

Leitfaden Digitale Verwaltung: KI, Ethik und Recht

Praxisleitfaden für die Verwaltung, Version 2.0



AI
ETHICS

- + Checkliste und Entscheidungsbaum NEU
- + generative künstliche Intelligenz
- + Quantum AI
- + KI in der Verwaltung ja / nein
- + Extra AI-Act – Berücksichtigung der KI-Verordnung
- + KI Übersicht (Governance-Struktur)

Leitfaden Digitale Verwaltung: KI, Ethik und Recht

Praxisleitfaden für die Verwaltung, Version 2.0

Wien, 2026



Impressum

Medieninhaber, Verleger und Herausgeber:

Bundeskanzleramt Österreich, Ballhausplatz 2, 1010 Wien
Sektion III – Öffentlicher Dienst und Verwaltungsinnovation

Version 1.0:

Autor und Autorinnen PD Mag. Dr. Peter Biegelbauer, S.M. (Projektleitung wissenschaftliches Team), Caroline Lackinger, BA, Dr. Sven Schlarb, Edgar Subak, BA, Pia Weinlinger, MA
Projektteam BKA: Mag.^a Ursula Rosenbichler (Leitung), Michael Huber, LL.B, MSc.,
Ralf M. Tatto, BA MA MA

Version 2.0:

Autor und Autorinnen: PD Mag. Dr. Peter Biegelbauer, S.M. (Projektleitung wissenschaftliches Team), Dr. Sven Schlarb, Pia Weinlinger, MA, Mag. Rodrigo Conde-Jimenez, BA, Dr.^a Heidi Scheichenbauer, Dr.^a Madeleine Müller, BA MU, DDr. Krisztina Rozgonyi, MBA, Dr. David M. Schneeberger, BA BA MA

Projektteam BKA: Ralf M. Tatto, BA MA MA (Leitung), Walter Winter, MA

Fotonachweis: Adobe Stock (Cover), BKA/Olivia Klonfar (S.3)

Layout: BKA Corporate Identity & Kommunikationsdesign

Copyright und Haftung:

Auszugsweiser Abdruck ist nur mit Quellenangabe gestattet, alle sonstigen Rechte sind vorbehalten. Es wird darauf verwiesen, dass alle Angaben in dieser Publikation trotz sorgfältiger Bearbeitung ohne Gewähr erfolgen und eine Haftung des Bundesministeriums für Kunst, Kultur, öffentlichen Dienst und Sport und der Autorinnen und Autoren ausgeschlossen ist. Rechtsausführungen stellen die unverbindliche Meinung der Autorinnen und Autoren dar und können der Rechtsprechung der unabhängigen Gerichte keinesfalls vorgreifen.

Kontakt und Rückmeldungen:

Bundeskanzleramt Österreich, Ballhausplatz 2, 1010 Wien
Sektion III – Öffentlicher Dienst und Verwaltungsinnovation

Ralf M. Tatto, BA MA MA, ralf.tatto@bka.gv.at bzw. digitaleinnovation@bka.gv.at

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

Hinweis zu Internetverweisen: Die im Leitfaden angeführten Links wurden zum Zeitpunkt der Erstveröffentlichung (**Dezember 2024**) geprüft und waren erreichbar. Aufgrund nachträglicher Änderungen externer Websites kann die dauerhafte Verfügbarkeit einzelner Links nicht gewährleistet werden.

ISBN 978-3-9505412-7-4

Wien, 2026

Ausgezeichnet mit dem 3. Platz beim „Digitaler Humanismus in der Praxis“-Award 2026.

Vorwort des Herrn Sektionsleiters

Liebe Mitarbeiterinnen und Mitarbeiter,
liebe Leserinnen und Leser,

Die rasante Digitalisierung verändert nicht nur die Art und Weise, wie wir kommunizieren, arbeiten und leben, sondern auch die Anforderungen an die öffentliche Verwaltung. Sie bietet uns Chancen, die Verwaltung effizienter und bürger- und bürgerinnennäher zu gestalten, stellt uns aber gleichzeitig vor große technologische, ethische und rechtliche Herausforderungen. In diesem Spannungsfeld aus Innovation und Verantwortung spielt die öffentliche Verwaltung eine zentrale Rolle.

Die nunmehr zweite Version des *Leitfadens Digitale Verwaltung* ist ein weiterer Schritt, um unsere Verwaltung fit für das digitale Zeitalter zu machen. Er baut auf den Erfolgen und Erfahrungen der ersten Version auf, geht aber noch einen Schritt weiter: Er nimmt aktuelle technologische Entwicklungen wie **generative Künstliche Intelligenz** und **Quantum AI** ebenso auf wie die regulatorischen Anforderungen der europäischen KI-Verordnung (**AI Act**). Gleichzeitig bleibt er ein praxisnahes Werkzeug, das Orientierung bietet und den Fokus auf einen menschenzentrierten Ansatz legt.

Unser Ziel ist klar: Die Digitalisierung der Verwaltung muss ethisch fundiert, rechtlich abgesichert und gleichzeitig innovativ sein. Vertrauen, Transparenz und der Schutz der Grundrechte sind dabei keine Option, sondern die Basis unseres Handelns.

Dieser Leitfaden ist nicht nur ein Dokument, sondern ein **lebendiges Instrument**. Er ist das Ergebnis eines breiten Dialogs zwischen Wissenschaft, Praxis und Verwaltungspartnerinnen und -partner. Auch in Zukunft werden wir ihn weiterentwickeln, um die rasanten technologischen Veränderungen kontinuierlich zu berücksichtigen.

Abschließend möchte ich mich bei allen Beteiligten bedanken, die mit ihrer Expertise und ihrem Engagement zur Weiterentwicklung dieses Leitfadens beigetragen haben. Gemeinsam schaffen wir eine öffentliche Verwaltung, die die Chancen der Digitalisierung nutzt und dabei stets den Menschen in den Mittelpunkt stellt.

Mag. Andreas Buchta-Kadanka

Leiter der Sektion III – Öffentlicher Dienst und Verwaltungsinnovation



Sektionsleiter
Mag. Andreas Buchta-
Kadanka

Dieses Vorwort ist ausnahmslos der einzige Inhalt in dieser Publikation, bei dem mit KI-Unterstützung (ChatGPT4o) gearbeitet wurde (Zusammenfassung und Vergleich der bisherigen Leitfäden sowie Entwurfsvorschlag zur Inspiration), die Ergebnisse wurden von einer Person mit Fachexpertise verifiziert und zur Anwendung freigegeben. Bei Fragen zu diesem Inhalt wenden Sie sich bitte an den im Impressum angeführten Kontakt.

Inhalt

1 Einleitung	7
1.1 Praxisleitfaden Ziele und Zielgruppen.....	8
2 Checkliste für ethische KI in der öffentlichen Verwaltung	12
3 Methoden und Vorgehen	20
4 Technik: Was ist KI?	22
4.1 Generative Künstliche Intelligenz.....	27
4.2 Quantum AI.....	30
5 Handlungsbedarf	32
6 Abwägung: KI in der Verwaltung, oder nicht?	38
7 Chancen und Herausforderungen beim Einsatz von KI in der öffentlichen Verwaltung	44
7.1 KI und Auswirkungen auf die Arbeitswelt der öffentlichen Verwaltung.....	44
7.2 KI in der öffentlichen Verwaltung und Auswirkungen auf die Bevölkerung.....	53
7.3 KI und Ökologie.....	58
7.4 Digitale Souveränität in der öffentlichen Verwaltung.....	64
7.5 Beschaffung.....	65
8 Rechtlicher Rahmen	71
8.1 Die Grundlage des Verwaltungshandelns.....	71
8.2 Datenschutzgrundverordnung.....	73
8.3 Die EU regelt KI: der AI Act.....	78
8.4 Weitere EU-Regulierungen.....	89
8.5 Artificial Intelligence Mission Austria 2030.....	90
9 Ethische KI: Prinzipien und Leitlinien	92
9.1 Ethische Leitlinien: Governance durch „Soft Law“.....	92
9.2 Ethische Leitlinien für die österreichische Verwaltung.....	94
10 KI-Folgenabschätzung	98
10.1 Grundrechte-Folgenabschätzung im AI Act.....	98

10.2 Kriterien- und Maßnahmenkatalog für ethische KI in der Verwaltung EKIV.....	102
10.3 EU-Bewertungsliste für vertrauenswürdige künstliche Intelligenz (ALTAI).....	107
10.4 VCIO-Modell.....	107
10.5 Folgenabschätzung für Grundrechte und Algorithmen (FRAIA).....	108
10.6 Data Ethics Decision Aid (DEDA).....	108
10.7 Weitere KI-Folgenabschätzungsinstrumente.....	109
11 KI Governance-Struktur.....	111
11.1 EU-Ebene.....	111
11.2 Nationale Ebene.....	112
12 Empfehlungen für mögliche weitere Schritte:	
Ziel menschenzentrierte KI-Governance.....	114
Quellenverzeichnis.....	123
Glossar für Fachbegriffe.....	133
Anhang.....	135
Abbildungsverzeichnis.....	144
Tabellenverzeichnis.....	144

1 Einleitung

Das Ziel dieses Leitfadens ist es, Verwaltungsbedienstete in Planung, Design, Erstellung bzw. Vergabe, Einsatz und Evaluierung digitaler, insbesondere aber Künstlicher Intelligenz (KI)-basierter Anwendungen zu unterstützen. Für die ethischen Überlegungen hierbei ist die Komplexität des eingesetzten Systems, also z. B. ob es als 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 Bürgerinnen und Bürger 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 nach wie vor rasant ausbreitet.

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

Die Anwendung des Praxisleitfadens ist freiwillig und nicht als verbindliche Vorschrift gedacht. Es ist jedoch möglich, dass er in Zukunft zu einer verbindlichen Richtlinie weiterentwickelt wird.

Der hier vorliegende Leitfaden versteht sich als Version 2.0. Die Version 1.0 wurde im Oktober 2023 herausgegeben (BMKÖS 2023). Mit der sich weiterentwickelnden Technik und der delegierten Gesetzgebung zum AI Act, der Tätigkeit der europäischen Institutionen und insbesondere des neu eingerichteten AI Office, also dem Europäischen Amt für künstliche Intelligenz wird in Zukunft eine weitere Überarbeitung notwendig sein. In diesem Sinne soll es sich also um ein lebendiges Dokument handeln. Vor diesem Hintergrund nehmen wir gerne unter folgenden E-Mail-Adressen Anregungen entgegen: digitaleinnovation@bka.gv.at und / oder office.isp@ait.ac.at. Viele Inhalte dieses Leitfadens basieren auf Rückmeldungen von Leserinnen und Leser der vorangegangenen Leitfadenversion bzw. der Teilnehmerinnen und Teilnehmer der dazugehörigen Weiterbildungsschiene auf der Verwaltungsakademie des Bundes. Vielen Dank dafür!

1.1 Praxisleitfaden Ziele und Zielgruppen

Die Digitalisierung verändert die Gesellschaft rasant. Mehr als drei Viertel der Österreicherinnen und Österreicher 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). Gleichzeitig geben 44% der Befragten einer Studie der Universität Salzburg an, KI im Alltag eher kritisch zu sehen, lediglich 16% reagierten euphorisch (RTR, 2024).

Ein großer Teil der Bewerberinnen und Bewerber 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 im November 2022 als erste generative KI, die der breiten Öffentlichkeit weitgehend ohne Zutrittschürden zugänglich ist, entstehen täglich tausende Anwendungen, die einer weltweiten Nutzung offenstehen.

Der Einsatz von KI eröffnet also umfangreiche Möglichkeiten, birgt aber ebenso Gefahren. Die in vielen beruflichen wie auch privaten Bereichen eingesetzten Anwendungen, beispielsweise zur Spracherkennung, Übersetzung oder Optimierung von Mobilitätsbedürfnissen, aber auch zur automatisierten Gesichtserkennung oder zur Analyse der Produktivität von Mitarbeiterinnen und Mitarbeiter (wie etwa durch Microsoft „Workplace Analytics“, Christl 2021) beruhen auf der (gegenwarts- oder zukunftsbezogenen) Analyse personenbezogener Daten. Viele Anwendungen versprechen Effizienzsteigerung und Arbeitserleichterung. Es sind jedoch auch verschiedene Problem-bereiche 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 etwaige 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 nach wie vor eine wichtige Rolle im gesellschaftlichen Diskurs über den öffentlichen Dienst. Dementsprechend findet sich hier ein Betätigungs-

feld für die Anwendung von digitalen Lösungen und insbesondere KI mit dem Ziel der Steigerung von Effizienz und der Entlastung von Mitarbeiterinnen und Mitarbeiter 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, Leitl-Staudinger et al. 2023).

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, Bias und (KI) Ethik besonders brisant. Es geht es hier um Kernbereiche staatlicher Aufgaben, die unter den zuvor angesprochenen besonderen rechtsstaatlichen und Demokratie unterstützenden Bedingungen erfüllt werden müssen.

Ethik und Recht sind dabei nicht immer deckungsgleich, denn ethisches Handeln verlangt in manchen Situationen mehr als die Einhaltung geltenden Rechts – dies gilt auch für einen demokratisch organisierten Rechtsstaat. Klug eingesetzt, können Ethik und Recht aber gemeinsam einen krisenfesten Rahmen für das Handeln der öffentlichen Verwaltung bilden.

Eine mit der EU (KI-)Ethikdiskussion übereinstimmende Grundhaltung erleichtert Entscheidungen auf allen Ebenen der öffentlichen Verwaltung durch die Herstellung eines einheitlichen Entscheidungsrahmens, in dem sich alle Verwaltungsangehörigen bewegen können. Ein entscheidender Bestandteil dieser ethischen Grundhaltung ist die Betonung der Menschenzentriertheit digitaler Technologien. Dies bedeutet, dass digitale Technologien, insbesondere KI, so gestaltet und eingesetzt werden sollen, dass sie den Menschen in den Mittelpunkt stellen und dessen Rechte und Würde wahren.

Auch die „*European Declaration of Digital Rights and Principles*“ aus dem Jahr 2022 formuliert eine Vision der EU für eine digitale Transformation, die den Menschen in den Vordergrund rückt und auf den Grundwerten und Rechten der EU basiert. Diese Erklärung definiert sechs Kernprinzipien, die die digitale Zukunft der EU prägen sollen (European Commission 2022a):

1. Den Menschen und seine Rechte in den Mittelpunkt der digitalen Transformation stellen,
2. Solidarität und Inklusion fördern,
3. Wahlfreiheit im digitalen Raum sicherstellen,
4. Teilhabe am digitalen öffentlichen Raum stärken,
5. Sicherheit, Schutz und Befähigung von Individuen – insbesondere von jungen Menschen – erhöhen,
6. Nachhaltigkeit in der digitalen Entwicklung fördern.

Mit der „*European Declaration of Digital Rights and Principles*“ verpflichten sich die EU und ihre Mitgliedstaaten, diese menschenzentrierte Vision digitaler Transformation zu unterstützen und zu fördern.

Daraus leitet sich auch das bereits in den ersten Zeilen umrissene primäre Ziel dieses Leitfadens ab: die Unterstützung von Verwaltungsbediensteten in Planung, Design, Entwicklung bzw. Vergabe, Einsatz und Evaluierung digitaler, insbesondere aber KI basierter Anwendungen. Dieses Ziel ist allerdings nur durch übereinstimmendes Handeln aller Verwaltungsangehörigen zu erreichen. Anwenderinnen und Anwender, Entwicklerinnen und Entwickler und Management in der öffentlichen Verwaltung müssen gleichermaßen grundlegende technische Prinzipien, Chancen und Herausforderungen, ethische Prinzipien und Standards sowie Möglichkeiten der Folgenabschätzung des Einsatzes von KI im öffentlichen Dienst verstehen.

In diesem Zusammenhang ist insbesondere der Schutz von Grundrechten von besonderer Bedeutung, denn auch 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 (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), dem AI Act (im deutschen Sprachraum auch KI-Verordnung (EU 2024)), 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, 2020; Deutscher Ethikrat 2023), Großbritannien (UK Parliament 2021, 2024) und den USA (The White House 2022, 2023) zur Regulierung von KI werden immer wieder ähnliche Prinzipien diskutiert: menschliche Letztentscheidungsverantwortung, (Daten-)Sicherheit, Einhaltung von Privatsphäre, Transparenz, Diversität, 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.

Diese Zielsetzungen teilt sich der vorliegende Praxisleitfaden mit dem Bericht „KI in der Bundesverwaltung“ des österreichischen Rechnungshofs, dessen Erscheinen für 2025 angekündigt wurde. Darin geht es um eine Bestandsaufnahme gegenwärtiger Bemühungen, einen Überblick über Vorgaben, Koordinationsmaßnahmen und Rahmenbedingungen sowie die strategische Ausrichtung des Einsatzes von KI in der Bundesverwaltung. Mit dem Ziel, einen robusten Rahmen für die Implementierung von KI in der öffentlichen Verwaltung zu schaffen, wird in diesem Praxisleitfaden ein spezielles Augenmerk auf ethische Prinzipien und regulatorische Diskussionen gelegt. Die Struktur des Leitfadens ist dabei wie folgt: zuerst wird der Hintergrund des Projektes, das zu diesem Leitfaden geführt hat, umrissen. Danach folgt, bevor auf die Inhalte des Leitfadens selbst eingegangen wird, die Checkliste für ethische KI in der Verwaltung. Die Checkliste ist ein zentrales Element des Leitfadens und dient als Orientierungs- und Ausgangspunkt für die Berücksichtigung ethischer und rechtlicher Aspekte bei der Beschaffung, Entwicklung und Implementierung von KI-Anwendungen. Anschließend wird der Begriff KI erklärt und der Handlungsbedarf in Bezug auf deren Einsatz in der öffentlichen Verwaltung umrissen. In einem weiteren Schritt geht es um die Abwägung, ob KI in einem gegebenen Fall in der Verwaltung eingesetzt werden soll oder nicht. Danach werden Chancen und Herausforderungen beim Einsatz von KI in der Verwaltung im Hinblick auf Gesellschaft, Ökologie und digitale Souveränität thematisiert. 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 Checkliste für ethische KI in der öffentlichen Verwaltung

Da die Ziele des Praxisleitfadens sowie die Zielgruppen, an die er sich richtet, nun bekannt sind, möchten wir zunächst die Checkliste für ethische KI in der öffentlichen Verwaltung vorstellen, bevor wir uns den einzelnen Abschnitten des Leitfadens im Detail widmen.

Ein zentraler Grundsatz im Umgang mit KI lautet, dass KI-Systeme ethisch und verantwortungsbewusst eingesetzt werden müssen. In der Praxis bedeutet dies, dass sowohl technische als auch organisatorische Maßnahmen erforderlich sind, um sicherzustellen, dass ethische und rechtliche Aspekte – wie der Schutz der Privatsphäre – gewahrt bleiben. Dabei steht die zuvor erwähnte Menschenzentriertheit im Vordergrund. Ben Shneiderman, ein führender Experte auf dem Gebiet der Mensch-Computer-Interaktion, definiert menschenzentrierte KI als ein Paradigma, das darauf abzielt, menschliche Fähigkeiten zu verstärken und zu verbessern, indem KI-Systeme entwickelt und eingesetzt werden, die zuverlässig, sicher und vertrauenswürdig sind. Diese Systeme sollen die Selbstwirksamkeit des Menschen unterstützen, Kreativität fördern, Verantwortlichkeiten festlegen und die soziale Teilhabe ermöglichen (Shneiderman 2020). Das Ziel von menschenzentrierter KI ist es also, KI so zu gestalten, dass sie den Menschen unterstützt, ohne ihn zu ersetzen. Auch der AI Act verfolgt eine derartige Perspektive bei der Regulierung von KI, indem klare Anforderungen an die Entwicklung und Nutzung von KI-Systemen in dem Gesetzestext formuliert wurden. Zu den Inhalten des AI Acts gehört u. a. die Förderung der Einführung vertrauenswürdiger und menschenorientierter KI-Anwendungen, die die Grundrechte schützen, Transparenz gewährleisten und ethische Werte wahren (siehe Abschnitt 8.3 AI Act).

In diesem Zusammenhang soll die Checkliste als Leitfaden für die Umsetzung ethischer und menschenzentrierter KI in der öffentlichen Verwaltung dienen. Sie bietet eine strukturierte Unterstützung, um sicherzustellen, dass die relevanten ethischen und rechtlichen Aspekte bei der Entwicklung, Beschaffung und Nutzung von KI-Systemen berücksichtigt wurden. Durch die Verwendung der Checkliste soll erreicht werden, dass ethische Prinzipien wie der Schutz der Privatsphäre, die Vermeidung von Diskriminierung und die Gewährleistung von Transparenz und Rechenschaftspflicht über den gesamten Lebenszyklus von KI-Anwendungen hinweg integriert werden.

Die Kriterien der Checkliste beruhen dabei auf jenen ethischen Prinzipien, die in der internationalen Debatte, insbesondere in der EU, diskutiert werden. Zudem erweitert die Checkliste die Ethik-Leitlinien der High Level Expert Group (HLEG 2018, 2019) der Europäischen Kommission für vertrauenswürdige KI („*trustworthy AI*“), indem sie sich auf ethische und rechtliche Herausforderungen stützt, die spezifisch für den öffentlichen Sektor von Bedeutung sind. Diese Erweiterungen wurden in Zusammenarbeit mit Verwaltungsbediensteten in gemeinsamen Workshops überprüft. Dabei flossen praxis-

relevante Aspekte ein, die von den Teilnehmenden als besonders entscheidend für den erfolgreichen und verantwortungsvollen Einsatz von KI in der öffentlichen Verwaltung bewertet wurden. Zudem wurde Recht als Kriterium in die Checkliste aufgenommen. Abbildung 1 ist eine Darstellung der 9 Prinzipien für ethisch und rechtlich vertretbare KI in der öffentlichen Verwaltung. Eine genaue Definition für die einzelnen Kriterien ist in Abschnitt 9.2 zu finden.



Abbildung 1: KI-Ethikprinzipien für die österreichische Verwaltung

Da sich sowohl regulatorische Rahmenbedingungen als auch technologische Entwicklungen im Bereich von KI stetig weiterentwickeln, werden die in der Checkliste angeführten Fragen im Zuge künftiger Aktualisierungen des Praxisleitfadens ebenfalls regelmäßig überprüft und angepasst. Dies bedeutet, dass sowohl die Checkliste als auch der Praxisleitfaden insgesamt als „living documents“ zu betrachten sind, die sich fortlaufend an neue Erkenntnisse und gesetzliche Rahmenbedingungen anpassen werden. Die Checkliste soll ebenfalls nicht als erschöpfende Aufzählung verstanden werden, sondern vielmehr als Orientierungspunkt für weiterführende Überlegungen und Handlungsansätze.

Anwendung der Checkliste

Die Checkliste dient dazu, spezifische rechtliche und ethische Anforderungen für den Einsatz eines KI-Systems zu überprüfen. Bevor jedoch die Checkliste zum Einsatz kommt, sollte eine grundlegende Prüfung vorgenommen werden, ob der Einsatz eines KI-Systems überhaupt sinnvoll und gerechtfertigt ist. Zu diesem Zweck gibt es den Entscheidungsbaum (siehe Abschnitt 6, Abbildung 5), der dabei hilft, die Eignung und die übergeordneten Rahmenbedingungen für den KI-Einsatz vorab zu klären. Der Entscheidungsbaum soll sicherstellen, dass der Einsatz nur dann weiterverfolgt wird, wenn zentrale Fragen zur Zweckmäßigkeit, Wirtschaftlichkeit und zu notwendigen Anwendungsbedingungen positiv beantwortet werden konnten. Liegt ein passender Anwendungsfall für den Einsatz von KI vor, empfiehlt es sich, die Checkliste für ethische KI in der Verwaltung zu nutzen.

Die Anwendung der Checkliste sollte idealerweise durch ein interdisziplinäres Team erfolgen. Dieses Team sollte Expertinnen und Experten aus verschiedenen Bereichen umfassen, darunter Recht, Ethik, Datenmanagement, IT sowie Domänenwissen aus dem spezifischen Fachbereich, der von der KI-Anwendung betroffen ist.

Primär richtet sich die Checkliste an:

- Verantwortliche und Fachpersonen im Bereich Beschaffung,
- Mitarbeitende, die das KI-System nutzen oder damit arbeiten werden,
- Rechts- und Compliance-Beauftragte,
- Führungskräfte und Entscheidungsträgerinnen und Entscheidungsträger.

Einige Fragen der Checkliste enthalten Verweise auf weiterführende Abschnitte im Praxisleitfaden, die Hintergrundinformationen und Detailwissen bieten. Diese Verweise sollen als Hilfestellung dienen, um die Fragen der Checkliste fundiert beantworten zu können. Insbesondere bei komplexeren Fragestellungen, die möglicherweise eine genauere Klärung des Kontextes erfordern, bieten die weiterführenden Abschnitte Erläuterungen.

Wenn die Checkliste vollständig ausgefüllt ist und alle Fragen mit „Ja“ beantwortet wurden, bedeutet dies, dass die KI-Anwendung die aktuellen ethischen und rechtlichen Anforderungen weitgehend beachtet. Es ist jedoch wichtig zu betonen, dass die Checkliste keine Kontrolle der Einhaltung der rechtlichen Vorschriften im Sinne des AI

Acts ersetzt. Die entsprechenden Pflichten für Anbieter und Betreiber für KI-Systemen werden dabei näher im Abschnitt 8.3 zum AI Act thematisiert.

Sollten einige Fragen der Checkliste mit „Nein“ beantwortet werden, deutet dies auf bestehende Defizite hin, die angesprochen werden müssen. In solchen Fällen sollten Maßnahmen ergriffen werden, um die entsprechenden Anforderungen zu erfüllen. Dies könnte etwa die technische Anpassung des KI-Systems oder organisatorische Änderungen umfassen. Zum Beispiel könnten technische Anpassungen notwendig sein, wenn das KI-System bestimmte Anforderungen, wie den Schutz der Privatsphäre oder die Vermeidung von Diskriminierung nicht erfüllt. Organisatorische Änderungen könnten bedeuten, dass interne Abläufe oder Verantwortlichkeiten überdacht werden müssen, um eine bessere Kontrolle und Überwachung der KI-Anwendungen zu gewährleisten. Für eine detaillierte Bewertung der einzelnen Fragen empfiehlt es sich außerdem, den Kriterien- und Maßnahmenkatalog (EKIV) in Abschnitt 10.2 des Leitfadens heranzuziehen.

Manchmal wird die Beantwortung einer Frage mit „Nein“ nicht vermeidbar sein, beispielsweise bei der Verwendung großer Sprachmodelle wie ChatGPT, die große Teile des frei zugänglichen Internets als Trainingsdaten verwenden. Das bedeutet insbesondere, dass dann ausdrückliche Vorsicht bei der Verwendung des jeweiligen KI-Systems angebracht ist – vor allem für die öffentliche Verwaltung.

Ein exemplarischer Anwendungsfall im Anhang des Leitfadens zeigt, wie die Fragen der Checkliste in der Praxis angesprochen und bewertet werden können.

Checkliste

Recht

Wird das KI-System im Rahmen des (schlicht) hoheitlichen Verwaltungshandelns eingesetzt? (Siehe 8.1 Die Grundlage des Verwaltungshandelns) Ja Nein

Sofern (schlicht) hoheitliches Verwaltungshandeln vorliegt, wurde abgeklärt, ob eine ausreichende Rechtsgrundlage für den Einsatz der KI besteht? Ja Nein

Verarbeitet die KI-Anwendung Daten im Einklang mit den Anforderungen der Rechtsnormen und -prinzipien, die im nationalen und EU-Rechtsrahmen festgelegt sind? (Siehe 8.2 Datenschutzgrundverordnung) Ja Nein

Gewährleistet der Einsatz des KI-Systems, dass die Grundrechte von Bürgerinnen und Bürger in keiner Weise beeinträchtigt werden? (Siehe 5 Anwendungsfall: Beispiele aus Australien und den Niederlanden) Ja Nein

Wurde die Risikoeinstufung der KI-Anwendung gemäß dem AI Act ermittelt? (Siehe 8.3 Die EU regelt KI: der AI Act) Ja Nein

Falls zutreffend: Ist ein Conformity Assessment laut dem AI Act bereits erfolgt? (Einschätzung als High-risk Anwendung) (Siehe 8.3 Die EU regelt KI: der AI Act, Konformitätsbewertungsverfahren) Ja Nein

Falls zutreffend: Wurden die Transparenzpflichten nach dem AI Act erfüllt? (Einschätzung als Medium-risk Anwendung) (Siehe 8.3 Die EU regelt KI: der AI Act, Konformitätsbewertungsverfahren, Transparenzanforderungen) Ja Nein

Wurde geprüft, ob eine Grundrechte-Folgenabschätzung (AI Act) durchgeführt werden muss? (Siehe 10.1 Grundrechte-Folgenabschätzung im AI Act) Ja Nein

Wurde geprüft, ob eine Datenschutz-Folgeabschätzung durchgeführt werden muss? (Siehe 8.2 Datenschutzgrundverordnung) Ja Nein

Transparenz

Sind die spezifischen Ziele und Zwecke des Einsatzes der KI-Anwendung identifiziert und dokumentiert? Ja Nein

Gibt es eine Dokumentation, die die technische Entwicklung des Modells erläutert? Ja Nein

Ist die Funktionsweise der KI-Anwendung nachvollziehbar? Ja Nein

Sind die Datensätze, die mit dem KI-System verbunden sind, bekannt? (Siehe 8.2 Datenschutzgrundverordnung) Ja Nein

Wird den Nutzerinnen und Nutzer, wann immer möglich, erklärt, wie das KI-System zu seinen Ausgaben, Inhalten, Empfehlungen oder Ergebnissen kommt und welche Logik dahintersteckt? Ja Nein

Werden Personen informiert, wann und auf welche Weise sie mit einer KI-Anwendung interagieren? Ja Nein

Unvoreingenommenheit und Fairness

Sind die Daten, die zum Training des KI-Systems verwendet werden, vielfältig und repräsentativ für den jeweiligen Kontext? (Siehe 7.1 Wissen: Bias; Anwendung: Geschlechterbias) Ja Nein

Gibt es einen Prozess, um verwendete Datenquellen auf mögliche Verzerrungen und Ungenauigkeiten zu prüfen? Ja Nein

Ist die KI-Anwendung so konzipiert, dass sie die Entmenschlichung, Diskriminierung, Stereotypisierung oder Manipulation von Menschen vermeidet? (Siehe 5 Anwendungsfall: Chatbot) Ja Nein

Gibt es ein Verfahren, mit dem Personen gegen den Einsatz bzw. die Ausgabe des KI-Systems Einspruch oder sonstige Rechtsmittel dagegen erheben können? Ja Nein

Effektivität und Effizienz

Ergeben sich konkrete Vorteile für die breite Öffentlichkeit durch den Einsatz dieses KI-Systems? (z. B. Zeitersparnis bei der Beantragung einer staatlichen Leistung) Ja Nein

Hat die KI-Anwendung das Potenzial die Arbeitssituation, der im öffentlichen Dienst tätigen Personen zu verbessern oder zumindest nicht zu verschlechtern? (Siehe 7.1 KI und Auswirkungen auf die Arbeitswelt der öffentlichen Verwaltung) Ja Nein

Gibt es eine Richtlinie zu Qualitäts- und Leistungszielen für das KI-System? Ja Nein

Werden die Verwaltungsbediensteten entsprechend geschult und unterstützt, um die KI-Anwendung wirkungsvoll einzusetzen? (Siehe 12 Kompetenzaufbau und Fortbildung) Ja Nein

Gibt es fortlaufende Unterstützung bei Problemen oder Herausforderungen? Ja Nein

Wurden die Umweltauswirkungen der KI-Anwendung berücksichtigt? (Siehe 7.3 KI und Ökologie) Ja Nein

Sicherheit

Falls zutreffend: Wurde ein Risikomanagementsystem für die KI-Anwendung geschaffen? (Hohes Risiko) (Siehe 8.3 Die EU regelt KI: der AI Act) Ja Nein

Werden Aufzeichnungen über die Betriebsleistung der KI-Anwendung und alle Vorfälle oder Störungen für einen bestimmten Zeitraum archiviert? Ja Nein

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

Zugänglichkeit und Inklusion

Ist die KI-Anwendung menschenzentriert konzipiert? Also so, dass sie von verschiedenen Endnutzerinnen und Endnutzer mit unterschiedlichen Kompetenzniveaus verwendet werden kann? Ja Nein

Wurde überprüft, ob Alternativen zur KI-Anwendung angeboten werden können, um einen gleichberechtigten nicht-KI-bezogenen Zugang zu gewährleisten? (Siehe 7.2 Digital Divide) Ja Nein

Menschliche Aufsicht

Wurde die KI-Anwendung so entwickelt, dass menschliche Aufsicht möglich ist (z.B. human-in-the-loop, human-on-the-loop)? (Siehe Wissen: Human in the Loop, Human on the Loop) Ja Nein

Wird das KI-System in regelmäßigen Abständen überprüft (zumindest in Bezug auf Leistung/Qualität, Sicherheit, Einhaltung der geltenden Gesetze und Vorschriften)? Ja Nein

Rechenschaftspflicht

Sind klare Verantwortlichkeiten für Entwicklerinnen und Entwickler, Betreiberinnen und Betreiber und Nutzerinnen und Nutzer der KI-Anwendung festgelegt? Ja Nein

Wurde festgelegt, wer die letztendliche Verantwortung und Rechenschaftspflicht für den KI-Einsatz sowie die Ausgaben des KI-Systems trägt? Ja Nein

Digitale Souveränität

Sind Maßnahmen zur Daten Governance vorhanden, um festzulegen, wie Daten im Zusammenhang mit der Nutzung von KI-Systemen gesammelt, verwendet, gespeichert, gepflegt und verbreitet werden? (Siehe 7.5 Beschaffung) Ja Nein

Wird die Datensouveränität der Verwaltung durch den Einsatz der KI gewahrt, insbesondere im Hinblick auf die Privatsphäre der Bürgerinnen und Bürger? (Siehe 7.4 Digitale Souveränität in der öffentlichen Verwaltung) Ja Nein

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

3 Methoden und Vorgehen

Das Projekt „Digitale Verwaltung und Ethik“ beschäftigte sich im Rahmen von zwei Projektteilen von Juni 2022 bis Juni 2023 sowie von April bis Dezember 2024 mit den ethikrelevanten Überlegungen bezüglich des Einsatzes 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.
- Rechtliche Diskussionen zum Thema KI zu reflektieren.
- Rahmenbedingungen für eine Folgenabschätzung zu schaffen.
- Standards für die Aus- und Weiterbildung von Verwaltungsbediensteten zu setzen.

