen von Anweisungen und Anforderungen an die Verwaltung zu sehen. Die öffentliche Verwaltung steht unter dem Druck in Qualität und Quantität wachsende Anforderungen mit zunehmend weniger Personal bewerkstelligen zu müssen. Gesellschaftliche Ansprüche an die öffentliche Verwaltung wandeln sich im Zuge der fortschreitenden Digitalisierung des Alltags ebenfalls. Daher setzt die Anwendung von

3 Also die praktische Anwendung der dem Staat verantworteten Entscheidungsgewalt gegenüber Bürger:innen in einem spezifischem Rechtsgegenstand.

KI dort ein, wo begrenzte Ressourcen effektiver zur Erfüllung von Verwaltungsaufgaben umgesetzt werden können.

7.2 Datenschutzgrundverordnung

Die europäische Datenschutz-Grundverordnung (DSGVO) ist seit dem 25. Mai 2018 als unmittelbar anwendbare EU-Verordnung in allen EU-Mitgliedsstaaten gültig. Diese hat jedoch sogenannte Öffnungsklauseln, die nationale Spielräume zulässt. Als Ergänzung der DSGVO wurde das österreichische Datenschutzgesetz 2000 durch das „Datenschutz-Anpassungsgesetz 2018“ und das „Datenschutz-Deregulierungsgesetz“ novelliert. Beide Novellierungen sind im nun gültigen „Datenschutzgesetz“ (DSG) enthalten.

Die DSGVO findet allgemein Anwendung, wenn personenbezogene Daten verarbeitet werden. Im Umkehrschluss heißt das, dass Daten, die an keine Person geknüpft oder einfach keinen Personenbezug haben, nicht in die DSGVO fallen.

Das Verarbeiten von Daten beinhaltet das Erheben, Erfassen, Organisieren, Ordnen, Speichern, Auslesen, Abfragen, Verwenden, Weitergeben, Verarbeiten, Bereitstellen, Abgleichen, Verknüpfen und Löschen von Daten. Als Verantwortliche werden Personen gesehen, die Daten sammeln und verarbeiten.

Hinsichtlich der Datenspeicherung sind einige zentrale Grundsätze einzuhalten:

- Der Grundsatz der Rechtmäßigkeit besagt, dass die Datenspeicherung eines gesetzlichen Auftrages oder einer vertragsrechtlichen Grundlage bedarf. Eine Alternative dazu ist die Unterzeichnung der Einverständniserklärung.
- Der Grundsatz der Datenminimierung gibt vor, dass nur jene Daten gespeichert werden dürfen, die auch hinsichtlich der Rechtsgrundlage als relevant erscheinen. Daher ist die Frage zu stellen: Welchem Zweck dient die Datenspeicherung?
- Sobald Zweck und Rechtsgrundlage der Speicherung wegfallen, sind auch die Daten zu löschen, z. B. wenn ein Vertragsverhältnis, zu dessen Zweck Daten gespeichert wurden, erlischt. Es gibt Ausnahmen, welche die Datenspeicherung auch nach Beendigung eines Vertrags über mehrere Jahre vorsehen.

Die Pflichten der Datensammler:innen sind:

- jederzeit soll eine betroffene Person, von der Daten gespeichert wurde, über eigene Daten Auskunft erhalten dürfen,
- für die Speicherung verwendete IT-Systeme sollen belastbar und vertraulich, z. B. mit Firewall, Virenschutz, Datenbackup ausgestattet sein,
- für den Datenzugriff gilt, die Mitarbeiter:innen müssen verantwortungsvoll und im Datenumgang geschult sein,
- im Falle eines Sicherheitslecks sind Betroffene und Datenschutzbehörde zu informieren und weiterer Schaden ist zu verhindern,

- eine etwaige Auslagerung der Datenverwaltung ist mittels Vertrags möglich, wobei Datensicherheit und Verschwiegenheitspflicht gewährleistet und seitens der Auftraggeber überprüft werden müssen.

In den Ministerien gibt es außerdem Datenschutzbeauftragte, die mit dem individuellen Anwendungsfall kontaktiert werden können und über konkret notwendige Maßnahmen im ministeriellen Kontext informieren können.

Eine Abteilung in einem Ministerium möchte Daten von Bürger:innen sammeln, welche die Internet-Seite einer Aktivität dieser Abteilung aufrufen, um eine Website-Aufmachung mit zielgruppenspezifischeren Inhalten zu gestalten. Um Daten über das Alter, Wohnort, Beruf und Zweck des Zugriffs auf die Website zu sammeln, erscheint beim Anklicken der Website eine Frage, am Bildschirm, bei der die Bürger:innen zustimmen können, ob sie teilnehmen oder nicht. Bei der Zustimmung muss einem Formblatt zugestimmt werden, das verspricht, die Datenschutz-Grundverordnung und das Datenschutzgesetz einzuhalten. Die Abteilung selbst hat keine Datenspeicher-Möglichkeit.

Ist das beschriebene Vorgehen DSGVO-konform?

Im Folgenden werden dazu relevante Orientierungspunkte diskutiert.

Wie soll mit personenbezogenen Daten umgegangen werden? Wenn das Ministerium personenbezogene Daten von Bürger:innen sammelt, diese aber selbst nicht speichern kann, müssen sie trotzdem sicher gespeichert werden können. Dazu kann das Ministerium die personenbezogenen Daten an ein Auftragsunternehmen zur Speicherung auslagern, wobei ein Auftragsverarbeitungsvertrag zwischen IT-Dienstleister und dem Ministerium geschlossen werden. Darüber sind betroffene Personen zu informieren und Einwilligungen einzuholen. Zu beachten sind dabei neben den Informationspflichten auch das Recht auf Widerruf der Einwilligung.

Verwaltungsbedienstete sind angehalten, besonders auf die Herstellung geeigneter technischer und organisatorischer Maßnahmen im Sinn des Artikels 24 Absatz 1 DSGVO zu achten.

Hervorstreichende Pflichten sind dabei unter anderem auch die Verpflichtung zur data breach notification (gem. Art. 33f DSGVO) sowie die Festlegung und Dokumentation sämtlicher hier genannter Maßnahmen im Sinn des Artikel 24 Absatz 1 DSGVO, damit einhergehend eine Einmeldung in das zentrale Datenverarbeitungsregister der Republik Österreich (DataReg) (Lachmayer 2018, 116).

Anwendungsfall: DSGVO

7.3 Die EU regelt KI: der AI Act

Der AI Act der EU (im deutschen Sprachraum teils auch „KI-Verordnung“) markiert einen wegweisenden Schritt zur gesetzlichen Regulierung von KI. Der Rechtsakt soll einen einheitlichen Rechtsrahmen für den Einsatz von KI in der EU schaffen und so sowohl Innovation fördern als auch Missbrauch verhindern. Mit Transparenz- und Überwachungsvorgaben, insbesondere für KI-Systeme mit hohem Risiko, soll der Markt für KI geregelt, das Vertrauen in KI gestärkt sowie der Nutzer:innenschutz gewährleistet werden. Die Auswirkungen des AI Acts werden weit über die Grenzen der EU hinaus spürbar sein, setzen neue Standards für die Regulierung von KI und erfordern eine Neubewertung der Rechtslage in zahlreichen Ländern, einschließlich Österreich.

Wissen: Warum ist der AI Act aktuell der wichtigste KI-Rechtsakt?

Der AI Act regelt die Entwicklung und Nutzung von KI-Systemen in der EU, indem er die Regeln für die Markteinführung, die Inbetriebnahme und die Nutzung von KI-Systemen harmonisiert. Der AI Act verbietet dabei bestimmte Praktiken im Bereich der KI und legt verbindliche Anforderungen für KI-Systeme mit hohem Risiko fest. Die EU folgt einem risiko-basierten Zugang: je höher das Risiko eines KI-Systems, desto strenger die Regulierung. Dies spiegelt sich in einer Risikopyramide wider, die das Risiko von KI-Systemen von niedrig bis inakzeptabel hoch einstuft (siehe Abbildung 3). Damit werden die Berichtspflichten und Folgenabschätzungen der jeweiligen KI-Systeme nach Risikolevel bestimmt.

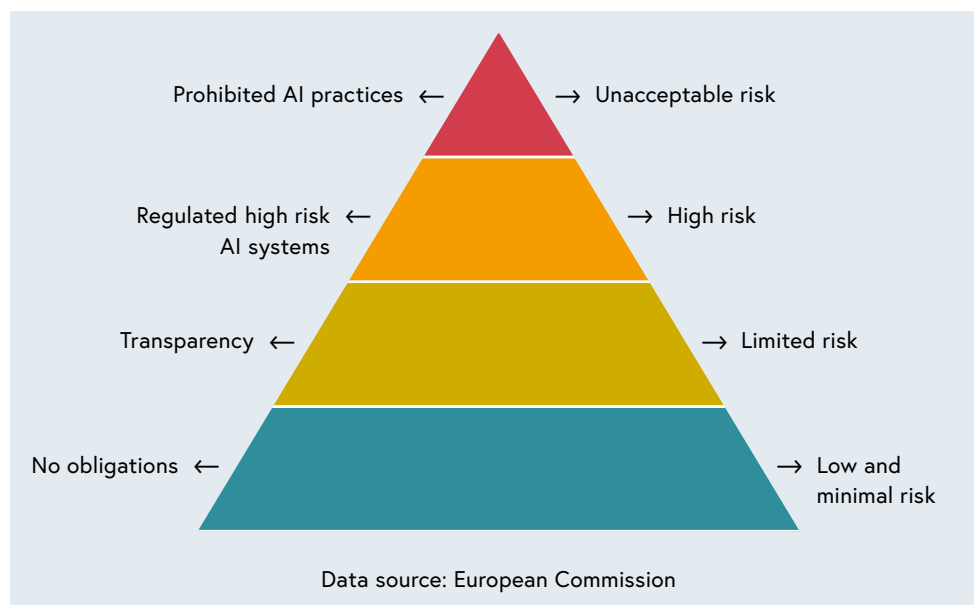


Abbildung 3: Risikopyramide AI Act – Bildquelle: Europäisches Parlament⁴

4 [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

Insbesondere Organisationen, die KI-Systeme mit hohem Risiko nutzen, werden strengen Auflagen unterworfen. Dazu gehören umfangreiche Risikobewertungen, das Führen von Aktivitätsprotokollen und die Bereitstellung von Daten zur Überprüfung durch Behörden. Sektoren wie Strafverfolgung, Migration, Infrastruktur, Produktsicherheit und Rechtspflege werden aller Voraussicht nach zu den Bereichen gehören, in denen der Einsatz von KI aufgrund ihres hohen Risikos streng reguliert wird (European Commission 2021c, o. J.).

Kritikpunkte an den vorliegenden Entwürfen zum AI Act kommen vor allem von drei Akteursgruppen (Wirtschaft, Wissenschaft, NGOs).

BusinessEurope, eine führende Vertretung von Wirtschaftsinteressen, warnte vor einer erheblichen administrativen Belastung für die Industrie, die von Investitionen in die Entwicklung von KI-Systemen abschrecken und die Wettbewerbsfähigkeit der EU langfristig beeinträchtigen könnte (BusinessEurope 2021).

Einige Wissenschaftler:innen sehen den Bereich „KI Systeme“ als zu breit definiert, befürchten Überregulierung und fordern Ausnahmen für wissenschaftliche Zwecke und Open Source Anwendungen, um als Forschende und nicht als Anbieter von KI-Systemen eingestuft zu werden. Andere fordern im Gegensatz dazu eine größere Geltungsbreite des AI Acts, der sehr wohl alle Anwendungen in Hochrisikodomänen zum Grundrechtsschutz regulieren sollte, also auch einfachere Algorithmen. Zentrale Risiken zeigen sich auch im Bereich Nachhaltigkeit und bei der auf einer Selbsteinschätzung des Anbieters basierenden Risikoklassifikation, die in den meisten Fällen keine externe Aufsicht verlangt und somit Hochrisikooanwendungen als akzeptabel eingeschätzt werden könnten (Ebers et al. 2021; Madiega 2022).

Das EDRi-Netzwerk (European Digital Rights), ein Zusammenschluss von Expert:innen und Nichtregierungsorganisationen (NGOs), fordert einen Fokus auf Grundrechte und den Schutz von Betroffenen von KI-Systemen. Access Now und andere NGOs kritisieren den risikobasierten Zugang des AI Acts und fordern eine rechtebasierte Regulierung sowie Folgenabschätzungen zu Menschenrechten, da sie argumentieren, dass KI unberechenbar sein kann und einige Systeme Menschenrechte grundsätzlich untergraben, zum Beispiel bei Emotionserkennung sowie teilweise in den Bereichen der Polizeiarbeit und Migration (Hidvegi et al. 2021; Madiega 2022; EDRi 2023).

Wissen: Kritik am AI Act

7.4 Weitere EU-Normierungen

Die Bedeutung der Festlegung von Normierungen zur Kontrolle des Digitalisierungsprozesses und der Anwendung von KI wird von der EU durch die Einführung unterschiedlicher Richtlinien, wie der EU-Produkthaftungsrichtlinie, der KI-Haftungsrichtlinie, des Daten-Governance-Gesetzes sowie anderer EU-weiter gesetzgeberischer Maßnahmen unterstrichen. Diese sind unerlässlich, um die Verantwortung, die Sicherheit und den Schutz von Menschen und Organisationen zu gewährleisten.

EU-Produkthaftungsrichtlinie und KI-Haftungsrichtlinie

Personen, die durch KI-Produkte oder -Dienstleistungen geschädigt werden, sollen in der EU den gleichen Schutz erhalten wie diejenigen, die durch andere Mittel zu Schaden kommen. Im Einklang mit den Zielen des AI Acts wurden 2022 zwei neue EU-Richtlinien festgelegt, die das gewährleisten sollen: Erstens wurde die EU-Produkthaftungsrichtlinie modernisiert, welche die verschuldensunabhängige Haftung der Hersteller für die Entschädigung von Personenschäden, Sachschäden oder Datenverlusten durch unsichere Produkte regelt.

Zweitens wurden durch die Richtlinie über KI-Haftung jene Fälle harmonisiert geregelt, die durch erstere Richtlinie nicht erfasst werden, zum Beispiel Verletzungen der Privatsphäre oder Schäden durch Sicherheitsprobleme. Es gibt dabei zwei wesentliche Elemente zur Vereinfachung rechtlicher Verfahren für Opfer von KI-Systemen. Zum einen wird die „Kausalitätsvermutung“ angeführt, bei der die Anforderungen an Betroffene, eine detaillierte Erklärung zur Schadensbegründung einzureichen, erleichtert wird. Zum anderen wird das Recht auf Zugang zu Beweismitteln (wenn Hochrisiko-KI-Systeme betroffen sind) erweitert (European Commission 2022b).