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 Auftraggeberinnen und Auftraggeber im BMKÖS, waren die Diskussionen in insgesamt sechs Workshops von September 2022 bis Juni 2024 wichtig. Die Verwaltungsbediensteten, die an den Workshops teilnahmen, formulierten dabei aktiv Ideen und brachten Vorschläge sowie Anliegen ein.

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

Die Workshops wurden von der Abteilung für Strategisches Performance Management und Verwaltungsinnovation (III/C/9, Leitung bis Juni 2023 Ursula Rosenbichler, ab Juli 2023 Ralf Tatto) des damaligen BMKÖS und dem AI Ethics Lab (Leitung Peter Biegelbauer) des AIT Austrian Institute of Technology gemeinsam organisiert und umfassten Teilnehmerinnen und Teilnehmer aus dem Bundeskanzleramt, dem Bundesministerium für Arbeit und Wirtschaft (BMAW), dem Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF), dem Bundesministerium für Finanzen (BMF), dem Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK), dem Bundesministerium für Inneres (BMI), dem Bundesministerium für Land- und Forstwirtschaft, Regionen, und Wasserwirtschaft (BML) und dem Bundesministerium für Landesverteidigung (BMLV).

Im letzten Workshop im Juni 2024 wurde ein Schwerpunkt auf die Erfahrungen mit der ersten Version des Praxisleitfadens gelegt, um Anregungen für dessen Überarbeitung aufzunehmen und zu diskutieren.

Feedback aus den Workshops, aus den vom AIT und ehem. BMKÖS gemeinsam abgehaltenen Seminaren an der Verwaltungsakademie des Bundes, Interaktionen insbesondere mit den Auftraggeberinnen und Auftraggeber des ehem. BMKÖS, aber auch mit dem AI Policy Forum sowie andere Kontakte mit Verwaltungsbediensteten waren in der zweiten Projektphase im Jahr 2024 von besonderer Bedeutung. Während die Autorinnen und Autoren 2023 ausschließlich vom AIT AI Ethics Lab gestellt wurden, setzte sich das Team 2024 aus Mitgliedern des AIT AI Ethics Lab und des Research Institute zusammen.

4 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.

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

Legg und Hutter erstellten eine Übersicht mit über 70 Definitionen (Legg und Hutter 2007). Russel und Norvig reduzierten 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 2023). 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.

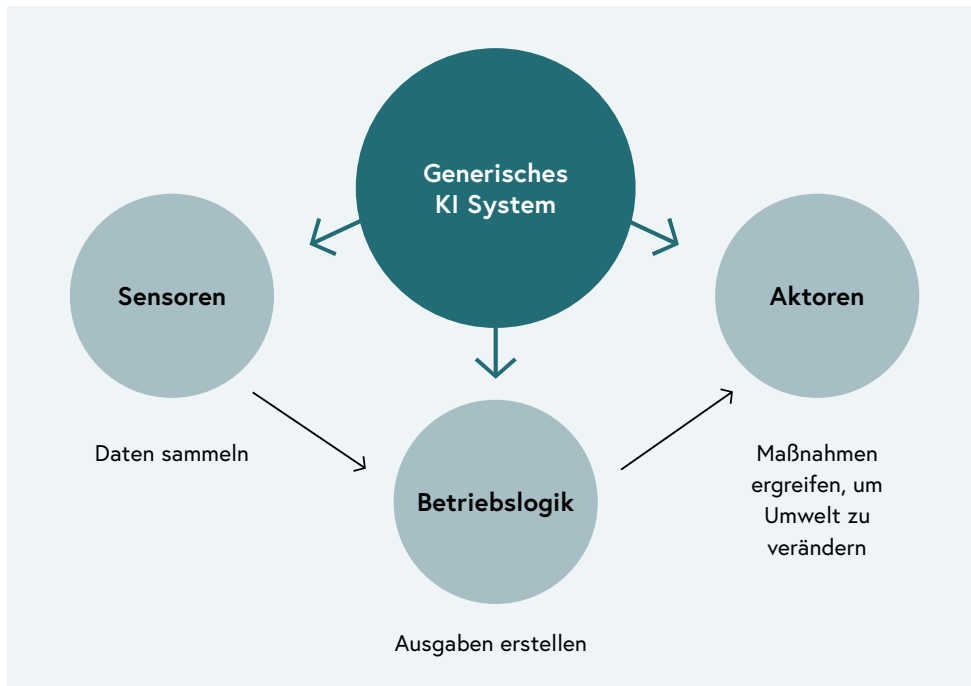


Abbildung 2: Generisches KI-System

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).



Um den Einsatz von KI in der Europäischen Union zu regeln, haben die europäischen (Co-)Gesetzgeber Rat und Europäisches Parlament auf Vorschlag der Europäischen Kommission den AI Act (Verordnung über Künstliche Intelligenz) verabschiedet. Dieser wurde am 12. Juli 2024 im Amtsblatt der Europäischen Union veröffentlicht und trat am 1. August 2024 in Kraft. In der Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz lautet die Definition eines KI-Systems wie folgt:

Für die Zwecke dieser Verordnung bezeichnet der Ausdruck „KI-System“ ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben, wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können.

Definition: KI-System (laut Art. 3 Z. 1 AI Act)

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

Tabelle 1: Unterschied starke und schwache KI

Starke KI 	Schwache KI 
<ul style="list-style-type: none"> • Systeme handeln, lernen und entwickeln sich selbständig weiter. • Können neue, bislang unbekannte Aufgabenstellungen ohne menschliche Intervention lösen. • Entwickeln eigene Lernstrategien und setzen Ziele eigenständig. • Verfügen über Fähigkeiten, die menschlicher Intelligenz ähneln, wie selbstständiges Planen und Problemlösen. 	<ul style="list-style-type: none"> • Systeme lösen spezifische Aufgaben in einem klar definierten Anwendungsbereich. • Verarbeiten Eingabedaten, um eine vordefinierte Ausgabe zu generieren. • Haben keine Fähigkeiten zur selbständigen Weiterentwicklung oder Problemlösung außerhalb ihres spezifizierten Anwendungsbereichs.

Im vergangenen Jahrzehnt haben vor allem KI-Anwendungen auf der Basis neuronaler Netze, die durch sehr große Datenmengen trainiert wurden, in verschiedensten Aufgabenbereichen, wie z. B. Bildklassifizierung, Gesichts-, Objekt- und Spracherkennung, etc., beindruckende Ergebnisse erzielt. Ein Beispiel, das 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 baldige Möglichkeit der Schaffung einer starken KI eher angezweifelt oder in Anbetracht des aktuellen Wissensstands sogar ganz ausgeschlossen (Humm et al. 2022).

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 sowie 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, da keine Nachvollziehbarkeit der Entscheidungsfindung möglich ist. 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, 2021):

- **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 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 die 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 Regelmäßigkeiten 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 und Nutzerinnen verwendet werden. KI stellt Schlüsseltechnologien für die Big Data-Analyse bereit. Die Verarbeitung der Daten muss nicht zwangsläufig 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 kann, 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 eingestuft wird 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 in Entscheidungsprozessen übernimmt, das heißt, entweder Entscheidungen unterstützt 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.

4.1 Generative Künstliche Intelligenz

Generative Künstliche Intelligenz beschreibt den Einsatz von KI-Technologien und -Methoden, die es ermöglichen, anhand von Eingaben – sogenannten „Prompts“ (siehe Infobox) – eine Vielzahl unterschiedlicher Inhalte zu erzeugen. Angefangen bei Texten, über fotorealistic Bilder, bis hin zu Videos, Computercode und musikalischen Kompositionen können unterschiedlichste Ausgabeformen generiert werden.

Prompting/Prompt-Engineering bezeichnet im Zusammenhang mit KI die Formulierung genauer Anweisungen, um von einem generativen KI-System die gewünschte Ausgabe zu erhalten. In Anlehnung an den englischen Begriff für „Eingabeaufforderung“ werden diese Anweisungen auch als „Prompts“ bezeichnet. Der „Engineering“-Zusatz soll zeigen, dass das Verfassen dieser teils umfangreichen und strukturierten Anweisungen selbst je nach Aufgabenstellung möglicherweise detaillierte Kenntnisse des Fachbereichs (z. B. technische Parameter für die Bildgenerierung bzw. Kenntnisse in einem Fachbereich) und eine gewisse Kunstfertigkeit verlangt.

Nehmen wir als einfaches Beispiel an, die Personalabteilung einer Organisation möchte generative KI dazu nutzen, eine E-Mail zur Erinnerung an das jährliche Feedbackgespräch zu generieren. Zu diesem Zweck könnte eine einfache Eingabe wie folgt lauten:

Erstelle eine Erinnerungs-E-Mail für das jährliche Feedbackgespräch.

Möglicherweise entspricht der generierte Text nicht den Vorstellungen des Auftraggebers bzw. der Auftraggeberin. Daher könnte eine verfeinerte Anweisung mit Techniken des Prompt-Engineering wie folgt aussehen:

Stell dir vor, du bist Personalleiter in einer Behörde und schreibst eine E-Mail an alle Mitarbeiter, um sie an ihre jährlichen Feedbackgespräche zu erinnern.

Die E-Mail sollte folgende Punkte enthalten:

- *Betreff: „Erinnerung: Ihr jährliches Feedbackgespräch steht an!“*
- *Einleitung: Eine freundliche Begrüßung und ein Hinweis auf die Bedeutung des Feedbackgesprächs.*
- *Details: Das Datum und die Uhrzeit des Gesprächs, wie man sich darauf vorbereitet und wo das Gespräch stattfindet.*
- *Vorbereitung: Eine Aufforderung, vorab über die eigenen Leistungen und Ziele nachzudenken.*
- *Abschluss: Ein freundlicher Abschluss mit Kontaktinformationen für Fragen.*

Die E-Mail sollte höflich, motivierend und professionell klingen.

Beispiel: Prompt-Engineering

Diese Formate lassen sich in Eingabe- und Ausgabeprozessen auch zunehmend miteinander verschränken. Diese Fähigkeit – auch Multi-Modalität genannt – bedeutet, dass Texte, Bilder, Audio und andere Datenarten auch zugleich verarbeitet und gegebenenfalls auch ausgegeben werden können.

Während die Forschung im Laufe der vergangenen zehn Jahre eine Reihe wichtiger Meilensteine zur Entwicklung dieser Technologie erreichte¹, wurde die Bedeutung und die Fähigkeiten generativer KI erst durch ChatGPT im November 2022 durch die Veröffentlichung benutzerfreundlicher Web-Anwendungen mit interaktiver Dialogführung (Chat) und optimierter Ausgabe wirklich in der breiten Öffentlichkeit wahrgenommen.

Grundlage für diese Fähigkeiten bildet die Verarbeitung großer Datenmengen mithilfe von Deep-Learning-Algorithmen. Diese Daten werden in sogenannten Grundlagen- oder Basismodellen (engl. „*General-purpose*“ bzw. „*Foundation Models*“) verarbeitet, die statistischen Zusammenhänge und Muster in den Daten erkennen und für die Generierung neuer Inhalte nutzen. Besonders bekannt sind die großen Sprachmodelle (engl. „*Large Language Models, LLMs*“), die beispielsweise in der von OpenAI veröffentlichten

1 Eine Auswahl grundlegender wissenschaftlicher Neuerungen und zugehöriger Publikationen sind: Variational Autoencoders (VAEs) (Kingma und Welling 2013), Generative Adversarial Networks (GANs) (Goodfellow et al. 2014), Flow-Based Generative Models (Rezende und Mohamed 2015), Transformer-Architekturen (Vaswani et al. 2017), Diffusion Models (Ho, Jain und Abbeel 2020),

Webanwendung ChatGPT zum Einsatz kommen. Andere Beispiele für hochentwickelte LLMs sind Modelle von Google AI (BERT), Meta (LLaMA) oder Anthropic (Claude). Diese Modelle ermöglichen es der KI Fragen zu beantworten, Texte zusammenzufassen und Inhalte zu erstellen.

Retrieval-unterstützte Generierung (engl. *Retrieval Augmented Generation*, kurz: RAG) kann die Fähigkeiten vortrainierter Sprachmodelle ergänzen, indem es externe Wissensquellen dynamisch einbindet, um präzisere und aktuellere Antworten zu liefern, besonders wenn das zugrundeliegende Modell auf älteren oder begrenzten Daten trainiert wurde. RAG funktioniert so, dass ein Sprachmodell bei einer Anfrage von NutzerInnen und Nutzer nicht sofort Ausgaben generiert, sondern zuvor Informationen aus einer speziellen Wissensdatenbank abrufen. Diese Datenbank ist so organisiert, dass zuerst inhaltlich (semantisch) zur Anfrage der NutzerInnen und Nutzer passende Textausschnitte gesucht werden. Die gefundenen Inhalte fließen dann in das Sprachmodell als zusätzliche Kontextinformationen für die Generierung von Ausgaben ein. Im Vergleich zu einer reinen Generierung durch das Sprachmodell kann dieser Ansatz die Qualität der erzeugten Inhalte verbessern. Besonders in fachspezifischen oder komplexen Anwendungsbereichen kann so eine deutliche Steigerung der inhaltlichen Genauigkeit und Relevanz der generierten Texte erreicht werden.

Wichtig: Obwohl die Hinzufügung kontextbezogener Informationen die Qualität und Relevanz mit Hilfe von RAG generierter Inhalte verbessern kann, lässt sich das Generieren „falscher“ Informationen (Halluzinationen, vgl. 7.1) und Verzerrungen (Bias) dadurch nicht ausschließen. Aus diesem Grund ist es wichtig, LLM-basierte Systeme vor dem produktiven Einsatz systematisch hinsichtlich rechtlicher Konformität (zum Beispiel Datenschutz, Haftung, Urheberrecht) und der Berücksichtigung ethischer Aspekte (zum Beispiel Bias, Transparenz, Verantwortlichkeit) zu überprüfen.

Ohne Zweifel birgt die generative KI das Potential, Text- und kreative Arbeit zu verbessern, indem sie kreative Schreibprozesse beschleunigt, personalisierte Inhalte auf der Grundlage von NutzerInnen- und Nutzerpräferenzen erstellt und Inspiration für eigene Ideen liefert. Ohne die sich daraus ergebenden Möglichkeiten in Abrede stellen zu wollen, ist es ebenfalls notwendig, sich mit den Herausforderungen und Bedenken zu befassen, die insbesondere im Hinblick auf die Integration von generativen KI-Systemen in Arbeitsprozesse zu berücksichtigen sind. Einige Problemstellungen und Herausforderungen sowie der Umgang mit generativer KI am Arbeitsplatz werden in Abschnitt 7.1 behandelt.

4.2 Quantum AI

Quantencomputing und KI zählen derzeit zu den am intensivsten diskutierten Technologien, die das Potential haben, einschneidende Veränderungen in verschiedenen Bereichen der Gesellschaft und Wirtschaft herbeizuführen.

Herkömmliche Computer basieren auf binären Bits, das heißt, die grundlegendsten Einheiten können nur zwei Zustände einnehmen: 1 für „Strom fließt“ und 0 für „Strom fließt nicht“. Jede nur denkbare Darstellung, ob Text, Bild, Audio oder Video wird auf maschinellem Niveau als eine lange Reihe von Nullen (0) und Einsen (1) dargestellt und verarbeitet. Im Gegensatz dazu haben die Quantum Bits (Qubits) den Vorteil, dass sie mehrere Zustände, und zwar gleichzeitig 0 und 1, verknüpft mit unterschiedlichen Wahrscheinlichkeiten annehmen können. Dadurch können Quantencomputer viele Zustände gleichzeitig darstellen und verarbeiten, was insbesondere die Fähigkeit zur Durchführung paralleler Berechnungen verbessert.

Die Überschneidung von Quantencomputing und KI hält Vorteile für beide Seiten bereit. KI, insbesondere maschinelles Lernen, kann effiziente Lösungen, optimierte Leistung und im besten Fall sogar neue Erkenntnisse im Bereich des Quantencomputing liefern (Ramon et al. 2023). Und umgekehrt, kann die Leistungsfähigkeit der Quantencomputer die Ausführung von Algorithmen beschleunigen. Um nur ein Beispiel zu nennen, gibt es möglicherweise Auswirkungen auf die Sicherheit kryptographischer Verfahren, die beim Verschlüsseln oder Signieren von Nachrichten oder Dokumenten zum Einsatz kommen. Digitale Signaturen enthalten ein Rätsel, das für herkömmliche Computersysteme sehr schwer zu lösen ist. Jedoch könnten solche Rätsel durch die Fähigkeit der Parallelisierung – in Abhängigkeit von den jeweils verwendeten kryptographischen Algorithmen – für Quantencomputer durchaus in vertretbarer Zeit lösbar sein. Dies stellt für kryptographisch abgesicherte Informationen und Systeme ein Sicherheitsrisiko dar.

Speziell in Bezug auf KI sind insbesondere die sogenannten quantenneuronalen Netzwerke relevant, da sie das Potential haben, die Leistungsfähigkeit von auf künstlichen neuronalen Netzen basierenden und sehr große Datenmengen verarbeitenden Deep Learning Verfahren erheblich zu steigern. Während klassische neuronale Netzwerke, die hinter vielen modernen KI-Systemen stehen, die Parameter des Netzwerks in der Binärrepräsentation speichern, machen sich Quantenneuronale Netzwerke die besonderen Eigenschaften von Qubits zu Nutze. Da Qubits überlagerte Quantenzustände annehmen können, sind Quantenneuronale Netzwerke in der Lage, unzählige Berechnungen parallel durchzuführen, was die Darstellung und Verarbeitung größerer Netzwerkstrukturen als bei klassischen neuronalen Netzwerken ermöglicht. Dies könnte beim Training sehr großer und komplexer Modelle, wie zum Beispiel bei der Erstellung großer Sprachmodelle in der Zukunft eine Rolle spielen.

Allerdings lassen sich die genauen Auswirkungen der erfolgreichen Anwendung von Quantencomputing im Bereich der KI-Systeme – außer der generellen Aussicht auf schnellere Verarbeitung, verbesserte Algorithmen und effizientere Problemlösungen – nur schwer vorhersagen. Schon aktuelle Sprachmodelle, die auf herkömmlichen Computer-

systemen mit GPUs (Grafikprozessoren) erstellt wurden, haben teils mehrere hundert Milliarden bis hin zu Billionen von Parametern und es ist nicht offensichtlich, welche Verbesserungen durch eine Vergrößerung – also eine Erhöhung der Anzahl der Parameter – genau zu erwarten sind. Aus dem gleichen Grund wird erst die Praxis zeigen, welche Verbesserungen sich durch den Einsatz von Quantencomputing für KI-Systeme tatsächlich ergeben werden. Zudem sind die Ergebnisse, die im Forschungsgebiet des Quantencomputings erreicht wurden, noch überwiegend theoretischer und experimenteller Natur. Die Funktionsfähigkeit, Zuverlässigkeit und Skalierbarkeit von Quantensystemen muss sich erst noch als praxistauglich erweisen und in realen Anwendungsszenarien unter Beweis stellen. Sollte dies gelingen, besteht die Möglichkeit, dass Quantenalgorithmen Muster und Zusammenhänge in Daten erkennen, die mit Hilfe klassischer Algorithmen nicht entdeckt werden können.

5 Handlungsbedarf

Angesichts der in der Einleitung beschriebenen spezifischen Aufgaben und Herausforderungen der öffentlichen Verwaltung beim Einsatz von KI-Systemen und der im Abschnitt 4 angesprochenen technischen Komplexität der Technologie muss hier besonders auf einen ethikkonformen 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 Bürgerinnen und Bürger 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 hervorgehoben.

In Australien wurde das Robodebt System eingesetzt, im Rahmen dessen die australische Steuerbehörde automatisiert jährliche Einkommensdaten mit dem von Sozialversicherungsempfängern 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 Rückzahlungen 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 zehntausend Euro, 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 Abschnitt 12).

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. KI-Literacy ist auch eine zentrale Anforderung des AI Acts an alle Entwicklerinnen und Entwickler, Anwenderinnen und Anwender und anderen Akteursgruppen, die sich mit dem Thema KI auseinandersetzen (siehe Abschnitt 8.3).

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 Bürgerinnen und Bürger. Die Autorinnen und Autoren 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, Human on the Loop

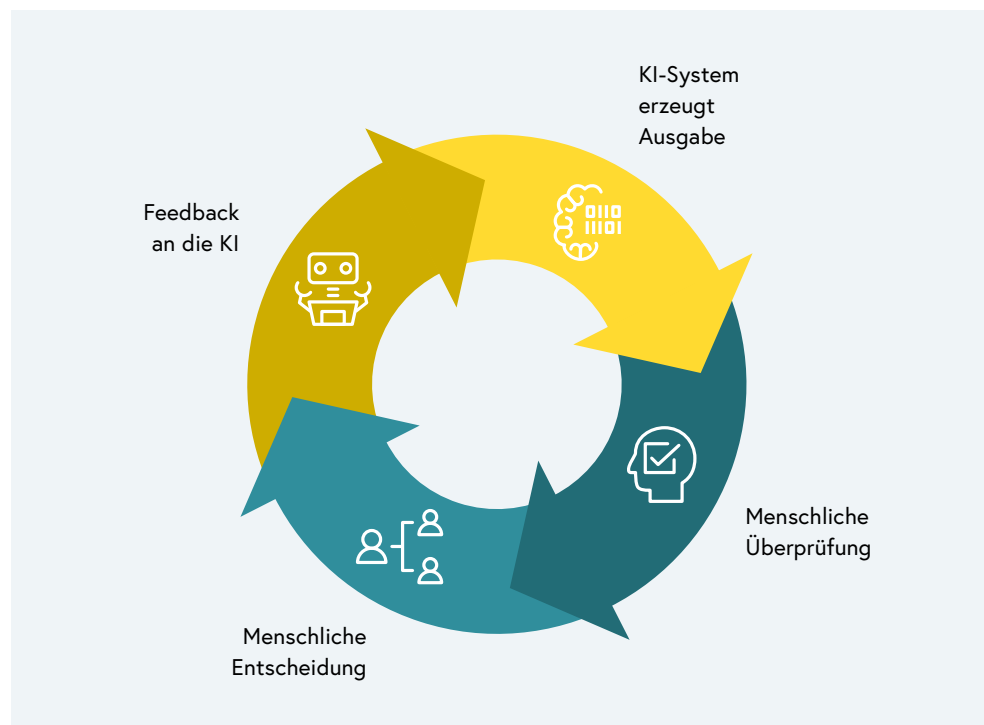


Abbildung 3: Beispiel für Human in the Loop

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 nicht unproblematisch – wie die vorher genannten Beispiele aus Australien und den Niederlanden zeigen. Noch kritischer sind allerdings Bereiche wie Rechtsetzung und Rechtsprechung, also etwa Tätigkeiten der Parlamentsdirektion und der Justizverwaltung.

Tatsächlich sind die Chancen und Risiken nicht in jedem Bereich gleich gelagert. So ist die Einführung des von Bürgerinnen und Bürger 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“, „WienMobil“) grundsätzlich weniger problematisch: im Regelfall halten sich potentielle negative Folgen einer beispielsweise gebiasten/verzerrten Empfehlung in Grenzen.

Ein Chatbot ermöglicht textbasierte Dialoge, um mit einem technischen System zu kommunizieren. Benutzerinnen und Benutzer können sich dabei üblicherweise in Alltagssprache mit dem System austauschen und Fragen stellen. Das System kann – muss aber nicht zwangsläufig – KI 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 darf der Chatbot keine menschenverachtenden oder diskriminierenden Nachrichten verbreiten.

Der im März 2016 von der Microsoft Corporation veröffentlichte selbstlernende Chatbot „Tay“ reagierte innerhalb von weniger als 24 Stunden mit sexistischen und rassistischen Kommentaren, so dass der Bot abgeschaltet wurde 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 aussendete (Wille 2021).

Anwendungsfall: Chatbot

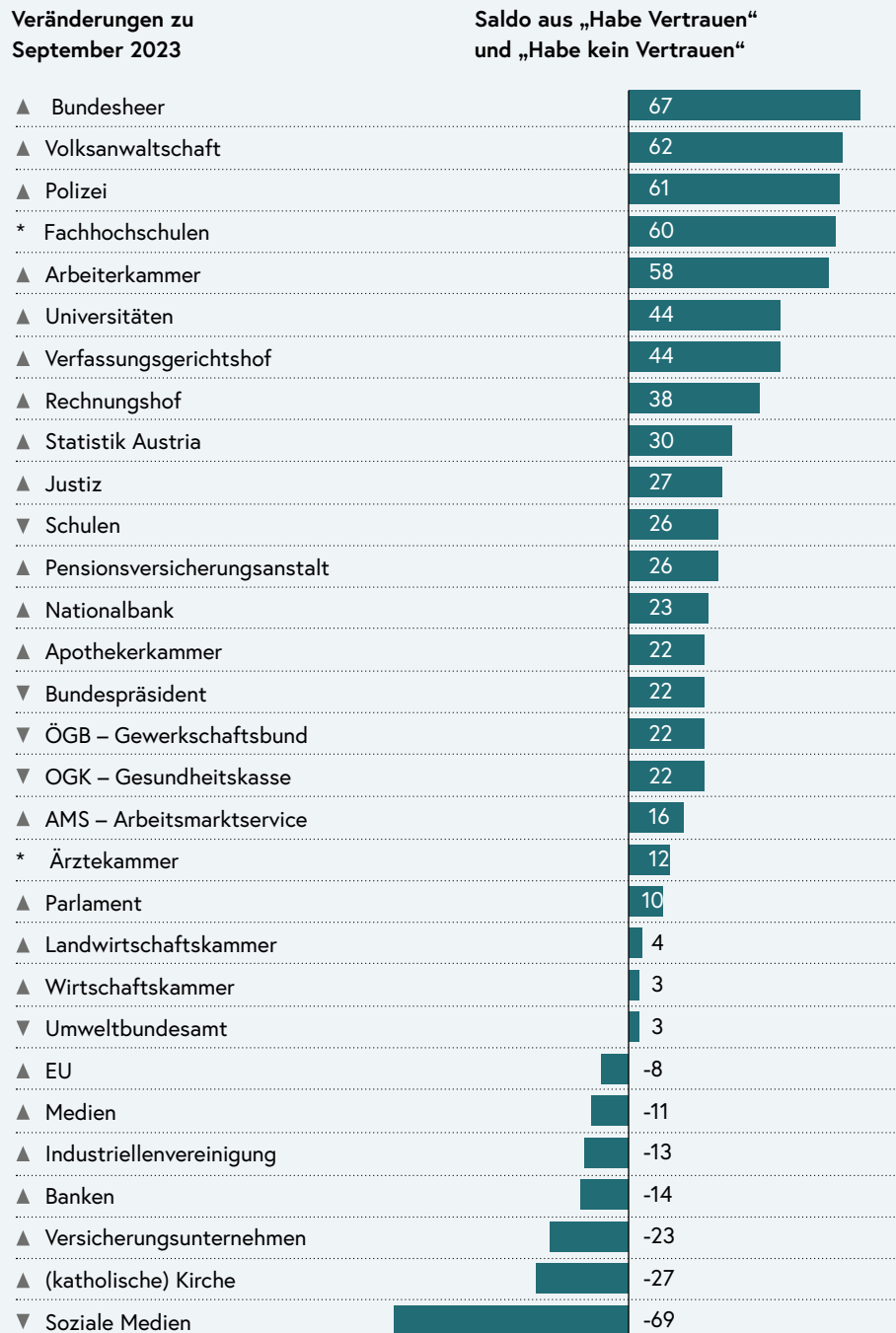
Gänzlich anders gelagert sind die Bereiche der Rechtsetzung und der Rechtsprechung. Hier sind die Gebote von Transparenz, Nachvollziehbarkeit, Rechenschaftspflicht und Verhinderung von Bias von zentraler Bedeutung. So erscheint der Einsatz von algorithmischen Entscheidungsunterstützungssystemen vor dem 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 Bürgerinnen und Bürger mit dem Staat darstellt und damit auch für den Staat steht. Somit spielt für das Vertrauen der Bürgerinnen und Bürger 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 2024).

Vertrauen in Institutionen

APA-OGM-Vertrauensindex – Befragung aus 1.068 Online-Interviews (Oktober 2024)



* Neu abgefragt

Quelle: OGM

Abbildung 4: Repräsentative Umfrage APA/OGM-Vertrauensindex 2024

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.

6 Abwägung: KI in der Verwaltung, oder nicht?

Dieser Abschnitt beschäftigt sich mit der Frage, ob und unter welchen Bedingungen KI-Technologie in der Verwaltung eingesetzt werden soll. Als Entscheidungshilfe soll dazu der in Abbildung 5 gezeigte Entscheidungsbaum dienen. 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 Entscheidungsbaum beginnt mit der grundsätzlichen Frage, ob ein geeigneter Anwendungsfall für den Einsatz dieser Technologien vorliegt. Der Einsatz von KI-Technologie ist im Prinzip nur dann sinnvoll, wenn der Anwendungsfall bestimmte Kriterien, wie zum Beispiel mindestens eines der folgenden, erfüllt:

1. Es gibt sich wiederholende Aufgaben, die sich automatisieren lassen. Eine Maschine könnte diese Aufgaben übernehmen bzw. erforderliche Kompetenzen mit Hilfe von KI-Technologie beitragen.
2. Die Aufgabenstellung, die mit Hilfe von KI-Technologie gelöst werden soll, ist nicht überwiegend auf menschliche Urteilsfähigkeit angewiesen. Bei komplexen Aufgaben, deren Lösungsansatz erst durch eine vielschichtige Problemanalyse ermittelt werden muss, sind die derzeitigen Methoden des maschinellen Lernens nach aktuellem Stand der Technik in der Regel nicht ausreichend (vgl. schwache KI, Abschnitt 4).
3. Es gibt einen Bedarf nach Entscheidungsunterstützung, das heißt, KI-Technologien können in unterstützender Form, zum Beispiel zur Aufbereitung und Visualisierung benötigter Informationen, einen Beitrag für bestimmte Aufgabenstellungen leisten.
4. Die KI-Technologie kann existierende Arbeitsprozesse, beispielsweise die Texterstellung mit Hilfe generativer KI, unterstützen (vgl. Abschnitt 4.1 „Generative KI“).

Die Entscheidung für den Einsatz von KI-Technologie sollte dabei aufgrund der Analyse existierender Arbeitsprozesse und Dienstleistungen mit Hinblick auf Verbesserungspotential, das sich möglicherweise mit Hilfe von KI ausschöpfen lässt, getroffen werden.

Im zweiten Schritt wird die Frage gestellt, ob der erwartete Nutzen umfassend gegen die anfallenden Kosten abgewogen wurde. Aus wirtschaftlicher Sicht stellt sich die grundlegende Frage, ob eine KI-Anwendung in der Verwaltung mit eigenen Mitteln aufgebaut werden soll oder ob sich mit Hilfe kommerzieller Anbieter die Ziele kostengünstiger erreichen lassen. Zu berücksichtigen sind hier die zu Projektbeginn anfallenden Entwicklungs- und Implementierungskosten, laufende Betriebskosten, erforderliche Expertise und Ressourcen, sowie die Risiken und Sicherheitsaspekte, die

bei der Eigenentwicklung gegeben sind. Demgegenüber stehen bei kommerziellen Anbietern möglicherweise geringere Einstiegskosten, schnellere Implementierung, regelmäßige Updates, Support, und geringere Anforderungen in Bezug auf den Aufbau internen technischen Know-hows. Jedoch sind bei Drittanbietern die langfristigen Kosten, mangelnde Flexibilität und Freiheiten in Bezug auf Features und Update-Zyklen sowie Einschränkungen hinsichtlich der Datensouveränität zu bedenken. Zusammenfassend ist hier also eine gründliche Kosten-Nutzen-Analyse notwendig. Allerdings sollten nicht nur die direkten finanziellen Aufwände, sondern auch die indirekten Kosten, wie etwa durch Umweltauswirkungen verursachte Belastungen, berücksichtigt werden.

Für den Fall, dass ein geeigneter Anwendungsfall vorliegt und sich dieser aus wirtschaftlicher Perspektive als tragfähig erweist, stellt sich die Frage, ob die rechtlichen Erfordernisse grundsätzlich erfüllt werden können. 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. Zentrale Aspekte sind (im hoheitlichen Bereich) das Legalitätsprinzip, die Risikoeinschätzung als Basis für die Bestimmungen nach dem AI Act (vgl. Abschnitt 8.3) und der Datenschutz nach der DSGVO (vgl. Abschnitt 8.2), einschließlich einer rechtmäßigen Grundlage für die Datenverarbeitung, Datensicherheit und die klare Information über die Art der Verarbeitung. Zudem müssen urheberrechtliche Vorschriften beachtet werden, insbesondere hinsichtlich der Verwendung geschützter Daten zum Training und der möglicherweise ebenfalls dem Urheberschutz unterliegenden Ausgabe der KI. Außerdem sollten Haftungsfragen, also die Frage, wer im Falle von durch die Anwendung der KI verursachten Schäden zur Verantwortung gezogen wird, vertraglich geregelt sein.

Der nächste Schritt „Ist eine Risikoeinstufung nach dem AI Act gegeben?“ befasst sich mit Unterteilung von KI-Systemen in vier Risikostufen durch den Anbieter eines solchen Systems: inakzeptables Risiko, hohes Risiko, eingeschränktes Risiko bzw. Transparenzpflichten und minimales Risiko (vgl. Abschnitt 8.3). Kurz zusammengefasst sind Systeme mit einem inakzeptablen Risiko verboten, da sie erhebliche Gefahren für die Sicherheit, die Grundrechte oder andere wichtige gesellschaftliche Werte darstellen. Systeme mit hohem Risiko, wie etwa solche, die in der Verwaltung für die Bewertung von Personen oder Entscheidungen eingesetzt werden, unterliegen strengen Vorschriften und müssen besondere Anforderungen hinsichtlich Transparenz, Sicherheit, Genauigkeit und Überwachung erfüllen. In Bezug auf gewisse KI-Systeme (z. B. Chatbots, Deepfakes) gibt es spezifische Transparenzpflichten, während Systeme mit minimalem Risiko weitgehend ohne zusätzliche Regulierungen eingesetzt werden können. Eine Risikoeinstufung durch den Anbieter ist notwendig, um sicherzustellen, dass die rechtlichen Anforderungen mit Bezug auf den AI Act eingehalten werden und dass der Einsatz von KI im Verwaltungsbereich verantwortungsvoll erfolgt. Eine erste Rollen- und Risikoeinschätzung kann derzeit mithilfe des „EU AI Act Compliance Checkers“² der NGO „Future of Life Institute“ vorgenommen werden.

2 <https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/>

Wesentlicher Teil des Datenschutzaspektes nach der DSGVO ist die Frage, inwieweit personenbezogene Daten für die Erstellung von Modellen beziehungsweise Optimierung der KI-Technologie involviert sind. Ist dies der Fall, so ist zu prüfen, ob datenschutzrechtliche Bedenken mit Verfahren der Anonymisierung, Pseudonymisierung (Ersetzung eines Identifikationsmerkmals durch ein Pseudonym) beziehungsweise durch Abstrahierung ausgeräumt bzw. abgemildert werden können. In jedem Fall ist hier 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. Es 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 KI-Anwendungen oft sogenannte Grundlagenmodelle (engl. „*Foundation Models*“) verwenden, die statistische Zusammenhänge aus sehr vielen Eingabedaten erkennen, die häufig aus Social Media- und Internet-Datenquellen gesammelt wurden. Algorithmen können dann mit dem Training an wenigen Beispielen für einen spezifischen Anwendungsbereich optimiert werden (sog. *fine-tuning*). Dabei besteht jedoch das Risiko, Verzerrungen, die in den ursprünglichen Daten der Grundlagenmodelle enthalten sind, zu übernehmen und dadurch Personengruppen bzw. Minderheiten 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.

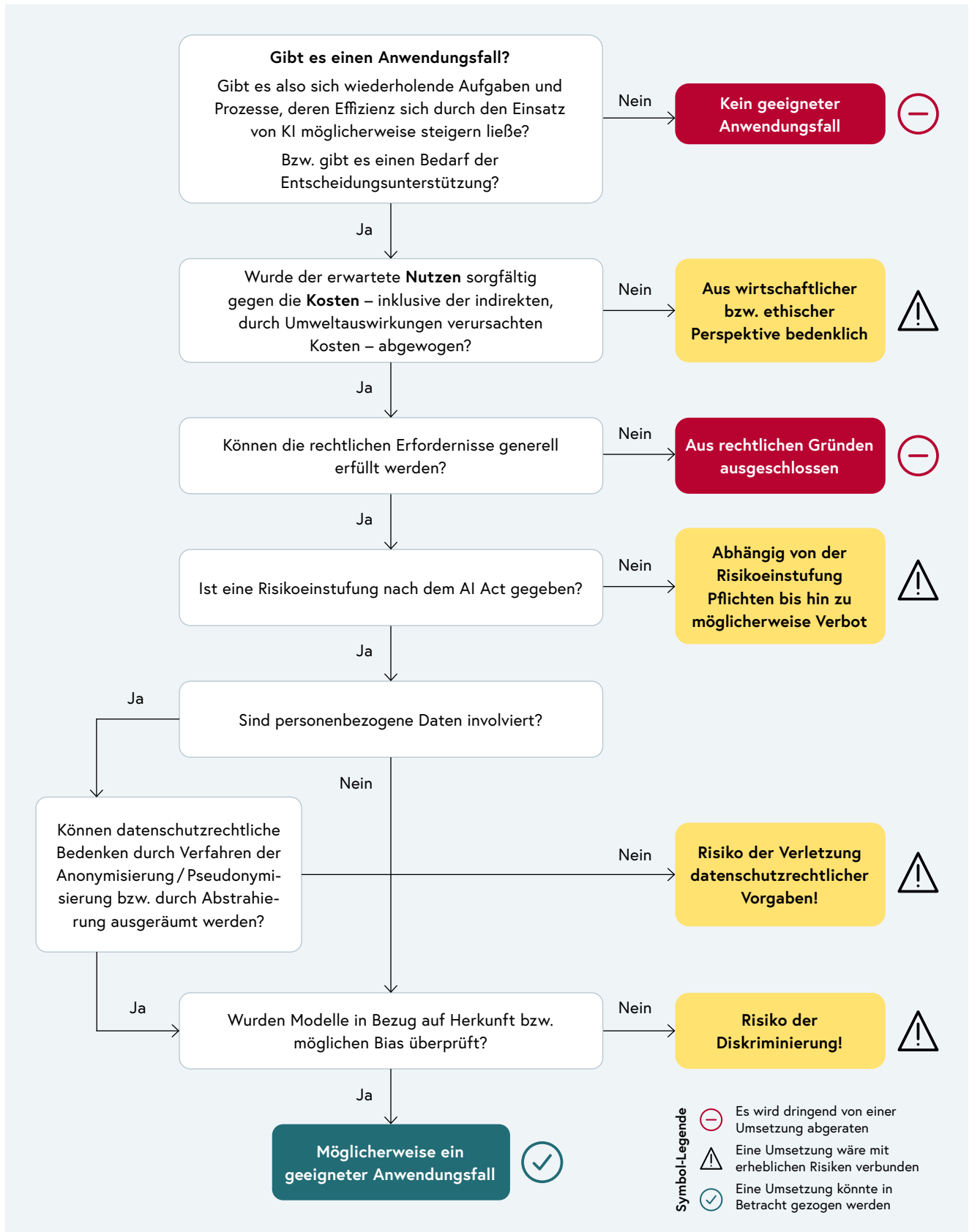


Abbildung 5: Entscheidungsbaum zur Verwendung von KI-Technologie

Beispiel für die Anwendung des Entscheidungsbaums³

Die öffentliche Verwaltung erhält täglich Anfragen zu verschiedenen Themen wie Meldeangelegenheiten, Beantragung von Dokumenten (z. B. Reisepässe, Personalausweise), Terminvergaben, Informationen zu lokalen Dienstleistungen und mehr. Die manuelle Bearbeitung dieser Anfragen kann zeitaufwändig sein. Um die zeitgerechte Beantwortung bei stetig steigender Anzahl von Anfragen zu gewährleisten, wird der Einsatz einer KI-Anwendung in Betracht gezogen, welche die Bearbeitung von Anfragen zumindest teilweise automatisieren bzw. unterstützen soll. Es soll dafür ein Chatbot zum Einsatz kommen. Dieser soll Anfragen an die zuständigen Abteilungen weiterleiten oder einfache, häufig gestellte Fragen sofort beantworten (Aschauer 2024).

Gibt es einen Anwendungsfall?

- Ja, es gibt sich wiederholende Aufgaben und Prozesse, nämlich die Bearbeitung und Weiterleitung von Anfragen, deren Effizienz durch den Einsatz von KI gesteigert werden könnte. Zudem gibt es einen klaren Bedarf an Entscheidungsunterstützung, um Verwaltungsmitarbeiterinnen und -mitarbeiter bei der Beantwortung der Anfragen mit passenden Vorlagen, Antworten und Entscheidungsunterstützungen zu helfen.

Wurde der erwartete Nutzen sorgfältig gegen die Kosten abgewogen?

- Es wurde ein Gutachten bei einer unabhängigen Beratungsorganisation in Auftrag gegeben, die unter anderem eine Kosten/Nutzen Analyse beinhaltete. Laut Gutachten sind durch den KI-Service erhebliche Zeitersparnis, verbesserte Effizienz und kürzere Reaktionszeiten für Bürgerinnen und Bürger zu erwarten. Um die Datenhoheit zu gewährleisten, wurde eine Eigenimplementierung den Lösungen von Drittanbietern vorgezogen. Die Vorteile überwiegen die höheren Kosten für die Implementierung und Wartung der KI. Umweltauswirkungen konnten durch gemeinsame Projekte zur Auswahl und Erstellung geeigneter Grundlagenmodelle minimiert werden. Umweltfreundliche Rechenzentren sollen durch die Verwendung alternativer Energien die negativen Auswirkungen für die Umwelt minimieren. Die Umsetzung der KI-Anwendung wird vor diesem Hintergrund als wirtschaftlich sinnvoll und umweltverträglich erachtet.

3 Dieses Anwendungsbeispiel erhebt selbstverständlich keinen Anspruch auf Vollständigkeit hinsichtlich aller praxisrelevanten Fragestellungen. Aspekte wie Haftungsfragen bzw. Haftungsausschluss, die Gewährleistung der Rechtsrichtigkeit der Chatbot-Ausgaben oder Vorgaben zur Zweckbindung und Nutzung des Chatbots werden hier nicht behandelt. Diese Themen sind jedoch in der Praxis von Bedeutung und sollten bei der tatsächlichen Implementierung eines solchen Systems gesondert betrachtet werden.

Ist eine Rechtsgrundlage gegeben?

- Aus dem Gutachten geht ebenfalls hervor, dass die rechtlichen Erfordernisse für den Einsatz der KI in der Verwaltung gegeben sind. Das Urheberrecht wird durch die Klärung der Rechte an Dokumenten, die für Training und Optimierung der Modelle zum Einsatz eingehalten und die Einhaltung der Datenschutz-Grundverordnung (DSGVO) ist laut Gutachten ebenfalls gegeben. Besondere Berücksichtigung erfuhren dabei die rechtlichen Anforderungen des Verfassungs- und Verwaltungsrechts (bspw. Grundrechte, Genehmigung von Erledigungen, rechtliches Gehör, Begründungspflicht, Legalitätsprinzip).

Sind personenbezogene Daten involviert?

- Ja, die Anfragen von Bürgerinnen und Bürger enthalten personenbezogene Daten (z. B. Namen, Adressen, Anliegen). Um eine bessere Kundenzufriedenheit zu erlangen, ermöglicht der Dienst das Anlegen eines eigenen Profils. Bürgerinnen und Bürger werden ausdrücklich gefragt, ob sie mit der Nutzung der eigenen Daten im Sinne einer Verbesserung der Dienstleistung einverstanden sind. Dabei werden ausschließlich diejenigen Daten verwendet, die einer personalisierten Antwort-Erstellung dienen. Es wurde dabei darauf geachtet, dass der Dienst vollumfänglich ohne Personalisierung genutzt werden kann. So können zum Beispiel personenbezogene Daten mit der Anfrage mitgegeben werden, ohne dass diese gespeichert werden. Auch auf die Speicherung der Anfrage – und Antworthistorie wird aufgrund von Datenschutz- und Sicherheitsbedenken verzichtet.

Können datenschutzrechtliche Bedenken durch Verfahren der Anonymisierung/Pseudonymisierung bzw. durch Abstrahierung ausgeräumt bzw. abgemildert werden?

- In einem gemeinsamen Projekt mit europäischen Verwaltungsorganisationen wurden optimierte Grundlagenmodelle erstellt. Die dafür benötigten Daten wurden von den Projektpartnern beigetragen. Mit Hilfe von Anonymisierungs- und Pseudonymisierungsverfahren wurden die Daten zur Verwendung als Trainingsdaten vorbereitet. Ortsbezogene Daten mit Personenbezug wurden anonymisiert und dergestalt abstrahiert, sodass keine Rückschlüsse auf Aufenthalt oder Bewegungen von Bürgerinnen und Bürger gezogen werden können.

Wurden Modelle in Bezug auf Herkunft bzw. möglichen Bias überprüft?

- KI-Modelle werden von unabhängigen Organisationen mit sozialwissenschaftlicher und technischer Kompetenz überprüft, um sicherzustellen, dass verwendete Modelle nicht voreingenommen (biased) sind und Anfragen von Bürgerinnen und Bürger fair und neutral behandelt werden. Dies wird durch eine Analyse der Trainingsdaten und durch kontinuierliche Überwachung der KI-Entscheidungen gewährleistet.

7 Chancen und Herausforderungen beim Einsatz von KI in der öffentlichen 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. Insbesondere wird der Einsatz von generativer KI am Arbeitsplatz thematisiert, wobei sowohl auf Potenziale als auch auf Risiken und offizielle Leitlinien (z. B. KI-Guidelines des Bundes) für den Einsatz von generativer KI in Arbeitsprozessen eingegangen wird. Ein weiterer Teil dieses Abschnitts widmet sich den Auswirkungen von KI auf die Bevölkerung. Hier wird dargelegt, wie KI die Interaktion mit Bürgerinnen und Bürger verändern kann und welche Chancen und Herausforderungen sich durch die Automatisierung von Dienstleistungen ergeben. 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. Danach rückt das Thema der digitalen Souveränität und Datensouveränität in der öffentlichen Verwaltung in den Fokus. Dieser Teil beleuchtet die Bedeutung der Unabhängigkeit der öffentlichen Verwaltung für die Gestaltung digitaler Dienstleistungen, den Schutz sensibler Daten und die Wahrung der Interessen der Bürgerinnen und Bürger. Der letzte Themenbereich, „Beschaffung“, verdeutlicht, wie die öffentliche Verwaltung durch die gezielte Vergabe von Aufträgen zur Stärkung der digitalen Souveränität und zur Umsetzung vertrauenswürdiger KI-Lösungen beitragen kann. Der Schwerpunkt liegt dabei auf der Gewährleistung der Qualität und Sicherheit von KI-Systemen durch klare Spezifikationen und Kriterien bereits während der Vorbereitung und Planung von Beschaffungsprozessen für KI-Anwendungen.

7.1 KI und Auswirkungen auf die Arbeitswelt 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 sparten-spezifisch sind, gibt es viele KI-Tools, die auch in der öffentlichen Verwaltung prinzipiell verwendet werden können (Haslinger, 2022, 61f):

- Recruiting Software, 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 Arbeitskolleginnen und -kollegen die jeweilige Kandidatinnen und Kandidaten 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 Fahrradbotinnen und -boten, Reinigungskräfte, Kreativschaffende, Clickworkerinnen und -worker und Fahrerinnen und Fahrer. 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 in der öffentlichen Verwaltung ist die erwartete Effizienzsteigerung von Arbeits- und Verwaltungsprozessen. Die Implementierung von KI-Anwendungen bietet grundsätzlich das Potenzial, den Arbeitsaufwand für Beschäftigte in der Verwaltung zu reduzieren. Dies trifft besonders auf Bereiche des Wissensmanagements und auf Routinetätigkeiten zu, wie beispielsweise die Ausstellung von Dokumenten oder die Beratung zu häufig nachgefragten Themen.

Gleichzeitig bringt der Einsatz von KI in der öffentlichen Verwaltung erhebliche Herausforderungen in Bezug auf die Akzeptanz der Mitarbeitenden mit sich. Ein zentrales Problemfeld ist die potenzielle Verlagerung von Entscheidungsprozessen von Menschen auf KI-Systeme, was Risiken wie Datenschutzverletzungen, Verlust an Autonomie sowie mögliche Benachteiligungen der Mitarbeitenden nach sich ziehen kann. Insbesondere die Erkennung von „abweichendem“ Verhalten im Sinne einer Optimierung der Arbeitseffizienz kann auf ungerechten Annahmen basieren und somit zur Diskriminierung und psychischen Belastung von Arbeitnehmerinnen und Arbeitnehmer 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 Eingabe- oder Trainingsdaten erfolgen kann. KI-Systeme, deren Entscheidungsvorschläge 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 (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 Bewerberinnen und Bewerber eingeführt hatte (Dastin 2022). Das System verwendete zur Bewertung aktueller Bewerberinnen und Bewerber historische Daten. Da sich jedoch in der Vergangenheit überwiegend männliche Kandidaten beworben hatten, wurden männliche gegenüber weiblichen Kandidatinnen und Kandidaten systematisch von der KI-Anwendung bevorzugt.

Wissen: Bias; Anwendung: Geschlechterbias

KI-gestützte Überwachung am Arbeitsplatz wirft große ethische Bedenken auf. Die Verletzung der Privatsphäre von Arbeitnehmerinnen und Arbeitnehmer, Diskriminierung aufgrund algorithmischer Entscheidungen und die Auswirkungen auf die psychische Gesundheit sind mögliche Probleme (Christl 2021). Praktiken der Überwachung, die die Menschenwürde berühren und durch keine Betriebsvereinbarung geregelt sind, sind unzulässig und insofern abzulehnen.

Um diesen Gefahren zu begegnen, empfiehlt die OECD folgende Strategien für den Einsatz von KI am Arbeitsplatz (OECD 2024):

1. KI-Strategien für den Arbeitsplatz sollten idealerweise von Arbeitnehmerinnen und Arbeitnehmer, Vorgesetzten und anderen relevanten Stakeholdern mitgestaltet werden.
2. Alle Betroffenen sollten über die mit KI verbundenen psychischen Gesundheitsrisiken (z. B. Stress durch höhere Leistungsanforderungen in kürzerer Zeit, steigender Arbeitsdruck, Überforderung mit der Technologie, Zukunftsängste) informiert werden und es sollten interne Regelungen getroffen werden, die diese Risiken adressieren.

3. Dem Einsatz von KI bei der Überwachung und Automatisierung von Entscheidungsprozessen müssen klare Grenzen gesetzt werden und es sind Maßnahmen zur menschlichen Aufsicht zu ergreifen, um den Schutz der Sicherheit, der Rechte und der Möglichkeiten der Arbeitnehmerinnen und Arbeitnehmer zu gewährleisten.
4. Es sind klare Verantwortlichkeiten festzulegen, um das Vertrauen in das KI-System zu gewährleisten.
5. Die Einführung von KI-Systemen sollte durch die Erstellung von Verhaltenskodizes (engl. „Codes of Conduct“) und/oder die Einsetzung von Ethikbeauftragten geregelt werden.
6. Die Arbeitnehmerinnen und Arbeitnehmer sollten über ausreichende Schutzmechanismen verfügen, um sich gegen voreingenommene oder unfaire Entscheidungen zu wehren, z. B. durch das Recht, eine KI-Entscheidung zu beanstanden.
7. Der Einsatz von Haftungsausschlüssen und/oder Warnsystemen ist erforderlich, um Arbeitnehmerinnen und Arbeitnehmer darauf hinzuweisen, dass sie mit einem KI-System interagiert haben oder interagieren werden.

Es sollten Konsultationsmechanismen zwischen Arbeitnehmerinnen und Arbeitnehmer, Vorgesetzten und anderen relevanten Akteuren eingerichtet werden, um KI-bezogenen Richtlinien, Strategien und Maßnahmen an die jeweiligen Bedürfnisse anzupassen. Außerdem ist die Entwicklung von KI- und technologischem Fachwissen sowie digitaler Bildung für alle relevanten Akteure von zentraler Bedeutung.

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 12). 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 Expertinnen und Experten oder Unternehmen verringert werden, was die Autonomie fördert. Ein weiterer wichtiger Punkt ist die ethische und strategische Abwägung, wann und wie KI-Systeme implementiert werden sollten. **KI sollte in erster Linie darauf abzielen, Arbeitsprozesse zu erleichtern und zu unterstützen.** Es geht nicht darum, menschliche Arbeitskräfte zu ersetzen, sondern ihnen (technologische) Werkzeuge an die Hand zu geben, die ihre Arbeit verbessern und sie selbst entlasten können. Diese Differenzierung ist entscheidend, um potenzielle Ängste vor Arbeitsplatzverlusten zu mindern und eine Akzeptanz für KI-gestützte Prozesse zu schaffen.

Umgang mit (generativer) KI am Arbeitsplatz

Wie in Abschnitt 4.1 erwähnt, kann KI und insbesondere generative KI, die Arbeit der öffentlichen Verwaltung durch weitreichende Automatisierung unterstützen. Der Einsatz generativer KI hat bereits vielversprechendes Potenzial gezeigt, insbesondere in Bereichen wie der Kommunikation mit Bürgerinnen und Bürger über Chatbots, der Generierung und Zusammenfassung von Inhalten sowie der Softwareentwicklung, die jeweils zu Effizienzsteigerungen geführt haben (Biegelbauer et al 2024, McKinsey & Company 2024).

Doch trotz vieler Vorteile bringt der Einsatz generativer KI auch Herausforderungen mit sich (Schneeberger 2024b). Insbesondere bei der Erstellung faktenbasierter Inhalte stößt die Technologie an ihre Grenzen, da generative KI mitunter Informationen verfälscht. Ein wesentlicher Aspekt dabei ist die mögliche Inkonsistenz generierter Inhalte. Selbst wenn große Sprachmodelle mit überprüften und vertrauenswürdigen Dokumenten trainiert wurden, können sie in verschiedenen Bereichen zweifelhafte oder sogar „falsche“, auch als „Halluzinationen“ bezeichnete, Ergebnisse liefern. Besonders problematisch ist, dass die oft sprachlich und grammatikalisch einwandfreien und plausibel scheinenden Texte erst bei genauerem Hinsehen Unstimmigkeiten erkennen lassen. Die nicht-verifizierte Übernahme von Inhalten birgt daher das Risiko der Verbreitung unzuverlässiger oder sogar nicht rechtsrichtiger Information.

Außerdem gilt es zu berücksichtigen, dass große Sprachmodelle allein keinen direkten Zugriff auf Echtzeitdaten haben und ihre Wissensbasis nicht kontinuierlich aktualisieren können. Das bedeutet, dass sie auf Informationen beschränkt sind, die bis zum Zeitpunkt der Modellerstellung verfügbar waren. Um dieses Defizit zu kompensieren, integrieren Dienste generativer KI zunehmend Datenbanken und Online-Datenquellen, die aktuelle und präzise Informationen bereitstellen. Außerdem ist die Qualität der Ausgabe auch davon abhängig, wie gut ein Wissensbereich in den Trainingsdaten repräsentiert ist, so dass die Ausgabequalität in speziellen Themengebieten möglicherweise nicht zufriedenstellend ist.

Ein besonders wichtiger Aspekt ist, dass Modelle, insbesondere solche zur Text- und Bildgenerierung, möglicherweise Verzerrungen enthalten, da sie auf umfangreichen Sammlungen von Webinhalten basieren. Diese Quellen können alle erdenklichen Formen von Vorurteilen, von Beschimpfungen und Hassrede bis hin zu Voreingenommenheit, Rassismus, Sexismus, ideologisch geprägten Ansichten und der Ausgrenzung von Minderheiten widerspiegeln. Problematisch ist dabei, dass die Suche nach Ursachen und Lösungsansätzen nur eingeschränkt möglich ist, da fast ausschließlich große Konzerne der Informationstechnologiebranche die notwendigen Ressourcen für die Erstellung der Grundlagenmodelle aufbringen können. Der Zugang zu den ursprünglichen Daten wird in der Regel verwehrt bzw. ist vielleicht auch gar nicht vollumfänglich nachvollziehbar, so dass nur anhand der Analyse des Ausgabeverhaltens Rückschlüsse auf die zu Grunde liegenden Verzerrungen gezogen werden können.

Außerdem wirft die Nutzung von Echtzeitdaten und die Verarbeitung großer Datenmengen auch ethische und datenschutzrechtliche Fragen auf. Oft beinhalten die

zur Modellerstellung verwendeten Daten vertrauliche Informationen. Mitunter wird durch gezielte Attacken versucht, personenbezogene Informationen, wie zum Beispiel Kreditkartennummern aus großen Sprachmodellen zu extrahieren (Panda et al. 2024). Zwar wird von Betreibern von Diensten, die auf großen Sprachmodellen basieren, versichert, dass Sicherheitsmaßnahmen getroffen werden, jedoch bleibt ein grundsätzliches Risiko für Informationslecks bestehen.

Die Arbeitssituation der Menschen, die funktionsfähige Dienste auf der Basis großer Sprachmodelle ermöglichen, sollte ebenfalls berücksichtigt werden. Zum einen basieren die Grundlagenmodelle auf den Eingabedaten, die automatisiert im Trainingsprozess zur Modellerstellung verarbeitet werden. Zum anderen benötigt gerade generative KI zusätzliche Prozessschritte, die wesentlich auf menschliche Arbeitskraft für die Annotation von Text- und Bilddaten benötigt werden. So setzte die Firma OpenAI zum Beispiel kenianische Arbeiter für weniger als 2 Dollar pro Stunde ein, um problematischen Ausgaben von ChatGPT entgegenzuwirken (Perrigo 2023).

Die KI-Guidelines des BKA

Der Einsatz von (generativer) KI soll daher mit Bedacht gehandhabt werden und bestimmten Leitlinien folgen, um einen verantwortungsvollen Gebrauch zu sichern. Das österreichische Bundeskanzleramt (BKA) hat in diesem Zusammenhang 8 Grundprinzipien für den Einsatz digitaler Informations- und Gestaltungsangebote im öffentlichen Dienst definiert, die auch für den Umgang mit generativer KI gelten (BKA 2026)⁴. Diese Grundsätze basierend auf diesem Praxisleitfaden und beinhalten:



1. Nur (datenschutz)rechtlich unbedenkliche und, sofern nicht gesetzlich bzw. stellenspezifisch definiert, freie Inhalte verwenden!
2. Bei der Verarbeitung jeglicher personenbezogenen Daten, die spezifische rechtliche Zulässigkeit des Vorhabens prüfen!
3. KI als Unterstützung verwenden, Informationen kritisch hinterfragen und Entscheidungen selbst treffen!
4. KI lediglich in jenen Bereichen einsetzen, in denen man über fachliche Expertise verfügt!
5. Den Einsatz von KI transparent machen und kennzeichnen!
6. Die Ergebnisse von KI auf Qualität prüfen und ob diese ethisch und moralisch unbedenklich sind und dadurch keine Personen und Personengruppen diskriminiert werden!
7. KI gezielt und in einem vertretbaren Ausmaß, möglichst ökologisch sowie nachhaltig nutzen!
8. Dienstliche und, sofern zulässig, private Nutzung stets trennen!

4 <https://oeffentlicherdienst.gv.at/verwaltungsinnovation/public-management-und-governance/digitale-verwaltung/ki-guidelines/>

Auch der Magistrat der Stadt Wien hat sich in seinem „Kompass für den behördlichen Umgang mit generativer KI“ (Stadt Wien 2024)⁵ mit der verantwortungsvollen Nutzung von generativer KI befasst. In diesen Leitlinien wird betont, wie wichtig es ist, neue Technologien auf der Grundlage eines wertebasierten und verantwortungsvollen Verwaltungshandelns zu nutzen und sich an den Prinzipien des digitalen Humanismus zu orientieren.

Tabelle 2 illustriert drei Beispiele für gute und weniger empfehlenswerte Praktiken für den Einsatz generativer KI. Die Beispiele umfassen (1) die Erstellung kreativer und visueller Inhalte, (2) die Analyse und/oder Zusammenfassung von Inhalten und (3) die Erstellung von Inhalten für Meetings und/oder Projekte.

Tabelle 2: Good und Bad Practice Beispiele für die Anwendung generativer KI

	Empfohlen 	Nicht empfohlen 
Kreative Inspiration und visuelle Inhalte generieren	<ul style="list-style-type: none"> • Erstellung nicht-sensibler visueller Inhalte für Präsentationen, Marketingmaterialien • Sicherstellen, dass KI-generierte die zugrundeliegenden Daten oder Informationen korrekt darstellen 	<ul style="list-style-type: none"> • KI zur Bilderstellung in Kontexten verwenden, in denen visuelle Genauigkeit für die Vermittlung einer Botschaft entscheidend ist (z. B. offizielle Berichte)
Analyse / Zusammenfassung von Inhalten	<ul style="list-style-type: none"> • Erstellung schneller Analysen oder Zusammenfassungen von nicht sensiblen, öffentlich zugänglichen Texten / Berichten / Publikationen • Die Ausgabe vor der Verwendung auf ihre Richtigkeit überprüfen 	<ul style="list-style-type: none"> • Vertrauliche oder geschützte Dokumente zur Analyse / Zusammenfassung durch KI verwenden • Sich ausschließlich auf KI-Analysen / Zusammenfassungen verlassen
Erstellen von Inhalten für Meetings / Projekte	<ul style="list-style-type: none"> • Erstellung von allgemeinen / generischen Vorlagen für z.B. Projektpläne oder wiederkehrende Meetings, ohne interne und nicht veröffentlichte Informationen preiszugeben 	<ul style="list-style-type: none"> • Angabe von sensiblen Informationen wie Projektnamen, Mitarbeiterdetails oder internen Inhalten

5 <https://digitales.wien.gv.at/ki-kompass-fuer-bedienstete-der-stadt-wien/>

KI-generierte Inhalte erkennen

Die aktuellen Fortschritte im Bereich der generativen KI ermöglichen die mühelose Erstellung von Texten, Bildern sowie Audio- und Videoclips. Doch gleichzeitig stellen diese technologischen Fortschritte auch neue Herausforderungen dar. Immer häufiger nutzen Akteure die Möglichkeiten der generativen KI für betrügerische Zwecke oder um gezielt Desinformationen zu verbreiten und so die öffentliche Meinung zu manipulieren oder das Vertrauen in legitime Informationsquellen zu untergraben. Die Fähigkeit, KI-generierte Inhalte zuverlässig zu erkennen, wird daher zunehmend zentral, um die Integrität der digitalen Kommunikation zu wahren und potenziellen Missbrauch frühzeitig zu identifizieren.