Data Governance Act

Um das Vertrauen in die gemeinsame Nutzung von Daten zu stärken, die Mechanismen für die Datenverfügbarkeit zu verbessern und technische Herausforderungen im Zusammenhang mit der Wiederverwendung von Daten zu bewältigen, hat die EU den Data Governance Act eingeführt. Der Rechtsakt soll außerdem die Einrichtung und das Wachstum gemeinsamer europäischer Datenräume in Schlüsselbereichen unter Beteiligung sowohl privater als auch öffentlicher Stellen erleichtern. Wo immer personenbezogene Daten betroffen sind, gilt nach wie vor die DSGVO (European Commission 2022a; 2023).

Weitere gesetzgeberische Aktivitäten auf EU-Ebene

Im Jahr 2022 wurden innerhalb kurzer Zeit mehrere EU-Verordnungen zur digitalen Transformation beschlossen, die teilweise direkt, teilweise indirekt auch für die öffentliche Verwaltung von Bedeutung sind, z. B. der Digital Markets Act zu fairem und marktkonformem Verhalten großer Onlineplattformen (European Commission 2022d), der Digital Services Act zu von großen Online-Plattformen und Suchmaschinen verwendeten algorithmischen Systemen (European Commission 2022e), die neue Machinery Regulation zur Sicherheit von Maschinen und Robotern (European Commission 2022c). Weitere Strategiedokumente, Gesetzesvorhaben und Gesetze zum Thema Digitalisierung auf EU-Ebene können auf der Website des Europäischen Parlaments⁵ nachgeschlagen werden.

7.5 Artificial Intelligence Mission Austria 2030

Die österreichische KI-Strategie (Artificial Intelligence Mission Austria 2030, AIM AT 2030) definiert Ziele für die Umsetzung von KI in Österreich und schlägt Maßnahmen zu deren Erreichung vor. Die AIM Austria bezieht sich dabei auch auf die öffentliche Verwaltung.

Wissen: AIM AT 2023

Österreichs nationale KI-Strategie mit dem Titel „Artificial Intelligence Mission Austria 2030“ (AIM AT 2030) wurde 2021 im Ministerrat verabschiedet (BMDW und BMK 2021). Im Dokument wurde aufgrund der dynamischen Entwicklung von KI die Strategie als „agile Strategie“ definiert (BMK und BMDW 2021: 20), was Offenheit für laufende Anpassung signalisieren soll. Der zeitliche Rahmen wurde auf 2021 bis 2030 festgelegt und sie hat drei zentrale Ziele (BMK und BMDW 2021, 9):

1. Es wird ein am Gemeinwohl orientierter, breiter Einsatz von KI angestrebt, der in verantwortungsvoller Weise auf Basis von Grund- und Menschenrechten, europäischen Grundwerten und des kommenden europäischen Rechtsrahmens erfolgt.
2. Österreich soll sich als Forschungs- und Innovationsstandort für KI in Schlüsselbereichen und Stärkefeldern positionieren und
3. mittels der Entwicklung und des Einsatzes von KI soll die Wettbewerbsfähigkeit des österreichischen Technologie- und Wirtschaftsstandorts gesichert werden.

5 <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age>

Die KI-Strategie nimmt auch auf den Einsatz von KI in der Verwaltung Bezug. Die Bundesregierung nimmt sich vor, Maßnahmen zu ergreifen, um einen sicheren Einsatz von KI in der Verwaltung zu gewährleisten. Dabei werden die gesetzlichen Grundlagen, insbesondere im Hinblick auf Datenschutz, unter Berücksichtigung von Nachvollziehbarkeit und Transparenz bei KI-basierten Entscheidungen evaluiert. Ziel ist es, Leitlinien für den Einsatz von KI in der Verwaltung zu definieren, die im Einklang mit den Grundrechten stehen. Des Weiteren strebt die Bundesregierung an, Verwaltungsprozesse im Hinblick auf ihre Eignung für KI zu evaluieren, um die Effizienz, Qualität und Treffsicherheit der Dienstleistungen für Bürger:innen zu verbessern. Ein weiterer Schwerpunkt liegt auf der Ausweitung der Bereitstellung und Nutzung von offenen und nicht personenbezogenen Verwaltungsdaten. Zudem plant die Bundesregierung die Erweiterung der Aus- und Weiterbildungsmodelle für öffentlich Bedienstete im Bereich der Digitalen Kompetenz, einschließlich spezifischer Schulungsprogramme, um den Mitarbeiter:innen die erforderlichen KI-relevanten Fähigkeiten zu vermitteln (BMDW und BMK 2021, 56–59).

Wesentlich ist im Zusammenhang mit der KI-Strategie auch das AI Policy Forum. Diese interministerielle Arbeitsgruppe wurde im November 2021 unter dem gemeinsamen Vorsitz von BMK und BMF eingerichtet und trifft sich in regelmäßigen Abständen, um die ressort-übergreifende Umsetzung der KI-Strategie AIM AT 2030 zu begleiten und diese auch weiterzuentwickeln. Als thematisches Forum der Bundesverwaltung soll es außerdem den Austausch über Erfahrungen und Herangehensweise zum Einsatz von KI in den Bundesministerien fördern und aktuelle Fragen zur Künstlichen Intelligenz diskutieren. Ein wesentliches Element des AI Policy Forums ist die Einrichtung von Ad-Hoc Arbeitsgruppen für einzelne Maßnahmen unter Einbindung einschlägiger Expert:innen aus Forschung, Wissenschaft, Wirtschaft, Sozialpartnern, NGOs und der Zivilgesellschaft.

8 Ethische KI: Prinzipien und Leitlinien

Abschnitt 9 befasst sich mit den ethischen Überlegungen im Zusammenhang mit KI und unterstreicht die Notwendigkeit, ethische Prinzipien und Leitlinien für die Implementierung von KI-Anwendungen aufzustellen. Der Abschnitt gibt einen Überblick über verschiedene Ansätze für ethische Leitlinien für KI und betont die Bedeutung von Kriterien wie Rechtmäßigkeit, Transparenz, Fairness, Effizienz, Sicherheit, Zugänglichkeit, Rechenschaftspflicht und digitale Souveränität. Diese Kriterien sind speziell auf die besonderen Anforderungen der öffentlichen Verwaltung zugeschnitten und dienen als Orientierungshilfe für Verwaltungsbedienstete bei der Durchführung ethischer Bewertungen von KI-Systemen. Ziel ist es, einen verantwortungsvollen und ethisch vertretbaren Einsatz von KI-Technologien in der öffentlichen Verwaltung zu gewährleisten.

8.1 Ethische Leitlinien: Governance durch „Soft Law“

Wie bereits in Abschnitt 8 „Rechtlicher Rahmen“ erwähnt, sind ethische Aspekte von KI bereits teilweise rechtsverbindlich geregelt (siehe DSGVO) und werden mit dem AI-Act weiter reguliert. Diese rechtlichen Rahmenbedingungen sind jedoch nur ein Teil der Lösung, um den ethischen Einsatz von KI zu gewährleisten. „Soft Law“ in Form von KI-Ethikrichtlinien kann eine wichtige Rolle bei der Ergänzung gesetzlicher Regelungen spielen. Derartige Richtlinien sind nicht rechtsverbindlich, bieten aber eine Reihe von Grundsätzen und Empfehlungen, welche die ethische Entscheidungsfindung im Zusammenhang mit der Entwicklung und Nutzung von KI beeinflussen können (vgl. Sigfrids et al. 2022).

Weltweit hat es bereits eine Reihe von Bestrebungen gegeben, ethische Leitlinien für die Entwicklung von KI zu definieren. Allein im AI Ethics Guidelines Global Inventory von AlgorithmWatch⁶ sind mehr als 100 derartiger Leitlinien von Unternehmen, NGOs, Regierungen, unterschiedlichster Organisationen und supranationalen Institutionen gesammelt.

Auf europäischer Ebene hat die EU ethische Leitlinien für die Entwicklung von KI herausgegeben, die sich vor allem darauf beziehen, Vertrauen in die Entwicklung, den Einsatz und die Nutzung von KI-Systemen zu schaffen. Die einflussreichen Ethik-Leitlinien für vertrauenswürdige KI wurden auch in den rechtlichen Anforderungen für risikoreiche KI-Systeme im AI-Act der EU herangezogen.

⁶ <https://algorithmwatch.org/de/ai-ethics-guidelines-global-inventory/>

Die EU High-Level Expert Group on Artificial Intelligence (AI-HLEG) definiert vertrauenswürdige KI anhand von sieben Prinzipien: (1) menschliches Handeln und Aufsicht, (2) technische Robustheit und Sicherheit, (3) Datenschutz und Data Governance, (4) Transparenz, (5) Vielfalt, Nichtdiskriminierung und Fairness, (6) soziales und ökologisches Wohlergehen und (7) Rechenschaftspflicht (AI HLEG 2019). Die Implementierung dieser Grundsätze soll erfolgen durch „Mechanismen zur Überwachung der von KI getroffenen Entscheidungen, damit sie vertrauenswürdige sind und den ethischen Richtlinien entsprechen“ (Kaur et al. 2021). Im Jahr 2020 erstellte die AI-HLEG dann darauf aufbauend eine Bewertungsliste für vertrauenswürdige KI (ALTAI, siehe Abschnitt 10.2).

Darüber hinaus haben die 193 UNESCO-Mitgliedsstaaten im November 2021 erstmals ein globales Abkommen zur KI-Ethik und ein internationales Standardinstrument verabschiedet, die „Recommendation on the Ethics of Artificial Intelligence“ (UNESCO 2022). Diese Empfehlungen bieten einen Rahmen, um sicherzustellen, dass die Entwicklung und Nutzung von KI im Einklang mit den Menschenrechten und der Menschenwürde sowie der Rechtsstaatlichkeit steht. Weitere supranationale ethische Leitlinien stammen von der OECD, die fünf komplementäre Grundsätze und Instrumente für ethische KI festlegt, nämlich zu Nachhaltigkeit, Fairness, Transparenz, Sicherheit und Verantwortlichkeit (OECD 2019).

Obwohl auf nationaler Ebene nur wenige Länder wie Australien und China (Beijing Academy of Artificial Intelligence, 2019, Australian Government Dept. Industry, Science and Resources, 2019) explizite ethische Leitlinien für KI haben, enthalten die meisten nationalen KI-Strategien ethische Grundsätze als normativen Rahmen, um die verantwortungsvolle Entwicklung von KI zu leiten. Vor allem die europäischen Länder haben in ihren KI-Strategien ethischen Überlegungen einen hohen Stellenwert eingeräumt. Die französische KI-Strategie „AI for Humanity“ ist ein Beispiel dafür. Die Strategie unterstreicht die Bedeutung einer ethischen und verantwortungsvollen KI-Entwicklung und zielt darauf ab, Frankreich als Vorreiter in diesem Bereich zu positionieren. Die Strategie betont, dass KI das menschliche Wohlergehen fördern sollte, und erkennt an, dass ethische und verantwortungsvolle KI der Schlüssel dazu ist, dass KI der Gesellschaft als Ganzes zugutekommt (General Secretary of the French Digital Council 2018). Ein Beispiel für einen Leitfaden, der sich auf die Verwaltung und auf ein bestimmtes Politikfeld konzentriert, kommt aus Deutschland. Die „Selbstverpflichtenden Leitlinien für den KI-Einsatz in der behördlichen Praxis der Arbeit und Sozialverwaltung“ des deutschen Bundesministeriums für Arbeit und Soziales (BMAS 2022) bieten hier Anhaltspunkte. Auch von anderen deutschen Institutionen gibt es relevante Dokumente, die sich allerdings nicht explizit auf die Verwaltung beziehen, wie die umfassende Stellungnahme zum Thema „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“

vom Deutschen Ethikrat veröffentlicht (Deutscher Ethikrat 2023) oder der „Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz“ des deutschen Fraunhofer IAIS (Fraunhofer IAIS 2021).

Wissen: Nationale Leitlinien und KI-Strategien

8.2 Ethische Leitlinien für die österreichische Verwaltung

Obwohl ethische Leitlinien wie jene der HLEG wertvolle Orientierungshilfen für einen ethisch vertretbaren Einsatz von KI bieten, sind diese nicht so einfach in der Praxis umsetzbar und außerdem nicht auf spezifische Bedürfnisse der gegebenen Kontexte der öffentlichen Verwaltung zugeschnitten. Da es außerdem noch keine ethischen Handlungsanleitungen für KI für die österreichische öffentliche Verwaltung gibt, wurden im Zuge des Projekts „Digitale Verwaltung und Ethik“ Kriterien und Maßnahmen erarbeitet, die sich speziell an Verwaltungsangehörige richten.

Als Grundlage für die vorgeschlagenen Kriterien wurden die für Europa zentralen HLEG-Ethikrichtlinien herangezogen und um weitere, für den öffentlichen Dienst besonders relevante Kriterien ergänzt.

Ziel der EU HLEG-Ethikrichtlinien ist es, die Einhaltung geltender Gesetze, Transparenz, Unparteilichkeit, Fairness, Effektivität, Effizienz, Sicherheit, Barrierefreiheit und Inklusion, Rechenschaftspflicht und digitale Souveränität bei dem Einsatz und der Nutzung von KI-Technologien zu gewährleisten.

Wissen: Ziele der EU HLEG-Ethikrichtlinien

Zusätzlich zu den Kriterien wurden Maßnahmenvorschläge entwickelt, um den ethischen und rechtmäßigen Einsatz und die Nutzung von KI-Technologien in der öffentlichen Verwaltung sicherzustellen. Zu diesen Maßnahmen gehören die Förderung der KI-Literacy von Verwaltungsbediensteten und der Öffentlichkeit, die Durchführung von Folgenabschätzungen, die Zertifizierung von KI-Anwendungen, die Einrichtung unabhängiger Aufsichtsgremien und die kontinuierliche Aktualisierung von KI-Kenntnissen. Diese Maßnahmen sollen dazu beitragen sicherzustellen, dass der Einsatz und die Nutzung von KI-Technologien zum Gemeinwohl einen Beitrag leisten oder wenigstens mit demselben vereinbar sind (siehe Abschnitt 11).

Im Folgenden werden die angeführten Kriterien kurz beschrieben.