Deepfakes werden seit einigen Jahren auf ihren potenziellen Missbrauch untersucht, der die Verbreitung von Fake News, die Manipulation von Wahlen, die Verleumdung von Personen, die Untergrabung des Vertrauens in die Medien und vieles mehr umfasst. Diese Risiken bedrohen den demokratischen Diskurs, die gesellschaftliche Stabilität und können außerdem psychische Schäden verursachen. Um diesen Bedenken Rechnung zu tragen, nimmt die EU im AI Act Stellung und definiert Deepfakes in Artikel 3 Ziffer 60 als synthetische oder manipulierte Bild-, Audio- oder Videoinhalte, die täuschend echt erscheinen und existierenden Personen, Orten, Objekten oder anderen Ereignissen oder Entitäten ähneln. Der AI Act schreibt vor, dass die Betreiber von KI-Systemen zur Erzeugung von Deepfakes die künstliche Natur durch eine klare Kennzeichnung der Öffentlichkeit gegenüber offenlegen müssen. Es gibt jedoch zwei Ausnahmen: Strafverfolgungsbehörden sind nicht verpflichtet, die Verwendung von Deepfakes bei der Untersuchung von Straftaten offenzulegen und kreative Werke (Bellettristik, Satire, Kunst) unterliegen ebenfalls nur begrenzten Offenlegungspflichten. Generell stellt sich hier die Frage, ob Akteure mit zweifelhaften Absichten solche Offenlegungsvorgaben tatsächlich befolgen würden. Umso wichtiger ist es daher, dass möglichst viele Menschen die Kompetenz besitzen, gefälschte Texte, Bilder und Videos anhand bestimmter Merkmale zu erkennen. Zudem wird das Hinterfragen von Information und die eigenständige Überprüfung der Quellen zunehmend unverzichtbar werden.

Wissen: Deepfakes

Insbesondere LLMs wie ChatGPT haben die Fähigkeit erlangt, Texte zu generieren, die in ihrer Qualität und Struktur so ausgereift sind, dass sie oft nicht mehr so einfach als maschinell erstellt identifiziert werden können. Um KI-generierten Inhalte zu erkennen, müssen daher bestimmte Merkmale verstanden werden, die die Modelle derzeit noch aufweisen. Folgend werden einige grundlegende Tipps zur Erkennung von KI-generierten Inhalten vorgestellt:

Texte

KI-generierte Texte weisen oft sich wiederholende Muster auf und haben keinen Tiefgang. Sie liefern oberflächliche Informationen ohne ein differenziertes Verständnis. Das liegt daran, dass die Modelle dazu neigen, ähnliche Inhalte zu wiederholen, ohne die Variabilität, die man bei menschlichen Texten findet. Darüber hinaus können KI-generierte Texte logische Ungereimtheiten oder abrupte Themenwechsel enthalten, da KI-Modelle Schwierigkeiten haben können, kohärente Erzählungen auszugeben. Der Stil der Inhalte kann oft auch übermäßig formal und einheitlich sein, ohne die für menschliche Autoren typische stilistische Vielfalt. Darüber hinaus sind generative KI-Modelle dafür bekannt, Fakten oder Quellen zu erfinden, die es in der Realität gar nicht gibt. Diese erfundenen Informationen können überzeugend klingen, sind jedoch bei genauer Überprüfung falsch oder irreführend.

Bilder

Mit KI erstellte Bilder lassen sich derzeit noch relativ einfach identifizieren, da sie häufig einen künstlichen oder comichaften Stil aufweisen. Die Bilder sind meist in Details ungenau, wie beispielsweise in der Darstellung von Texturen, Lichtverhältnissen oder anderen natürlichen Elementen. Ein charakteristisches Merkmal, das auf KI-generierte Bilder hinweisen kann, ist die Darstellung von Händen. KI-Modelle haben nach wie vor Schwierigkeiten, menschliche Finger und Handstellungen realistisch darzustellen. Es lohnt sich auch, die Proportionen der abgebildeten Objekte sowie Schattierungen und mögliche Verzerrungen im Hintergrund genauer zu betrachten, da diese bei KI-generierten Bildern oft unnatürlich wirken.

Video

Bei KI-generierten Videos sollte man auf die Bewegungsabläufe von Personen und Objekten achten. Häufig treten in KI-generierten Videos Unregelmäßigkeiten auf, wie etwa ruckartige Bewegungen, unnatürliche Übergänge oder inkonsistente Animationen. Auch der Hintergrund kann verräterisch sein. Unstimmigkeiten wie flackernde oder verschwommene Bereiche, unpassende Schatten oder sich verändernde Objekte im Hintergrund können auf ein KI-generiertes Video hindeuten.

Audio

KI-generierte Stimmen haben mittlerweile ein sehr hohes Maß an Authentizität erreicht, was es besonders schwierig macht, solche Audio-Fälschungen zu erkennen. Dennoch gibt es bestimmte akustische Merkmale, die auf eine KI-Erzeugung hinweisen können. Ein metallischer, monotoner Klang oder eine unnatürlich gleichmäßige Sprachmelodie kann ein Indiz dafür sein. Ebenso können falsche Aussprache, ungewöhnliche Betonungen oder eine mechanisch wirkende Sprechweise Hinweise auf eine KI-generierte Stimme sein. Auch unnatürliche Geräusche oder Verzögerungen in der Artikulation, die den natürlichen Fluss der Sprache stören, können darauf hindeuten, dass es sich um eine maschinell erzeugte Aufnahme handelt.

Weitere hilfreiche Tipps zur Erkennung von KI-generierten Inhalten finden sich auf dem österreichischen IKT-Sicherheitsportal.⁶ Dort werden zudem nützliche digitale Tools vorgestellt, die dabei helfen können, KI-generierte Texte, Bilder und Videos zu identifizieren

Mit der laufenden Weiterentwicklung von KI wird es immer schwieriger, Fälschungen zuverlässig zu erkennen. Ein besonderer Fall sind sogenannte Hybridfälschungen, bei denen teilweise echte Elemente verwendet werden, die in einem neuen, verfälschten Kontext präsentiert werden. Solche Fälschungen sind besonders schwer zu durchschauen, weil sie auf realen Ereignissen oder Bildern basieren, die mit neuen, falschen Informationen vermischt werden. Daher wird es in Zukunft immer wichtiger, den Inhalten im Internet nicht blind zu vertrauen. Stattdessen sollte man die Plausibilität und den Kontext der Informationen stets kritisch hinterfragen und überprüfen, ob die dargestellten Fakten in ihrer Gesamtheit stimmig sind.

Wissen: Hybride Bedrohungen durch KI-generierte Fake News

7.2 KI in der öffentlichen Verwaltung und Auswirkungen auf die Bevölkerung

Der Einsatz von KI-Anwendungen in der öffentlichen Verwaltung hat nicht nur Auswirkungen auf die Verwaltungsbediensteten, sondern auch auf die breite Öffentlichkeit. Ein Grund dafür ist, dass der Einsatz von KI in der öffentlichen Verwaltung zur Einführung neuartiger oder automatisierter Dienstleistungen bzw. Bearbeitungsformen für Bürgerinnen und Bürger 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 Bürgerinnen und Bürger 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 Bürgerinnen und Bürger, die Ermittlung des Mehrwerts für die Bürgerinnen und Bürger 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 Bürgerinnen und Bürger lassen sich im Zusammenhang mit dem Einsatz von KI in der öffentlichen Verwaltung vor allem die folgenden Vorteile hervorheben:

⁶ <https://www.onlinesicherheit.gv.at/Services/News/KI-Inhalte-erkennen.html>

(1) Interaktion und Kommunikation mit den Bürgerinnen und Bürger

Die Einführung von KI in der öffentlichen Verwaltung kann die Art und Weise, in der Verwaltungsbedienstete mit den Bürgerinnen und Bürger interagieren, verändern und möglicherweise verbessern. Insbesondere können virtuelle Assistenten den Bürgerinnen und Bürger 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 Bürgerinnen und Bürger zu erleichtern.⁷ Wenn Kamu eine bestimmte Frage nicht versteht, wird den Benutzerinnen und Benutzer angeboten, sie direkt an eine/n Mitarbeiterinnen und Mitarbeiter weiterzuleiten, welche/r dann ebenfalls über das Chat-Protokoll verfügt. Der Chatbot ist in Englisch und Finnisch verfügbar.

Anwendungsfall: Chatbot Kamu in Finnland

(2) Effizienz und Personalisierung

KI kann Routineaufgaben automatisieren und so die manuelle Arbeit von Verwaltungsbediensteten reduzieren, was zu einer schnelleren Bereitstellung von Dienstleistungen für die Bürgerinnen und Bürger führen kann. Außerdem kann KI dazu beitragen, die Bedürfnisse der einzelnen Bürgerinnen und Bürger schneller zu erkennen, um 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 helfen diejenigen zu identifizieren, die Unterstützung benötigen.

(3) Zugänglichkeit und Chancengleichheit

KI kann dazu beitragen, öffentliche Dienstleistungen für Bürgerinnen und Bürger, die beispielsweise in abgelegenen Gebieten leben, durch die Erbringung von verwaltungsbezogenen Dienstleistungen aus der Ferne zugänglicher zu machen. Außerdem können KI-Anwendungen Bürgerinnen und Bürger unterstützen, etwa beim Ausfüllen von Formularen oder beim Überwinden von Sprachbarrieren bei Anfragen.

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

Neben ihren vielfältigen Vorteilen birgt die fortschreitende Entwicklung von KI auch das Potenzial, eine digitale Kluft (engl. „*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 Bürgerinnen und Bürger in öffentliche KI-Anwendungen besteht darin, dass es der KI an demokratischer Rechenschaftspflicht mangelt (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 Bürgerinnen und Bürger angemessene Begründungen für die Entscheidungen der öffentlichen Verwaltung erhalten. Darüber hinaus verringert die Einführung algorithmischer Entscheidungsunterstützung 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 (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 Bürgerinnen und Bürger, 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 Bürgerinnen und Bürger 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 in der öffentlichen Verwaltung 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-in- bzw. Opt-out-Möglichkeiten*, welche Bürgerinnen und Bürger Möglichkeit geben, KI-basierte Dienste an- oder abzulehnen.

Co-Kreation und Partizipation

Die Zusammenarbeit mit möglichst vielen Interessensgruppen wird als Schlüssel für Akzeptanz von KI-Anwendungen betrachtet (Madan und Ashok 2022; Gesk und Leyer 2022). Vor allem durch die Einbeziehung der breiten Öffentlichkeit kann die öffentliche Verwaltung sicherstellen, dass die KI-Systeme gesellschaftlichen Bedürfnissen entsprechen und mit den öffentlichen Werten und Zielen in Einklang stehen. Pilotprojekte oder Reallabore/Sandboxes (mehr dazu in Abschnitt 12) 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 Bürgerinnen und Bürger in den Entwicklungsprozess einzubeziehen. Auf diese Weise können gleichzeitig Fähigkeiten und Wissen für den Umgang mit KI-Anwendungen aufgebaut werden, um das Bewusstsein für potenzielle ethische und rechtliche Fragen zu schärfen.

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 Bürgerinnen und Bürger 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 und Bürgerinnen 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 Bundeskanzleramt (BKA), 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 Bürgerinnen und Bürger befasst und beschreibt, was digitale Kompetenzen konkret beinhalten.

Auf dieser Basis wurde im November 2024 der Nationale Referenzrahmen für Digitale Kompetenzen in Österreich veröffentlicht. In 2025 liegt der Schwerpunkt der Weiterentwicklung der Kompetenzmodelle sowohl in Österreich als auch in Europa auf KI Kompetenzen. Es sollen Lernergebnisse entwickelt werden, die KI Kompetenzen besser greif- und messbar machen.

KI Kompetenzen bei Bürgerinnen und Bürgern aufzubauen, ist außerdem ein wichtiger Schwerpunkt der Digitalen Kompetenzoffensive für die kommenden Jahre. Dafür wird eben ein KI Kompetenz-Programm entwickelt, das auf bereits etablierten Workshops für digitale Basiskompetenzen für die Bevölkerung aufbaut. Das Ziel ist es, die Bürgerinnen und Bürger dazu zu befähigen, fundierte Entscheidungen über KI Systeme zu treffen und u.a. auch eine Basis für KI-Systeme in E-Government Anwendungen zu schaffen.

Opt-in- beziehungsweise Opt-out Möglichkeiten

Eine weitere wichtige Maßnahme ist die Bereitstellung von Opt-in- beziehungsweise Opt-out-Möglichkeiten für Bürgerinnen und Bürger bei der Nutzung von KI-basierten Diensten. Opt-in- beziehungsweise 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 Bürgerinnen und Bürger die Kontrolle über ihre Privatsphäre

behalten und bei Bedarf personalisierte Unterstützung erhalten können. Die öffentliche Verwaltung sollte darüber hinaus darauf achten, dass die Bürgerinnen und Bürger über ihr Recht auf Ablehnung von KI-basierten Diensten informiert sind. Dies kann durch die Bereitstellung klarer Informationen über das Opt-in beziehungsweise Opt-out erreicht werden, beispielsweise durch benutzerfreundliche Schnittstellen.

7.3 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 um die Frage, welche Auswirkungen und ethische Fragestellungen mit der Entwicklung und Nutzung von KI-Systemen verbunden sind. Im vergangenen Jahrzehnt haben insbesondere datengetriebene Methoden des Maschinellen Lernens in der Künstlichen Intelligenz weite Verbreitung erfahren. Die Erfolge in verschiedenen Anwendungsbereichen, wie der medizinischen Diagnose, und Entwicklung neuer Medikamente im Gesundheitswesen, autonomen Fahrsystemen, Bilderkennung und automatischer Übersetzung, haben hohe Erwartungen geweckt und zu signifikanten privaten Investitionen geführt.⁸ Aufgrund des erheblichen Ressourcenaufwands, der durch die Herstellung und Nutzung dieser Systeme verursacht wird, sind jedoch auch ökologische Aspekte in den letzten Jahren zunehmend in den Fokus der öffentlichen Wahrnehmung gerückt (Hacker 2024).

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

8 Allein in den USA wurden im Jahr 2023 private Investitionen im Umfang von 62,5 Mrd. EUR in KI getätigt, gefolgt von China mit 7,3 Mrd. EUR, und die EU und das Vereinigte Königreich (UK) zogen im Jahr 2023 zusammen private Investitionen im Wert von 9 Mrd. EUR an (Madiega et al. 2024).

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.

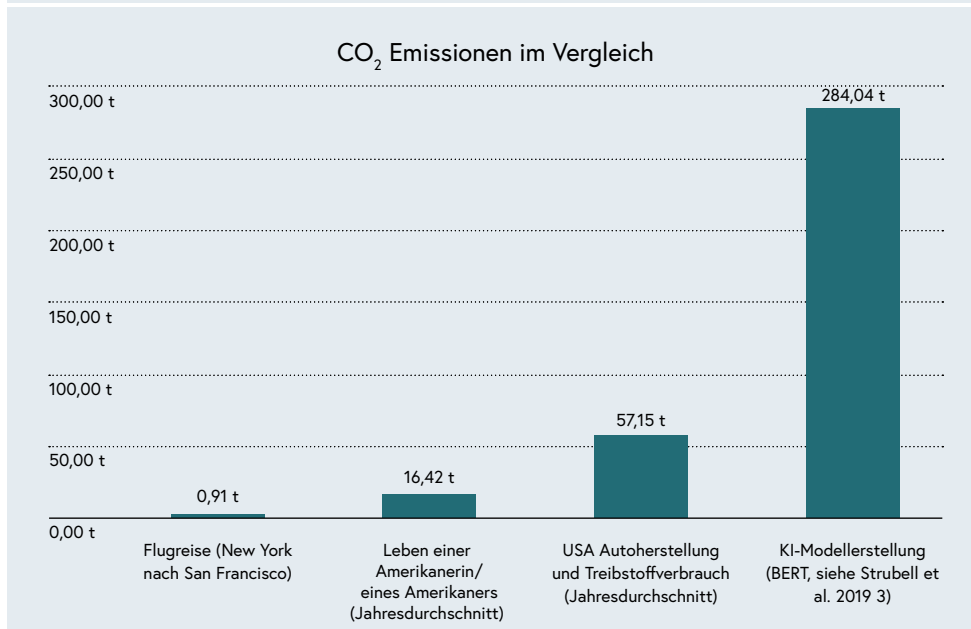


Abbildung 6: Vergleich der Treibhausgasemissionen einer Flugreise, des Lebens eines Menschen in den USA im Jahresdurchschnitt, eines Autos während der Lebensdauer und der Erstellung eines KI-Modells

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

Die zur Erstellung von Modellen benötigte Energie repräsentiert jedoch nur einen Teil der insgesamt durch die Entwicklung und Verwendung von KI-Systemen verursachten Ressourcenverbrauchs.

Für den Betrieb eines KI-Systems fallen vom Zeitpunkt der Entwicklung über den Betrieb bis hin zur Aussonderung des Dienstes Energie- und Ressourcen die für Hardwareinfrastruktur (Wartung, Kühlung) an (Rohde et al. 2021, 43). Darüber hinaus ist auch die Rohstoffgewinnung und Herstellung benötigter Hardware mit erheblichen Umweltauswirkungen verbunden.

Um die Auswirkungen auf die Umwelt einzuschätzen, ist also eine umfassende Bewertung der Dienste notwendig, welche sowohl die energie- und ressourcenintensive Erstellung der Modelle im Rechenzentrum als auch die Server-Kapazitäten für die Nutzung der Dienste berücksichtigt. Im Einzelnen betrifft dies die folgenden Aspekte:

- Ressourcen, die für mehrere Trainingsläufe erforderlich sind, d. h. die Rechenleistung, die insbesondere bei komplexen und großen Datensätzen für das Training von KI-Modellen benötigt wird,
- Prozesse für die Generierung von Antworten und die Erledigung komplexer Aufgaben, die GPUs (Grafikprozessoren) benötigen,
- Benutzeroberflächen für Nutzerinnen und Nutzer, die häufig über eine Webhosting-Infrastruktur bereitgestellt werden,
- die Übertragung von Daten über Netzwerke, insbesondere bei datenintensiven Anwendungen,
- von Benutzerinnen und Benutzer verwendete Endgeräte und deren Energieverbrauch,
- durch Aussonderung eines KI-Dienstes anfallender Aufwand, z. B. für die Entsorgung und Wiederverwertung,
- der Lebenszyklus der Hardware, also die gesamte Lebensdauer von IT-Hardware, einschließlich ihrer Herstellung, Nutzung und Entsorgung,
- die Art der Energiequellen, die zur Stromversorgung und zur Kühlung in Rechenzentren genutzt werden.

Grafik-Prozessoren (engl. „*Graphics Processing Units [GPUs]*“): GPUs sind spezialisierte Computerchips, die für die schnelle Bearbeitung und Darstellung von Grafiken entwickelt wurden. Sie verfügen über viele parallele Recheneinheiten, die es ihnen ermöglichen, große Datenmengen und komplexe Berechnungen parallel zu verarbeiten. Sowohl für die Erstellung als auch während der Verwendung benötigt generative KI eine sehr große Anzahl von Berechnungen durchführen, für die in der Regel spezielle Grafikprozessoren (GPUs) verwendet werden.

Wissen: Grafik-Prozessoren (GPUs)

Das Unternehmen Hugging Face hat in einer Veröffentlichung den Ressourcenverbrauch seines Modells BLOOM mit 25 Tonnen und unter Einbeziehung der Infrastruktur mit 50 Tonnen CO₂ angegeben (Luccioni et al. 2022). Und in einer aktuellen Studie hat ein französisches Forschungsteam den Ressourcenverbrauch über den gesamten Lebenszyklus eines generativen KI-Dienstes am Beispiel des KI-Bildgenerators „Stable Diffusion“ durchgeführt (Stable Diffusion wurde von Forschern der Ludwig-Maximilians-Universität München unter Verwendung von Trainingsdaten von gemeinnützigen Organisationen

als Open Source Projekt entwickelt, www.diffus.me). Laut dieser Studie verbraucht der Dienst im Laufe eines Jahres 360 Tonnen CO₂-Äquivalente (CO₂e). Darüber hinaus wird der Ressourcenverbrauch in Bezug auf die benötigten seltenen Rohmaterialien mit demjenigen der Produktion von 5659 Smartphones verglichen und der Energieverbrauch auf 2,48 Gigawattstunden geschätzt (Berthelot et al. 2024), was dem durchschnittlichen Jahresverbrauch von 71 Wiener Haushalten entspricht.⁹ Zum Zwecke der umfassenden Bewertung der durch KI-Services verursachten Umweltauswirkungen schlägt das Forschungsteam daher eine Lebenszyklusanalyse (engl. „*Life Cycle Assessment, LCA*“) vor, welche die gesamte Bandbreite der Umweltauswirkungen einbezieht, die einem KI-Produkt bzw. einer KI-Dienstleistung zuzuschreiben sind (Berthelot et al. 2024).

Forscher des MIT haben in einem aktuellen Bericht über Klima-Auswirkungen und Nachhaltigkeit von KI-Services bemängelt, dass die aktuellen Bemühungen um Entwicklung und Betrieb nachhaltiger KI-Services sich hauptsächlich auf die Verbesserung der Effizienz der Rechenzentren und die dadurch erwartete Reduzierung des Energieverbrauchs konzentrieren, sich jedoch unter dem Strich eine stetig expandierende Nachfrage nach Rechenleistung inklusive höhere Ressourcen-Nachfrage ergibt (Bashir et al. 2024). Dies deckt sich mit der Argumentation, dass möglicherweise gerade die Effizienzsteigerungen durch Verbesserung von Hardware, Algorithmen oder Software-Architektur eine höhere Nachfrage generieren, was auch mit dem Jevons-Paradoxon in Verbindung gebracht wird, welchem zufolge technischer Fortschritt, der die effizientere Nutzung eines Rohstoffes erlaubt, letztlich zu einer erhöhten Nutzung dieses Rohstoffes führt, anstatt sie zu senken (de Vries 2023).

⁹ Quelle: Statistik Austria, <https://www.wien.gv.at/spezial/energiebericht/energie-von-der-gewinnung-bis-zur-nutzung/energieverbrauch-eines-wiener-haushalts/>

Die Entwicklung und Verwendung von KI-Services hängt in hohem Maße von der Verfügbarkeit leistungsfähiger Rechenzentren ab. Masanet et al. führen dazu an, dass Schätzungen zufolge der weltweite Energiebedarf von Rechenzentren von 194 TWh im Jahr 2010 auf 204 TWh im Jahr 2018, und schließlich auf 460 TWh im Jahr 2022 gestiegen ist (Masanet et al. 2020). Laut Bericht der internationalen Energiebehörde wird der weltweite Stromverbrauch von Rechenzentren im Jahr 2026 zwischen 620 und 1.050 TWh (konkrete Schätzung: 800 TWh) liegen (IEA 2024).

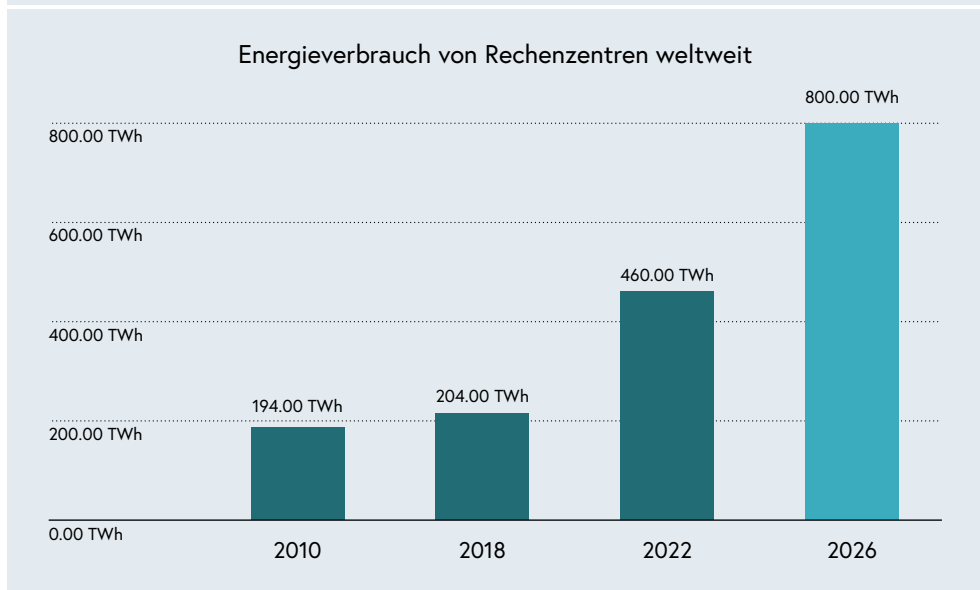


Abbildung 7: Die Entwicklung des weltweiten Energiebedarfs von Rechenzentren mit einer Prognose für 2026 laut der internationalen Energieagentur (IEA)

Es ist anzunehmen, dass die zunehmende Nachfrage nach KI diesen Trend maßgeblich fördern wird.

Anwendungsfall: Die Entwicklung des Energieverbrauchs von Rechenzentren

Um angemessene und zielgerichtete Regelungen in Bezug auf KI-Dienste zu ermöglichen, schlagen die Forscherinnen und Forscher eine umfassende und vergleichende Bewertungsmöglichkeit vor, die den potenziellen gesellschaftlichen Nutzen von generativer KI gegen die Kosten eines ungebremsten Wachstums derselben abwägt (Bashir et al. 2024, 5).

Insbesondere dem Energieverbrauch von Diensten generativer KI wird angesichts der zu erwartenden Steigerung der Nachfrage vermehrt Aufmerksamkeit geschenkt (siehe z.B. Berthelot et al. 2024). Eine vollumfängliche Integration LLM-basierter generativer KI zur Beantwortung natürlichsprachlich formulierter Anfragen würde den Energieverbrauch signifikant steigern. In einem Interview sagte der Vorsitzende John

Hennessy von Alphabet – die Muttergesellschaft zu welcher unter anderem Google gehört – gegenüber Reuters, dass die Integration generativer KI ungefähr zehnmal so viel Kosten verursachen könnte wie eine Standard-Schlüsselwortsuche (Dustin und Nellis 2023). Dabei ist zu berücksichtigen, dass bei Google derzeit ca. 9 Milliarden Suchanfragen täglich verarbeitet werden, und dass KI in der Art wie sie von OpenAIs ChatGPT bereitgestellt wird, für die Integration in Google über 500.000 NVIDIA A100 HGX Server¹⁰ mit insgesamt über 4 Millionen GPU-Einheiten benötigen würde (de Vries 2024, 2192) – eine Größenordnung, die derzeit selbst angesichts der Investitionsmacht großer IT-Konzerne nicht zu realisieren ist. Hinzu kommt, dass die dafür erforderliche Hardware, die derzeit überwiegend von NVIDIA in Form der GPUs mit einem Marktanteil von über 90% bereitgestellt wird, in einem solchen Ausmaß nicht kurzfristig geliefert werden kann (de Vries 2024, 2193).

Es gibt nur grobe Schätzungen dazu, welchen Anteil am Energiebedarf der Erstellung und Verwendung von KI zuzuschreiben ist. Angaben dazu basieren meist auf der Selbstauskunft der Unternehmen, wie zum Beispiel, dass Google durch eigene Energieeffizienzmaßnahmen den Anteil von Machine Learning – bzw. im weiteren Sinne von KI – am Gesamtenergieverbrauch unter 15% halten konnte (Patterson et al. 2022).

Unabhängig von diesen Maßnahmen ist jedoch 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 technologische Neuerung und Steigerung der Effizienz im Vordergrund stehen.

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

10 Der Einzelstückpreis eines solchen Servers variiert je nach Konfiguration, Anbieter, Region und spezifischen Anforderungen, kann jedoch mit mehreren A100-GPUs, Kühlung, CPU, Speicher und weiteren Komponenten derzeit zwischen 100.000,- und 300.000,- Euro kosten.

- 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.

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 ethischen Nutzung von KI 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).

7.4 Digitale Souveränität in der öffentlichen 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 Bürgerinnen und Bürger und der öffentlichen Institutionen zu schützen. Sie ist die Grundlage für kompetentes und vertrauenswürdiges 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 Bürgerinnen und Bürger 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“.

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 12). Die Zertifizierung von Daten und KI-Modellen kann in Bezug auf digitale bzw. Datensouveränität unterstützend wirken (mehr zu Zertifizierungen ist ebenfalls in Abschnitt 12 zu finden).

7.5 Beschaffung

Die größte Wirkung kann bei der Stärkung der digitalen Souveränität durch die Vorgabe von Bedingungen und die Definition zu erfüllender Kriterien im Rahmen von Beschaffungsvorgängen erzielt werden.

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 Klein-

betrieben davon profitieren (BMK und BMDW 2021, 56). Die öffentliche Verwaltung kann dabei durch ihre Vergabetätigkeit eine Vorbildwirkung entfalten.

Um KI verantwortungsvoll zu beschaffen, müssen die Verwaltungsbediensteten jedoch über das Wissen und die Ressourcen verfügen, die für einen solchen Beschaffungsprozess notwendig sind. Bei der Beschaffung von KI-Anwendungen für KI-Projekte ist die Innovationsfördernde Öffentliche Beschaffung–Servicestelle (IÖB) ein Ansprechpartner für die Verwaltungsbediensteten. Hier bietet 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 10.2) ausgeführt werden, schon zu Beginn von Entwicklungsprozessen, wie z. B. im Rahmen des „Ethics by Design“-Ansatzes vorgeschlagen, bedacht werden.

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 Entwicklerinnen und Entwickler, Ethikerinnen und Ethiker, Sozialwissenschaftlerinnen und -wissenschaftler und Auftraggeberinnen und Auftraggeber bzw. Anwenderinnen und Anwender 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“

Ein Good-Practice-Beispiel hinsichtlich des Beschaffungswesens durch die öffentliche Verwaltung stammt aus Amsterdam. Die Stadt Amsterdam legt beim Zukauf von KI-Anwendungen vertraglich einen Rahmen für die Informationen fest, die Anbieter bereitstellen müssen. Auf diese Weise kann die Stadt 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: Einblicke in die Funktionsweise der KI-Anwendung.
- Verfahrenstransparenz: Informationen über den Zweck, Entwicklungsprozess einschließlich verwendeter Daten und Maßnahmen zur Qualitätssicherung und Risikominderung

- Erklärbarkeit: Transparente Darstellung der KI-Entscheidungsprozesse, sodass z. B. Bürgerinnen und Bürger gegebenenfalls rechtlich gegen KI-generierte Entscheidungsvorschläge vorgehen können.

Welche Art von Informationen bereitgestellt und angefordert werden können, hängt aber auch davon ab, welche Art von KI-System beschafft werden soll. Das Digital Regulation Cooperation Forum (DRCF), eine Kooperation von vier britischen Regulatoren im digitalen Bereich, zählt in seinem 2023 veröffentlichten Bericht über Transparenz bei der Beschaffung algorithmischer Systeme (kurz „KI-Beschaffungsbericht“)¹¹ folgende Formen der Beschaffung von algorithmischen Systemen auf:

- Erwerb eines vorgefertigten und einsatzbereiten Systems („Standard-System“),
- Verwendung eines Anbietermodells, bei dem Daten zur Verfügung gestellt werden, um eigene Ausgaben zu generieren,
- Bereitstellung von Tools durch den Anbieter, die es ermöglichen, eigene algorithmische Systeme zu entwickeln (z. B. „no code“ oder „low code“ Lösungen), die auch Personen mit wenig Programmierkenntnissen die Entwicklung eigener Systeme erlauben,
- Erwerb von Datensätzen vom Anbieter, die zum Training eigener algorithmischer Systeme genutzt werden können,
- Verwendung eines kostenlosen, quelloffenen algorithmischen Systems (Open Source).

Wenn „General Purpose AI (GPAI)“ Modelle beschafft werden, die in KI-Systeme integriert werden können, verlangt der AI Act von den Anbietern eine technische Dokumentation, die den Behörden zur Verfügung gestellt werden muss sowie eine Zusammenfassung der für das Training verwendeten Daten (Artikel 53 Absatz 1 AI Act). Anbieter von Hochrisiko-KI-Systeme müssen den Betreibern wiederum detaillierte Gebrauchsanweisungen zur Verfügung stellen (Artikel 13 AI Act). Diese Anleitung soll den Betreibern unter anderem dabei helfen, das für sie passende System zu wählen, die beabsichtigten und ausgeschlossenen Anwendungszwecke zu verstehen sowie das System richtig zu nutzen. Um die Lesbarkeit und Zugänglichkeit der in den Gebrauchsanweisungen enthaltenen Informationen zu verbessern, sollten ggf. erläuternde Beispiele – beispielsweise zu den Einschränkungen und den beabsichtigten und ausgeschlossenen Verwendungszwecken des KI-Systems – aufgenommen werden.

11 <https://www.gov.uk/government/publications/transparency-in-the-procurement-of-algorithmic-systems-findings-from-our-workshops>

Allgemeine Empfehlungen für die Vorbereitung und Planung von Beschaffungsprozessen von KI-Anwendungen

Die Beschaffung von KI-Systemen stellt für die öffentliche Verwaltung eine besondere Herausforderung dar, da vielfältige rechtliche, ethische und technische Aspekte berücksichtigt werden müssen.

Beispiele für Beschaffungsvorgänge von KI-Lösungen finden sich in den Niederlanden, wie zuvor geschildert, den USA (Hickok 2024) oder im Vereinigten Königreich, wo die Cabinet Office, die Office for Artificial Intelligence (OAI) und das World Economic Forum¹² einen gemeinsamen Leitfaden veröffentlicht haben, der eine Reihe von Leitprinzipien für die Beschaffung von KI-Technologien sowie Einblicke in die Bewältigung von Herausforderungen bietet, die bei der Beschaffung auftreten können. Basierend auf diesen Vorschlägen lassen sich vier zentrale Empfehlungen ableiten, die bei der Planung und Vorbereitung von Beschaffungsprozessen für KI-Anwendungen in der öffentlichen Verwaltung zu beachten sind.

1. Zielgerichtete Anforderungen für den Einsatz der KI-Anwendung ableiten

Der erste Schritt im Beschaffungsprozess besteht darin, den Bedarf an dem KI-System klar zu definieren und Anforderungen präzise zu formulieren. Eine exakte Problemdefinition bildet die Basis für erfolgreiche KI-Projekte. Es muss geklärt werden, welche Probleme gelöst und welche Ziele durch den Einsatz der KI-Anwendung erreicht werden sollen. Beispielsweise könnte eine Abteilung ein KI-System benötigen, das große Datenmengen analysiert, um Entscheidungsprozesse zu unterstützen. Anforderungen können aus der Problemstellung abgeleitet und detailliert beschrieben werden, um Missverständnisse zu vermeiden und die Grundlage für die Auswahl des richtigen Systems zu schaffen. Eine ungenaue oder unvollständige Anforderungsanalyse erhöht das Risiko von Fehlbeschaffungen. Außerdem muss klar sein, welche Anforderungen an Systemintegrationen notwendig wären, damit das KI-System in bestehende interne IT-Strukturen eingebettet werden kann. Die gesammelten Anforderungen sollten am besten in einem Anforderungskatalog festgehalten werden.

2. Ethische Designkriterien festlegen

Bei der Beschaffung von KI-Systemen, die extern entwickelt werden, besteht das Risiko der sogenannten Anbieterbindung (engl. „Vendor Lock-in“). Dieses Risiko entsteht, wenn die Nutzung eines KI-System stark von einem Anbieter abhängig ist und mit intransparenten „Black-Box“-Algorithmen arbeitet. „Black-Box“-Algorithmen führen Entscheidungen auf Grundlage von Prozessen und/oder Daten durch, die für Nutzer und Nutzerinnennicht vollständig einsehbar oder verständlich sind. Eine solche Situation führt dazu, dass man nur schwer nachvollziehen kann, wie das KI-System zu seinen Ergebnissen gelangt oder wie es mit den eingegebenen Daten arbeitet. Um dieses Risiko zu

12 https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf

reduzieren, ist es notwendig, klare ethische Designkriterien zu definieren. Es muss für den Auftraggeber und die Nutzerinnen und Nutzer möglich sein, die Funktionsweise des KI-Systems zu verstehen. Beispielsweise sollte bei einer automatisierten Entscheidungsunterstützung durch die KI klar sein, welche Faktoren zu dieser Ausgabe des KI-Systems beigetragen haben.

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. Designkriterien wie Transparenz und Nachvollziehbarkeit lassen sich zum Beispiel aus den ethischen Prinzipien der High-Level Expert Group ableiten (HLEG 2018, 2019). Für die Ableitung von Designspezifikationen kann auch der „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“, oder die Checkliste für ethische KI in der Verwaltung herangezogen werden.

Außerdem sollten ethische Prüfungen der in der Verwaltung verwendeten KI-Systeme von Organisationen mit relevanter Expertise gemeinsam mit den jeweiligen Dienststellen durchgeführt werden können, ohne dem anbietendem Unternehmen Schaden durch eine etwaige Lüftung von Betriebsgeheimnissen zuzufügen.

3. Rechtliche Rahmenbedingungen klären

Das Vergaberecht umfasst alle Regeln und Vorschriften, die die öffentliche Hand nun beim Einkauf von Gütern und Leistungen und bei der Vergabe von Konzessionen befolgen muss. Wenn eine öffentliche Stelle KI Software/Systeme beschaffen will, die sie später einsetzen möchte, sind vergaberechtliche Anforderungen zu beachten. Hinzu kommt, dass bereits seit Inkrafttreten der DSGVO die Berücksichtigung des Datenschutzes (wie etwa „*privacy by design*“) bei der Gestaltung der Vergabeunterlagen zur Vorbereitung der späteren Auftragsausführung von wesentlicher Bedeutung ist. Künftig wird dies auch die Anforderungen, die sich aus dem AI Act ergeben, betreffen.

Die Beachtung der Anforderungen die sich aus der DSGVO bzw. dem AI Act ergeben, sollten somit Bestandteil der Leistungsbeschreibung sein.

4. Einen Daten-Governance-Plan erstellen

Daten sind die Grundlage der allermeisten KI-Systeme. Daher ist ein Plan für das Datenmanagement ein wesentlicher Bestandteil der Vorbereitung auf Beschaffung und Einsatz eines KI-Systems. Ein Daten-Governance-Plan sollte festlegen, welche Datenquellen für das KI-System genutzt werden und welche Anforderungen an die Datenqualität gestellt werden. Daten müssen auf Verzerrungen oder Lücken geprüft werden, da sie direkten Einfluss auf die Entscheidungen oder Ausgaben des KI-Systems haben. Verzerrte Daten führen zu fehlerhaften oder unfairen Ausgaben/Entscheidungen, weshalb Maßnahmen zur Sicherstellung der Datenqualität ein zentraler Bestandteil des Beschaffungsprozesses sind. Dabei ist zu beachten, dass Daten immer einen gewissen Grad an Verzerrungen aufweisen, daher sollte gemeinsam mit dem Anbieter klar definiert werden, welcher Verzerrungsgrad akzeptabel ist. Es sollte ebenfalls festgelegt werden, wer Zugang

zu den Daten hat und wie der Zugriff geregelt wird. Dies betrifft nicht nur den Schutz vor unbefugtem Zugriff, sondern auch die Frage, wie die Daten intern und gegenüber Anbietern gehandhabt werden. In diesem Zusammenhang müssen datenschutzrechtliche Bestimmungen beachtet werden, insbesondere die DSGVO (siehe Abschnitt 8.2). Ein gut durchdachter Daten-Governance-Plan trägt also dazu bei, dass das KI-System auf verlässliche und qualitativ hochwertige Daten zugreift und datenschutzkonforme Prozesse gewährleistet sind.

Obwohl die Zusammenarbeit KI-Entwicklerinnen und Entwickler und der öffentlichen Verwaltung noch relativ neu ist, gibt es beispielsweise in Deutschland Beispiele für erfolgreiche Kooperationen. Zum Beispiel hat das KI-Startup Aleph Alpha gemeinsam mit der baden-württembergischen Landesregierung über die Plattform „InnoLab_bw“ einen KI-basierten Text-Assistenten „F13“ entwickelt, der die Verwaltungsbediensteten bei täglichen Aufgaben wie assistierten Recherchefunktionen, Textzusammenfassungen bzw. Herstellung bestimmter Textsorten unterstützt und entlasten soll. Aleph Alpha hat außerdem eine Rahmenvereinbarung mit der bayerischen Landesregierung unterzeichnet, um gemeinsam KI-Systeme für Verwaltungszwecke zu entwickeln.

Anwendungsfall: Kooperation KI-Entwickler und öffentliche Verwaltung

8 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 (ausführlich Mayrhofer und Parycek 2022; Scheichenbauer und Rothmund-Burgwall 2024) 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 und der „*Data Governance Act*“, die beschlossen wurden, um Verantwortlichkeit, Sicherheit und Schutz in den Mitgliedstaaten 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 – aber häufig dennoch für die Interpretation und Implementierung relevante – Leitlinien, Leitfäden, Strategien und Absichtserklärungen zu verstehen sind (siehe Abschnitt 9).

Definition: „*Soft Law*“ und „*Hard Law*“

8.1 Die Grundlage des Verwaltungshandelns

Unabhängig davon, dass dieses Wissen sehr wahrscheinlich bei Leser- bzw. Anwenderinnen und Anwender aus der öffentlichen Verwaltung keine Neuigkeit darstellt, wird der Vollständigkeit halber nachfolgend auf die Grundlagen des Verwaltungshandelns eingegangen.

Verwaltungshandeln hat seine Basis in der Rechtsstaatlichkeit und baut immer auf einer Rechtsgrundlage 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. Artikel 18 Absatz 1 Bundes-Verfassungsgesetz, 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 2021).

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 2021).

Das Verwaltungshandeln lässt sich darüber hinaus in Hoheitsakte im engeren Sinn, das heißt typengebundenes Verwaltungshandeln, z. B. Bescheid, Weisung, Verordnung (also die praktische Anwendung der dem Staat verantworteten Entscheidungsgewalt gegenüber Bürgerinnen und Bürger in einem spezifischem Rechtsgegenstand), 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 Artikel 18 Absatz 1 B-VG unterliegt, da dies in der einschlägigen Literatur teilweise bejaht (Stöger 2014; Feik 2007; Karkulik 2014), verneint (Adamovich-Funk 1987) oder abhängig von der konkreten Ausprägung der schlichten Hoheitsverwaltung differenziert gesehen wird (Mörth 2020). 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.

Sofern also der Einsatz von KI-Systemen durch Behörden im Schutzbereich eines Grundrechts wirkt, muss er am Maßstab der Grundrechte gemessen werden. Dies gilt auch, wenn es sich um schlicht-hoheitliches Verwaltungshandeln bzw. Verwaltungshandeln handelt, das einem außenwirksamen Verwaltungshandeln vor- oder nachgelagert ist (Scheichenbauer und Seidl 2022) worunter KI-Systeme ebenso fallen können und dadurch zum schlicht-hoheitlichen Verwaltungshandeln zählen (Schneeberger 2024a). In diesen Fällen muss also in Verbindung mit Artikel 18 B-VG und dem Rechtsstaatsprinzip Rechtsschutz gewährleistet werden und es braucht eine klar gesetzlich determinierte Rechtsgrundlage (Scheichenbauer und Seidl 2022).

Anders kann es sich verhalten, wenn eine Behörde im Rahmen der Privatwirtschaftsverwaltung agiert und damit weder das Legalitätsprinzip anwendbar ist noch erhöhte Begründungspflichten gelten. Unter Privatwirtschaftsverwaltung wird die Erfüllung staatlicher Aufgaben durch privatrechtliche Handlungen, beispielsweise im Zuge des Abschlusses von zivilrechtlichen Verträgen mit Privatfirmen verstanden.

Es ist dennoch davon auszugehen, dass KI-Systeme in der öffentlichen Verwaltung generell dem (unionsrechtlichen) Rechtsstaatsprinzip unterliegen, selbst wenn das spezifische österreichische Legalitätsprinzip nicht anwendbar sein sollte, wie beispielsweise den Ethik-Leitlinien für eine vertrauenswürdige KI (AI HLEG 2019) zu entnehmen ist.

Damit wird die Bedeutung des Rechtsstaatsprinzips im Unionsrecht unterstrichen, worunter der Grundsatz der Gesetzmäßigkeit der öffentlichen Verwaltung, das Prinzip der Rechtsklarheit, das Verhältnismäßigkeitsprinzip und das Rechtssicherheitsprinzip fallen und damit jedenfalls einzuhalten sind. Dabei kann es erforderlich sein, auf über das Recht hinausgehende, ethische Grundsätze und Normen zurückzugreifen, um etwaige rechtsfreie Räume oder Regelungslücken im Hinblick auf die Verwendung von neuartiger KI-Technologie in der Verwaltung abzudecken.

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 österreichische KI-Strategie oder auch den vorliegenden Praxisleitfaden.

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

8.2 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 Datenverarbeitungen, die an keine Person geknüpft bzw. einfach keinen Personenbezug haben, nicht unter die Regelungen der 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. Grundsätzlich kann davon ausgegangen werden, dass jede Verarbeitung personenbezogener Daten durch Mitarbeiterinnen und Mitarbeiter im Tätigkeitsbereich einer Organisation unter der Kontrolle dieser Organisation erfolgt. In Ausnahmesituationen kann es jedoch vorkommen, dass Beschäftigte beschließen, personenbezogene Daten für eigene Zwecke zu verwenden, wodurch die erteilte Befugnis unrechtmäßig überschritten wird und somit die Mitarbeiterinnen und Mitarbeiter dadurch selbst zu Verantwortlichen werden.¹³

Wichtig ist zu betonen, dass primärer Adressat der datenschutzrechtlichen Pflichten der jeweilige „Verantwortliche“ ist. Verantwortlicher ist gemäß DSGVO die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet (vgl. Artikel 4 Ziffer 7 DSGVO).

Hinsichtlich der Datenverarbeitung sind einige zentrale Grundsätze einzuhalten, wie etwa die Grundsätze der Rechtmäßigkeit und der Datenminimierung.

Der Grundsatz der Rechtmäßigkeit besagt, dass die Datenspeicherung etwa eines gesetzlichen Auftrages oder einer vertragsrechtlichen Grundlage bedarf. Eine Alternative dazu ist die Einholung von Einwilligungen, die freiwillig abgegeben werden müssen.

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 des Verantwortlichen (das ist in der Regel eine Behörde, Einrichtung oder andere Stelle, die über die Verarbeitung der Daten entscheidet) sind:

- Jederzeit soll eine betroffene Person, von der Daten gespeichert wurde, über eigene Daten Auskunft erhalten dürfen. Zudem ist er auch dafür zuständig die Betroffenen über die Verarbeitungen zu informieren und ihre anderen Datenschutzrechte (falls relevant etwa auf Löschung, Berichtigung, Einschränkung der Verarbeitung und Widerruf der Einwilligung sicherzustellen).
- Für die Speicherung verwendete IT-Systeme sollen belastbar und vertraulich, z. B. mit Firewall, Virenschutz, Datenbackup ausgestattet sein.
- Für den Datenzugriff gilt, die Mitarbeiterinnen und Mitarbeiter müssen verantwortungsvoll und im Datenumgang geschult sein.

13 https://www.edpb.europa.eu/system/files/2023-10/edpb_guidelines_202007_controllerprocessor_final_de.pdf

- Im Falle einer Verletzung des Schutzes personenbezogener Daten sind gegebenenfalls Betroffene bzw. die Datenschutzbehörde zu informieren und weiterer Schaden ist zu verhindern.
- Eine etwaige Auslagerung der Datenverwaltung ist mittels Vertrags möglich, wobei Datenschutz und Datensicherheit, insbesondere im Hinblick auf das Datengeheimnis sowie Verschwiegenheitspflichten sowie andere Geheimhaltungspflichten, gewährleistet und seitens der Auftraggeber überprüft werden müssen.

In den Ministerien gibt es außerdem Datenschutzbeauftragte, die für individuelle Anwendungsfälle kontaktiert werden sollten und über die notwendigen Maßnahmen im ministeriellen Kontext Auskunft geben können.

Eine Abteilung in einem Ministerium möchte Daten von Bürgerinnen und Bürger 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 Bürgerinnen und Bürger zustimmen können, ob sie teilnehmen oder nicht. Bei der Einwilligung muss einem Formblatt zugestimmt werden, das verspricht, die Datenschutz-Grundverordnung, das Datenschutzgesetz sowie etwaige Materiegesetze 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 Bürgerinnen und Bürger 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 muss. Unter anderem sind betroffene Personen darüber 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*“ (Benachrichtigung der von einer Verletzung des Schutzes personenbezogener Daten betroffenen Person gem. Artikel 33 folgende 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

Im Rahmen von KI-Reallaboren (auch „*Sandboxes*“, vgl. Europäische Kommission 2023, Bauknecht und Kubezko 2024) können bestimmte KI-Systeme vor dem Inverkehrbringen oder der Inbetriebnahme unter behördlicher Einbeziehung und damit innerhalb eines kontrollierten Umfelds erprobt und ihre Risiken gemildert werden (mehr zu Real-laboren/Sandboxes ist im Abschnitt 12 zu finden). Anbieter bzw. zukünftige Anbieter von KI-Systemen können diesen kontrollierten Rahmen für einen begrenzten Zeitraum und unter regulatorischer Aufsicht nutzen, um innovative KI-Systeme zu entwickeln, zu trainieren, zu validieren und – unter Umständen auch unter realen Bedingungen zu testen. In solchen Reallaboren kann es auch zu einer Verarbeitung von personenbezogenen Daten kommen. Hier enthält der AI Act eine datenschutzrechtliche Erleichterung, indem er den Grundsatz der Zweckbindung aufweicht und – bei Einhaltung der (durchaus anspruchsvollen) sonstigen Bedingungen – eine Befugnis zur zweckändernden Weiterverarbeitung von personenbezogenen Daten normiert.

Prüfung der Sicherstellung von bestehenden Betroffenenrechten

Vor der Verarbeitung von personenbezogenen Daten ist sicherzustellen, dass die damit verbundenen Betroffenenrechte wahrgenommen werden können bzw. ist das Vorliegen von Ausnahmen gemäß DSGVO bzw. DSG oder von Beschränkungen durch spezielle Rechtsvorschriften zu prüfen. Dies umfasst insb. folgende Punkte:

- Informationspflichten gegenüber der betroffenen Person.
- Das Recht der betroffenen Person auf Auskunft, auf Berichtigung, auf Löschung und auf Einschränkung der Verarbeitung (Artikel 15 bis 18 DSGVO und §§ 43 bis 45 DSG).
- Mitteilungspflichten gegenüber Empfängerinnen und Empfänger von personenbezogenen Daten über erfolgte Berichtigungen, Löschungen oder Einschränkungen der Verarbeitung (Artikel 19 DSGVO).
- Das Recht der betroffenen Person auf Datenübertragbarkeit (Artikel 20 DSGVO).
- Das Recht der betroffenen Person auf Widerspruch (Artikel 21 DSGVO).
- Die Rechte der betroffenen Person in Zusammenhang mit der automatisierten Entscheidungsfindung im Einzelfall einschließlich Profiling (Artikel 22 DSGVO).

Dokumentation der Verarbeitungstätigkeiten

Vor der Verarbeitung von personenbezogenen Daten sind die Verarbeitungstätigkeiten vom zuständigen Verantwortlichen im Verzeichnis der Verarbeitungstätigkeiten zu dokumentieren. Sofern und soweit die Verarbeitungstätigkeiten nicht bereits durch Rechtsvorschriften festgelegt werden, ist bei der Festlegung der Verarbeitungstätigkeiten das Einvernehmen mit der zuständigen Fachabteilung herzustellen.

Datenschutzrechtliche Vereinbarungen und Verträge

Sofern und soweit dies nicht bereits durch Rechtsvorschriften festgelegt wird, ist im Fall einer Verarbeitung von personenbezogenen Daten durch gemeinsam für die Verarbeitung Verantwortliche mittels einer Vereinbarung in transparenter Form festzulegen, durch welchen Verantwortlichen welche datenschutzrechtlichen Verpflichtungen und Aufgaben erfüllt werden.

Erfolgt die Einbeziehung von Auftraggeberinnen und Auftraggeber ist – soweit dies nicht bereits durch Rechtsvorschriften festgelegt wird – durch die zuständigen Verantwortlichen die Verarbeitung von personenbezogenen Daten durch Auftraggeberinnen und Auftraggeber mittels einer Vereinbarung über eine Auftragsverarbeitung nach Artikel 28 DSGVO bzw. § 48 DSG schriftlich zu regeln.

Die vertraglichen Regelungen mit den Auftraggeberinnen und Auftraggeber sind jedenfalls vor der Aufnahme des IT-Betriebs bzw. vor der erstmaligen Datenverarbeitung abzuschließen.

Zudem ist auf Basis von Artikel 35 DSGVO bzw. § 52 DSG, der einschlägigen Leitlinie der Artikel-29-Datenschutzgruppe und der nationalen Verordnung (DSFA-V, BGBl II 278/2018) zu prüfen, ob eine Datenschutz-Folgenabschätzung durchzuführen ist und eine solche – sofern erforderlich – vor Aufnahme der Verarbeitungstätigkeiten durchzuführen.

Rahmenbedingungen für den Einsatz von Cloud-Services

Erfolgt im Rahmen der Verwendung von Cloud Services auch die Verarbeitung von personenbezogenen Daten, ist eine Prüfung der Zulässigkeit der Verarbeitung durchzuführen. Die Zulässigkeitsprüfung umfasst insbesondere folgende Punkte:

- Die auf die Verarbeitung anzuwendende datenschutzrechtliche Vorschrift (DSGVO bzw. 3. Hauptstück des DSG).
- Die Rechtmäßigkeit der beabsichtigten Verarbeitung von personenbezogenen Daten besonderer Kategorien personenbezogener Daten und personenbezogener Daten über strafrechtliche Verurteilungen und Straftaten.
- Die Einhaltung von Löschfristen bzw. von Löschzyklen.
- Sofern relevant die Bedingungen für die Einwilligung der betroffenen Person (ausdrückliche Einwilligung bei besonderen Kategorien personenbezogener Daten).
- Sofern relevant die Rechtmäßigkeit einer beabsichtigten Übermittlung einschließlich der Übermittlung an Drittländer oder an internationale Organisationen.
- Sofern relevant die Rechtmäßigkeit einer beabsichtigten Zweckänderung.

- Sofern relevant die Rechtmäßigkeit einer beabsichtigten automatisierten Entscheidungsfindung einschließlich Profiling.
- Sofern relevant die Rechtmäßigkeit einer beabsichtigten Verarbeitung von personenbezogenen Daten zu spezifischen Zwecken (wie die Verarbeitung für im öffentlichen Interesse liegende Archivzwecke, wissenschaftliche oder historische Forschungszwecke oder statistische Zwecke, zur Zurverfügungstellung von Adressen zur Benachrichtigung und Befragung von betroffenen Personen in Zusammenhang mit der Freiheit der Meinungsäußerung und der Informationsfreiheit, für die Verarbeitung im Katastrophenfall, im Beschäftigungskontext).

Die Zulässigkeitsprüfung soll dabei die Verarbeitung von personenbezogenen Daten im Rahmen des operativen IT-Betriebs in Produktivumgebungen sowie eine allenfalls vorgesehene Verarbeitung außerhalb von Produktivumgebungen, wie beispielsweise in Schulungs-, Qualitätssicherungs- und Integrationsumgebungen umfassen.

Zum Thema Cloud Einsatz und Anforderungen der DSGVO sollten auch die entsprechenden Ausführungen im Österreichischen Informationssicherheitshandbuch beachtet werden.¹⁴

8.3 Die EU regelt KI: der AI Act

Der Artificial Intelligence Act (AI Act), im Deutschen die Verordnung über Künstliche Intelligenz (KI-VO), 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. Dabei folgt er einem risikobasierten Ansatz, mit dem eine verhältnismäßige und den Risiken angemessene Regulierung sichergestellt werden soll. Dadurch soll der Markt für KI geregelt, das Vertrauen in KI gestärkt sowie der Schutz von Nutzerinnen und Nutzer 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?

¹⁴ Österreichisches Informationssicherheitshandbuch, Version 4.4.0 vom 6.11.2023 abrufbar unter A.3.2 <https://www.sicherheitshandbuch.gv.at/downloads/sicherheitshandbuch.pdf>

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.

Die EU verfolgt einen risikobasierten Ansatz, bei dem die Vorschriften strenger werden, je höher das Risiko eines KI-Systems ist. Der AI Act verbietet dabei bestimmte Praktiken im Bereich der KI, legt verbindliche Anforderungen für KI-Systeme mit hohem Risiko und – parallel dazu – Transparenzpflichten (siehe Abschnitt 9.3) für gewisse KI-Systeme fest und fordert die Sicherstellung von AI Literacy durch alle Anbieter und Betreiber von KI-Systemen. Dies spiegelt sich in einer Risikopyramide wider, die das Risiko von KI-Systemen von niedrig bis inakzeptabel hoch einstuft (siehe Abbildung 8). Damit werden die Berichtspflichten und Folgenabschätzungen der jeweiligen KI-Systeme nach Risikolevel bestimmt.

Der Großteil der Anforderungen richtet sich dabei an sog. Anbieter, das heißt eine *„Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System [...] entwickelt oder entwickeln lässt und es unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringt oder das KI-System unter ihrem eigenen Namen oder ihrer Handelsmarke in Betrieb nimmt“* (Artikel 3 Ziffer 3 AI Act). Diese Rolle würde bspw. vorliegen, wenn eine Behörde selbst ein KI-System entwickelt oder entwickeln lässt und nicht ein bereits vorliegendes System ankauft. Häufiger ist jedoch anzunehmen, dass eine Behörde als Betreiber auftreten wird. Dies ist eine *„Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet“* (Artikel 3 Ziffer 4 AI Act) (Pilniok 2024). Der Betreiber, etwa eine Behörde, darf damit nicht mit dem:r konkreten Endnutzerinnen und Endnutzer (z. B. einem bzw. einer Verwaltungsbediensteten) gleichgesetzt werden. Die Betreiberpflichten richten sich somit in dieser Konstellation an die Behörde, wobei die konkreten Endnutzerinnen und Endnutzer durch den Betreiber beispielsweise entsprechend im Umgang mit dem KI-System geschult werden müssen.

Besondere Relevanz für die öffentliche Verwaltung hat dabei die Sicherstellung der AI-Literacy oder auch KI-Kompetenz (Artikel 4 AI Act). Dieser Begriff beschreibt *„die Fähigkeiten, die Kenntnisse und das Verständnis, die es [...] ermöglichen, KI-Systeme sachkundig einzusetzen sowie sich der Chancen und Risiken von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden“* (Artikel 3 Ziffer 56 AI Act). Daher haben Anbieter und Betreiber von KI-Systemen unabhängig von der Risikoeinstufung Maßnahmen zu ergreifen, um nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen. Diesem Ziel dienen u. a. dieser Leitfaden und entsprechende Schulungen, die in Österreich unter anderem von der Verwaltungsakademie des Bundes angeboten werden (siehe Abschnitt 12).

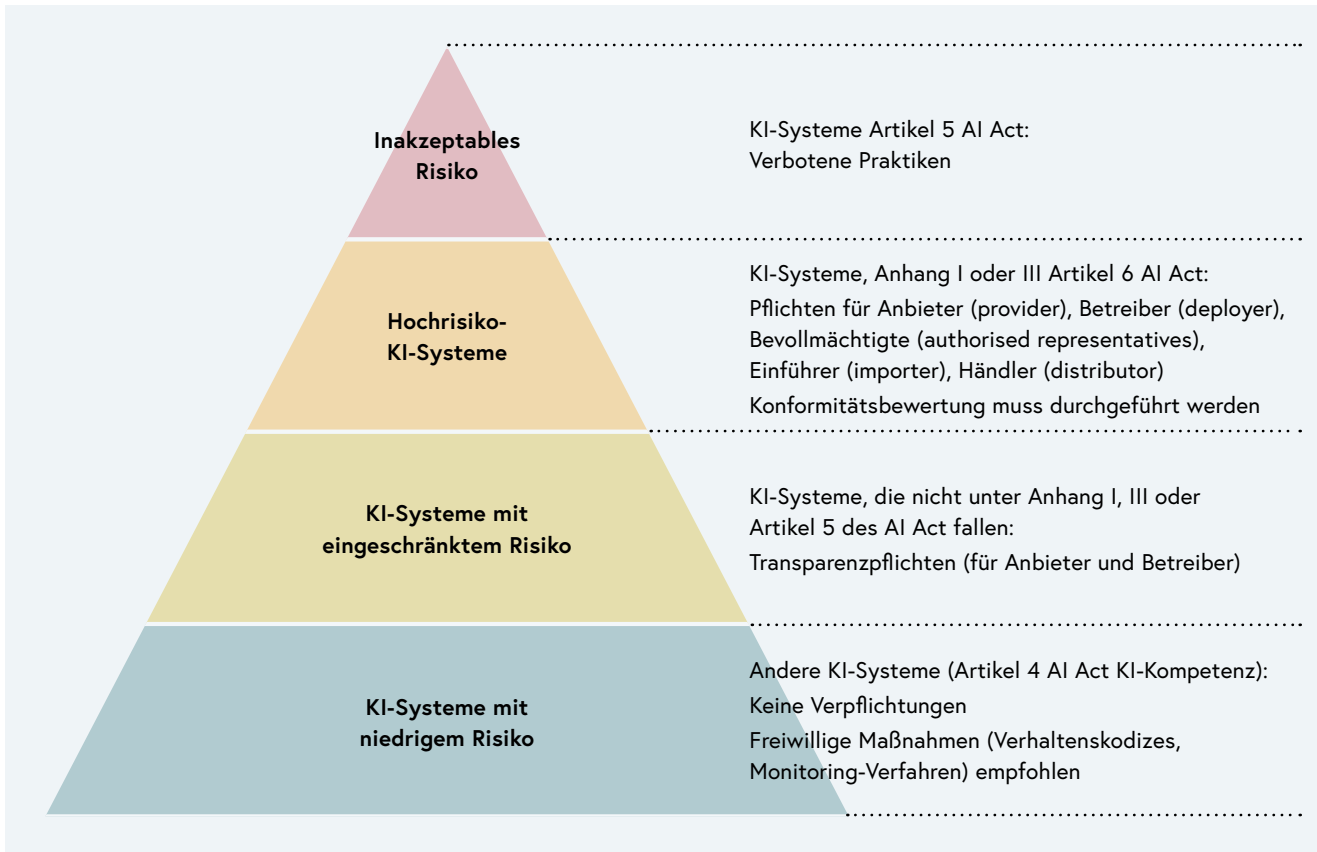


Abbildung 8: Risikopyramide AI Act (eigene Darstellung)

Verbotene Praktiken

Der AI Act enthält in Artikel 5 eine Liste von Praktiken, die aufgrund ihres Gefahrenpotentials prinzipiell verboten sind. Diese KI-Systeme dürfen somit nicht in Verkehr gebracht, in Betrieb genommen oder verwendet werden. Für die Verwaltung von besonderer Relevanz sind dabei folgende verbotene Praktiken:

- Sogenanntes „**social scoring**“, beispielsweise KI-Systeme zur Bewertung oder Einstufung von natürlichen Personen oder Gruppen von Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens oder bekannter, abgeleiteter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale, wobei daraus eine Schlechterstellung oder Benachteiligung erfolgen muss,
- Formen von „**predictive policing**“, etwa KI-Systeme zur Durchführung von Risikobewertungen in Bezug auf natürliche Personen, um das Risiko, dass eine natürliche Person eine Straftat begeht, ausschließlich auf der Grundlage des Profiling einer natürlichen Person oder der Bewertung ihrer persönlichen Merkmale und Eigenschaften zu bewerten oder vorherzusagen,
- **Emotionserkennungssysteme** am Arbeitsplatz, zum Beispiel in Bezug auf Verwaltungsmitarbeiterinnen und -mitarbeiter, und in Bildungseinrichtungen,

- **Biometrische Echtzeit-Fernidentifizierungssysteme** in öffentlich zugänglichen Räumen. Ausnahmen, beispielsweise für die Suche nach einer Person, die besonders schwere Straftaten begangen haben soll, relativieren dieses Verbot.

Hochrisiko-KI-Systeme

Im Zentrum des risikobasierten Ansatzes des AI Acts steht die Regulierung von Hochrisiko-KI-Systemen. Diesen widmet der AI Act als Herzstück einen großen Teil seiner Artikel (Artikel 6 bis 49 AI Act). Der AI Act sieht dabei zwei Wege zur Klassifizierung als Hochrisiko-KI-System vor:

- Erstens „eingebettete Hochrisiko-KI-Systeme“, das heißt Produkte oder Sicherheitskomponenten gemäß Artikel 6 Absatz 1 AI Act, die bereits in den Geltungsbereich des EU-Produktsicherheitsrechts fallen (beispielsweise Medizinprodukte). Diese Kategorie spielt für die Verwaltung nur eine eingeschränkte Rolle.
- Zweitens „eigenständige Hochrisiko-KI-Systeme“ gemäß Artikel 6 Absatz 2 AI Act in Verbindung mit Anhang III, der acht risikoreiche Anwendungsbereiche und konkrete Anwendungsfälle enthält.