Kriterien

- **Recht:** Eines der wichtigsten Kriterien ist die Einhaltung des geltenden Rechts. Das bedeutet, dass die KI-Anwendung die einschlägigen Gesetze und Vorschriften einhalten muss, einschließlich der Grundrechte wie Gleichheit und Meinungsfreiheit sowie Schutz der Privatsphäre und Datenschutz.
- **Transparenz:** Transparenz bedeutet, Informationen über die KI-Anwendung verfügbar und zugänglich zu machen und zu erläutern, wie die KI-Anwendung in der öffentlichen Verwaltung entwickelt, trainiert und eingesetzt wird. Die öffentliche Verwaltung sollte außerdem technische und nicht-technische Mittel einsetzen, um Transparenz in KI-Entscheidungsprozessen zu fördern. Dazu gehört, dass die Öffentlichkeit und die Verwaltungsbediensteten über die Ziele der KI-Anwendung informiert werden und dass die Faktoren und Überlegungen, die die Entscheidungsergebnisse beeinflussen, offengelegt werden.
- **Unparteilichkeit und Fairness:** Die KI-Anwendung muss unvoreingenommene und vielfältige Daten und Modelle verwenden, um zu vermeiden, dass bestehende Vorurteile aufrechterhalten werden. Die Fairness der KI-Anwendung muss darüber hinaus im Kontext der öffentlichen Verwaltung bewertet werden.
- **Effektivität und Effizienz:** Der Einsatz von KI-Anwendungen in der Verwaltung muss deren Effektivität und Effizienz nachhaltig verbessern, ohne die Arbeitssituation der im öffentlichen Dienst tätigen Menschen zu verschlechtern. Es sollten Kriterien festgelegt werden, anhand derer festgestellt werden kann, wann der Einsatz von KI die Effektivität und Effizienz der Verwaltung und ihrer Dienstleistungen verbessert und wann er die Arbeitssituation der Beschäftigten im öffentlichen Dienst verschlechtert.
- **Sicherheit:** Die KI-Anwendung muss sicher eingesetzt werden, um sensible Informationen zu schützen und unbefugten Zugriff zu verhindern. Unabhängige Aufsichtsbehörden sollten die Sicherheit der KI-Anwendung überwachen und gewährleisten und die Bedenken der Bürger:innen hinsichtlich des Missbrauchs von KI-Technologien müssen berücksichtigt werden.
- **Barrierefreiheit und Inklusion:** Die KI-Anwendung muss für Menschen mit unterschiedlichen Fähigkeiten, Hintergründen und Kulturen zugänglich und integrativ sein. Die Bevölkerung muss in die Lage versetzt werden, die KI-Anwendung zu nutzen und davon zu profitieren. Außerdem sollten Alternativen zur KI-Technologie angeboten werden, um einen gleichberechtigten Zugang zu öffentlichen Dienstleistungen zu gewährleisten.
- **Rechenschaftspflicht:** Die Verantwortlichen für den Einsatz und die Nutzung von KI müssen klare Zuständigkeiten und Verantwortlichkeiten haben. Es sollten Mechanismen vorhanden sein, die sicherstellen, dass die Verantwortlichen sich ihrer Verantwortung auch bewusst sind.
- **Digitale Souveränität:** Die Verwaltung muss in der Lage sein, die Entwicklung von KI-Lösungen zu beeinflussen, unabhängig anzuwenden und vertrauliche Daten in ihrem eigenen Einflussbereich zu halten. Es müssen Maßnahmen ergriffen werden, um sensible Daten zu schützen und den Zugriff Dritter zu verhindern, wenn die Entwicklung oder der Betrieb von KI ausgelagert wird.

9 KI-Folgenabschätzung

In diesem Abschnitt werden verschiedene Instrumente vorgestellt, mit denen die öffentliche Verwaltung KI Ethik Prinzipien und Leitlinien in der Praxis umsetzen kann. Sie dienen der Folgenabschätzung bei der Entwicklung und dem Einsatz von KI-Technologien. Während einige auf die öffentliche Verwaltung zugeschnitten sind, sind andere in der internationalen Diskussion zentral.

Zur Übersicht sind die im Folgenden genauer vorgestellten Instrumente hier aufgezählt:

- der DVuE „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ wurde speziell auf die österreichische Verwaltung zugeschnitten,
- die Bewertungsliste ALTAI ist eine zentrale Grundlage der Diskussion auf EU-Ebene,
- der KI-Forschungsleitfaden der EU ist die Basis für EU-Projekte, die an oder mit KI arbeiten,
- das VCIO-Modell ist ein besonders komplettes Beispiel einer KI-Folgenabschätzung,
- die Instrumente FRAIA und DEDA werden in den Niederlanden aktuell flächendeckend in der Verwaltung eingeführt.

Zuletzt werden einige weitere Beispiele der zahlreichen Tools mit verschiedenen Schwerpunkten angeführt.

9.1 Kriterien- und Maßnahmenkatalog für ethische KI in der Verwaltung EKIV

Die Einführung und Nutzung von KI-Technologien in der öffentlichen Verwaltung erfordern einen Rahmen, der Orientierungs- und Anhaltspunkte bietet, um zentrale ethische Fragestellungen zu behandeln und geeignete Maßnahmen festzulegen. Im vom BMKÖS beauftragten Projekt „Digitale Verwaltung und Ethik“ wurde daher ein „Kriterien- und Maßnahmenkatalog für ethische KI in der Verwaltung (EKIV)“ erstellt, der speziell auf die Bedürfnisse der öffentlichen Verwaltung zugeschnitten ist. Der Inhalt dieses Katalogs ist im Anhang ausgeführt. Dort findet sich auch eine Checkliste für KI-Projekte auf Basis des EKIV.

Der EKIV dient als dialogisches Instrument zur Folgenabschätzung und erleichtert die frühzeitige Erkundung und ethische Bewertung einer KI-Anwendung. Er eignet sich dabei besonders für ein Überdenken einzelner Bereiche, alleine oder im Rahmen von Workshops.

Darüber hinaus wurde ebenso eine Checkliste für KI-Projekte entwickelt, die auf dem EKIV beruht. Sie ist als Instrument gedacht, um einerseits am Anfang eines Planungs-, Beschaffungs- oder Evaluationsprozesses zu prüfen, ob ein spezifischer Bereich eines genaueren Augenmerks bedarf. Andererseits kann die Checkliste auch am Ende eines derartigen Prozesses stehen, um nochmals zu prüfen, ob alle Kriterien bedacht wurden.

Im Vergleich dazu bietet der EKIV die Möglichkeit für eine vertiefende Betrachtung eines defizitären Bereiches auf der Basis von detaillierten Fragestellungen, welche die Formulierung wirksamer Maßnahmen zur Bewältigung ethischer Herausforderungen im Zusammenhang mit der KI-Anwendung ermöglichen. Beide, EKIV und Checkliste, sollen als unterstützende Instrumente verstanden werden, um ethische KI in der Verwaltung zur Anwendung zu bringen. Die Beantwortung einzelner Fragen mit „Ja“ oder „Nein“ bedeutet somit nicht notwendigerweise das Ende eines KI-Projektes, kann aber Änderungsbedarf im Hinblick auf eine KI-Anwendung indizieren.

Da es sich beim gesamten Leitfaden um ein „living document“ handelt, soll auch der EKIV kontinuierlich verfeinert und erweitert bzw. angepasst werden, um neuen Überlegungen und aufkommenden ethischen Herausforderungen Rechnung zu tragen. Regelmäßige Aktualisierungen und ergänzende Beiträge von Expert:innen aus Wissenschaft und Verwaltung sollen sicherstellen, dass der Katalog relevant und aktuell bleibt und die neuesten Erkenntnisse und Einsichten einbezieht.

Ziel des EKIV ist somit die Unterstützung von Verwaltungsangehörigen bei der ethischen Bewertung einer spezifischen KI-Anwendung. Außerdem soll damit ein generelles Verständnis der potenziellen Auswirkungen von KI-Anwendungen auf verschiedene Aspekte der Gesellschaft gefördert werden. Ziel ist es auch, Transparenz, Verantwortlichkeit und ethische Entscheidungsfindung während des gesamten KI Planungs-, Design und Implementierungsprozesses zu fördern.

Kriterien- und Maßnahmenkatalog für ethische KI in der Verwaltung (EKIV)

Kriterien

Recht

- Welche rechtlichen Rahmenbedingungen werden zu Beginn des KI-Projekts herangezogen?
- Ist beim Einsatz der KI-Anwendung die Einhaltung geltenden Rechts gewährleistet?
- Wird beim Einsatz der KI-Anwendung der Schutz von Grundrechten wie der Privatsphäre und der Menschenwürde gewährleistet?
- Wie wird gewährleistet, dass die KI-Anwendung an potenzielle Veränderungen geltenden Rechts angepasst wird?

Transparenz

- Wie werden die Öffentlichkeit und die Verwaltungsbediensteten über die spezifischen Ziele und Zwecke der KI-Anwendung in der öffentlichen Verwaltung informiert?
- Wie werden die Auswirkungen des Einsatzes der KI-Anwendung der Öffentlichkeit und den Verwaltungsbediensteten vermittelt?

Unvoreingenommenheit und Fairness

- Wie wird die Fairness der KI-Anwendung im Kontext der öffentlichen Verwaltung bewertet?
- Welche Kriterien werden verwendet, um festzustellen, ob die Anwendung fair ist?
- Wie wird sichergestellt, dass die zum Training der KI-Anwendung verwendeten Daten vielfältig und repräsentativ (z. B. für Mitarbeiter:innen oder die Bevölkerung) sind?
- Wie werden die potenziellen Verzerrungen der KI-Anwendung denjenigen vermittelt, die von ihrem Einsatz betroffen sind, z. B. den Bürger:innen oder den Bediensteten der öffentlichen Verwaltung?
- Ist es notwendig oder sinnvoll, Bürger:innen und /oder Organisationen der Zivilgesellschaft einzubeziehen, um Vorurteile bei der Anwendung der KI-Technologie zu erkennen und zu beseitigen?

Effektivität und Effizienz

- Welche Kriterien werden verwendet, um zu bestimmen, wann der Einsatz der KI-Anwendung die Effektivität und Effizienz öffentlicher Dienstleistungen verbessert?
- Welche Kriterien werden verwendet, um zu bestimmen, ob sich die Arbeitssituation der im öffentlichen Dienst tätigen Personen durch den Einsatz einer KI-Anwendung verschlechtert?

Sicherheit

- Welches sind die größten Sicherheitsrisiken, die mit dem Einsatz der KI-Anwendung verbunden sind, und wie können diese Risiken wirksam gemindert werden?
- Welche Sicherheitsvorkehrungen gibt es zum Schutz vor dem Missbrauch oder der böswilligen Nutzung der KI-Anwendung?

Barrierefreiheit und Inklusivität

- Wie wird die KI-Anwendung für Menschen mit unterschiedlichen Fähigkeiten, Hintergründen und Kulturen zugänglich und integrativ gestaltet?
- Wie wird die Bevölkerung dazu befähigt, die KI-Anwendung zu nutzen und davon zu profitieren?
- Welche Alternativen zur KI-Technologie werden angeboten, um einen gleichberechtigten Zugang zu öffentlichen Dienstleistungen zu gewährleisten (insbesondere persönliche Ansprechpartner)?

Rechenschaftspflicht

- Welche klaren Verantwortlichkeiten und Rechenschaftspflichten werden für Entwickler:innen, Manager:innen und Nutzer:innen der KI-Technologie eingeführt?
- Welche Mechanismen gibt es, um sicherzustellen, dass die für den Einsatz und die Nutzung von KI Verantwortlichen sich ihrer Verantwortung bewusst sind?
- Welche Mechanismen oder Strategien gibt es, um sicherzustellen, dass Schäden, die durch die KI-Anwendung entstehen, angemessen entschädigt werden?

Digitale Souveränität

- Wie wird sichergestellt, dass ausreichend Wissen um KI-Technologien in der Verwaltung vorhanden ist, um bei Beschaffungsvorgängen mit Dienstleistern und bei Kontrollen im Rahmen der Regulation von KI mit Herstellern auf Augenhöhe umgehen zu können?
- Welche (ethischen) Aspekte werden bei der Entscheidung über die Auslagerung der KI-Entwicklung und -Implementierung an externe Dienstleister berücksichtigt?
- Wenn Entwicklung oder Betrieb von KI-Anwendungen ausgelagert wird: Welche Maßnahmen gibt es zum Schutz sensibler Daten und zur Verhinderung des Zugriffs durch dritte Organisationen?

Maßnahmen

KI Literacy

- Durch welche Maßnahmen wird die KI-Kompetenz der Verwaltungsbediensteten gefördert, um ein Grundverständnis für KI-Technologien, ihre Voraussetzungen und Anwendungen sowie ihre Auswirkungen auf Verwaltung, Staat und Gesellschaft zu gewährleisten?
- Wie wird die KI-Kompetenz der breiten Öffentlichkeit, insbesondere im Hinblick auf die Notwendigkeit eines grundsätzlichen Verständnisses für eine qualifizierte Diskussion um den Einsatz von KI-Anwendungen durch die Verwaltung, gefördert?
- Gibt es spezielle Initiativen, um die KI-Kompetenz von unterrepräsentierten Bevölkerungsgruppen zu erhöhen?
- Wie wird sichergestellt, dass die KI-Kenntnisse jeweils aktualisiert werden, wenn sich die Technologie weiterentwickelt?

Kontinuierliche Bewertung / Folgenabschätzung und Verbesserung

- Welche Maßnahmen werden ergriffen, um die Auswirkungen von KI-Technologien kontinuierlich, also im Planungsstadium, bei der Entwicklung und beim Einsatz, zu bewerten und zu beurteilen und ihre Leistung und Wirkung zu verbessern?
- Welche Kriterien werden verwendet, um zu bestimmen, welche KI-Anwendungen als besonders risikoreich gelten und eine kontinuierliche Bewertung und Beurteilung erfordern?
- Wer ist für die Durchführung der Folgenabschätzung verantwortlich?
- Welche Ressourcen und Fachkenntnisse sind für die Durchführung einer gründlichen Folgenabschätzung erforderlich, und wie können sie beschafft werden?
- Wie werden die Ergebnisse der Folgenabschätzung der Öffentlichkeit und den Entscheidungsträger:innen wirksam mitgeteilt?

Zertifizierung

- Von wem werden KI-Modelle und -Daten zertifiziert, um sicherzustellen, dass sie ethisch und rechtlich einwandfrei sind, insbesondere bei riskanten Anwendungen?
- Wie sieht das Verfahren für die Zertifizierung eines in der öffentlichen Verwaltung eingesetzten KI-Modells aus?
- Wer ist für die Überwachung des Zertifizierungsprozesses für KI-Modelle, die in der öffentlichen Verwaltung eingesetzt werden, zuständig?
- Wie oft werden zertifizierte KI-Modelle neu bewertet, um sicherzustellen, dass sie weiterhin mit ethischen und rechtlichen Standards übereinstimmen?

Aufsicht

- Welche Rechte und Pflichten haben die unabhängigen Aufsichtsgremien?
- Welche Monitoringmechanismen werden eingerichtet, um den ethischen und rechtmäßigen Einsatz und die Nutzung von KI-Technologien zu überwachen und sicherzustellen?
- Über welche Qualifikationen und Fachkenntnisse verfügen die Mitglieder der Aufsichtsgremien?

9.2 EU-Bewertungsliste für vertrauenswürdige künstliche Intelligenz (ALTAI)

Die EU-Bewertungsliste für vertrauenswürdige künstliche Intelligenz (Assessment List for Trustworthy Artificial Intelligence, ALTAI) und das zugehörige webbasierte Tool⁷ sind eine wichtige Grundlage der internationalen Diskussion zum ethischen Einsatz von KI, jedoch fokussieren sie nicht auf den Kontext der Verwaltung. Sie sollen Unternehmen und Organisationen praxisnah dabei helfen, die Vertrauenswürdigkeit ihrer in der Entwicklung befindlichen KI-Systeme selbst einzuschätzen. Vertrauenswürdige KI basiert auf sieben Schlüsselanforderungen, die von der AI High-Level Expert Group on Artificial Intelligence (HLEG) in den Ethikrichtlinien für eine vertrauenswürdige KI eingeführt wurden.

⁷ <https://altai.insight-centre.org/>

9.3 Ethics Self-Assessment für EU-Forschungsförderung und EU „Ethics by Design“ KI-Forschungsleitfaden

Ein wichtiges und instruktives Instrument, jedoch ohne Fokus auf die Verwaltung, ist das für die Vergabe von Forschungsförderungen verpflichtende Ethics Self-Assessment der EU, das Teil des Förderungsvertrages wird und geprüft werden kann. Neben anderen Bereichen wird auch KI im Self-Assessment abgedeckt und es wird empfohlen, dabei die „Ethics by Design“ Methodologie einzusetzen. Der Teil zu KI verweist auf die Bewertungsliste ALTAI der EU HLEG (siehe Abschnitt 10.2) und den KI-Forschungsleitfaden der EU für mehr Details (European Commission 2021b).

Bei der „Ethics by Design“ Methodologie geht es darum, mögliche ethische Bedenken von Anfang an und in allen Stadien von KI miteinzubeziehen. Damit soll verhindert werden, dass, z. B. Entwickler:innen, Ethiker:innen, Sozialwissenschaftler:innen und Auftraggeber:innen bzw. Anwender:innen isoliert voneinander arbeiten und nur am Ende einer KI-Systementwicklung eine Ethik-Checkliste durchgegangen wird. Ethik wird damit ein zentraler Bestandteil der Planung, Entwicklung und des Einsatzes von KI. Mehr Details zu dieser Methodologie können im EU KI-Forschungsleitfaden „Ethics By Design and Ethics of Use Approaches for Artificial Intelligence“ nachgelesen werden (European Commission 2021a).

Wissen: „Ethics by Design“

Der EU KI-Forschungsleitfaden „Ethics By Design and Ethics of Use Approaches for Artificial Intelligence“ bezieht sich speziell auf Forschungstätigkeiten zur Entwicklung und/oder unter Einsatz von KI-basierten Systemen oder Techniken, folgt dem oben erwähnten „Ethics by Design“ Konzept und beschreibt, wie dieses im KI-Entwicklungsprozess berücksichtigt werden kann. Die daraus entstandene Checkliste („Checklist: Specification of Objectives against Ethical Requirements“) dient als Hilfsmittel für die Implementierung. Um einen ethischen Ansatz in der KI(-basierten) Forschung sicherzustellen, sollen KI-Systeme sechs ethischen Prinzipien entsprechen und Merkmale, wie technische Robustheit, transparenten Umgang mit Ungenauigkeiten und Fehlern, Berücksichtigung des sozialen Kontextes, Zuverlässigkeit, Verhinderung von Schäden, und Erklärungsmöglichkeit von wichtigen Entscheidungsprozessen, aufweisen, um die Prinzipien zu bewahren und zu fördern. Außerdem weist der Leitfaden darauf hin, wie in den Bereichen Projektmanagement, Beschaffung, Implementierung und Monitoring auf einen ethischen Einsatz und eine ethische Nutzung von KI geachtet werden kann (European Commission 2021a).

9.4 VCIO-Modell

Das VCIO-Modell (Values-Criteria-Indicators-Observables) ist ebenfalls relevant in der internationalen Diskussion zu KI-Ethik und dient unter anderem der Risikoklassifikation von KI-Systemen. Das Modell ähnelt der Risikopyramide des AI Act der EU (siehe Abschnitt 8.3) und wurde von der AI Ethics Impact Group und der Bertelsmann-Stiftung entwickelt. Um die Umsetzung allgemeiner Werte messbar und bewertbar zu machen, schlüsselt es sie in Kriterien, Indikatoren und beobachtbare Faktoren auf. Dies wird mit der Verwendung einer Risikomatrix zur Klassifizierung verschiedener Anwendungsfälle von KI verbunden.

Dadurch werden die Fälle in fünf Risikoklassen eingeteilt, wobei Klasse 0 für KI-Systeme steht, die keiner Regulierung bedürfen, und Klasse 4 für Situationen, in denen KI-Systeme aufgrund des hohen Risikos überhaupt nicht angewendet werden sollten. Darüber hinaus wird ein Ethics Label analog zu den Energieeffizienzklassen von Haushaltsgeräten vorgeschlagen, das eine rasche Einschätzung einer KI-Anwendung im Hinblick auf ethische Verwendbarkeit ermöglicht (AI Ethics Impact Group et al. 2020).

9.5 Folgenabschätzung für Grundrechte und Algorithmen (FRAIA)

Die Folgenabschätzung für Grundrechte und Algorithmen (Fundamental Rights and Algorithm Impact Assessment, FRAIA) der Utrecht Data School ist ein Diskussions- und Entscheidungsfindungsinstrument für Regierungsorganisationen, das dazu dient, die potenziellen Risiken für die Menschenrechte im Zusammenhang mit dem Einsatz von Algorithmen zu ermitteln und zu mindern. FRAIA schafft eine Plattform für einen interdisziplinären Dialog zwischen Entwickler:innen und denjenigen, die KI-Systeme einsetzen.

Durch den Einsatz von FRAIA kann die Verwaltung alle relevanten Aspekte des Einsatzes von Algorithmen rechtzeitig und strukturiert bearbeiten. Es umfasst eine Vielzahl von Fragen zu den Themen, die erörtert werden sollten, wenn eine Regierungsorganisation die Entwicklung, die Beauftragung mit der Entwicklung, den Kauf, die Anpassung und / oder die Verwendung eines KI-Systems in Betracht zieht. Das Instrument trägt dazu bei, Risiken wie Nachlässigkeit, Ineffizienz oder Verletzungen der Rechte der Bürger:innen zu verringern. In naher Zukunft soll diese Art der Folgenabschätzung in den Niederlanden verpflichtend eingeführt werden (Gerards et al. 2022; Utrecht University 2022; Clausen und Schäfer 2023).

9.6 Data Ethics Decision Aid (DEDA)

Das Toolkit Data Ethics Decision Aid (DEDA) der Utrecht Data School bietet einen dialogischen Rahmen in einem partizipativen Modell mit Workshops für die ethische Untersuchung bzw. Unterstützung der Durchführung von KI- und Datenprojekten. Es ermöglicht der Verwaltung, eine vorausschauende Haltung einzunehmen und ihre Rechenschaftspflicht wahrzunehmen, indem untersucht wird, inwiefern die vorliegenden Daten Risiken bergen und welche Auswirkungen ihr Einsatz haben kann. Das Tool ist einfach zu verstehen und wird für Brainstorming-Sitzungen, für die Dokumentation des Entscheidungsprozesses und für die Erfüllung der Rechenschaftspflicht eingesetzt (Franzke et al. 2021; Utrecht Data School o. J.).

DEDA wurde in einem partizipativen Prozess mit niederländischen Verwaltungsbediensteten entwickelt und im Anschluss seit 2017 in mehreren Gemeinden, dem Verband der niederländischen Gemeinden und im Ministerium für allgemeine Angelegenheiten (NL) eingesetzt (Franzke et al. 2021). Eine praktische Auswirkung von DEDA war der Beschluss einer Regionalregierung in den Niederlanden, keine Daten durch WiFi-Tracking zu erfassen, das während der Pandemie die Besucherzahlen in Freizeitbereichen überwachte. Um die Coronarichtlinien einzuhalten, wurde beschlossen, die Besucherzahlen nur durch die Anzahl der Autos und Fahrräder zu erfassen (Schäfer und Clausen 2021).

Anwendungsfall: Einsatz von DEDA in den Niederlanden

9.7 Weitere KI-Folgenabschätzungsinstrumente

Während hier einige Instrumente beschrieben wurden, gibt es noch zahlreiche weitere Tools, die interessant sind, auch wenn sie sich meist nicht an die öffentliche Verwaltung richten. Einige Beispiele sind hier angeführt.

- **AlgorithmWatch Checklisten:** Die Triage Checkliste prüft, welche ethischen Transparenz-Themen es wert sind, während der Projektdurchführung detailliert dokumentiert zu werden und ob es notwendig ist, einen Transparenzbericht zu schreiben. Die Checkliste für den Transparenzbericht ist eine detaillierte Anleitung zum Schreiben eines solchen Transparenzberichtes (Loi et al. 2021).
- **VERA (Verantwortung und Algorithmen):** Das interaktive Werkzeug der Arbeiterkammer prüft die Verantwortlichkeiten bei der Einführung von Algorithmen und zeigt Kompetenzkonflikte und Verantwortungslücken auf. Es stellt eine Ergänzung zum Leitfaden „Algorithmen in der Entscheidungsfindung“ dar, worin detailliertere

Fragen und mehr Themen zu finden sind (Adensamer et al. 2021; Adensamer und Klausner 2021; Bundesarbeiterkammer 2021).

- Examining the Black Box: Der Bericht klärt die Unterschiede zwischen verschiedenen Arten von Instrumenten für die Bewertung von algorithmischen Systemen und hilft dadurch bei der Auswahl einer passenden Evaluierungsform (Ada Lovelace Institute und DataKind UK 2020).
- National Institute of Standards and Technology AI Risk Management Framework (NIST AI RMF): Das Instrument der US-Behörde dient dazu mit den KI-bezogenen Risiken für Einzelpersonen, Organisationen und die Gesellschaft besser umzugehen. Es wird durch verschiedene Hilfsmittel, z. B. ein Erklärvideo und eine Roadmap, ergänzt, ist für die freiwillige Nutzung gedacht und soll Überlegungen zur Vertrauenswürdigkeit in Entwurf, Entwicklung, Nutzung und Bewertung von KI einbeziehen (NIST 2023).
- Normenserie 7000 des Institute of Electrical and Electronics Engineers: Diese Norm „IEEE 7000-2021: IEEE Standard Model Process for Addressing Ethical Concerns during System Design“ legt eine Reihe von Prozessen fest, mit denen Organisationen ethische Werte in allen Phasen der Softwareentwicklung einbeziehen können (IEEE SA 2021).
- ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML): Die ISO Norm schafft einen Rahmen für die Beschreibung eines generischen KI-Systems, das Maschinelles Lernen einsetzt (ISO 2022).
- „Audit Framework for Algorithms“ des niederländischen Rechnungshofs: Das Instrument dient der Bewertung der Qualität und des verantwortungsvollen Einsatzes von Algorithmen in der Praxis und soll Schwachstellen der Algorithmen aufdecken (Netherlands Court of Audit 2021).

10 Empfehlungen für mögliche weitere Schritte: Ziel menschenzentrierte KI-Governance

Im letzten Abschnitt dieses Leitfadens werden Handlungsmöglichkeiten in unterschiedlichen Bereichen, wie Kompetenzaufbau und Fortbildung, KI-Management Entscheidungshilfen, Experimente, Zertifizierungen, Nutzungsbedingungen und Kontrolle, Folgenabschätzung und Risikomanagement sowie Kommunikation und Stakeholdereinbindung vorgeschlagen. Das sich in zahlreichen europäischen (AI HLEG 2019; 2020) und nationalen Dokumenten (BMK und BMDW 2021) widerspiegelnde Ziel ist dabei eine menschenzentrierte KI-Governance, die den zuvor wiedergegebenen ethischen Prinzipien und Standards und dem rechtlichen Rahmen entspricht. Die digitale Transformation, insbesondere im Hinblick auf den Einsatz von KI, sollte nicht als gegeben hingenommen, sondern als Chance verstanden und gesellschaftlichen und organisationsbezogenen Anforderungen entsprechend geformt werden.

Wie aus der folgenden Abbildung ersichtlich, empfehlen wir Tools bzw. Institutionalisierungsmaßnahmen, also strukturelle Anpassungen, mit sechs verschiedenen Zielsetzungen in den entsprechenden Bereichen.

Ziel (Funktion)	Kompetenz- aufbau & Fortbildung	KI-Management Entscheidungs- hilfen	Experimentation	Zertifizierungen, Nutzungsbe- dingungen & Kontrolle	Folgenabschät- zung & Risiko Management	Kommunikation & Stakeholder- einbindung
Tools (Prozess)	Bildungsstan- dards für die Beschaffung und Verwendung von KI-Anwen- dungen / interne Kompetenzen zu technischen Verfahren und Entwicklung ethischer KI ("Ethics by De- sign", De-biasing)	KI-Einsatz-Ent- scheidungsbaum, AI RMF (NIST), Selbstverpflich- tende Leitlinien (DE BMAS), Risk Assessment Tools und Entwick- lungsstandards („Ethics by Design“; IEEE), Entscheidungs- kriterien für interne / externe Beschaffungsvor- gänge (Amster- dam Clauses)	Diskussionen zu Good Practices und Heraus- forderungen, Experimente zu verschiedenen Vorgehensweisen und Tools	Zertifizierungen von ISO, IEEE, TÜV, in Entste- hung begriffene Tools wie data. nutrition, data. hazards	Risk Assess- ment Tools EKIV, DEDA (UDS), VCIO (VDE et al.), FRAIA (UDS), ALTAI (EC HLEG), Fraunhofer KI-Prüfkatalog, NL Rechnungshof "Audit Framework for Algorithms"	DEDA (UDS), Workshops mit Stakeholdern, Diskussionen zu Good Practices und Herausforde- rungen
Institutionalisierung (Struktur)	Curriculum ab Juni 2023, VAB KI-Ethikseminar und Führungs- kräfte Lehrgang ab WS 2023, the- menspezifische interne Kom- petenzstellen, Informations- und Diskussionsver- anstaltungen für und mit Verwal- tung, Politik und Öffentlichkeit	Verwaltungs- Ethikrat mit inter- nen und externen Expert:innen (vgl. FI Aurora bzw. Etairos Ethikrat), KI-Observatorium (DE BMAS)	Interministe- rielles KI-Ethik Lab (vgl. AIT AI Ethics Lab, vgl. FI Projekt Aurora), Regulatory KI-Sandbox / Reallabor, AI Policy Forum	Freiwillig / nach AIA Implemen- tierung KI-Beh- örde, Daten- repositorien, welche konform zu rechtlichen und ethischen Vorgaben sind (Compliance), Richtlinien und Vorgaben für Ver- wendung (bspw. Durch Firmen)	Transparenzre- gister, Verarbei- tungsverzeichnis, KI-Behörde, NL Algorithm Register	AI Policy Forum, PIAZZA Format (Algorithm Watch), Inter- ministerielles KI-Ethik Lab

Tabelle 2: Empfehlungen KI-Governance

Kompetenzaufbau und Fortbildung

Das Ziel von Kompetenzaufbau und Fortbildungen in der Verwaltung ist hier von besonderer Bedeutung, weil KI-Literacy die Basis für alle anderen Maßnahmen darstellt. KI-Literacy bedeutet ein grundlegendes Verständnis von Funktion, Arbeitsweise und Möglichkeiten von KI-Anwendungen und ist die Voraussetzung für Planung, Vergabe bzw. Design, Entwicklung, Einsatz und Evaluation einer derartigen technischen Lösung. Darüber hinaus ist der erfolgreiche Einsatz von KI-Management, Entscheidungshilfen, Experimentation mit KI, Zertifizierung, Anwendung von Instrumenten zur Folgenabschätzung und Kommunikation von KI in der öffentlichen Verwaltung ebenso von KI-Literacy abhängig.