Für den Einsatz von KI in der Verwaltung von Relevanz sind dabei folgende in Anhang III genannten Bereiche:

- **Biometrie:** beispielsweise biometrische Fernidentifizierungssysteme, insbesondere für die Sicherheitsbehörden,
- **Kritische Infrastruktur:** KI-Systeme, die als Sicherheitsbauteile im Rahmen der Verwaltung und des Betriebes von kritischer Infrastruktur verwendet werden, etwa im Bereich digitale Infrastruktur, Straßenverkehr, Wasser-, Gas-, Wärme-, Stromversorgung,
- **Allgemeine und berufliche Bildung:** zum Beispiel KI-Systeme zur Feststellung des Zugangs oder der Zulassung oder zur Zuweisung natürlicher Personen zu Einrichtungen aller Ebenen der allgemeinen und beruflichen Bildung, Systeme zur Bewertung von Lernergebnissen, Systeme zur Bewertung des angemessenen Bildungsniveaus, Systeme zur Überwachung und Erkennung von verbotenen Verhalten von Schülern bei Prüfungen,
- **Beschäftigung, Personalmanagement und Zugang zur Selbstständigkeit:** beispielsweise KI-Systeme, die für die Einstellung oder Auswahl natürlicher Personen verwendet werden sollen, KI-Systeme, die für Entscheidungen, die Bedingungen von Arbeitsverhältnissen, Beförderungen und Kündigungen von Arbeitsvertragsverhältnissen beeinflussen, KI-Systeme für die Zuweisung von Aufgaben aufgrund des individuellen Verhaltens oder persönlicher Merkmale oder Eigenschaften oder für die Beobachtung und Bewertung der Leistung und des Verhaltens von Personen,

- **Zugänglichkeit und Inanspruchnahme grundlegender privater und grundlegender öffentlicher Dienste und Leistungen:** insbesondere KI-Systeme, die von Behörden oder im Namen von Behörden verwendet werden sollen, um zu beurteilen, ob natürliche Personen Anspruch auf grundlegende öffentliche Unterstützungsleistungen und -dienste haben und ob solche Leistungen und Dienste zu gewähren, einzuschränken, zu widerrufen oder zurückzufordern sind,
- **Strafverfolgung:** etwa KI-Systeme zur Bewertung des Risikos, dass eine natürliche Person Opfer von Straftaten wird; Lügendetektoren; KI-Systeme zur Bewertung der Verlässlichkeit von Beweismitteln; KI-Systeme zur Unterstützung von Strafverfolgungsbehörden zur Bewertung des Risikos, dass eine natürliche Person eine Straftat begeht oder erneut begeht; KI-Systeme zur Bewertung persönlicher Merkmale und Eigenschaften oder vergangenen kriminellen Verhaltens von natürlichen Personen oder Gruppen; KI-Systeme zur Erstellung von Profilen natürlicher Personen im Zuge der Aufdeckung, Ermittlung oder Verfolgung von Straftaten,
- **Migration, Asyl und Grenzkontrolle:** beispielsweise Lügendetektoren; KI-Systeme zur Bewertung eines Risikos, KI-Systeme zur Unterstützung der Prüfung von Asyl- und Visumanträgen sowie Aufenthaltstiteln und damit verbundenen Beschwerden, Systeme im Zusammenhang mit Migration, Asyl oder Grenzkontrolle zum Zwecke der Aufdeckung, Anerkennung oder Identifizierung natürlicher Personen,
- **Rechtspflege und demokratische Prozesse:** etwa KI-Systeme, die von einer oder im Namen einer Justizbehörde verwendet werden sollen, um eine Justizbehörde bei der Ermittlung und Auslegung von Sachverhalten und Rechtsvorschriften und bei der Anwendung des Rechts auf konkrete Sachverhalte zu unterstützen. Dem Wortlaut nach bezieht sich dies nur auf Justizbehörden wie Verwaltungsgerichte, nicht aber Verwaltungsbehörden.

Würde ein KI-System von österreichischen Behörden eingesetzt, um die Anspruchsprüfung auf den Erhalt von Arbeitslosengeld zu automatisieren, dürfte ein Hochrisiko-KI-System im Sinne von Anhang III Ziffer 5 litera a AI Act vorliegen, da es sich dabei um eine grundlegende öffentliche Unterstützungsleistung handelt.

Sofern keine Ausnahme im Rahmen von Artikel 6 Absatz 3 AI Act vorliegt, das heißt dargelegt wird, dass dem System kein hohes Risiko zukommt, müssen Anbieter und Betreiber eines solchen Hochrisiko-KI-Systems die unten näher dargelegten Pflichten erfüllen.

Anwendungsfall: Vergabe grundlegender öffentlicher Leistungen

Ausnahmen von der Einstufung als Hochrisiko-KI-System

Auch wenn ein KI-System in einem dieser Bereiche eingesetzt wird, kann eine Ausnahme von der Einstufung als Hochrisiko-KI-System vorliegen, „wenn es kein erhebliches Risiko der Beeinträchtigung in Bezug auf die Gesundheit, Sicherheit oder Grundrechte natürlicher Personen birgt, indem es unter anderem nicht das Ergebnis der Entscheidungsfindung wesentlich beeinflusst“ (Artikel 6 Absatz 3 AI Act).

Diesbezüglich werden vier alternative Bedingungen genannt:

- Das KI-System ist dazu bestimmt, eine eng gefasste Verfahrensaufgabe durchzuführen, beispielsweise die Umwandlung von unstrukturierten Daten in strukturierte Daten, Einordnung von eingehenden Dokumenten in Kategorien, Erkennung von Duplikaten.
- Das KI-System ist dazu bestimmt, das Ergebnis einer zuvor abgeschlossenen menschlichen Tätigkeit zu verbessern, etwa die Verbesserung der Sprache von Dokumenten, zum Beispiel professioneller Ton oder wissenschaftlicher Sprachstil.
- Das KI-System ist dazu bestimmt, Entscheidungsmuster oder Abweichungen von früheren Entscheidungsmustern zu erkennen, und ist nicht dazu gedacht, die zuvor abgeschlossene menschliche Bewertung, ohne eine angemessene menschliche Überprüfung zu ersetzen oder zu beeinflussen, zum Beispiel, um in Bezug auf Benotungsmuster eines Lehrers oder einer Lehrerin nachträglich zu prüfen, ob der Lehrer oder die Lehrerin möglicherweise von dem Benotungsmuster abgewichen ist, um so auf mögliche Unstimmigkeiten oder Unregelmäßigkeiten aufmerksam zu machen.
- Das KI-System ist dazu bestimmt, eine vorbereitende Aufgabe für eine Bewertung durchzuführen, beispielsweise intelligente Lösungen für die Bearbeitung von Dossiers, etwa verschiedene Funktionen wie Indexierung, Suche, Text- und Sprachverarbeitung oder Verknüpfung von Daten mit anderen Datenquellen.

Anforderungen an Hochrisiko-KI-Systeme

Anbieter müssen sicherstellen, dass KI-Systeme mit hohem Risiko eine Reihe von Anforderungen erfüllen (Schwartzmann, Keber und Zenner 2024):

- **Risikomanagementsystem:** kontinuierlicher und iterativer Prozess, um Risiken zu ermitteln, zu analysieren, abzuschätzen, zu bewerten und geeignete Risikomanagementmaßnahmen zu ergreifen,
- **Daten- und Daten-Governance:** Anforderungen für Trainings- Validierungs- und Testdatensätze zum Beispiel Qualitätskriterien, etwa Datenbereinigung, Vollständigkeit, Vermeidung von Bias, Repräsentativität,
- **Technische Dokumentation:** Anfertigung einer detaillierten technischen Dokumentation; Grundlage für das Konformitätsbewertungsverfahren,
- **Aufzeichnungspflichten:** Implementierung einer automatischen Protokollierung von Ereignissen,

- **Transparenz und Bereitstellung von Informationen für die Betreiber:** Sicherstellung, dass das KI-System „hinreichend transparent ist, damit die Betreiber die Ausgaben eines Systems angemessen interpretieren und verwenden können“, Bereitstellung von detaillierten Betriebsanleitungen,
- **Menschliche Aufsicht:** Berücksichtigung einer wirksamen menschlichen Aufsicht zur Verhinderung/Minimierung von Risiken, zum Beispiel die Möglichkeit, das System mit einer „Stoptaste“ zu unterbrechen,
- **Genauigkeit, Robustheit und Cybersicherheit:** Garantien für ein angemessenes Maß an Genauigkeit, Robustheit und Cybersicherheit während des gesamten Lebenszyklus.

Konformitätsbewertungsverfahren

Diese Anforderungen werden in einem sogenannten Konformitätsbewertungsverfahren überprüft, bevor ein solches System in Verkehr gebracht oder in Betrieb genommen wird. In Bezug auf Hochrisiko-KI-Systeme gemäß Artikel 6 Absatz 2 AI Act ist dabei (größtenteils) keine Einbeziehung einer externen Konformitätsbewertungsstelle vorgesehen (Artikel 43 Absatz 2 AI Act). Die Bewertung auf Grundlage einer internen Kontrolle erfolgt somit durch „Eigenzertifizierung“ durch den Anbieter des KI-Systems selbst. Dieser Aspekt wurde in der Literatur teilweise kritisiert (Wachter 2024, Schneeberger 2024a).

Das Konformitätsbewertungsverfahren endet dabei, wenn die oben beschriebenen Anforderungen erfüllt werden, mit der Ausstellung einer „Konformitätserklärung“ durch den Anbieter, der Anbringung einer sog „CE-Kennzeichnung“ und der Registrierung des Systems in einer Datenbank (Artikel 47-49 AI Act).

Vor dem Hintergrund der besonderen Verantwortung des Staates der Öffentlichkeit gegenüber sowie der weiter oben geschilderten Fallbeispiele gescheiterter Einsätze von KI in den öffentlichen Verwaltungen verschiedener Länder, erscheint für die Verwaltung ein besonders sensibles Vorgehen aus ethischen Gründen angebracht. Über die rein rechtlichen Verpflichtungen hinaus wäre im Fall der Verwendung von KI-Anwendungen in der öffentlichen Verwaltung diesbezüglich ein vollständiges Konformitätsbewertungsverfahren von den Software-Anbietern im Zuge eines Beschaffungsverfahrens zu verlangen. Das Ergebnis dieses Verfahrens sollte darüber hinaus seitens entsprechend geschulter Verwaltungsbediensteter einem Plausibilitätstest unterzogen werden.

Rolle von (harmonisierten) Normen

Eine bedeutende Rolle spielen dabei auch (harmonisierte) Normen, die von Standardisierungsorganisationen ausgearbeitet und im Amtsblatt der EU veröffentlicht werden. Diese Normen konkretisieren die (stellenweise vagen) Anforderungen des AI Acts. Stimmt ein KI-System mit harmonisierten Normen überein, wird seine Konformität mit dem AI Act vermutet (Artikel 40 Absatz 1 AI Act). Dies trägt zur Rechtssicherheit für Anbieter bei, wurde aber primär aufgrund mangelnder demokratischer Beteiligung am Normsetzungsverfahren in der Literatur auch kritisiert (Ebers 2021, Wachter 2024).

Unabhängig davon werden sich wohl – analog zu existierenden Zertifizierungssystemen, etwa der International Organization for Standardization (ISO), des Institute of Electrical and Electronics Engineers (IEEE) oder des Technische Überwachungs-Vereins (TÜV) – in Abstimmung mit nationalen und europäischen Behörden (vgl. Abschnitt 11) neue Zertifizierungsverfahren herausbilden.

Pflichten der Betreiber von Hochrisiko-KI-Systemen

Im Vergleich zu den Anforderungen an Anbieter von KI-Systemen, primär die Etablierung eines Qualitätsmanagementsystems und die Durchführung eines Konformitätsbewertungsverfahrens (Artikel 16), kennt der AI Act nur eingeschränkte Anforderungen an die Betreiber, das heißt die Nutzerinnen und Nutzer, von Hochrisiko-KI-Systemen (Artikel 26 AI Act). Dazu zählen die Verwendung des KI-Systems gemäß der Betriebsanleitung, Maßnahmen, um eine effektive menschliche Aufsicht sicherzustellen, Sicherstellung, dass die Eingabedaten der Zweckbestimmung entsprechen und repräsentativ sind, Überwachung des Betriebes und die Aufbewahrung von Protokollen (Schwartzmann, Keber und Zenner 2024). Da gerade der Einsatz von hochriskanten KI-basierten Systemen negative Auswirkungen auf Grundrechte entfalten kann, sieht der AI Act außerdem unter gewissen Voraussetzungen die verpflichtende Durchführung einer Grundrechte-Folgenabschätzung (GRFA) vor (siehe dazu Abschnitt 10.1).

Zu beachten ist auch, dass ein Betreiber unter den in Artikel 25 Absatz 1 genannten Umständen zum neuen Anbieter werden kann und damit auch Anbieterpflichten übernimmt. Für die Verwaltung sind dabei primär zwei Szenarien relevant:

- Die Vornahme einer wesentlichen Veränderung eines Hochrisiko-KI-Systems, sodass es weiterhin ein Hochrisiko-KI-System bleibt,
- die Veränderung der Zweckbestimmung eines KI-Systems, das nicht als hochriskant eingestuft wurde, sodass das betreffende KI-System zu einem Hochrisiko-KI-System wird.

Transparenzpflichten für bestimmte KI-Systeme

Unabhängig von der Risikoeinstufung sieht der AI Act in Artikel 50 Transparenzansforderungen für gewisse KI-Systeme vor. So müssen die Anbieter von KI-Systemen, die direkt mit natürlichen Personen interagieren (beispielsweise Chatbots), sicherstellen, dass die Personen darüber informiert werden, dass sie mit einem KI-System interagieren. Weitere Kennzeichnungs- und Informationspflichten gelten für die Anbieter von generativen Systemen in Form einer Kennzeichnung von synthetisch erzeugten Inhalten sowie für Betreiber von Emotionserkennungssystemen und Systemen, die Deepfakes erzeugen.

Von Relevanz ist auch, dass Betreiber eines KI-Systems, das Text erzeugt oder manipuliert, „der veröffentlicht wird, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren“, offenlegen müssen, dass der Text künstlich erzeugt bzw. manipuliert wurde, sofern diese keiner redaktionellen Kontrolle unterliegen.

Finden Chatbots, beispielsweise der vom Finanzministerium verwendete Bot FRED, Einsatz im Behördenalltag, muss der Anbieter, das heißt (typischerweise) der Entwickler, solche KI-Systeme mit einem Hinweis dergestalt kennzeichnen, dass keine Interaktion mit einer natürlichen Person, sondern mit einem KI-System stattfindet.

Wenn KI-Systeme zur Erstellung und Veröffentlichung von Texten eingesetzt werden, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse, etwa Nachrichtenmeldungen, zu informieren, hat die Verwaltung – auch in ihrer Rolle als Betreiber beziehungsweise professioneller Nutzer und Nutzerinnen – ebenfalls entsprechende Transparenzpflichten zu erfüllen.

Anwendungsfall: Transparenzpflicht

KI-Systeme mit minimalem Risiko

Auf unterster Ebene der Risikopyramide des AI Acts stehen KI-Systeme mit minimalem Risiko. Es handelt sich dabei um jene KI-Systeme, die nicht in die zuvor besprochenen Kategorien der verbotenen Praktiken, der Hochrisiko-KI-Systeme oder unter Transparenzpflichten fallen. Diese Systeme werden durch den AI Act keinen weiteren Anforderungen unterworfen. Eine freiwillige Anwendung bestimmter Anforderungen ist jedoch im Wege von Verhaltenskodizes möglich – und im Fall der öffentlichen Verwaltung vor dem Hintergrund ihrer speziellen Verantwortung der Öffentlichkeit gegenüber durchaus sinnvoll (Artikel 95 AI Act). Darüber hinaus bleiben weitere (potenzielle) nationale und internationale Anforderungen, die beispielsweise auf der DSGVO, dem Verfassungs- und Verwaltungsrecht beruhen, anwendbar.

Zeitlicher Anwendungsbereich des AI Acts

Die Bestimmungen des AI Acts erlangen dabei nach Artikel 113 gestaffelt Gültigkeit:

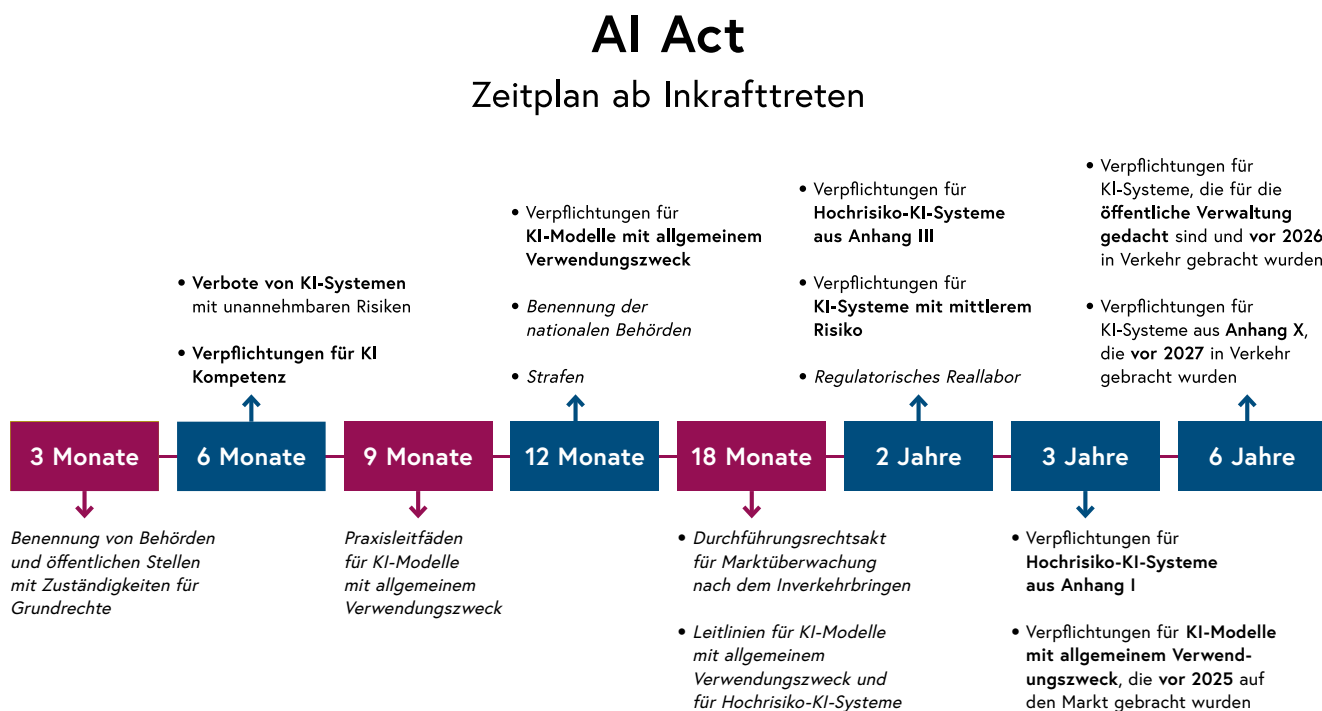


Abbildung 9: Zeitplan AI Act © BKA/Digital Austria

Quelle: <https://www.digitalaustria.gv.at/Themen/KI/AI-Act.html>

Kritikpunkte am AI Act und seinen vorhergehenden Entwürfen kamen und kommen vor allem von drei Akteursgruppen, nämlich Wirtschaft, Wissenschaft und Nichtregierungsorganisationen (NGOs).

Organisationen wie die International Association of Privacy Professionals (IAPP) kritisieren die Breite der KI-Definition und weisen darauf hin, dass sie vage und unklar sein könnte (Caroli 2024), während andere, wie das Beratungsunternehmen KPMG, die Notwendigkeit einer solch breiten Definition unterstreichen, um möglichst viele Branchen in die Einhaltung des AIA einzubeziehen (KPMG International 2024).

BusinessEurope (2021), eine führende Vertretung von Wirtschaftsinteressen, und Autoren wie Friedl und Casola (2024), warnen 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.

Im Gegensatz dazu haben das European Center for Not-for-Profit Law oder NGOs wie Open Future oder Human Rights Watch kritisiert, dass Unter-

nehmen primär auf Basis von Eigenzertifizierung im Konformitätsbewertungsverfahren den Nachweis für die Einhaltung von Bestimmungen einbringen sollen. Diese Vorgangsweise könnte den Zweck der Verordnung gefährden und den AI Act in seiner Wirksamkeit stark einschränken (European Center for Not-for-Profit Law, 2024; Human Rights Watch, 2021; Keller & Warso, 2023; Wachter 2024).

In Bezug auf die zuvor angesprochene essenzielle Rolle von (harmonisierten) Normen, die die technische Implementierung vieler Anforderungen konkretisieren (werden), wurde der Einfluss von, nicht demokratisch zusammengesetzten oder legitimierten, Standardisierungsorganisationen sowie die fehlenden Rechtsmittel und Zugänglichkeit solcher Standards kritisiert (Ebers, 2021). In ähnlicher Weise haben mehrere Akteure den AI Act wegen seiner potenziell negativen Auswirkungen auf die Menschenrechte kritisiert. So fordert beispielsweise das EDRI-Netzwerk (European Digital Rights), ein Zusammenschluss von Expertinnen und Experten und NGOs, einen Fokus auf Grundrechte und den Schutz von Betroffenen von KI-Systemen (EDRI, 2023). Andere NGOs wie Article 19 oder Access Now kritisieren die rechtlichen Schlupflöcher im AI Act, die Menschenrechtsverletzungen oder die Massenüberwachung von schutzbedürftigen Personen aus Gründen der nationalen Sicherheit ermöglichen, wie die Ausnahmen in Artikel 6 Absatz 3 des AI Acts zeigen (Access Now, 2024; Article 19, 2024).

Besonders zum Missbrauch einladend erscheint demnach der erwähnte Artikel 6 Absatz 3 AI Act, der Anbietern die Möglichkeit gibt, obwohl das KI-System als „eigenständiges“ Hochrisiko-KI-System in einem der in Anhang III genannten Anwendungsbereiche verwendet wird, sich darauf zu berufen, dass eine Ausnahme vorliegen würde, weil es kein erhebliches Risiko der Beeinträchtigung in Bezug auf die Gesundheit, Sicherheit oder Grundrechte natürlicher Personen birgt. Eine solche Einschätzung ist vom Anbieter nur zu dokumentieren und das System unterliegt einer Registrierungspflicht (Artikel 6 Absatz 4 AI Act). Eine behördliche Zustimmung zur Berufung auf diese Ausnahme muss jedoch nicht vorliegen. Bis konkretisierende Leitlinien zu dieser Ausnahme vorliegen, ist zu erwarten, dass sich zahlreiche Anbieter darauf berufen werden, dass von ihren KI-Systemen kein hohes Risiko ausgeht. Beispielsweise könnte ein Anbieter, der ein KI-System in Verkehr bringt, das im Kontext der Vergabe grundlegender öffentlicher Unterstützungsleistungen (etwa Arbeitslosengeld) verwendet wird, behaupten, dass dieses nur eine vorbereitende Aufgabe für eine Bewertung durchführt, die Letztentscheidung von Menschen getroffen wird und das System somit nicht hochriskant wäre.

Wissen: Kritik am AI Act

8.4 Weitere EU-Regulierungen

Die Bedeutung der Festlegung von Normierungen zur Kontrolle des Digitalisierungsprozesses und der Anwendung von KI wird von der EU durch die Einführung von bzw. die Diskussion um unterschiedliche Richtlinien, wie die EU-Produkthaftungsrichtlinie, die KI-Haftungsrichtlinie, das Daten-Governance-Gesetz 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 Vorschläge für zwei neue EU-Richtlinien ausgearbeitet, 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. Der Anwendungsbereich der überarbeiteten Produkthaftungsrichtlinie wird dabei auf KI und Software erstreckt. Künftig sind Beweis erleichterungen für komplexe Konstellationen wie beim Einsatz von KI vorgesehen. Zwar werden Geschädigte weiterhin die Beweislast für die Haftungsvoraussetzungen tragen, jedoch greifen Offenlegungspflichten und Kausalitätsvermutungen, um Geschädigten die Durchsetzung von Ersatzansprüchen beim Einsatz von KI zu erleichtern.

Zweitens soll durch den Entwurf der Richtlinie über KI-Haftung („*AI Liability Directive, AILD*“) jene Fälle harmonisiert geregelt werden, die durch erstere Richtlinie nicht erfasst werden, zum Beispiel Verletzungen der Privatsphäre oder Schäden durch Sicherheitsprobleme. Die AILD soll indes Fragestellungen rund um die deliktische Haftung von Anbietern und Betreibern lösen. Dabei zielt auch die AILD auf die prozesualen Anforderungen der Geltendmachung von Ansprüchen gegenüber Anbietern oder Betreibern einer KI nach nationalem Recht ab.

Es gibt somit 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 (dies soll im Anwendungsbereich der AILD jedoch nur dann gelten, wenn Hochrisiko-KI-Systeme betroffen sind) erweitert (European Commission 2022c).

Während die überarbeitete Produkthaftungs-Richtlinie von den europäischen Institutionen bereits beschlossen und im Amtsblatt veröffentlicht wurde, scheint die Zukunft der AILD im Moment noch ungewiss (Wendt und Wendt 2024, 157f).

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“ verabschiedet. 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 2022b; 2023).

Weitere gesetzgeberische Aktivitäten auf EU-Ebene

In den letzten Jahren 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 Online-Plattformen (European Commission 2022d), der „Digital Services Act“ zu von allgemeinen Vermittlungsdiensten, Hosting-Diensten und (sehr großen) Online-Plattformen und -Suchmaschinen erbrachten digitalen Diensten (European Commission 2022e), die „Machinery Regulation“ zur Sicherheit von Maschinen und Robotern (European Commission 2022e). Weitere Strategiedokumente, Gesetzesvorhaben und Gesetze zum Thema Digitalisierung auf EU-Ebene können auf der Website des Europäischen Parlaments¹⁵ nachgeschlagen werden.

8.5 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):

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

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.

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 Bürgerinnen und Bürger 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 Mitarbeitern die erforderlichen KI-relevanten Fähigkeiten zu vermitteln (BMDW und BMK 2021, 56-59).

Abschnitt 11 gibt einen detaillierten Überblick über die Governance-Struktur von KI in Österreich und inkludiert auch die die Umsetzung der AIM AT 2030.

9 Ethische KI: Prinzipien und Leitlinien

Dieser Abschnitt 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.

9.1 Ethische Leitlinien: Governance durch „Soft Law“

Wie bereits in Abschnitt 8 „Rechtlicher Rahmen“ erwähnt, sind ethische Aspekte von KI bereits teilweise rechtsverbindlich geregelt (siehe Abschnitte DSGVO, AI Act). 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 (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*“ der deutschen NGO AlgorithmWatch¹⁶ sind mehr als 100 derartiger Leitlinien von Unternehmen, NGOs, Regierungen, unterschiedlichster Organisationen und supranationalen Institutionen gesammelt.

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

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 (AI HLEG 2019). Auf diese einflussreichen Ethik-Leitlinien für vertrauenswürdige KI wird auch in den Erwägungsgründen des AI Act Bezug genommen. 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 vertrauenswürdig 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 10.3).

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, aktualisiert 2024).

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

Wissen: Nationale Leitlinien und KI-Strategien

9.2 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.

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

Um dem Mangel an umfassenden Handlungsanleitungen für den Einsatz von KI in der öffentlichen Verwaltung entgegenzusteuern, wurden 2023 im Zuge des Projekts „Digitale Verwaltung und Ethik“ im Rahmen der Erstellung dieses Praxisleitfadens unter anderem auch Kriterien und Maßnahmen erarbeitet, die sich speziell an Verwaltungsangehörige richten (BMKÖS 2023). Als Grundlage für die vorgeschlagenen Kriterien wurden die für den AI Act zentralen HLEG-Ethikrichtlinien herangezogen und um weitere, für den öffentlichen Dienst besonders relevante Kriterien ergänzt. Die letztendlich ausgewählten Kriterien sind in Abbildung 10 wiedergegeben und umfassen: Recht, Transparenz, Unvoreingenommenheit und Fairness, Effektivität und Effizienz, Sicherheit, Zugänglichkeit und Inklusion, Menschliche Aufsicht, Rechenschaftspflicht, Digitale Souveränität.

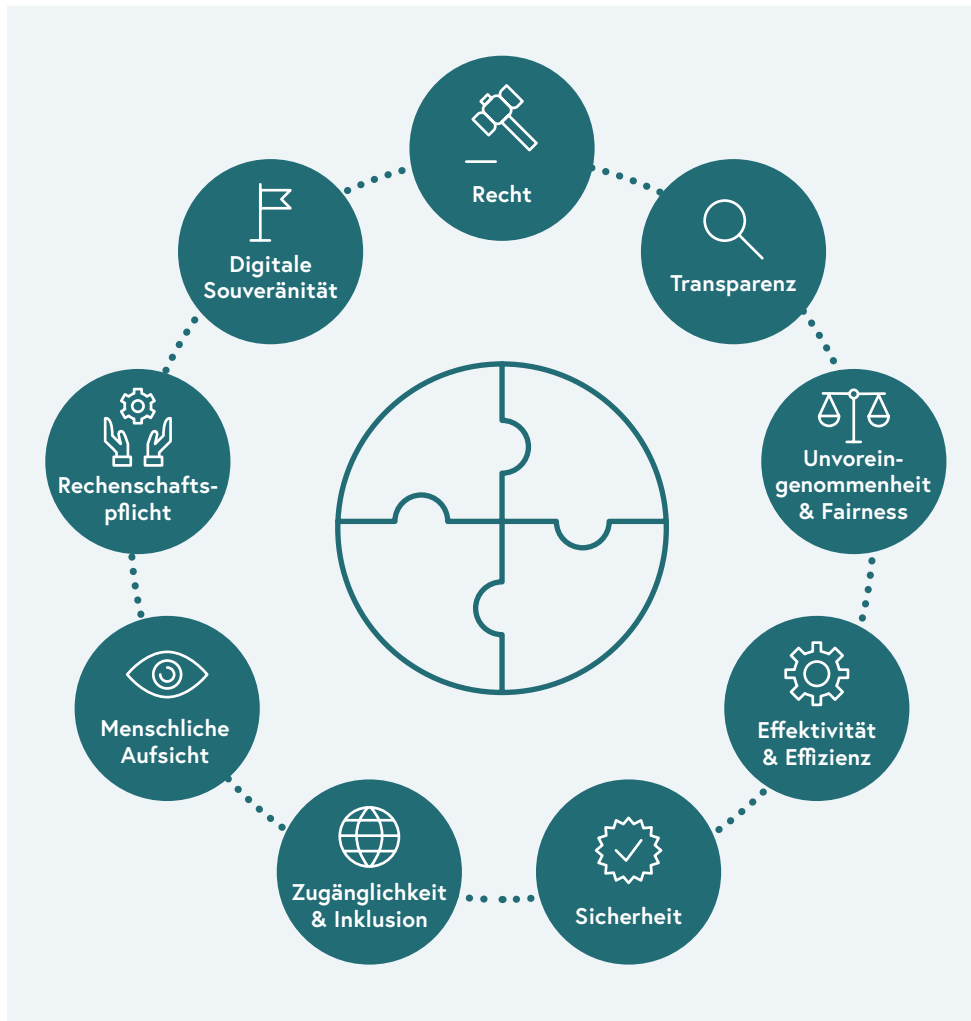


Abbildung 10: Kriterien für ethische KI-Anwendungen in der öffentlichen Verwaltung

Kriterien

Recht

Rechtliche Rahmenbedingungen und Vorschriften legen fest, unter welchen Bedingungen KI-Systeme in verschiedenen Anwendungsbereichen eingesetzt werden dürfen. Für den rechtmäßigen Einsatz ist es erforderlich, dass alle nationalen und europäischen gesetzlichen Vorgaben eingehalten werden, insbesondere in Bezug auf den AI Act, den Schutz personenbezogener Daten und die Wahrung von Grundrechten. Diese rechtlichen Vorgaben bilden den rechtlich verbindlichen Rahmen für die ethische und verantwortungsvolle Nutzung von KI in der öffentlichen Verwaltung .

Transparenz

Gemäß dem AI Act wird Transparenz in Bezug auf KI so definiert, dass die Entwicklung und der Einsatz von KI-Systemen nachvollziehbar und erklärbar sein soll. Dies umfasst die Pflicht, Menschen darüber zu informieren, wenn sie mit einem KI-System interagieren, Nutzerinnen und Nutzer über die Fähigkeiten und Grenzen des Systems aufzuklären

und betroffene Personen über ihre Rechte zu informieren (Erwägungsgrund 27 AI Act). Transparenz bedeutet aber auch, Informationen über die KI-Anwendung verfügbar und zugänglich zu machen und zu erläutern, wie die KI-Anwendung entwickelt, trainiert und zu welchem Zweck sie eingesetzt wird. Wenn die KI-Anwendung als nicht risikoreich eingestuft wurde, sollten Anbieter außerdem eine Dokumentation der Bewertung der KI-Anwendung erstellen und den zuständigen nationalen Behörden auf Anfrage vorlegen können, bevor diese in Betrieb genommen wird (Erwägungsgrund 53 AI Act).