Vor diesem Hintergrund ist die Schaffung von Bildungsstandards für die Beschaffung und Verwendung von KI-Anwendungen besonders wichtig. Das existierende „Digitale Kompetenzmodell für Österreich“, das sich auf Umgang mit Informationen und Daten, Kommunikation und Zusammenarbeit, Kreation digitaler Inhalte, Sicherheit, Problemlösen und Weiterlernen bezieht, stellt hier eine wertvolle Grundlage dar (BMDW 2021). Diese sollte weiterhin regelmäßig aktualisiert werden, insbesondere in Bezug auf ethisch relevante Kompetenzen und im Einklang mit dem kommenden AI Act (siehe Abschnitt 8.3, 9).

Da die öffentliche Verwaltung eine zentrale Bedeutung für den Staat hat, sind hier interne Kompetenzen zu technischen Verfahren und Entwicklung ethischer KI, etwa im Sinne von „Ethics by Design“ (siehe Abschnitt 10.3, European Commission 2021a), sowie zu Problemstellungen, wie dem Umgang mit Bias, Transparenz, Rechenschaftspflicht, Recht auf Privatheit, bzw. allgemeiner die Einhaltung von Menschenrechten, von großer Bedeutung.

Im Hinblick auf strukturelle Maßnahmen zum Themenbereich Kompetenzaufbau und Fortbildung eröffnen themenspezifische Kompetenzstellen in der Verwaltung die Möglichkeit sektorspezifisches Wissen der jeweiligen Dienststellen mit technischem Fachwissen zu verbinden. Dabei könnten die in den Kompetenzstellen angesiedelten Expert:innen als Ansprechpersonen des jeweiligen Fachbereichs eine niederschwellige erste Unterstützung in Bezug auf ethische KI-Entwicklung, Anwendung und Evaluation übernehmen.

Darüber hinaus wären Weiterbildungsmöglichkeiten für diejenigen Verwaltungsbediensteten sinnvoll, die sich mit der Planung, Anwendung oder dem Management von KI-Anwendungen auseinandersetzen. Hier könnte – eventuell in Abstimmung mit anderen Bildungseinrichtungen – die Verwaltungsakademie des Bundes (VAB) eine wichtige Rolle spielen, die ab dem Wintersemester 2023 in einer Pilotphase über ein erstes derartiges Angebot verfügen wird.

Vor dem Hintergrund der besonderen Rolle des Vertrauens in die Verwaltung (siehe Abschnitt 5), auch bezogen auf die Verwendung von KI-Anwendungen, erscheinen der Aufbau von KI-Kompetenzen einer breiten Öffentlichkeit und eine möglichst offene Informationspolitik im Hinblick auf in der Verwaltung verwendete Anwendungen sinnvoll. Kompetenzen und eine umfassende Information der Öffentlichkeit können Misstrauen vorbeugen bzw. im Fall auftretender Problemlagen mit spezifischen (KI-)Anwendungen die Basis für eine Krisenkommunikation darstellen.

KI-Management Entscheidungshilfen

Die Planung, die Beschaffung und der Einsatz von KI bedürfen einer Reihe von Managemententscheidungen, die durch entsprechende Entscheidungshilfen unterstützt werden können. Dazu dient auf einer allgemeineren Ebene dieser Leitfaden, vergleichbar dazu (wenn auch eingeschränkt auf ein bestimmtes Politikfeld) wären die „Selbstverpflichtenden Leitlinien für den KI-Einsatz in der behördlichen Praxis der Arbeit und Sozialverwaltung“ des deutschen Bundesministeriums für Arbeit und Soziales (BMAS 2022). Noch wesentlich grundsätzlicher und umfangreicher ist die Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ des Deutschen Ethikrats (Deutscher Ethikrat 2023).

Unmittelbar handlungsbezogen findet sich im Abschnitt ein Entscheidungsbaum zum Einsatz von KI bzw. der Beurteilung, ob eine KI-Anwendung im Hinblick auf ihren ethischen Einsatz untersucht werden sollte oder nicht. Daran schließt die in Abschnitt 10.1 vorgestellte Entscheidungshilfe „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ an.

Andere Tools mit vergleichbaren, aber nicht auf die Verwaltung fokussierenden, Zielsetzungen umfassen den bereits erwähnten und in der EU ausgiebig diskutierten „Ethics by Design“ Ansatz (European Commission 2021a), das AI Risk Management Framework des National Institute of Standards and Technology (AI RMF; NIST 2023), die umfangreiche Normenserie 7000 des Institute of Electrical and Electronics Engineers (IEEE SA 2021) oder die Norm ISO/IEC 23053 der International Organization for Standardization (ISO 2022).

Für eine erste Annäherung an Beschaffungsvorgänge zu KI-Anwendungen liegt ein White Paper der österreichischen IÖB-Servicestelle vor (IÖB 2021). Eine Art Triage zum Thema Ethik wird durch den in diesem Leitfaden vorgeschlagenen Entscheidungsbaum ermöglicht (für einen umfangreicheren Alternativvorschlag von der NGO AlgorithmWatch siehe Loi et al. 2021). Zudem lassen sich Entscheidungskriterien für Beschaffungsvorgänge aus den ethischen Prinzipien und Standards (siehe Abschnitt 9) ableiten bzw. mit der daraus hervorgegangenen und hier im Leitfaden vorgestellten Entscheidungshilfe „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ (siehe Abschnitt 10.1) bearbeiten. In den Niederlanden hat die Stadt Amsterdam „Standard Clauses for Procurement of Trustworthy Algorithmic Systems“ entwickelt (City of Amsterdam 2021).

Auf der strukturellen Ebene könnte eine Beratungsstruktur im Hinblick auf den Einsatz von KI in der öffentlichen Verwaltung angedacht werden, dem Beispiel Finnlands folgend. Erfahrungen dazu wurden dort im Rahmen des Projektes Aurora (Ministry of Finance Finland o.J.) gemacht, wo KI-Anwendungen flächendeckend in der gesamten Verwaltung zum Einsatz gebracht werden. Dieser Vorgang wird durch das Projekt Etairos (Etairos o.J.) begleitet, das der gesamten Verwaltung einen Ethikrat zur Verfügung stellt. Dieser Ethikrat kann mit einer Aufgabe betraut oder von sich aus tätig werden (FCAI 2022). Ein anderes Modell wird im deutschen Bundesministerium für Arbeit und Soziales angewendet, wo ein KI-Observatorium, bestehend aus einer Reihe von Projekten und universitären sowie außeruniversitären Forschungsinstitutionen, dem Ministerium laufend Expertise zur Verfügung stellt.⁸

8 <https://www.ki-observatorium.de/>

Experimentation

Eine erfolgreiche Einführung und Weiterentwicklung von KI-Anwendungen in der öffentlichen Verwaltung bedarf der Möglichkeit von Experimentation. Idealerweise sollte eine Technologie vor dem flächendeckenden Einsatz in einer Pilotphase erprobt werden. Verschiedene Politikfelder haben hier eine reichhaltige Erfahrung vorzuweisen, beispielsweise Bildung und Arbeitsmarkt. Aus den ersten Erfahrungen in einem spezifischen Sektor oder in einer Region kann dann für den Einsatz in anderen Bereichen gelernt werden (Biegelbauer 2013). Im bereits erwähnten die gesamte nationale finnische Verwaltung umfassenden Projekt Aurora gibt es tatsächlich auch systematische Vergleiche und Schlussfolgerungen aus der Einführung von KI in den jeweiligen Politikfeldern (Ministry of Finance Finland o.J.).

Unterstützend wirken hier Diskussionen zu nationalen und internationalen Good Practices, aber auch Herausforderungen, wie sie beispielsweise im AI Policy Forum stattfinden. Experimente zu unterschiedlichen Vorgehensweisen und Tools können dabei die Lernerfahrung verbessern. Der Erfahrungsaustausch lässt sich durch ein entsprechendes Forum oder Netzwerk (vgl. das finnische Projekt Etairos, Etairos o.J.) unterstützen, das auch unterschiedliche Instrumente der Erprobung von Anwendungen wie Reallabore bzw. Sandboxes begleiten könnte (für KI Ethik Labs: Biegelbauer et al. 2022). In der europäischen Diskussion werden derartige Instrumentarien intensiv diskutiert und voraussichtlich auch im AI Act verankert (Stand 2023).

Zertifizierungen, Nutzungsbedingungen & Kontrolle

Um Sicherheit für die Nutzung bestimmter Anwendungen oder Datensätze zu erhalten, empfiehlt sich die Erarbeitung bzw. Verwendung von Zertifizierungen. Diese liegen, für verschiedene Zwecke, beispielsweise vom TÜV Austria, dem IEEE und der ISO bereits vor (IEEE SA 2021; TÜV Austria, Institute for Machine Learning 2021; ISO 2022). Andere Tools, wie beispielsweise im Fall der Zertifizierung von Datensätzen, data.nutrition oder data.hazards, sind in Entstehung begriffen.⁹

Auch Repositorien zur sicheren Speicherung von Daten sind für die Digitalisierung der Verwaltung ebenso wie die Governance von KI von zentraler Bedeutung. Für derartige Datenrepositorien ist es natürlich wichtig, den rechtlichen und ethischen Vorgaben zu entsprechen. Diese sollten in Form von Richtlinien und Vorgaben für deren Verwendung vorliegen. Dies ist einerseits für den Prozess der Erstellung von KI-Anwendungen wichtig, andererseits für die Nutzung derartiger Repositorien, insbesondere durch Anwender:innen ohne IT-Ausbildung.

Für die Kontrolle der Einhaltung verschiedener Normierungen sieht der AI Act nationale Behörden vor, die ihrerseits voraussichtlich durch eine eigens einzurichtende EU-Institution beobachtet werden sollen. Eine derartige KI-Behörde könnte natürlich bereits vor dem Inkrafttreten des AI Acts eingerichtet werden, um wertvolle Erfahrung

⁹ Siehe <https://datanutrition.org/> und <https://datahazards.com/>

zu sammeln. Diese Vorgangsweise wurde in Spanien gewählt, wo der Beschluss für die Etablierung einer derartigen Behörde bereits 2022 gefallen ist.

Die Wirksamkeit der Regulierung von KI wird stark von der Ausgestaltung der österreichischen KI-Behörde abhängen. Grundsätzlich ist eine Bandbreite, von einer passiv agierenden Behörde, an die Anfragen und Beschwerden herangetragen werden, bis hin zu einer Institution, die auch ohne konkreten Verdacht einzelne Anwendungen von sich aus überprüft, denkbar. Insbesondere im Hochrisikobereich würde ein erhöhtes Aktivitätsniveau der neuen Behörde die Sicherheit für Staat, Wirtschaft und Gesellschaft gleichermaßen erhöhen. Unter anderem würde sich eine raschere Ausjudizierung von Technologieanwendungen positiv auswirken, indem wie bei der DSGVO Unsicherheiten verringert werden.

Folgenabschätzung & Risikomanagement

Damit im Zusammenhang steht auch die Abschätzung der Folgen von KI-Anwendungen, die vor der In-Verkehr-Setzung sowie, insbesondere bei Hochrisikoanwendungen, auch während des Einsatzes erfolgen sollte. Vor dem Hintergrund der großen Anzahl von KI-Anwendungen in Wirtschaft und Gesellschaft, erscheint ein Risikomanagement durch eine Einteilung in Risikoklassen – wie auch beim AI Act vorgesehen – sinnvoll. Damit wird vor allem verhindert, dass auch für weitgehend unbedenkliche Anwendungen Berichtspflichten entstehen.

Für die vorausschauende wie rückwirkende Abschätzung gesellschaftlicher Folgen kann einerseits auf eine mehrere Jahrzehnte umfassende Erfahrung mit Instrumentarien von Technikfolgenabschätzung, Impact Assessment und Foresight-Methoden zurückgeblendet werden. Andererseits gibt es eine Reihe konkret auf KI abgestimmter Tools, die den Umgang mit Risiken erleichtern. Zuerst soll hier auf den im Abschnitt 10.1 vorgestellten „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ verwiesen werden, der Teil dieses Leitfadens ist. Dort werden KI-Anwendungen in den Bereichen Recht, Transparenz, Unvoreingenommenheit und Fairness, Effektivität und Effizienz, Sicherheit, Barrierefreiheit und Inklusivität, Rechenschaftspflicht sowie digitale Souveränität untersucht. Andere Tools und Leitfäden, mit teils sehr unterschiedlichen Schwerpunktsetzungen, wurden in den Abschnitten 9 und 10 ausgeführt. Hier sollen zusätzlich der „Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz“ des deutschen Fraunhofer IAIS (Fraunhofer IAIS 2021) und der „Audit Framework for Algorithms“ des niederländischen Rechnungshofs (Netherlands Court of Audit 2021) hervorgehoben werden.

Auf der strukturellen Ebene sollte sich der Umgang mit Folgenabschätzung und Risiken auch in den Tätigkeiten der zu schaffenden KI-Behörde widerspiegeln. Ein wichtiger Arbeitsbereich dieser Institution wird zweifellos die Überprüfung der jeweiligen Risikoabschätzungen für die einzelnen Anwendungen darstellen. Darüber hinaus sollten aber auch Transparenzregister (vgl. Algorithmus Register der niederländischen

Regierung¹⁰) und Verarbeitungsverzeichnisse angelegt bzw. geprüft werden, um insbesondere im Bereich des Staates jederzeit darüber Auskunft geben zu können, in welchen Bereichen, für welche Zwecke und auf welche Art und Weise KI-Anwendungen zum Einsatz kommen.

Kommunikation & Stakeholdereinbindung

Die Einführung einer neuen Technologie sollte auch durch Kommunikation und die Einbindung von Stakeholdern begleitet werden. Dies scheint aus verschiedenen Gründen sinnvoll. In Bezug auf die Kommunikation gilt es die Gesellschaft auf den Einsatz von KI auf staatlicher Ebene vorzubereiten, um kein Vertrauen zu verspielen. Die Einbindung von Stakeholdern innerhalb und außerhalb der Verwaltung erleichtert den Zugang zu sektorenspezifischer Expertise bei Design, Kreation, Implementierung und Evaluation von KI-basierten Anwendungen.

Hier gibt es einerseits umfangreiche Erfahrung mit partizipativen Prozessen (BKA und BML 2008; BMKÖS 2020), aber auch mit Tools, die Stakeholdereinbindung spezifisch zum Thema KI-Projekte organisieren. Beispiele sind hier DEDA von der Utrecht Data School für die Beurteilung von KI-Projekten (siehe Abschnitt 10.6) und die PIAZZA Konferenz¹¹ für die Kommunikation zwischen Staat und Zivilgesellschaft.

Im Zuge der Einführung von KI-Anwendungen sind jedoch auch Diskussionsformate innerhalb der Verwaltung unabdingbar. Einerseits könnte es hier um den Austausch von Good Practices zwischen verschiedenen Verwaltungsressorts, Verwaltungsebenen, aber auch unterschiedlichen Ländern gehen. Auch der kreative Umgang mit Herausforderungen sollte hier seinen Platz finden, beispielsweise von Akzeptanzproblemen, limitierten Budgets, öffentlichem Druck, der Spannung zwischen Transparenz- und Sicherheitserfordernissen sowie verschiedenen technischen und organisationalen Lösungen.

Die Institutionalisierung einer derartigen Plattform könnte in einem ausgebauten AI Policy Forum bestehen, das sich über die Umsetzung und Weiterentwicklung der AIM AT 2030 Strategie hinaus auch als institutionalisierte Plattform mit dem oben angeführten Portfolio versteht. Für eine stärkere Einbeziehung externer Expertise bei der Bearbeitung eines derartigen Aufgabenpaketes könnte auch ein interministerielles KI-Ethik Lab eingerichtet werden, in dem im Austausch zwischen internen und externen Expert:innen gemeinsames Lernen und Lösungsfindung zu alltäglichen Problemstellungen bei KI-Entwicklung und -Einsatz betrieben werden könnte.

10 <https://algoritmes.overheid.nl/en>

11 <https://piazza-konferenz.de/>

Quellenverzeichnis

- Ada Lovelace Institute, und DataKind UK.** 2020. Examining the Black Box: Tools for assessing algorithmic systems. Report. <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf>. Zugegriffen: 19. Mai 2023.
- Adamovich-Funk, Ludwig Karl, und Bernd-Christian Funk.** 1987. *Allgemeines Verwaltungsrecht*. Wien: Verlag Österreich.
- Adensamer, Angelika, Rita Gsenger, und Lukas Daniel Klausner.** 2021. „Computer Says No“: Algorithmic Decision Support and Organisational Responsibility. *Journal of Responsible Technology* 7–8. <https://doi.org/10.1016/j.jrt.2021.100014>.
- Adensamer, Angelika, und Lukas Daniel Klausner.** 2021. Algorithmen in der Entscheidungsfindung: Leitfaden zu Verantwortlichkeit und Rechenschaft. Leitfaden. Arbeiterkammer Wien. https://wien.arbeiterkammer.at/interessenvertretung/arbeitdigital/DataPolitics/VerA_Leitfaden_Final.pdf. Zugegriffen: 19. Mai 2023.
- AI Ethics Impact Group, VDE, und Bertelsmann Stiftung.** 2020. From Principles to Practice: An interdisciplinary framework to operationalise AI ethics. Report. <https://www.ai-ethics-impact.org/en>. Zugegriffen: 07. Juni 2023.
- Alon-Barkat, Saar und Madalina Busuioc.** 2023. Human–AI Interactions in Public Sector Decision Making: „Automation Bias“ and „Selective Adherence“ to Algorithmic Advice. *Journal of Public Administration Research and Theory* 33 (1): 153–169. <https://doi.org/10.1093/jopart/muac007>.
- Aoki, Naomi.** 2020. An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly* 37 (4). <https://doi.org/10.1016/j.giq.2020.101490>.
- Australian Government Dept. Industry, Science and Resources.** 2019. Australia’s Artificial Intelligence Ethics Framework. <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework>. Zugegriffen: 19. Mai 2023.
- Beijing Academy of Artificial Intelligence.** 2019. Beijing AI principles. International Research Center for AI Ethics and Governance Website. <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/>. Zugegriffen: 19. Mai 2023.
- Berka, Walter.** 2021. *Verfassungsrecht*. 8. Auflage. Wien: Verlag Österreich.
- Biegelbauer, Peter.** 2013. *Wie lernt die Politik – Lernen aus Erfahrung in Politik und Verwaltung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Biegelbauer, Peter, Anahid Jalali, Sven Schlarb, und Michela Vignoli.** 2022. Ethical AI: Why and How? October 2022. *ERCIM News* 131: 9–10. <https://ercim-news.ercim.eu/images/stories/EN131/EN131-web.pdf>. Zugegriffen: 19. Mai 2023.
- BMDW.** 2021. Digitales Kompetenzmodell für Österreich. DigComp 2.2 AT. https://www.bmaw.gv.at/dam/jcr:54bbe103-7164-494e-bb30-cd152d9e9b33/DigComp2.2_V33-barrierefrei.pdf. Zugegriffen: 23. Juni 2023.
- BMDW, und BMK.** 2021. Vortrag an den Ministerrat (70/16) Strategie der Bundesregierung für Künstliche Intelligenz: Artificial Intelligence Mission Austria 2030 (AIM AT 2030). https://www.ris.bka.gv.at/Dokumente/Mrp/MRP_20210915_70/011_000.pdf. Zugegriffen: 05. Mai 2023.
- BMVIT, BMDW, BMBWF, und BMASGK.** 2018. Ergebnisbericht: Zusammenfassung der Ergebnisse der Expertinnen und Experten zur Erarbeitung eines Strategieplans für Künstliche Intelligenz. https://www.bundeskanzleramt.gv.at/dam/jcr:094fa5af-1acc-4238-8d7e-e27351005d45/15_13_bei_NB.pdf. Zugegriffen: 07. Juni 2023.
- BRZ.** 2020. Forum Alpbach: Bundesrechenzentrum stellt Prüfkatalog für vertrauenswürdige KI-Systeme vor. BRZ Website. <https://www.brz.gv.at/presse/BRZ-Breakout-Session-beim-Europaeischen-Forum-Alpbach-2020.html>. Zugegriffen: 24. Juni 2023.
- Bundesarbeiterkammer.** 2021. VERA. <https://vera.arbeiterkammer.at/#/>. Zugegriffen: 07. Juni 2023.

- BJA, und BML.** 2008. Standards der Öffentlichkeitbeteiligung: Empfehlungen für die gute Praxis. <https://partizipation.at/wp-content/uploads/2022/09/standards-der-oeffentlichkeitsbeteiligung-2008-druck.pdf>. Zugegriffen: 20. Juni 2023.
- BMAS.** 2022. Selbstverpflichtende Leitlinien für den KI-Einsatz in der behördlichen Praxis der Arbeits- und Sozialverwaltung. https://www.bmas.de/SharedDocs/Downloads/DE/Publikationen/a862-01-leitlinien-ki-einsatz-behoerdliche-praxis-arbeits-sozialverwaltung.pdf?__blob=publicationFile&v=2. Zugegriffen: 20. Juni 2023.
- BMK, und BMDW.** 2021. Strategie der Bundesregierung für Künstliche Intelligenz: Artificial Intelligence Mission Austria 2030 (AIM AT 2030). <https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/strategie-bundesregierung.html>. Zugegriffen: 05. Mai 2023.
- BMKÖS.** 2020. Grünbuch: Partizipation im digitalen Zeitalter. <https://oeffentlicherdienst.gv.at/publikationen/gruenbuch-partizipation-im-digitalen-zeitalter/>. Zugegriffen: 20. Juni 2023.
- BusinessEurope.** 2021. The Artificial Intelligence Act (AI Act) – a BusinessEurope position paper. <https://www.businesseurope.eu/publications/artificial-intelligence-act-ai-act-businesseurope-position-paper>. Zugegriffen: 19. Mai 2023.
- BusinessEurope.** 2023. Joint industry statement on the EU Artificial Intelligence (AI) Act. <https://www.businesseurope.eu/publications/joint-industry-statement-eu-artificial-intelligence-ai-act>. Zugegriffen: 07. Juni 2023.
- Buxmann, Peter, und Holger Schmidt.** 2021. Grundlagen der künstlichen Intelligenz und des maschinellen Lernens. In *Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg*, Hrsg. Peter Buxmann und Holger Schmidt, 3–25. Berlin: Springer Gabler.
- Christl, Wolfie.** 2021. Digitale Überwachung und Kontrolle am Arbeitsplatz. Von der Ausweitung betrieblicher Datenerfassung zum algorithmischen Management. Studie gefördert durch Digifonds, Arbeiterkammer Wien. Cracked Labs. <https://crackedlabs.org/daten-arbeitsplatz/info>. Zugegriffen: 20. Juni 2023.
- City of Amsterdam.** 2021. Standard Clauses for Procurement of Trustworthy Algorithmic Systems. Amsterdam. https://assets.amsterdam.nl/publish/pages/1017896/standard_clauses_for_procurement_of_trustworthy_algorithmic_systems_1.docx. Zugegriffen: 20. Juni 2023.
- Clausen, Nelly, und Mirko Tobias Schäfer.** 2023. Angewandte Ethik für Daten- und KI-Projekte in der öffentlichen Verwaltung. In *Handbuch Digitalisierung der Verwaltung*, Hrsg. Tobias Krause, Christian Schachtner, Basanta Thapa, 233–251. Bielefeld: transcript.
- Coeckelbergh, Mark.** 2021. AI for climate: freedom, justice, and other ethical and political challenges. *AI and Ethics* 1: 67–72. <https://doi.org/10.1007/s43681-020-00007-2>.
- Dastin, Jeffrey.** 2022. Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women. In *Ethics of Data and Analytics: Concepts and Cases*, Hrsg. Kirsten Martin, 296–299. New York: Auerbach Publications.
- Datenethikkommission.** 2019. Gutachten der Datenethikkommission der Bundesregierung. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=6B37F4B2D6F0875D6DB26D190B85F5C0.2_cid373?__blob=publicationFile&v=6. Zugegriffen: 15. Juni 2023.
- Deutscher Bundestag, Projektgruppe „KI und Staat“.** 2019. Zusammenfassung der vorläufigen Ergebnisse, Stand: 18. Dezember 2019. <https://www.bundestag.de/dokumente/textarchiv/2020/kw44-pa-enquete-ki-abschlussbericht-801192>. Zugegriffen: 15. Juni 2023.
- Deutscher Ethikrat.** 2023. Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz: Stellungnahme. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>. Zugegriffen: 20. Juni 2023.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, und Kristina Toutanova.** 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1810.04805>.
- Dilmegani, Cem.** 2022. Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2023. AI Multiple Website. <https://research.aimultiple.com/ai-bias/>. Zugegriffen: 15. Juni 2023.

- Döbel, Inga, Miriam Leis, Manuel Molina Vogelsang, Dmitry Neustroev, Henning Petzka, Annamaria Riemer, Stefan Rüping, Angelika Voss, Martin Wegele, Juliane Welz.** 2018. Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung. Studie. Fraunhofer-Gesellschaft. https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf. Zugegriffen: 15. Juni 2023.
- Ebers, Martin, Veronica R. S. Hoch, Frank Rosenkranz, Hannah Ruschemeier, und Björn Steinrötter.** 2021. The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J — Multidisciplinary Scientific Journal* 4 (4): 589–603. <https://doi.org/10.3390/j4040043>.
- Eidler, Jakob, Knut Blind, Rainer Frietsch, Simone Kimpeler, Henning Kroll, Christian Lerch, Thomas Reiss, Florian Roth, Torben Schubert, Johanna Schuler and Rainer Walz.** 2020. Technologiesouveränität: von der Forderung zum Konzept. Perspectives – Policy Briefs 02/2020. Policy Brief. Fraunhofer Institute for Systems and Innovation Research (ISI). <https://www.isi.fraunhofer.de/content/dam/isi/dokumente/publikationen/technologiesouveraenitaet.pdf>. Zugegriffen: 07. Juni 2023.
- Ertel, Wolfgang, und Nathanael T. Black.** 2016. *Grundkurs Künstliche Intelligenz*. 4. Auflage. Wiesbaden: Springer.
- Etairos. o. J. ETAIROS: Towards Ethical Use of AI.** Etairos Website. <https://etairos.fi/en/front-page/>. Zugegriffen: 17. Mai 2023.
- European Commission.** 2018. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>. Zugegriffen: 19. Mai 2023.
- European Commission.** 2021a. Ethics By Design and Ethics of Use Approaches for Artificial Intelligence. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf. Zugegriffen: 19. Mai 2023.
- European Commission.** 2021b. EU Grants: How to complete your ethics self-assessment. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf. Zugegriffen: 19. Mai 2023.
- European Commission.** 2021c. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022a. Data Governance Act Explained. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022b. New Liability Rules on Products and AI to Protect Consumers. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5807. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022c. New Rules to Ensure the Safety of Machinery and Robots. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7741. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022d. The Digital Markets Act: Ensuring Fair and Open Digital Markets. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022e. Digital Services Act: Commission Is Setting up New European Centre for Algorithmic Transparency. <https://digital-strategy.ec.europa.eu/en/news/digital-services-act-commission-setting-new-european-centre-algorithmic-transparency>. Zugegriffen: 07. Juni 2023.
- European Commission.** 2023. European Data Governance Act. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>. Zugegriffen: 07. Juni 2023.
- European Commission.** o. J. Regulatory framework proposal on artificial intelligence. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>. Zugegriffen: 30. April 2023.

- European Digital Rights (EDRI).** 2023. Civil society urges European Parliament to protect people's rights in the AI Act. <https://edri.org/our-work/civil-society-urges-european-parliament-to-protect-peoples-rights-in-the-ai-act/>. Zugegriffen: 19. Mai 2023.
- Feik, Rudolf.** 2007. *Öffentliche Verwaltungskommunikation: Öffentlichkeitsarbeit, Aufklärung, Empfehlung, Warnung*. Wien: Springer.
- Finnish Center for Artificial Intelligence (FCAI).** 2022. FCAI Ethics Advisory Board: Ethics Matters. FCAI Website. <https://fcai.fi/ethics-advisory-board>. Zugegriffen: 07. Juni 2023.
- Franzke, Aline Shakti, Iris Muis, und Mirko Tobias Schäfer.** 2021. Data Ethics Decision Aid (DEDA): A Dialogical Framework for Ethical Inquiry of AI and Data Projects in the Netherlands. *Ethics and Information Technology* 23 (3): 551–67. <https://doi.org/10.1007/s10676-020-09577-5>.
- Fraunhofer IAIS.** 2021. Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz: KI-Prüfkatalog. https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf. Zugegriffen: 21. Mai 2023.
- General Secretary of the French Digital Council.** 2018. AI for Humanity: French Strategy for Artificial Intelligence. President of the French Republic. <https://www.aiforhumanity.fr/en/>. Zugegriffen: 19. Mai 2023.
- Gerards, Janneke, Mirko Tobias Schäfer, Arthur Vankan, und Iris Muis.** 2022. Impact Assessment Fundamental Rights and Algorithms. Report. Ministry of the Interior and Kingdom Relations. <https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms>. Zugegriffen: 07. Juni 2023.
- Gesk, Tanja Sophie, und Michael Leyer.** 2022. Artificial intelligence in public services: When and why citizens accept its usage. *Government Information Quarterly* 39 (3). <https://doi.org/10.1016/j.giq.2022.101704>.
- Goddard, Kate, Abdul Roudsari, und Jeremy C. Wyatt.** 2012. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19 (1): 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>.
- Goldacker, Gabriele.** 2017. Digitale Souveränität. Publikation. Kompetenzzentrum Öffentliche IT, Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS. <https://doi.org/10.24406/publica-fhg-298824>.
- Haslinger, Susanne.** 2022. Gesetzliche Rahmenbedingungen für den Einsatz von KI-Assistenzsystemen am Arbeitsplatz. In *Verantwortungsvolle Einbindung von KI-Assistenzsystemen am Arbeitsplatz. Ein Handbuch für Arbeitnehmende und ihre Vertretungen*, Hrsg. Julian Anslinger, Jaroslava Huber, Michael Haslgrüber, Anita Thaler, 61–73. Graz: IFZ – Interdisziplinäres Forschungszentrum für Technik, Arbeit und Kultur. <https://doi.org/10.17605/OSF.IO/98B4H>.
- Hidvegi, Fanny, Daniel Leufer, und Estelle Massé.** 2021. The EU Should Regulate AI on the Basis of Rights, Not Risks. Access Now Website. <https://www.accessnow.org/eu-regulation-ai-risk-based-approach/>. Zugegriffen: 07. Juni 2023.
- High-Level Expert Group on Artificial Intelligence (AI HLEG).** 2019. Ethik-leitlinien für eine vertrauenswürdige KI. Amt für Veröffentlichungen der Europäischen Union. <https://data.europa.eu/doi/10.2759/22710>. Zugegriffen: 07. Juni 2023.
- High-Level Expert Group on Artificial Intelligence (AI HLEG).** 2020. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment. <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>. Zugegriffen: 07. Juni 2023.
- Holzinger, Gerhart, Peter Oberndorfer, und Bernhard Raschauer.** 2013. *Österreichische Verwaltungslehre*. Wien: Verlag Österreich.
- Humm, Bernhard G., Peter Buxmann und Jan C. Schmidt.** 2022. Grundlagen und Anwendungen von KI. In *Künstliche Intelligenz in der Forschung: Neue Möglichkeiten und Herausforderungen für die Wissenschaft*, Hrsg. Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt, Indra Spiecker genannt Döhmann, 13–42. Berlin: Springer Nature.