Unvoreingenommenheit und Fairness

Unvoreingenommenheit und Fairness bedeutet nach dem AI Act, dass KI-Systeme so entwickelt und eingesetzt werden, dass verschiedene Akteure einbezogen werden und der gleichberechtigte Zugang, die Gleichstellung der Geschlechter und die kulturelle Vielfalt gefördert werden, während diskriminierende Auswirkungen und unfaire Voreingenommenheit vermieden werden, die nach Unionsrecht oder nationalem Recht verboten sind (Erwägungsgrund 27 AI Act). Für die KI-Anwendung sollten außerdem vielfältige Datenquellen für den jeweiligen Kontext verwendet werden, um zu vermeiden, dass historisch-bedingte Vorurteile aufrechterhalten werden.

Effektivität und Effizienz

Der Einsatz von KI-Anwendungen in der Verwaltung muss deren Effektivität und Effizienz nachhaltig verbessern, ohne die Arbeitssituation der Verwaltungsbediensteten 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 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 Bürgerinnen und Bürger 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 sind grundsätzlich Alternativen zur KI-Technologie anzubieten, um einen gleichberechtigten Zugang zu öffentlichen Dienstleistungen zu gewährleisten.

Menschliche Aufsicht

Für KI-Systeme, insbesondere mit hohem Risiko, ist es notwendig, dass menschliche Kontrollmaßnahmen eingeführt werden, bevor die Systeme eingesetzt werden. Diese Maßnahmen sollen sicherstellen, dass das System auf menschliche Bedienerinnen und Bediener reagiert, dass es innerhalb ethischer Rahmenbedingungen funktioniert und von Personen mit entsprechender Schulung und Befugnis überwacht wird. Für KI-Systeme mit hohem Risiko sollten Mechanismen vorhanden sein, die Nutzerinnen und Nutzer anleiten, wann und wie sie eingreifen müssen, um negative Folgen oder Risiken zu verhindern (Erwägungsgrund 73 AI Act). Außerdem ist es ratsam, dass regelmäßig externe oder interne Audits und Bewertungen des KI-Systems durchgeführt werden, um mögliche Verzerrungen der Datenquellen oder sonstige Mängel zu erkennen und zu beheben.

Rechenschaftspflicht

Rechenschaftspflicht für KI-Systeme bezieht sich darauf, dass Verantwortlichkeiten sowohl für die Handlungen als auch die Ergebnisse des KI-Systems klar zugewiesen werden. Verantwortlichkeit umfasst in diesem Zusammenhang sowohl ethische als auch rechtliche Pflichten. Entwicklerinnen und Entwickler und Betreiberinnen und Betreiber, die mit KI-Systemen arbeiten, tragen die moralische Verantwortung dafür, dass die Technologie ethisch vertretbar funktioniert sowie Schäden vermeidet. Dazu gehört die Einhaltung der einschlägigen Gesetze und Vorschriften, die Gewährleistung der Einhaltung von Standards in Bezug auf Ethik, Datenschutz und Sicherheit und die potenzielle finanzielle Haftung für Schäden, die das KI-System verursachen kann. Wenn ein KI-System versagt oder unbeabsichtigte negative Folgen hat, müssen die entsprechenden Verantwortlichen zur Verantwortung gezogen werden können.

Digitale Souveränität

Die öffentliche 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.

Maßnahmen

Zusätzlich zu den Kriterien wurden Maßnahmenvorschläge entwickelt. Zu diesen Maßnahmen gehören die Förderung der KI-Literacy von Verwaltungsbediensteten und der Öffentlichkeit, die Durchführung von Folgenabschätzungen, die Zertifizierung von KI-Anwendungen und die Einrichtung von unabhängigen Aufsichtsgremien für die Überwachung des Betriebs der KI-Anwendung. Diese Maßnahmen sollen dazu beitragen, dass der Einsatz und die Nutzung von KI-Technologien zum Gemeinwohl einen Beitrag leisten oder wenigstens mit demselben vereinbar sind.

10 KI-Folgenabschätzung

In diesem Abschnitt werden verschiedene Instrumente vorgestellt, mit denen die öffentliche Verwaltung KI Ethik Prinzipien und Leitlinien in der Praxis umsetzen sowie die Auswirkungen von KI auf Grund- und Menschenrechte untersuchen 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. Die vorgestellten Instrumente unterscheiden sich auch dahingehend, ob sie rechtlich bindend oder aus ethischer Perspektive empfehlenswert sind. Zur Übersicht sind die im Folgenden genauer vorgestellten Instrumente hier aufgezählt:

- die Grundrechte-Folgenabschätzung nach Artikel 27 AI Act ist ein rechtlich verbindliches Instrument für Betreiber bestimmter Hochrisiko-KI-Systeme,
- 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,
- 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.

10.1 Grundrechte-Folgenabschätzung im AI Act

Der AI Act vertritt einen Regulierungsansatz, der auf den Menschen ausgerichtete und vertrauenswürdige KI fördern sowie garantieren möchte, dass der Einsatz von KI im Einklang mit Grund- und Menschenrechten steht (Artikel 1 AI Act). Da gerade der Einsatz von hochriskanten KI-basierten Systemen negative Auswirkungen auf Grundrechte entfalten kann, sieht der AI Act unter gewissen Voraussetzungen die verpflichtende Durchführung einer Grundrechte-Folgenabschätzung (GRFA) vor (Wendt und Wendt 2024). Folgenabschätzungen sind deshalb von großer Bedeutung, da sie Vorsichtsmaßnahmen als Alternative zu rechtlichen Einschränkungen darstellen, Verantwortlichkeit sowie Rechenschaft fördern und eine Eliminierung von Risiken im Vorhinein erlauben (Mantelero 2022, Selbst 2021 und Müller und Schneeberger 2024). Durch das besondere Instrument der GRFA lassen sich außerdem die Grundrechte von Menschen durch vorausschauende Maßnahmen schützen.

In der finalen Fassung des AI Act ist die GRFA in Artikel 27 verankert mit dem Ziel, die spezifischen Risiken für die Rechte von Einzelpersonen oder Gruppen von Einzelpersonen, die wahrscheinlich von einem bestimmten KI-System betroffen sein werden,

zu ermitteln und Maßnahmen festzusetzen, die im Falle eines Eintretens dieser Risiken zu ergreifen sind (siehe auch Erwägungsgrund 96 AI Act).

Diese Verpflichtung betrifft jedoch nicht alle KI-Systeme, sondern die GRFA ist vor der Inbetriebnahme eines Hochrisiko-KI-Systems gemäß Artikel 6 Absatz 2 AI Act (außer die in Anhang III unter Punkt 2 genannte kritische Infrastruktur) durchzuführen. Dies schließt KI-Systeme ein, die für folgende Zwecke oder in folgenden Bereichen zum Einsatz gelangen:

- Biometrie inklusive Emotionserkennung,
- Allgemeine und berufliche Bildung,
- Beschäftigung und Personalmanagement,
- Zugänglichkeit und Inanspruchnahme öffentlicher Dienste und Leistungen,
- Strafverfolgung,
- Migration, Asyl und Grenzkontrolle,
- Rechtspflege und demokratische Prozesse.

Die Verpflichtung zur Durchführung der GRFA trifft den Betreiber des Hochrisiko-KI-Systems, sofern es sich bei diesem um eine Einrichtung des öffentlichen Rechts handelt. Damit wird die öffentliche Verwaltung direkt angesprochen. Wichtige Dienstleistungen öffentlicher Art, etwa im Bereich Bildung, Gesundheitsversorgung, Sozialdienste oder Wohnungswesen, können auch von privaten Einrichtungen erbracht werden, weshalb die Verpflichtung nach Artikel 27 AI Act auch für private Einrichtungen, die solche öffentlichen Dienste erbringen, gilt. Mitumfasst sind außerdem Betreiber von KI-Systemen zur Kreditwürdigkeitsprüfung und zur Risikobewertung und Preisbildung bei Lebens- und Krankenversicherungen. Gemäß der gestaffelten Gültigkeit der einzelnen Bestimmungen des AI Acts (siehe die Ausführungen zum zeitlichen Anwendungsbereich des AI Acts in Abschnitt 8.3) ist Artikel 27 ab dem 2. August 2026 gültig und ab diesem Zeitpunkt für alle genannten Betreiber verpflichtend.

Die GRFA muss folgende Aspekte umfassen:

- Eine Beschreibung der Verfahren, bei denen der Betreiber das KI-System im Einklang mit seiner Zweckbestimmung einsetzt,
- Eine Beschreibung des Zeitraums und der Häufigkeit des Einsatzes des KI-Systems,
- Die Kategorien der natürlichen Personen und Personengruppen, die von seiner Verwendung im spezifischen Kontext betroffen sein könnten,
- Die spezifischen Schadensrisiken, die sich auf diese Personengruppen auswirken könnten,
- Eine Beschreibung der Umsetzung von Maßnahmen der menschlichen Aufsicht entsprechend der Gebrauchsanweisung des KI-Systems,
- Die Maßnahmen, die im Falle des Eintretens dieser Risiken zu ergreifen sind, einschließlich der Regelungen für die interne Unternehmensführung und Beschwerdemechanismen.

In den Erläuterungen zur GRFA wird überdies angeregt, dass der Betreiber bei der Durchführung der GRFA auch relevante Interessensträger, Vertreter von potenziell betroffenen Personengruppen, Sachverständige und zivilgesellschaftliche Organisationen einbeziehen kann, um einschlägige Informationen einzuholen. Auch, wenn die Einbeziehung dieser Gruppen und Organisationen keinen Niederschlag in Art 27 AI Act gefunden hat und somit nicht rechtlich verpflichtend ist, wird die Zusammenarbeit mit diesen jedenfalls empfohlen. Die Ergebnisse der durchgeführten GRFA sind dann der zuständigen Marktüberwachungsbehörde mitzuteilen. Als Marktüberwachungsbehörde gilt jene Behörde, die gemäß Artikel 10 der EU-Marktüberwachungsverordnung 2019 als solche von einem Mitgliedstaat benannt wird. In Österreich fungiert seit 2022 das Bundesamt für Eich- und Vermessungswesen als zentrale Verbindungsstelle für Marktüberwachung. Weitere Marktüberwachungsbehörden in Österreich wurden seit November 2024 benannt.

Gemäß den gesetzlichen Vorgaben ist die GRFA vor der ersten Inbetriebnahme des Systems durchzuführen. Dabei kann auch auf bereits durchgeführte oder vorhandene Folgenabschätzungen zurückgegriffen werden. Sofern bereits eine Datenschutz-Folgenabschätzung (DSFA) durchgeführt wurde, kann diese als Grundlage herangezogen werden, wobei lediglich die GRFA-spezifischen Inhalte zu ergänzen sind. Sollten sich im Laufe des Betriebs des Systems Änderungen ergeben, sind die Informationen der GRFA zu aktualisieren.

Wie eine GRFA durchzuführen ist und welche Elemente sie in Konkretisierung der gesetzlichen Vorgaben beinhalten muss, ist noch nicht abschließend geklärt. Daher hat die AI Office, die von der EU-Kommission eingerichtet wurde und diese bei der Umsetzung der Bestimmungen des AI Act unterstützt, ein Muster für einen Fragebogen auszuarbeiten, das von den Betreibern bei der Durchführung der GRFA verwendet werden kann.

Selbst wenn die Durchführung einer GRFA in einem konkreten Fall nicht verpflichtend ist, da es sich beispielsweise um kein Hochrisiko-KI-System handelt, kann es sich dennoch in Einzelfällen als nützlich erweisen, die Folgen des Einsatzes eines KI-Systems für die Grund- und Menschenrechte betroffener Personen zu ermitteln und entsprechende Maßnahmen zu setzen. So empfiehlt beispielsweise der Europarat ausdrücklich die Durchführung einer Grund- und Menschenrechtsfolgenabschätzung (*Human Rights Impact Assessment – „HRIA“*) beim Erwerb, der Entwicklung oder des Einsatzes eines KI-Systems durch Behörden, ohne dabei auf die genauen Anforderungen an das System einzugehen (CoE Commissioner for Human Rights 2019). Außerdem gibt es Aspekte des Einsatzes von KI, die von herkömmlichen Grund- und Menschenrechtsfolgenabschätzungen nicht abgedeckt werden, wie beispielsweise gesamtgesellschaftliche, demokratiepolitische oder umweltbezogene Auswirkungen von Künstlicher Intelligenz. An dieser Stelle setzt die Ethik an, weshalb im Folgenden weitere Folgenabschätzungen vorgestellt werden, die über die klassische GRFA hinausgehen und auch ethische Aspekte miteinbeziehen.

Grundrechte-Folgenabschätzung (Art 27 AI Act)

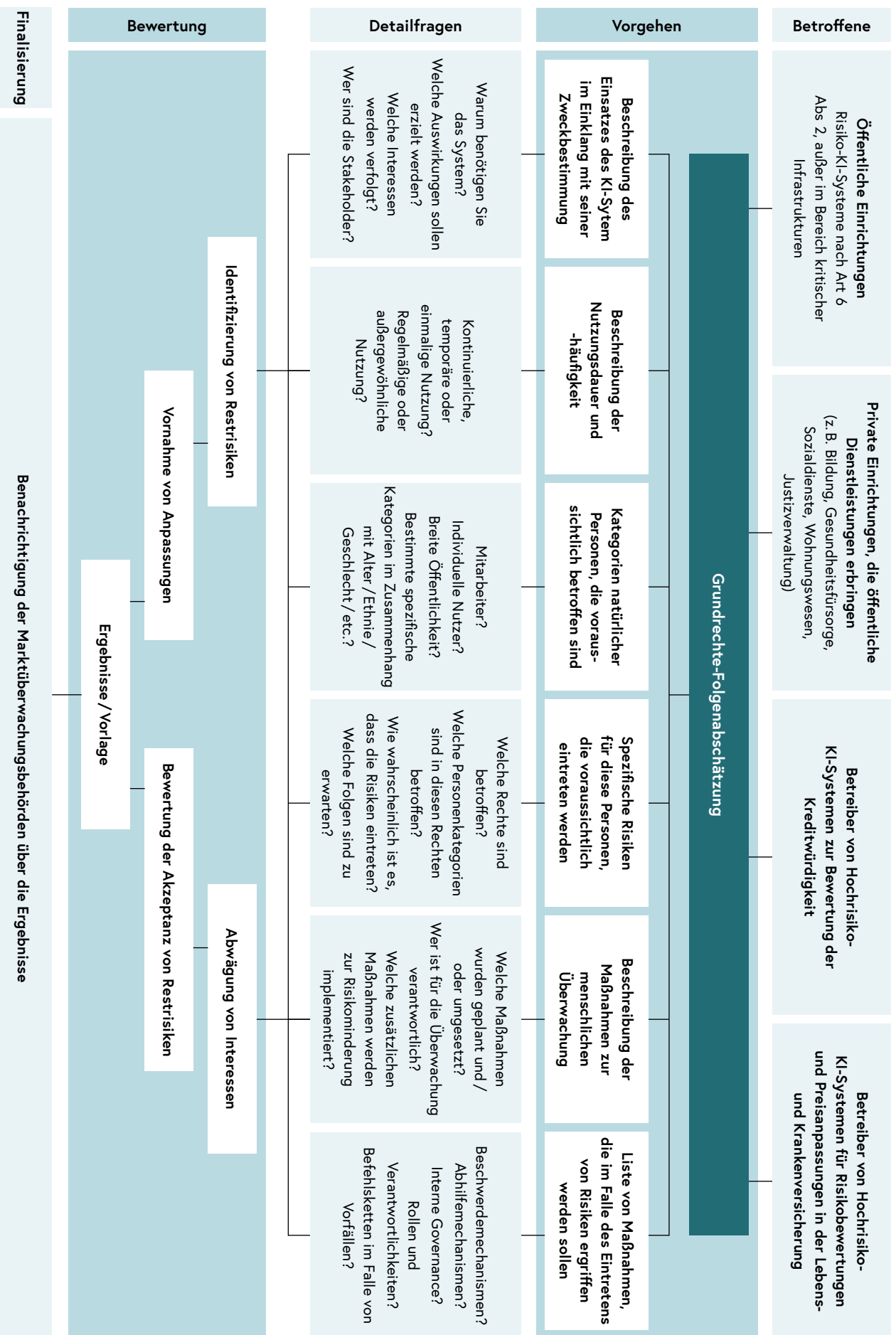


Abbildung 11: Grundrechte-Folgenabschätzung (Artikel 27 AI Act)
Eigene Darstellung basierend auf der Abbildung von Tea Mustać 2024

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

Die Einführung und Nutzung von KI-Technologien in der öffentlichen Verwaltung erforderten einen Rahmen, der Orientierungs- und Anhaltspunkte bietet, um zentrale ethische Fragestellungen zu behandeln und geeignete Maßnahmen festzulegen. Im vom damaligen 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 EKIV dient als dialogisches Instrument zur Folgenabschätzung und erleichtert die ethische Bewertung einer KI-Anwendung. Er eignet sich dabei besonders für ein Überdenken einzelner Bereiche, alleine, in einer kleinen Gruppe oder im Rahmen von größeren Workshops.

Der EKIV greift die Fragestellungen der Checkliste (Abschnitt 2) auf und überführt sie in offene Fragen, um eine tiefere Auseinandersetzung mit den Themen zu ermöglichen und Raum für ausführliche und reflektierte Antworten zu schaffen. Ein weiterer Vorteil dieses Prozesses ist, dass die Antworten auf offene Fragen zugleich als Dokumentation dienen können. Das heißt, während die Fragen beantwortet werden, kann gleichzeitig eine Aufzeichnung oder eine Art Protokoll über das KI-System erstellt werden. Dies kann später nützlich sein, beispielsweise, um ein Pflichtenheft für Anbieter zu erstellen oder in Folge die Nutzung der jeweiligen KI-Anwendung zu evaluieren.

Außerdem ermöglicht der EKIV eine vertiefende Betrachtung eines oder mehrerer defizitärer Bereiche aus der Checkliste. Beide, EKIV und Checkliste, sollen als unterstützende Instrumente verstanden werden, um ethische KI in der Verwaltung zur Anwendung zu bringen.

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. Eine ausführliche Beschreibung der aufgelisteten Kriterien finden sich in Abschnitt 9.2.

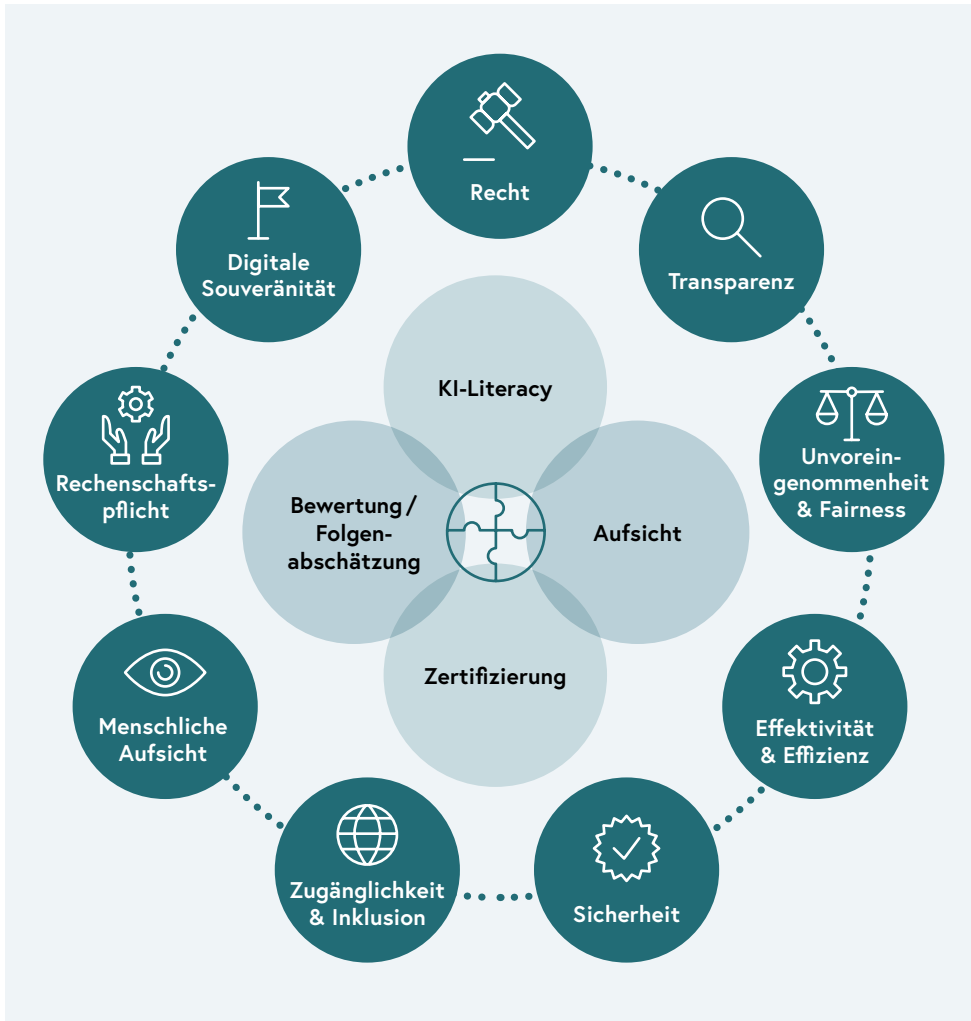


Abbildung 12: Kriterien (Außenkreis) und Maßnahmen (Innenkreis) für ethische KI in der Verwaltung

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

Kriterien

Recht

- In welchem Kontext wird das KI-System eingesetzt und handelt es sich dabei um ein hoheitliches Verwaltungshandeln?
- Inwiefern nutzt oder verarbeitet das KI-System personenbezogene Daten und welche Schutzmaßnahmen werden ergriffen, um die Einhaltung der Datenschutzverpflichtungen zu gewährleisten? Liegt eine gesetzliche Verpflichtung zur Durchführung einer Datenschutz-Folgeabschätzung nach der DSGVO vor?
- In welche Risikoklasse fällt die KI-Anwendung gemäß dem AI Act?

- Welche internen Konformitätsbewertungen werden durchgeführt, um zu überprüfen, ob das KI-System vor der Einführung alle gesetzlichen Anforderungen erfüllt?
- Welche Grundrechte könnten durch die Nutzung des KI-Systems beeinträchtigt oder gefährdet werden? Muss eine Grundrechte-Folgenabschätzung nach dem AI Act gemacht werden?

Transparenz

- Welche Ziele und Zwecke muss die KI-Anwendung für den gegebenen Kontext erfüllen?
- Wie tragen die Ergebnisse der KI-Anwendung zur Erreichung der Ziele bei?
- Welche Grenzen hat die KI-Anwendung? (Wo liegen die Limits der Nutzung, welche Aufgaben oder Funktionen kann die KI-Anwendung nicht erfüllen?)
- Welche Interessensgruppen sind von der Einführung der KI-Anwendung betroffen?
- Wie werden die Personen darüber informiert, wann und wie sie mit der KI-Anwendung interagieren?
- Wie wird den Nutzerinnen und Nutzer oder Betroffenen vermittelt, wie das KI-System zu seinen Ausgaben/Entscheidungen kommt und welche Logik dahintersteckt?
- Wie werden die Datensätze, die mit dem KI-System verbunden sind, verwaltet?

Unvoreingenommenheit und Fairness

- Welche Daten werden verwendet, damit die KI-Anwendung Inhalte ausgeben kann und wo kommen diese her?
- Wie wird sichergestellt, dass die Daten vielfältig und repräsentativ für den Kontext sind?
- Welche Verfahren gibt es, um die Datenquellen auf mögliche Verzerrungen und Ungenauigkeiten im KI-System zu überprüfen?
- Welcher Grad und welche Art von möglichen Verzerrungen (bias) in den Daten ist für die angestrebte Anwendung akzeptabel?

Effektivität und Effizienz

- Wie trägt die KI-Anwendung dazu bei, die Arbeitssituation der Verwaltungsbediensteten zu verbessern oder zumindest zu erhalten?
- Inwiefern wird das System die relevanten Verwaltungsaufgaben im Vergleich zum aktuellen Stand effektiver ausführen?
- Welche Qualitäts- und Leistungsziele werden für das KI-System festgelegt?
- Wie werden die Verwaltungsbediensteten geschult und unterstützt, um eine effektive Nutzung der KI-Anwendung zu ermöglichen? (Welche neuen (digitalen) Kompetenzen werden benötigt?)
- Wie werden die Umweltauswirkungen der KI-Anwendung bewertet und berücksichtigt?

Sicherheit

- Welche Notfallpläne werden etabliert, um etwaige Systemfehler oder Störungen zu beheben?
- Wie werden Aufzeichnungen über die Leistung der KI-Anwendung, Vorfälle und Fehlfunktionen archiviert und wie lange?
- Welche technischen und organisatorischen Maßnahmen werden ergriffen, um negative Auswirkungen von KI zu verhindern oder zu minimieren (Risikomanagement)?
- Welche Sicherheitsvorkehrungen gibt es zum Schutz vor dem Missbrauch oder der böswilligen Nutzung der KI-Anwendung?

Menschliche Aufsicht

- Welche Monitoringmechanismen werden eingerichtet, um den ethischen und rechtmäßigen Einsatz und die Nutzung von KI-Technologien zu überwachen und sicherzustellen?
- Wie können die Ausgaben, Ergebnisse oder Entscheidungen der KI-Anwendung überprüft werden?
- Über welche Qualifikationen und Fachkenntnisse verfügen die Monitoring-Verantwortlichen?

Barrierefreiheit und Inklusion

- 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/oder davon zu profitieren?
- Welche Alternativen zur KI-Technologie werden angeboten, um einen gleichberechtigten Zugang zu öffentlichen Dienstleistungen zu gewährleisten (insbesondere persönliche Ansprechpartner)?

Rechenschaftspflicht

- Welche Verantwortlichkeiten und Zuständigkeiten werden für Entwicklerinnen und Entwickler und Nutzerinnen und Nutzer der KI-Anwendung festgelegt? (z. B. Wer ist für die Implementierung, die Aufsicht und die Bearbeitung von Beschwerden verantwortlich?)
- Wer trägt die letztendliche Verantwortung und Rechenschaftspflicht für den Einsatz und die Ausgaben des KI-Systems?
- Welche Verfahren gibt es für Endnutzerinnen und Endnutzer und betroffene Personen oder Gruppen, um Probleme zu melden und/oder KI-Entscheidungen anzufechten?

Digitale Souveränität

- Wer hat das KI-System entwickelt und von wem wurde es vertrieben? Sind nicht-demokratische Länder in Herstellung und Vertrieb involviert?

- Wie wurde der Entwickler/Hersteller ausgewählt? (Wo lagen die Vorteile gegenüber anderen Anbietern im Sinne des Bestbieterprinzips?)
- Welche Designspezifikationen wurden für die KI-Anwendung vor der Beschaffung festgelegt?
- 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 und Vertriebspartnern auf Augenhöhe umgehen zu können?
- Welche ethischen und sicherheitspolitischen 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?
- Wie wird sichergestellt, dass die KI-Kenntnisse jeweils aktualisiert werden, wenn sich die Technologie weiterentwickelt?

Bewertung/Folgenabschätzung

- Wer ist für die Initiierung und Durchführung einer Folgenabschätzung verantwortlich?
- Welchen potenziellen individuellen und gesellschaftlichen Schaden kann das KI-System verursachen? (Wer ist betroffen und inwiefern?)
- Wie sind die möglichen negativen Auswirkungen des KI-Systems einzuschätzen? (geringes, mittleres oder hohes Schadenspotenzial)
- Wie oft werden zertifizierte KI-Modelle neu bewertet, um sicherzustellen, dass sie weiterhin mit ethischen und rechtlichen Standards übereinstimmen?

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 seitens der Verwaltung für die Überwachung des Zertifizierungsprozesses für KI-Modelle, die in der öffentlichen Verwaltung eingesetzt werden, zuständig? Wurde ein Plausibilitätscheck der Zertifizierung seitens der Verwaltung durchgeführt?

Aufsicht

- Welche Vorkehrungen werden eingerichtet, die eine externe Überprüfbarkeit des KI-Systems erleichtern (z. B. Dokumentation des Entwicklungsprozesses)?
- Werden unabhängige Aufsichtsgremien für das KI-System eingerichtet? Wenn ja: Welche Rechte und Pflichten haben die unabhängigen Aufsichtsgremien?
- Über welche Qualifikationen und Fachkenntnisse verfügen die Mitglieder der Aufsichtsgremien?

10.3 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 Tool¹⁷ 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.

10.4 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

17 <https://altai.insight-centre.org/>

sollten. Darüber hinaus wird ein Ethik 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).

10.5 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 Entwicklerinnen und Entwickler 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 zu verringern. Diese Art der Folgenabschätzung wird in den Niederlanden in verschiedenen Teilen der Verwaltung eingesetzt (Gerards et al. 2022; Utrecht University 2022; Clausen und Schäfer 2023).

10.6 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

10.7 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).

- „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).
- „Ethical Impact Assessment“ der UNESCO: Bei diesem Tool handelt es sich um ein Instrument der Bewertung der Vorteile und Risiken eines KI-Systems, das den gesamten Lebenszyklus (Design, Entwicklung, Einsatz) eines KI-Systems berücksichtigt. Dieses Instrument soll in erster Linie die an der Beschaffung von KI-Systemen beteiligten Mitarbeiterinnen und Mitarbeiter des öffentlichen Sektors in der Evaluierung unterstützen, ob das erworbene KI-System den von der UNESCO veröffentlichten KI-Ethik-Empfehlungen (UNESCO 2022) entspricht, kann aber auch von anderen Personen im privaten Sektor eingesetzt werden, um die ethische Gestaltung, Entwicklung und den Betrieb von KI-Systemen zu erleichtern (UNESCO 2023).

11 KI Governance-Struktur

11.1 EU-Ebene

AI Office („KI-Büro“, Artikel 64 AI Act)

Das AI Office ist für die Überwachung der Umsetzung und des Inkrafttretens des AI Acts sowie für die Formulierung der Regeln für KI-Modelle mit allgemeinem Verwendungszweck („*general purpose AI [GPAI] models*“, also generative KI wie LLMs) zuständig. Es ist für die Schaffung von Normen, Kodizes und Standards für die Steuerung von KI auf EU-Ebene und für die Übernahme der Verwaltungsaufgaben der Europäischen Kommission in diesem Bereich verantwortlich.

AI Board („KI-Gremium“, Artikel 65, 66 AI Act)

Das AI Board ist ein Gremium, dem neben dem AI Office (ohne Stimmrecht) und dem Europäischen Datenschutzbeauftragten als Beobachter je eine Vertreterin oder ein Vertreter pro Mitgliedstaat angehören. Die Aufgaben des AI Boards drehen sich um die Erleichterung der einheitlichen und wirksamen Anwendung des AI Acts, beispielsweise die Koordinierung der nationalen Behörden, die Beratung in Bezug auf die Vorschriften zu GPAI, die Unterstützung bei der Steigerung der KI-Kompetenz, das Liefern von Beiträgen zu Leitfäden und die Beratung der Kommission. Gemäß dem AI Act richtet das Board zwei ständige Untergruppen ein, die sich mit Marktüberwachung und dem Austausch von Behörden befassen. Bei Bedarf können auch temporäre Untergruppen, die sich auf spezifische Themen der KI-Politik fokussieren, eingerichtet werden.

Advisory Forum („Beratungsforum“, Artikel 67 AI Act)

Das Advisory Forum stellt technisches Fachwissen bereit und berät AI Board und Europäische Kommission. Ihm gehören die Agentur für Grundrechte (FRA), die Agentur der Europäischen Union für Cybersicherheit (ENISA), das Europäische Komitee für Normung (CEN), das Europäische Komitee für elektrotechnische Normung (CENELEC) und das Europäische Institut für Telekommunikationsnormen (ETSI) als ständige Mitglieder an. Darüber hinaus sind eine Auswahl von Interessensträgern, darunter Start-ups, Industrie, KMU, Zivilgesellschaft und Wissenschaft mit anerkanntem Expertinnen- und Expertenwissen im Bereich der KI vertreten. Jedes Mitglied wird für einen Zeitraum von 2 Jahren gewählt.

Scientific Panel („Wissenschaftliches Gremium“, Artikel 68 AI Act)

Bei dem Scientific Panel handelt sich um ein Gremium unabhängiger Sachverständiger, die von der Kommission auf der Grundlage ihres aktuellen wissenschaftlichen und technischen Fachwissens auf dem Gebiet der KI ausgewählt werden und die auch von Anbietern von KI-Systemen unabhängig sind. Es hat die Aufgabe, die Umsetzung des

AI Acts in Bezug auf GPAI zu unterstützen, beispielsweise indem es bei der Risikoeinstufung Hilfestellung leistet und Instrumente, Methoden und Benchmarks entwickelt, um die Fähigkeiten von GPAI zu bewerten, und die Arbeit der Marktüberwachungsbehörden zu unterstützen.

11.2 Nationale Ebene

AI Policy Forum

Diese interministerielle Arbeitsgruppe wurde im November 2021 unter dem gemeinsamen Vorsitz von BMK und BKA 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 Herangehensweisen zum Einsatz von KI in den Bundesministerien fördern und aktuelle Fragen zu KI diskutieren. Ein wesentliches Element des AI Policy Forums ist die Einrichtung von Ad-Hoc Arbeitsgruppen zu verschiedenen KI relevanten Themen. Dies geschieht zum Beispiel im Zusammenhang mit der Umsetzung des AI Acts, wobei der Schwerpunkt auf innovationsfördernden Maßnahmen und den spezifischen Bestimmungen für die Umsetzung von KI in der öffentlichen Verwaltung liegt. Dabei werden einschlägige Expertinnen aus Forschung, Wissenschaft, Wirtschaft, Sozialpartnern, NGOs und Zivilgesellschaft eingebunden.

KI-Servicestelle bei der RTR

Die bei der Rundfunk und Telekom Regulierungs-GmbH angesiedelte KI-Servicestelle ist Anlaufstelle und Ansprechpartner für eine breitere Öffentlichkeit zum Thema KI. Sie betreibt unter <https://ki.rtr.at> ein breit gefächertes Informationsportal, insbesondere zu regulatorischen Rahmenbedingungen beim Einsatz von KI. Ihre Aufgaben sind:

- als niedrigschwellige, kompetente und zentrale Anlaufstelle für KI-Projekte mit einem vielfältigen Informations- und Beratungsangebot rund um KI zu dienen,
- den Wissensaufbau und Wissensaustausch rund um KI zu fördern, und zu diesem Zweck Veranstaltungen und Studien durchzuführen,
- den KI-Beirat zu betreuen,
- die Marktentwicklung zu beobachten und bei Bedarf durch Leitlinien und Kommunikation zu unterstützen.

KI-Beirat

Der österreichische KI-Beirat wurde als weiteres Gremium innerhalb der RTR zur Unterstützung der österreichischen KI-Politik geschaffen. Er besteht aus 3 Mitgliedern, die vom Bundeskanzler gewählt werden, und acht Mitgliedern, die vom Finanzminister gewählt werden. Alle 11 Mitglieder sind Expertinnen und Experten in den Bereichen Ethik, Forschung, Wirtschaft, Recht oder technisches Wissen über KI. Die Aufgaben des Beirats sind:

- Beratung und Information der Mitglieder der österreichischen Regierung und der RTR über die neuesten Entwicklungen auf dem Gebiet der KI, einschließlich der technischen und ethischen Aspekte,
- Beobachtung der technologischen Entwicklung von KI innerhalb und außerhalb der Europäischen Union, einschließlich der Identifizierung von Chancen und Möglichkeiten für Österreich in diesem Bereich
- Unterstützung der für KI-Politik verantwortlichen Regierungsmitglieder durch Identifizierung der vielfältigen Themen im Zusammenhang mit KI-Politik und Priorisierung und Fokussierung auf die dringendsten Fragen,
- strategische Planung und Beratung im Rahmen des AI Policy Forums zur Entwicklung und Umsetzung der österreichischen KI-Strategie, einschließlich ihrer Ziele, Prioritäten und Maßnahmen.

Chief Digital Officer Taskforce

Die „Chief Digital Officer Taskforce (CDO-Taskforce)“ ist ein vom Bundeskanzleramt einberufenes Gremium. In ihr vereint sind die CDOs der Ressorts und deren Koordinierungsauftrag in Sachen Digitalisierung. Das Ziel der Taskforce ist der Informationsaustausch zu digitalisierungsrelevanten Maßnahmen, die Erstellung und Kommunikation der Digitalisierungsstrategie und deren Verfolgung.

AI-Stakeholder-Forum

Das AI Stakeholder Forum ist eine Initiative, die von Bundeskanzleramt (BKA) und Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK) organisiert und koordiniert wird. Es versammelt Akteure und Institutionen aus Wissenschaft, Industrie, Wirtschaft und Gesellschaft, wie z.B. die Sozialpartner, die außerhalb der öffentlichen Verwaltung stehen. Der Fokus liegt auf dem Austausch zwischen der Bundesregierung und verschiedenen Stakeholdern außerhalb der Regierung, die sich auf unterschiedliche Art und Weise mit dem Thema KI befassen, oder von dem Thema betroffen sind. Durch das AI Stakeholder Forum kann unter anderem im Kontext der Umsetzung des AI Acts oder der KI-Innovationsmöglichkeiten in Österreich eine bessere Abstimmung und bedarfsgerechte Gestaltung erreicht werden.

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

Im abschließenden 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 Tabelle ersichtlich, empfehlen wir Tools bzw. Institutionalisierungsmaßnahmen, also strukturelle Anpassungen, mit sechs verschiedenen Zielsetzungen in den entsprechenden Bereichen.

Tabelle 3: Empfehlungen KI-Governance

Ziel (Funktion)	Kompetenz- aufbau & Fortbildung	KI-Management Entscheidungshilfen	Experimentation	Zertifizierungen, Nutzungsbedingungen & Kontrolle	Folgenabschätzung & Risiko Management	Kommunikation & Stakeholder- einbindung
Tools (Prozess)	Bildungsstandards für die Beschaffung und Verwendung von KI-Anwendungen /interne Kompetenzen zu technischen Verfahren und Entwicklung ethischer KI („Ethics by Design“, De-biasing)	KI-Einsatz-Entscheidungsbaum, AI RMF (US NIST), Selbstverpflichtende Leitlinien (DE BMAS), Risk Assessment Tools und Entwicklungsstandards („Ethics by Design“; IEEE), Entscheidungskriterien für interne /externe Beschaffungsvorgänge (Amsterdam Klauseln)	Diskussionen zu Good Practices und Herausforderungen, Experimente zu verschiedenen Vorgehensweisen und Tools	Zertifizierungen von ISO, IEEE, TÜV, in Entstehung begriffene Tools wie data. nutrition, data. hazards	Risk Assessment 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 Herausforderungen
Institutionalisierung (Struktur)	VAB KI-Ethikseminar und Führungskräfte Lehrgang, Entwicklung KI Literacy Curriculum differenziert nach Kompetenzniveaus (vgl. DE KI-Campus), themenspezifische interne Kompetenzstellen, Informations- und Diskussionsveranstaltungen	Verwaltungsethikrat mit internen und externen Expertinnen und Experten (vgl. FI Aurora bzw. Etairos Ethikrat), KI-Observatorium (DE BMAS)	Interministerielles KI-Ethik Lab (vgl. AIT AI Ethics Lab, vgl. FI Projekt Aurora), Regulatory KI-Sandbox / Real-labor, AI Policy Forum	Freiwillig / nach AIA Implementierung KI-Behörde, Datenrepositorien, welche konform zu rechtlichen und ethischen Vorgaben sind (Compliance), Richtlinien und Vorgaben für Verwendung (bspw. durch Firmen)	Transparenzregister, Verarbeitungsverzeichnis, KI-Servicestelle (RTR) bzw. KI-Behörde, NL Algorithm Register	AI Policy Forum, PIAZZA Format (Algorithm Watch), Interministerielles KI-Ethik Lab

Kompetenzaufbau und Fortbildung

Das Ziel von Kompetenzaufbau und Fortbildung in der Verwaltung ist von besonderer Bedeutung, weil KI-Literacy die Basis für alle anderen Maßnahmen darstellt. 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 AI Act (siehe Abschnitt 8.3).

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 7.5, 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 Expertinnen und Experten als Ansprechpersonen des jeweiligen Fachbereichs eine niederschwellige erste Unterstützung in Bezug auf ethische KI-Entwicklung, Anwendung und Evaluation übernehmen. Auf jeden Fall sollte aber eine Person eine designierte Rolle als Ansprechpartnerinnen und Ansprechpartner haben, ähnlich den Datenschutzbeauftragten.

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

Hier gibt es international verschiedene Vorbilder, vom deutschen KI-Campus bis zum australischen National AI Center. Der KI-Campus¹⁸ ist eine Lernplattform für KI mit kostenlosen Online-Kursen, Videos und Podcasts, der vom Bundesministerium für

18 <https://ki-campus.org/>

Bildung und Forschung (BMBF) gefördert wird. Das australische National AI Center¹⁹ (NAIC) bietet Konferenzen, Workshops, Schulungsvideos sowie Informationsmaterial aller Art für verschiedene Zielgruppen an. Teil davon ist das Responsible AI Network RAIN, ähnlich dem NAIC eine Kooperation von Unternehmen, Forschung und zivilgesellschaftlichen Organisationen.

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 ebenfalls 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 6 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 schließen die in Abschnitt 2 und Abschnitt 10.2 vorgestellten Bewertungshilfen „Checkliste für ethische KI in der Verwaltung“ und „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-Serviceestelle vor (IÖB 2021) bzw. ein umfangreicher US-amerikanischer Beitrag (Hickok 2024). Eine Art Triage zum Thema Ethik wird durch den in diesem Leitfaden vorgeschlagenen Entscheidungsbaum ermöglicht (für einen

19 <https://www.industry.gov.au/science-technology-and-innovation/technology/national-artificial-intelligence-centre>

umfangreicheren Alternativvorschlag der NGO AlgorithmWatch siehe Loi et al. 2021). Zudem lassen sich Entscheidungskriterien für Beschaffungsvorgänge aus den ethischen Prinzipien und Standards ableiten bzw. mit der daraus hervorgegangenen und hier im Leitfaden vorgestellten Checkliste und Bewertungshilfe „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)“ bearbeiten. In den Niederlanden hat die Stadt Amsterdam „*Standard Clauses for Procurement of Trustworthy Algorithmic Systems*“ entwickelt (City of Amsterdam 2021).

Erprobung

Eine erfolgreiche Einführung und Weiterentwicklung von KI-Anwendungen in der öffentlichen Verwaltung bedürfen der Möglichkeit der Erprobung. Idealerweise sollte eine Technologie vor dem 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 gegenseitige Information und Abstimmung von Verwaltungsakteuren, 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 die im AI Act vorgeschlagenen Reallabore bzw. Sandboxes begleiten könnte (für KI Ethik Labs: Biegelbauer et al. 2022).

Regulatory Sandboxes („regulatorische Sandkästen“)

Regulatory Sandboxes, ein im Fintech Sektor bereits sehr populäres Governance Instrument, sind reale Testumgebungen, in denen Entwickler einer digitalen Innovation Funktionalitäten, Merkmale und Leistung in einer kontrollierten Umgebung testen können, die von Regulierungsbehörden überwacht wird (Allen 2019; Makarov und Davydova 2021). Regulatory Sandboxes gelten als Formen der „intelligenten“ und „lebendigen“ Regulierung, da sie es den Regulierungsbehörden ermöglichen, Probleme zu beobachten, die sich in frühen Entwicklungsphasen ergeben, sowie potenzielle Probleme, die für die Zielgruppe(n) und die Verwendungszwecke, für die die Anwendung entwickelt wurde, auftreten könnten (Zetzsche et al. 2017). Im Zusammenhang mit KI fordert der AI Act die Einrichtung zumindest einer nationalen regulatorischen Sandbox, wobei auch grenzüberschreitende Zusammenarbeit und Beteiligung gefördert wird (Artikel 57 Absatz 1 AI Act). Ihre Aufgabe ist es, die Entwicklung, das Training, das Testen und die Validierung von KI-Innovationen für einen begrenzten Zeitraum vor ihrer Marktfreigabe zu erleichtern (Artikel 57 Absatz 5 AI Act) (Buocz et al. 2023). Daneben sieht der AI Act auch außer-

halb von Reallaboren die Möglichkeit vor, Tests unter Realbedingungen durchzuführen (Artikel 60 AI Act). Während es in Österreich und anderen europäischen Ländern schon einige Beispiele für regulatorische Sandboxes für FinTechs gibt, befinden sich diese im Kontext von KI noch in der Entwicklung. Einige EU-Länder, beispielsweise Spanien²⁰, Dänemark²¹ und Schweden, haben Sandboxes für KI eingerichtet und sind dabei, diese in Betrieb zu nehmen. Außerhalb der EU gibt es dazu ebenso eine Reihe von Beispielen, etwa in Norwegen²², dem UK²³ oder Singapur (letztere richtet sich speziell an KMU).²⁴

Monitoring und Kontrolle

Monitoring ist ein wichtiger Aspekt, den man bei der Entwicklung und Implementierung von KI-Systemen im Auge behalten muss, da es unter anderem ermöglicht, die Leistung, Sicherheit, Wirksamkeit und Funktionsweise des Systems im Laufe der Zeit zu erkennen. Im Rahmen des AI Acts wird von den KI-Anbietern ein Monitoringplan für die Zeit nach dem Inverkehrbringen eines KI-Systems als wesentlicher Teil der erforderlichen technischen Dokumentation verlangt (Artikel 72, Anhang IV Ziffer 7 AI Act). Während die spezifischen Richtlinien für diesen Prozess von der AI Office der EU noch definiert werden müssen, gibt es verschiedene Arten des Monitorings, die man je nach Zweck durchführen kann. Es gibt unter anderem die folgenden (Yampolskiy, 2024):

1. Funktional – Verfolgung der Leistung des Systems im Hinblick auf seine vorgesehenen Aufgaben und Ziele. Dabei können Aspekte wie die Genauigkeit, Effizienz und Zuverlässigkeit des Systems gemessen werden.
2. Sicherheit – Schwerpunkt auf der Ermittlung von Schäden, die den Benutzer und Benutzerinnen entstehen können. Zu den Aspekten könnten die Sicherheit und Robustheit des Systems oder seine Übereinstimmung mit den einschlägigen Gesetzen gehören.
3. Ethische und soziale Aspekte – die Auswirkungen des Systems auf den Einzelnen und die Gesellschaft.
4. Entscheidungsmonitoring – dieses stellt sicher, dass die getroffenen Entscheidungen am besten geeignet sind, die gesteckten Ziele zu erreichen.

Der AI Act sieht vor, dass die nationalen Behörden die Einhaltung verschiedener Standards überwachen, die wiederum von der europäischen AI Office festgelegt und überwacht werden sollen. Einige nationale Behörden wurden bereits eingerichtet, noch bevor der AI Act in Kraft getreten ist, um frühzeitig Erfahrungen zu sammeln. Diese Vorgangsweise wurde in Spanien gewählt, wo der Beschluss für die Etablierung einer derartigen Behörde

20 Siehe <https://portal.mineco.gob.es/es-es/digitalizacionIA/sandbox-IA/Paginas/sandbox-IA.aspx>

21 Siehe <https://www.datatilsynet.dk/hvad-siger-reglerne/vejledning/regulatorisk-sandkasse/udvaelgelse-af-projekter>

22 Siehe <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/>

23 Siehe <https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/>

24 Siehe <https://go.gov.sg/sme-gen-ai>

bereits 2022 gefallen ist. Nach dem AI Act müssen die Behörden in den EU-Mitgliedsstaaten 12 Monate nach Inkrafttreten des AI Acts ihre Arbeit aufgenommen haben.

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 hier eine raschere Entscheidungspraxis zu Technologieanwendungen, als das bei der DSGVO der Fall war, positiv auswirken.

Zertifizierungen

Um Sicherheit für die Nutzung bestimmter Anwendungen oder Datensätze zu erhalten, empfiehlt sich die Erarbeitung bzw. Verwendung von Zertifizierungen. Zwar gibt es derzeit nur wenige Zertifizierungsprogramme, die spezifisch auf die Bewertung ethischer KI ausgerichtet sind, doch gibt es mehrere Ansätze, die zu diesem Zweck genutzt werden könnten. Zum Beispiel:

- **Audit-basierte Zertifizierung.** Dabei handelt es sich um Audits eines KI-Systems, um zu beurteilen, ob es eine Reihe von Spezifikationen einhält. Sie erfordern in der Regel eine Prüfung der durch das Verfahren vorgegebenen Dokumentation durch einen Dritten. Beispiele für Zertifizierungsprogramme sind IEEE's CertifAIed, das KI-Zertifizierungsprogramm von TÜV SÜD oder das Algorithmic Transparency Certificate der spanischen Vereinigung für digitale Wirtschaft.²⁵
- **Normen.** Es handelt sich dabei um Spezifikationen und Normen, die für die Massenanwendung in der Branche geschaffen wurden. Beispiele für Standards, die sich mit ethischen KI-Systemen befassen, sind die Normen ISO/IEC 23053:2022 und die IEEE-Serie 7000 (DIN und DKE 2022).
- **Labels, Qualitätszeichen.** Dabei handelt es sich um Gütesiegel und Qualitätsmarken, die von privaten oder öffentlichen Akteuren entwickelt wurden, um die Qualität eines KI-Systems zu gewährleisten. Sie können verwendet werden, um bestimmte Eigenschaften eines KI-Systems oder seine Gesamtleistung in einem oder mehreren Bereichen zu bewerten. Beispiele hierfür ist das Digital Trust Label der Swiss Digital Initiative. Andere Tools, wie beispielsweise im Fall der Zertifizierung von Datensätzen, data.nutrition oder data.hazards, sind in Entstehung begriffen.²⁶

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 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, anderer-

25 Siehe <https://www.algorithmictransparency.io/>

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

seits für die Nutzung derartiger Repositorien, insbesondere durch Anwenderinnen und Anwender ohne IT-Ausbildung.

Folgenabschätzung & Risikomanagement

Damit in Zusammenhang steht auch die Abschätzung der Folgen von KI-Anwendungen, die vor der In-Verkehr-Bringung 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 10.2 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 Inklusion, 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. Algorithmen Register der niederländischen Regierung²⁷) 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 (vgl. auch OECD 2024). Dies scheint aus verschiedenen Gründen sinnvoll. In Bezug auf die Kommunikation gilt es die Gesellschaft auf

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

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 Konferenz²⁸ 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 und Next 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 Expertinnen und Experten gemeinsames Lernen und Lösungsfindung zu alltäglichen Problemstellungen bei KI-Entwicklung und -Einsatz betrieben werden könnte.

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

Quellenverzeichnis

- Access Now.** (2024, März 13). The EU AI Act: A failure for human rights, a victory for industry and law enforcement. Access Now. <https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement/>. Zugegriffen: 13. September 2024.
- Ada Lovelace Institute, und DataKind UK.** 2020. Examining the Black Box: Tools for assessing algorithmic systems. Bericht. <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. Bericht. <https://www.ai-ethics-impact.org/en>. Zugegriffen: 07. Juni 2023.
- Allen, Hilary J.** 2019. Regulatory Sandboxes. *George Washington Law Review* 87:579.
- 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>.
- Article 19.** (2024, April 4). EU: AI Act fails to set gold standard for human rights. ARTICLE 19. <https://www.article19.org/resources/eu-ai-act-fails-to-set-gold-standard-for-human-rights/>. Zugegriffen: 12. September 2024.
- Aschauer, Ricarda.** 2024. ChatGPT, Gemini & Co. Anwendungsfelder von Large Language Models im rechtlichen Kontext. In Mayrhofer, Michael, Nessler, Bernhard, Bieber, Thomas, Fister, Mathis, Homar, Philipp, und Tumpel, Michael. Hrsg. *ChatGPT, Gemini&Co. Große Sprachmodelle und Recht*. Wien: Manz Verlag. 1–14.
- 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.
- Bashir, Noman, Priya Donti, James Cuff, Sydney Sroka, Marija Ilic, Vivienne Sze, Christina Delimitrou, und Elsa Olivetti.** 2024. The Climate and Sustainability Implications of Generative AI. An MIT Exploration of Generative AI, <https://doi.org/10.21428/e4baedd9.9070dfe7>.
- Bauknecht, Dierk, and Klaus Kubeczko.** 2024. Regulatory Experiments and Real-World Labs: A Fruitful Combination for Sustainability. *GAIA – Ecological Perspectives for Science and Society*, vol. 33, no. 1, Mar. 2024, pp. 44–50. DOI.org (Crossref), <https://doi.org/10.14512/gaia.33.S1.7>.
- 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.
- Berthelot, Adrien, Caron, Eddy, Jay, Mathilde und Lefèvre, Laurent.** 2024. Estimating the environmental impact of Generative-AI services using an LCA-based methodology. *Procedia CIRP* 122, 707-712. <https://doi.org/10.1016/j.procir.2024.01.098>.

- 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.
- Biegelbauer, Peter, Alexander Schindler, Rodrigo Conde-Jiménez und Pia Weinlinger.** 2024. Exciting Opportunities and Necessary Safeguards for Large Language Models in the Public Sector, *ERCIM News*, Vol. 136, 25-26
- 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.
- Brown, Tom B., Mann, Benjamin, Ryder, Nick, Subbiah, Melanie, Kaplan, Jared, Dhariwal, Prafulla, Neelakantan, Arvind, Shyam, Pranav, Sastry, Girish, Askell, Amanda, Agarwal, Sandhini, Herbert-Voss, Ariel, Krueger, Gretchen, Henighan, Tom, Child, Rewon, Ramesh, Aditya, Ziegler, Daniel M., Wu, Jeffrey, Winter, Clemens, Hesse, Christopher, Chen, Mark, Sigler, Eric, Litwin, Mateusz, Gray, Scott, Chess, Benjamin, Clark, Jack, Berner, Christopher, McCandlish, Sam, Radford, Alec, Sutskever, Ilya und Amodei, Dario.** 2020. Language Models are Few-Shot Learners. *arXiv*.
- 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.
- BKA, und BML.** 2008. Standards der Öffentlichkeitsbeteiligung: 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.
- Buocz, Thomas, Pfothner, Sebastian, und Eisenberger, Iris.** 2023. Regulatory sandboxes in the AI Act: reconciling innovation and safety? *Law, Innovation and Technology* 15 (2): 357–389.
- BusinessEurope.** 2021. The Artificial Intelligence Act (AI Act) – a BusinessEurope position paper. <https://www.busesseurope.eu/publications/artificial-intelligence-act-ai-act-busesseurope-position-paper>. Zugegriffen: 19. Mai 2023.
- BusinessEurope.** 2023. Joint industry statement on the EU Artificial Intelligence (AI) Act. <https://www.busesseurope.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.

- Caroli, Laura.** 2024, April 19. Will the EU AI Act work? Lessons learned from past legislative initiatives, future challenges. IAPP: International Association of Privacy Professionals. <https://iapp.org/news/a/will-the-eu-ai-act-work-lessons-learned-from-past-legislative-initiatives-future-challenges>. Zugegriffen: 12. September 2024.
- 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. *Handbuch Digitalisierung der Verwaltung*, Hrsg. Tobias Krause, Christian Schachtner, Basanta Thapa, 233–251. Bielefeld: transcript.
- CoE Commissioner for Human Rights.** 2019. Unboxing Artificial Intelligence: 10 steps to protect Human Rights. <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>. Zugegriffen: 21. August 2024.
- 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>.
- Croxtton, John, Robusto, David, Thallam, Satya und Calidas, Dough.** 25. Juni 2024. *How to Create an AI Incident Reporting System*. Federation of American Scientists. <https://fas.org/publication/establishing-an-ai-incident-reporting-system/>. Zugegriffen: 09. September 2024.
- 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.
- Dastin, Jeffrey und Nellis, Stephen.** 2023. Focus: For tech giants, AI like Bing and Bard poses billion-dollar search problem. Verfügbar unter: <https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/>. Zugegriffen: 05. September 2024.
- 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.
- de Vries, Alex.** 2023. The growing energy footprint of artificial intelligence. *Joule* 7(10), S. 2191–2194.
- 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 Bundestag.** 2020. Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. <https://dserver.bundestag.de/btd/19/237/1923700.pdf>. Zugegriffen: 16. September 2024.
- 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.
- DIN und DKE.** 2022. Deutsche Normungsroadmap Künstliche Intelligenz. 2. Auflage. <https://www.din.de/resource/blob/916792/20bf33d405710a703aa26f81362493bb/nrm-ki-deutsch-2022-final-web-250-data.pdf>. Zugegriffen: 15. Oktober 2024.

- 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.** 2021. Standardisierung Künstlicher Intelligenz und KI-Verordnungsvorschlag, *Recht Digital (RDJ)* 2 (12): 588–597.
- 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 Center for Not-for-Profit Law.** 2024, April 3. Packed with loopholes: Why the AI Act fails to protect civic space and the rule of law | ECNL. <https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law>. Zugegriffen: 12. September 2024.
- 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. European Declaration of Digital Rights and Principles in the Digital Decade’ COM. Zugegriffen: 10. Oktober 2024.
- European Commission.** 2022b. Data Governance Act Explained. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>. Zugegriffen: 07. Juni 2023.
- European Commission.** 2022c. 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.** 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 Sopic, 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>.
- Goodfellow, Ian J., Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron und Bengio, Yoshua.** 2014. Generative Adversarial Networks. *arXiv*, 1–9.
- Hacker, Philipp.** 2024. Sustainable AI Regulation. *Common Market Law Review* 61 (2): 345–386.
- Haslinger, Susanne.** 2022. Gesetzliche Rahmenbedingungen für den Einsatz von KI-Assistenzsystemen am Arbeitsplatz. *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>.
- Hickok, Merve.** 2024. From Trustworthy AI Principles to Public Procurement Practices. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783111250182>
- 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>.

- eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment. Zugegriffen: 07. Juni 2023.
- Ho, Jonathan, Jain, Ajay und Abbeel, Pieter.** 2020. Denoising Diffusion Probabilistic Models. *arXiv*, 1–25.
- Holzinger, Gerhart, Peter Oberndorfer und Bernhard Raschauer (Eds.).** 2013. *Österreichische Verwaltungslehre*. Wien, Verlag Österreich.
- Human Rights Watch.** 2021. How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers. Human Rights Watch. <https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net>. Zugegriffen: 12. September 2024.
- Humm, Bernhard G., Buxmann, Peter und Schmidt Jan C..** 2022. Grundlagen und Anwendungen von KI. *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.
- IEA (2024), Electricity 2024, IEA, Paris** <https://www.iea.org/reports/electricity-2024>, Licence: CC BY 4.0
- 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 Durresi.** 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.
- Keller, Paul und Warso, Zuzanna** (2023, März 28). The EU should not trust AI companies to self-regulate. Open Future. <https://openfuture.eu/blog/the-eu-should-not-trust-ai-companies-to-self-regulate>. Zugegriffen: 12. September 2024.
- KPMG International.** 2024. Decoding the EU AI Act: Understanding the AI Act's Impact and How You Can Respond. <https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2024/02/decoding-the-eu-artificial-intelligence-act.pdf>. Zugegriffen: 12. September 2024.
- Kingma, Diederik P. und Welling, Max.** (2013). Auto-Encoding Variational Bayes. *arXiv*, 1–14.
- 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>.

Leitl-Staudinger, Barbara, Pabel, Katharina, und Steiner, Wolfgang. 2023. *Österreichische Verwaltungslehre. Lehrbuch*. 4. Auflage. Wien: Verlag Österreich.

Lobe, Adrian. 2022. Diskriminierung durch und von KI. *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.

Luccioni, Alexandra Sasha, Viguier, Sylvain, Ligozat, Anne-Laure 2022. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *arXiv*.

Madan, Rohit, und Mona Ashok. 2022. A Public Values Perspective on the Application of Artificial Intelligence in Government Practices: A Synthesis of Case Studies. *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.

Madiega, Tambiama, Ilnicki, Rafał. 2024. AI investment: EU and global indicators, European Commission, Joint Research Centre. Von: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2024\)760392](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2024)760392)

Makarov, Vladislav O., und Marina L. Davydova. 2021. On the Concept of Regulatory Sandboxes. „*Smart Technologies“ for Society, State and Economy*. Springer International Publishing.

Mantelero, Alessandro. 2022. Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI. In *Information Technology and Law Series Volume 36*, Hrsg. Simone van der Hof, Bibi van den Berg, Gloria González Fuster, Eva Lievens und Bendert Zevenbergen. Den Haag: T.M.C Asser Press. <https://doi.org/10.1007/978-94-6265-531-7>. Zugegriffen: 29. August 2024.

Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. Recalibrating Global Data Center Energy-Use Estimates. *Science* 367, Nr. 6481, S984–986.

Mayrhofer, Michael, und Parycek, Peter. 2022. Digitalisierung des Rechts – Herausforderungen und Voraussetzungen. *Verhandlungen des 21. Österreichischen Juristentages*. Wien: Manz Verlag.

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.

McKinsey & Company. 2024. Mit Mut und Augenmaß, bitte! Wie GenAI die Arbeit der öffentlichen Verwaltung unterstützen und den Fachkräftemangel abfedern kann. Bericht. Zugegriffen: 20. August 2024

Ministry of Finance Finland. o. J. Implementation of the National AuroraAI Programme. Valtiovarainministeriö. <https://vm.fi/en/auroraai-en>. Zugegriffen: 17. Mai 2023.

Mörth, Philipp. 2020. Das Legalitätsprinzip. Gesetzesvorbehalt und Determinierungsgebot im österreichischen Recht. Wien: Verlag Österreich.

Müller, Madeleine, und Schneeberger, David M. 2024. Menschenrechtsfolgenabschätzungen im Artificial Intelligence Act. Ein Instrument zum Schutz von vulnerablen Gruppen oder bloße Pflichtübung? *juridikum* 2: 265–274.

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 Filipova, Niklas Gudowsky-Blatakes, Barbara Heller-Schuh, Manuela Kienecker, 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.** 2024. OGM/APA-Vertrauensindex Institutionen Oktober 2024. OGM Website. <https://www.ogm.at/2024/11/01/ogm-apa-vertrauensindex-institutionen-oktober-2024/>. Zugegriffen: 11.11.2024.
- Panda, Ashwinee/Choquette-Choo, Christopher A./Zhang, Zhengming/Yang, Yaoqing/Mittal, Prateek.** 2024. „Teach LLMs to Phish: Stealing Private Information from Language Models.“ *arXiv*. <https://arxiv.org/abs/2403.00871>.
- Patterson, David, Gonzalez, Joseph, Hölzle, Urs, Le, Quoc, Liang, Chen, Munguia, Luke-Ming, Rothchild, David, So, David R., Texier, Maud und Dean, Jeffrey.** 2022. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer*, 55(7), 18–28. <https://doi.org/10.1109/MC.2022.3148714>.
- Perrigo, Billy.** 2023. Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. Time. <https://time.com/6247678/openai-chatgpt-kenya-workers/> Zugegriffen: 09.August.2024.
- Pichal, Sundar.** 2018. AI at Google: our principles. Google Website. <https://blog.google/technology/ai/ai-principles/>. Zugegriffen: 07. Juni 2023.
- Pilniok, Arne.** 2024. Unionsrechtliche Regulierung des Einsatzes von KI-Systemen in der öffentlichen Verwaltung. *Die Öffentliche Verwaltung (DÖV)* 77 (14): 581–592.
- Ramon et al.** (2023). Experimentally testable noncontextuality inequalities beyond the Peres-Mermin square. In: *Physical Review A* 107(1), 010101. <https://doi.org/10.1103/PhysRevA.107.010101>.
- Raschauer, Bernhard.** 2021. Allgemeines Verwaltungsrecht. Lehrbuch. 6. Auflage. Wien: Verlag Österreich.
- 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.
- Rezende, Danilo J. und Mohamed, Shakir.** 2015. Variational Inference with Normalizing Flows. *arXiv*, 1–10.
- 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.** 2023. Künstliche Intelligenz. Ein moderner Ansatz. 4. Auflage. München: Pearson Studium.
- RTR.** 2024. Medienkompetenz-Bericht 2024: Medienkompetenz in der journalistischen Praxis. Wien.
- Schäfer, Mirko Tobias, und Nelly M. Clausen.** 2021. Participatory Data Ethics. A practical approach. *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 Rothmund-Burgwall, Moritz.** 2024. Einsatz von KI in der Verwaltung – Anforderungen aus den Blickwinkeln des Legalitätsprinzips und des Datenschutzrechts. *Digitalisierung und Recht Jahrbuch 2024*. Hrsg. Hoffberger-Pippan, Elisabeth, Ladeck, Ruth, und Ivankovics, Peter. Wien: Verlag Österreich. 199–222.
- Scheichenbauer, Heidi, und Seidl, Lisa.** 2022. Die verwaltungsrechtliche Einordnung von Internet-Recherchen durch Verwaltungsbehörden. *Datenschutzrecht Jahrbuch 2022*, Hrsg. Dietmar Jähnel, 49–62. Wien: Verlag Österreich. <https://doi.org/10.37942/9783708341194>.
- Schneeberger, David M.** 2024a. Machine Learning in der Verwaltung. Rechtsfragen der Black-Box-Problematik. Wien: Verlag