- IEEE Standards Association (IEEE SA).** 2021. IEEE 7000-2021: IEEE Standard Model Process for Addressing Ethical Concerns during System Design. Standard. <https://standards.ieee.org/ieee/7000/6781/>. Zugegriffen: 07. Juni 2023
- Initiative D21.** 2019. #ALGOMON: 9 Leitlinien zum ethischen Umgang mit Algorithmen-Monitoring. Leitlinien. https://initiated21.de/app/uploads/2019/12/algomon_leitlinien_191216.pdf. Zugegriffen: 07. Juni 2023.
- Innovationsfördernde Öffentliche Beschaffung (IÖB).** 2021. Künstliche Intelligenz – Wie kann die öffentliche Verwaltung KI nutzen und beschaffen. IÖB White Paper. https://www.ioeb.at/fileadmin/ioeb/Dokumente/Infothek/IOEB_White_Paper_-_Kuenstliche_Intelligenz.pdf. Zugegriffen: 12. Juni 2023.
- International Organization for Standardization (ISO).** 2022. ISO/IEC 23053:2022: Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML). Standard. <https://www.iso.org/standard/74438.html>. Zugegriffen: 20. Juni 2023.
- Jobin, Anna, Marcello Lenca, und Effy Vayena.** 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1 (9): 389–399. <https://doi.org/10.1038/s42256-019-0088-2>.
- Karkulik, Stefan.** 2014. Rechtsschutz gegen die Öffentlichkeitsarbeit der Verwaltung nach der Verwaltungsgerichtsbarkeits-Novelle 2012. *Journal für Rechtspolitik* 22 (3): 169–187.
- Kaur, Davinder, Suleyman Uslu, and Arjan Duresi.** 2021. Requirements for Trustworthy Artificial Intelligence – A Review. In *Advances in Networked-Based Information Systems: The 23rd International Conference on Network-Based Information Systems (NBIS-2020)*, Hrsg. Leonard Barolli, Kin Fun Li, Tomoya Enokido, Makoto Takizawa, 105–115. Cham: Springer. https://doi.org/10.1007/978-3-030-57811-4_11.
- Lachmayer, Konrad.** 2018. Die DSGVO im öffentlichen Bereich. *Österreichische Juristenzeitschrift* 03: 112–120.
- Latonero, Mark.** 2018. Governing Artificial Intelligence: Upholding Human Rights & Dignity. Publikation. Data&Society Research Institute. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf. Zugegriffen: 07. Juni 2023.
- Legg, Shane, und Marcus Hutter.** 2007. A Collection of Definitions of Intelligence. *Frontiers in Artificial Intelligence and applications* 157: 17–24. <https://doi.org/10.48550/arXiv.0706.3639>.
- Lobe, Adrian.** 2022. Diskriminierung durch und von KI. In *Vom richtigen Umgang mit den „Anderen“: Diskriminierung, Rassismus und Recht heute*, Hrsg. Eric Hilgendorf, und Enis Tiz, 147–160. Baden-Baden: Ergon Verlag. <https://doi.org/10.5771/9783956509346-147>.
- Loi, Michele, Anna Mätzener, Angela Müller, und Matthias Spielkamp.** 2021. Automated Decision-Making Systems in the Public Sector. Tool. Algorithm Watch. <https://algorithmwatch.org/en/adms-impact-assessment-public-sector-algorithmwatch/>. Zugegriffen: 19. Mai 2023.
- Madan, Rohit, und Mona Ashok.** 2022. A Public Values Perspective on the Application of Artificial Intelligence in Government Practices: A Synthesis of Case Studies. In *Handbook of Research on Artificial Intelligence in Government Practices and Processes*, Hrsg. José Ramon Saura und Felipe Debasa, 162–189. Hershey: IGI Global Publishing. <https://doi.org/10.4018/978-1-7998-9609-8>.
- Madiega, Tambiama.** 2022. Briefing EU Legislation in Progress: Artificial Intelligence Act. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf). Zugegriffen: 19. Mai 2023.
- McCarthy, John, Marvin Lee Minsky, Nathaniel Rochester, und Claude E. Shannon.** 2006 [1955]. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI magazine* 27 (4): 12–12.
- Ministry of Finance Finland.** o. J. Implementation of the National AuroraAI Programme. Valtiovarainministeriö. <https://vm.fi/en/auroraai-en>. Zugegriffen: 17. Mai 2023.
- National Institute of Standards and Technology (NIST).** 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>. Zugegriffen: 16. Juni 2023.

Nentwich, Michael, Matthias Weber, Dennis Appelt, Eva Buchinger, Leo Capari, Evgeniia Filippova, Niklas Gudowsky-Blatakes, Barbara Heller-Schuh, Manuela Kienegger, Klaus Kubeczko, Wenzel Mehnert, Michael Ornetzeder, Walter Peissl, Petra Schaper-Rinkel, Anna Wang, Dana Wasserbacher. 2021. Foresight und Technikfolgenabschätzung: Monitoring von Zukunftsthemen für das Österreichische Parlament. Report. ÖAW & AIT, 44–46. <https://publications.ait.ac.at/en/publications/foresight-und-technikfolgenabsch%C3%A4tzung-monitoring-f%C3%BCr-das-%C3%B6sterre-2>. Zugegriffen: 07. Juni 2023.

Netherlands Court of Audit. 2021. Audit framework for algorithms. <https://english.rekenkamer.nl/publications/publications/2021/01/26/audit-framework-for-algorithms>. Zugegriffen: 22. Juni 2023.

OECD. 2019. Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://oecd.ai/en/ai-principles>. Zugegriffen: 22. Juni 2023.

OGM research & communication. 2022. OGM/APA-Vertrauensindex Institutionen Juli 2022. OGM Website. <https://www.ogm.at/2022/07/29/ogm-apa-vertrauensindex-institutionen-juli-2022/>. Zugegriffen: 22. Juni 2023.

Pichal, Sundar. 2018. AI at Google: our principles. Google Website. <https://blog.google/technology/ai/ai-principles/>. Zugegriffen: 07. Juni 2023.

Rat für Forschung und Technologieentwicklung. 2021. Wie souverän kann und muss ein Staat bei system-relevanten Technologien sein?. OTS Website. https://www.ots.at/presseaussendung/OTS_20210119_OTS0168/wie-souveraen-kann-und-muss-ein-staat-bei-system-relevanten-technologien-sein. Zugegriffen: 3. Mai 2023.

Rohde, Friederike, Josephin Wagner, Philipp Reinhard, Ulrich Petschow, Andreas Meyer, Marcus Voß, und Anne Mollen. 2021. Nachhaltigkeitskriterien für künstliche Intelligenz: Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus. Publikation in der Schriftenreihe des IÖW 220/21. Berlin: IÖW. https://www.ioew.de/fileadmin/user_upload/BILDER_und_Downloaddateien/Publikationen/2021/IOEW_SR_220_Nachhaltigkeitskriterien_fuer_Kuenstliche_Intelligenz.pdf. Zugegriffen: 3. Mai 2023.

Russell, Stuart J., und Peter Norvig. 2012. *Künstliche Intelligenz. Ein moderner Ansatz*. 3. Auflage. München: Pearson Studium.

Schäfer, Mirko Tobias, und Nelly M. Clausen. 2021. Participatory Data Ethics. A practical approach. In *Mensch und Computer 2021 – Workshopband*. Hrsg. Carolin Wienrich, Philipp Wintersberger, und Benjamin Weyers. Bonn: Gesellschaft für Informatik e.V. <https://doi.org/10.18420/muc2021-mci-ws06-316>.

Scheichenbauer, Heidi, und Lisa Seidl. 2023. Die verwaltungsrechtliche Einordnung von Internet-Recherchen durch Verwaltungsbehörden. In *Datenschutzrecht Jahrbuch 2022*, Hrsg. Dietmar Jahnel, 49–62. Wien: Verlag Österreich. <https://doi.org/10.37942/9783708341194>.

Sigfrids, Anton, Nieminen Mika, Leikas Jaana, und Pikkuahto Pietari. 2022. How Should Public Administrations Foster the Ethical Development and Use of Artificial Intelligence? A Review of Proposals for Developing Governance of AI. *Frontiers In Human Dynamics* 4. <https://doi.org/10.3389/fhumd.2022.858108>.

Stahl, Bernd Carsten. 2021. *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*. New York: Springer.

Starke, Christopher, und Marco Lünich. 2020. Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy* 2 (16): 1–17. <https://doi.org/10.1017/dap.2020.19>.

Statistik Austria. 2021. IKT-Einsatz in Haushalten. Statistik Austria Website. <https://www.statistik.at/statistiken/forschung-innovation-digitalisierung/digitale-wirtschaft-und-gesellschaft/ikt-einsatz-in-haushalten>. Zugegriffen: 07. Juni 2023.

Stöger, Karl. 2014. Verhaltensökonomische Steuerungsinstrumente und Verfassungsrecht – Einige Gedanken. *Austrian Law Journal* 1: 89–98. <https://doi.org/10.25364/1.1:2014.1.8>.

Strubell, Emma, Ananya Ganesh, und Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. In the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. July 2019. <https://doi.org/10.48550/arXiv.1906.02243>.

- The White House.** 2022. Blueprint for an AI Bill of Rights. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>. Zugegriffen: 16. Juni 2023.
- TÜV Austria, Institute for Machine Learning.** 2021. Trusted Artificial Intelligence. White Paper. https://en.tuv.at/wp-content/uploads/sites/12/2022/03/Whitepaper_Trusted-AI_TUeV-AUSTRIA_JKU.pdf. Zugegriffen: 19. Mai 2023.
- UK Parliament, House auf Lords.** 2021. Public Authority Algorithm Bill. <https://hansard.parliament.uk/Lords/2021-11-29/debates/E07A5CBD-A767-4D35-9261-37B35BA086BB/PublicAuthorityAlgorithmBill%28HL%29>. Zugegriffen: 07. Juni 2023.
- UNESCO.** 2022. Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Zugegriffen: 19. Mai 2023.
- Utrecht Data School.** o. J. Data Ethics Decision Aid (DEDA). Tool. Utrecht Data School. <https://dataschool.nl/en/deda/>. Zugegriffen: 10. April 2023.
- Utrecht University.** 2022. Utrecht Data School Was Invited to the European Parliament to Talk about FRAIA. Utrecht University Website. <https://www.uu.nl/en/news/utrecht-data-school-was-invited-to-the-european-parliament-to-talk-about-fraia>. Zugegriffen: 07. Juni 2023.
- Valmeekam, Karthik, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati.** 2023. Large Language Models Still Can't Plan (A Benchmark for LLMs on Planning and Reasoning about Change). *arXiv preprint*. <https://doi.org/10.48550/arXiv.2206.10498>.
- Wille, Matt.** 2021. South Korean chatbot 'Lee Luda' killed off for spewing hate. Inverse/Input Website. <https://www.inverse.com/input/culture/south-korean-chatbot-lee-luda-killed-off-for-spewing-hate>. Zugegriffen: 18. Juni 2023.
- Winston, Patrick Henry.** 1992. *Artificial intelligence*. Addison-Wesley: Longman Publishing Co. Inc.
- World Wide Web Foundation.** 2018. Policy Brief W20 Argentina: Artificial Intelligence: open questions about gender inclusion. <https://webfoundation.org/docs/2018/06/AI-Gender.pdf>. Zugegriffen: 07. Juni 2023.

Anhang

Checkliste für ethische KI in der Verwaltung

Recht

Verarbeitet die KI-Anwendung Daten im Einklang mit den Anforderungen der Rechtsnormen und -prinzipien, die im nationalen und EU-Rechtsrahmen festgelegt ist? Ja Nein

Ist der Schutz von Grundrechten wie der Privatsphäre und der Menschenwürde während des Einsatzes der KI-Anwendung gewährleistet? Ja Nein

Ist die KI-Anwendung so konzipiert, dass sie sich an mögliche Änderungen des geltenden Rechts anpassen lässt? Ja Nein

Transparenz

Werden der Öffentlichkeit und den Verwaltungsbediensteten die spezifischen Ziele und Zwecke der KI-Anwendung vermittelt? Ja Nein

Werden der Öffentlichkeit und den Verwaltungsbediensteten die Auswirkungen des Einsatzes der KI-Anwendung nach der Implementierung ausreichend vermittelt? Ja Nein

Unvoreingenommenheit und Fairness

Können potenziell ethisch und sozial unerwünschte Auswirkungen der KI-Anwendung auf die Gesellschaft erkannt, gestoppt und ihre Wiederholung verhindert werden? Ja Nein

Wurde sichergestellt, dass die KI-Anwendung keine Personen stigmatisieren oder diskriminieren kann? (Z.B. aufgrund von Geschlecht, ethnischer oder sozialer Herkunft, Alter, sexueller Ausrichtung, Religion oder Weltanschauung) Ja Nein

Wurde sichergestellt, dass die KI-Anwendung keine Personen stigmatisieren oder diskriminieren kann? (Z.B. aufgrund von Geschlecht, ethnischer oder sozialer Herkunft, Alter, sexueller Ausrichtung, Religion oder Weltanschauung) Ja Nein

Effektivität und Effizienz

Hat die KI-Anwendung das Potenzial die Arbeitssituation, der im öffentlichen Dienst tätigen Personen zu verbessern oder zumindest nicht zu verschlechtern? Ja Nein

Werden Kriterien verwendet, um festzustellen, ob der Einsatz der KI-Anwendung die Effektivität und Effizienz der öffentlichen Dienste verbessert? Ja Nein

Sicherheit

Wurden relevante Sicherheitsrisiken im Zusammenhang mit der KI-Anwendung ermittelt? Ja Nein

Gibt es Risikominderungsmaßnahmen? Ja Nein

Gibt es Sicherheitsvorkehrungen zum Schutz vor Missbrauch oder böswilliger Nutzung der KI-Anwendung? Ja Nein

Zugänglichkeit und Inklusion

Ist die KI-Anwendung so konzipiert, dass sie von verschiedenen Arten von Endnutzer:innen mit unterschiedlichen Kompetenzniveaus verwendet werden kann? Ja Nein

Werden Alternativen zur KI-Anwendung angeboten, um einen gleichberechtigten nicht-KI-bezogenen Zugang zu öffentlichen Dienstleistungen zu gewährleisten? Ja Nein

Rechenschaftspflicht

Sind klare Verantwortlichkeiten und Rechenschaftspflichten für Entwickler:innen und Nutzer:innen der KI-Anwendung festgelegt? Ja Nein

Gibt es Mechanismen, die sicherstellen, dass die für den Einsatz und die Nutzung von KI-Verantwortlichen sich ihrer Verantwortung bewusst sind? Ja Nein

Gibt es Mechanismen, die eine angemessene Entschädigung für Schäden infolge des Einsatzes von KI sicherstellen? Ja Nein

Digitale Souveränität

Wenn die Entwicklung oder der Betrieb von KI-Anwendungen ausgelagert wird, gibt es Maßnahmen zum Schutz sensibler Daten und zur Verhinderung des Zugriffs durch Drittorganisationen? Ja Nein

Gibt es bezogen auf die KI-Anwendung in der Verwaltung ausreichende Kenntnisse über KI-Technologien, um mit Dienstleistern und Herstellern auf Augenhöhe zu interagieren? Ja Nein

Kriterien und Maßnahmenkatalog für ethische KI in der Verwaltung (EKIV)

Kriterien

1. **Recht:** Einhaltung geltenden Rechts als eine Minimalbedingung in Bezug auf die ethische Anwendung von KI, die unter anderem vor dem Hintergrund der raschen Entwicklung der Technologie für eine tatsächliche ethische Anwendung als nicht ausreichend erkannt wird.
 - Welche rechtlichen Rahmenbedingungen werden zu Beginn des KI-Projekts herangezogen?
 - Ist beim Einsatz der KI-Anwendung die Einhaltung geltenden Rechts gewährleistet?
 - Wird beim Einsatz der KI-Anwendung der Schutz von Grundrechten wie der Privatsphäre und der Menschenwürde gewährleistet?
 - Wie wird gewährleistet, dass die KI-Anwendung an potenzielle Veränderungen geltenden Rechts angepasst wird?

2. **Transparenz:** Ein für alle Beteiligten (Entwickler:in, Anwender:in, betroffene/r Bürger:in) nachvollziehbarer Entscheidungsfindungsprozess.
 - Wie werden die Öffentlichkeit und die Verwaltungsbediensteten über die spezifischen Ziele und Zwecke der KI-Anwendung in der öffentlichen Verwaltung informiert?
 - Wie werden die Auswirkungen des Einsatzes der KI-Anwendung der Öffentlichkeit und den Verwaltungsbediensteten vermittelt?

3. **Unvoreingenommenheit und Fairness:** Verwendung vielfältiger Daten und Modelle, um zu vermeiden, dass bestehende Vorurteile fortbestehen bzw. in der KI-Anwendung implizit mitwirken.
 - Wie wird die Fairness der KI-Anwendung im Kontext der öffentlichen Verwaltung bewertet?
 - Welche Kriterien werden verwendet, um festzustellen, ob die Anwendung fair ist?
 - Wie wird sichergestellt, dass die zum Training der KI-Anwendung verwendeten Daten vielfältig und repräsentativ (z. B. für Mitarbeiter:innen oder die Bevölkerung) sind?
 - Wie werden die potenziellen Verzerrungen der KI-Anwendung denjenigen vermittelt, die von ihrem Einsatz betroffen sind, z. B. den Bürger:innen oder den Bediensteten der öffentlichen Verwaltung?
 - Ist es notwendig oder sinnvoll, Bürger:innen und /oder Organisationen der Zivilgesellschaft einzubeziehen, um Vorurteile bei der Anwendung der KI-Technologie zu erkennen und zu beseitigen?

4. **Effektivität und Effizienz:** Einsatz der KI-Anwendungen nur dann, wenn die Effektivität und Effizienz von Dienstleistungen der öffentlichen Verwaltung nachhaltig verbessert werden, ohne die Arbeitssituation der im öffentlichen Dienst tätigen Personen zu verschlechtern.
 - Welche Kriterien werden verwendet, um zu bestimmen, wann der Einsatz der KI-Anwendung die Effektivität und Effizienz von Dienstleistungen der Verwaltung verbessert?
 - Welche Kriterien werden verwendet, um zu bestimmen, ob sich die Arbeitssituation, der im öffentlichen Dienst tätigen Personen durch den Einsatz einer KI-Anwendung verschlechtert?

5. **Sicherheit:** Sicherer Einsatz der KI-Technologie, um sensible Informationen zu schützen und unbefugten Zugriff zu verhindern.
 - Welches sind die größten Sicherheitsrisiken, die mit dem Einsatz der KI-Anwendung verbunden sind, und wie können diese Risiken wirksam gemindert werden?
 - Welche Sicherheitsvorkehrungen gibt es zum Schutz vor dem Missbrauch oder der böswilligen Nutzung der KI-Anwendung?

6. **Barrierefreiheit und Inklusivität:** Möglichkeit der Nutzung und Verfügbarkeit von KI-Technologien für Personen mit unterschiedlichen Fähigkeiten, Hintergründen und Kulturen.
 - Wie wird die KI-Anwendung für Menschen mit unterschiedlichen Fähigkeiten, Hintergründen und Kulturen zugänglich und integrativ gestaltet?
 - Wie wird die Bevölkerung dazu befähigt, die KI-Anwendung zu nutzen und davon zu profitieren?
 - Welche Alternativen zur KI-Technologie werden angeboten, um einen gleichberechtigten Zugang zu öffentlichen Dienstleistungen zu gewährleisten (insbesondere persönliche Ansprechpartner)?

7. **Rechenschaftspflicht:** Klare Verantwortlichkeiten (Funktionen wie Entwickler:in, Manager:in, Anwender:in etc.) und Rechenschaftspflichten.
 - Welche klaren Verantwortlichkeiten und Rechenschaftspflichten werden für Entwickler:innen, Manager:innen und Nutzer:innen der KI-Technologie eingeführt?
 - Welche Mechanismen gibt es, um sicherzustellen, dass die für den Einsatz und die Nutzung von KI-Verantwortlichen sich ihrer Verantwortung bewusst sind?
 - Welche Mechanismen oder Strategien gibt es, um sicherzustellen, dass Schäden, die durch die KI-Anwendung entstehen, angemessen entschädigt werden?

8. **Digitale Souveränität:** Fähigkeit der Verwaltung, die Entwicklung von KI-Lösungen beeinflussen, eigenständig anwenden zu können sowie vertrauliche Daten im eigenen Kontrollbereich zu belassen.
- Wie wird sichergestellt, dass ausreichend Wissen um KI-Technologien in der Verwaltung vorhanden ist, um bei Beschaffungsvorgängen mit Dienstleistern und bei Kontrollen im Rahmen der Regulation von KI mit Herstellern auf Augenhöhe umgehen zu können?
 - Welche (ethischen) Aspekte werden bei der Entscheidung über die Auslagerung der KI-Entwicklung und -Implementierung an externe Dienstleister berücksichtigt?
 - Wenn Entwicklung oder Betrieb von KI-Anwendungen ausgelagert wird: Welche Maßnahmen gibt es zum Schutz sensibler Daten und zur Verhinderung des Zugriffs durch dritte Organisationen?

Maßnahmen

1. **KI Literacy:** Förderung eines Grundverständnisses von KI-Anwendungen, deren Voraussetzungen und Anwendungen sowie deren Auswirkung auf Verwaltung, Staat und Gesellschaft als Grundbedingung für alle anderen Themen.
 - Durch welche Maßnahmen wird die KI-Kompetenz der Verwaltungsbediensteten gefördert, um ein Grundverständnis für KI-Technologien, ihre Voraussetzungen und Anwendungen sowie ihre Auswirkungen auf Verwaltung, Staat und Gesellschaft zu gewährleisten?
 - Wie wird die KI-Kompetenz der breiten Öffentlichkeit, insbesondere im Hinblick auf die Notwendigkeit eines grundsätzlichen Verständnisses für eine qualifizierte Diskussion um den Einsatz von KI-Anwendungen durch die Verwaltung, gefördert?
 - Gibt es spezielle Initiativen, um die KI-Kompetenz von unterrepräsentierten Bevölkerungsgruppen zu erhöhen?
 - Wie wird sichergestellt, dass die KI-Kenntnisse jeweils aktualisiert werden, wenn sich die Technologie weiterentwickelt?
2. **Kontinuierliche Bewertung/Folgenabschätzung und Verbesserung:** Durchführung einer Folgenabschätzung vor dem Einsatz einer KI-Anwendung, um negativen gesellschaftlichen Auswirkungen vorzubeugen; Kontinuierliche Evaluation insbesondere risikobehafteter KI-Anwendungen, um sicherzustellen, dass ihre Leistung und Wirkung weiterhin mit dem öffentlichen Wohl in Einklang stehen.
 - Welche Maßnahmen werden ergriffen, um die Auswirkungen von KI-Technologien kontinuierlich, also im Planungsstadium, bei der Entwicklung und beim Einsatz, zu bewerten und zu beurteilen und ihre Leistung und Wirkung zu verbessern?
 - Welche Kriterien werden verwendet, um zu bestimmen, welche KI-Anwendungen als besonders risikoreich gelten und eine kontinuierliche Bewertung und Beurteilung erfordern?
 - Wer ist für die Durchführung der Folgenabschätzung verantwortlich?

- Welche Ressourcen und Fachkenntnisse sind für die Durchführung einer gründlichen Folgenabschätzung erforderlich, und wie können sie beschafft werden?
 - Wie werden die Ergebnisse der Folgenabschätzung der Öffentlichkeit und den Entscheidungsträger:innen wirksam mitgeteilt?
3. **Zertifizierung:** Zertifizierung von KI-Modellen inklusive gegebenenfalls verwendeter Daten (insbesondere im Fall risikobehafteter Anwendungen).
- Von wem werden KI-Modelle und -Daten zertifiziert, um sicherzustellen, dass sie ethisch und rechtlich einwandfrei sind, insbesondere bei riskanten Anwendungen?
 - Wie sieht das Verfahren für die Zertifizierung eines in der öffentlichen Verwaltung eingesetzten KI-Modells aus?
 - Wer ist für die Überwachung des Zertifizierungsprozesses für KI-Modelle, die in der öffentlichen Verwaltung eingesetzt werden, zuständig?
 - Wie oft werden zertifizierte KI-Modelle neu bewertet, um sicherzustellen, dass sie weiterhin mit ethischen und rechtlichen Standards übereinstimmen?
4. **Aufsicht:** Einrichtung unabhängiger Aufsichtsgremien, die den Einsatz und die Nutzung von KI überwachen und sicherstellen, sodass KI-Anwendungen im Einklang mit ethischen und rechtlichen Standards eingesetzt werden.
- Welche Rechte und Pflichten haben die unabhängigen Aufsichtsgremien?
 - Welche Monitoringmechanismen werden eingerichtet, um den ethischen und rechtmäßigen Einsatz und die Nutzung von KI-Technologien zu überwachen und sicherzustellen?
 - Über welche Qualifikationen und Fachkenntnisse verfügen die Mitglieder der Aufsichtsgremien?

Glossar für Fachbegriffe

AI Act: auch KI-Verordnung genannt. Ein EU-Gesetzesvorhaben, das KI-Anwendungen regelt und einem risiko-basiertem Ansatz folgt (siehe Abschnitt 8.3)

AIM AT 2030: Artificial Intelligence Mission Austria, nationale KI-Strategie Österreichs (siehe Abschnitt 8.5)

Bias: ein verzerrter Output, z. B. von maschinenlernenden Algorithmen, in anderen Kontexten hat Bias andere Bedeutungen (siehe Abschnitt 5, 6.1)

Big Data: große Datenmengen, die aus unterschiedlichen Quellen gesammelt und auch mit KI bearbeitet werden können (siehe Abschnitt 4)

Chatbot: ein Computerprogramm, das Fragen z. B. von Bürger:innen im Chatformat automatisiert zu beantworten versucht (siehe Abschnitt 4, 5, 6.2, 6.4)

„Ethics by Design“: methodischer, oft interdisziplinärer Zugang bei der Entwicklung von KI, bei dem ethische Reflexion in allen Stadien berücksichtigt wird (siehe Abschnitt 10.3)

Fairness: keine allgemein akzeptierte Definition im Kontext von KI. Grundsätzlich als Abwesenheit von Vorurteilen, Bias oder Präferenz für ein Individuum oder eine Gruppe zu verstehen (siehe Abschnitt 10.1)

Governance: viele mögliche Definitionen, hier: durch verschiedene Politikinstrumente (z. B. hard bzw. soft law) herbeigeführter Versuch des Staates gemeinsam mit anderen Akteursgruppen aus Wirtschaft und Zivilgesellschaft gesellschaftliche Problemstellungen zu lösen (siehe Abschnitt 9.1, 11)

Hard Law: rechtlich bindende Normen wie Gesetze, Verordnungen und Verträge (siehe Abschnitt 8)

KI-Literacy: Grundlegendes Wissen und Kompetenzen zu KI, um in Mensch-KI-Interaktionen selbstbestimmt handeln zu können (siehe Abschnitt 5)

Neuronale Netze: auch künstliche neuronale Netze (KNN) oder simulierte neuronale Netze (SNN) genannt, sind Teil des maschinellen Lernens (ML), bilden das Kernstück von Deep-Learning-Algorithmen und dienen dazu, Informationen zu verarbeiten und komplexe Muster zu erkennen (siehe Abschnitt 4)

Soft Law: rechtlich nicht bindende Instrumente, beispielsweise Leitfäden, Leitlinien, Strategien und Absichtserklärungen (siehe Abschnitt 9.1)

Abbildungsverzeichnis

Abbildung 1: Vergleich der Treibhausgasemissionen einer Flugreise, des Lebens eines Menschen in den USA im Jahresdurchschnitt, eines Autos während der Lebensdauer und der Erstellung eines KI-Modells.....	31
Abbildung 2: Entscheidungsbaum zur Verwendung daten-getriebener KI-Technologie.....	38
Abbildung 3: Risikopyramide AI Act – Bildquelle: Europäisches Parlament.....	43

Tabellenverzeichnis

Tabelle 1: Workshop-Zeitverlauf.....	12
Tabelle 2: Empfehlungen KI-Governance	62
Tabelle 3: Checkliste für eine ethische KI.....	76

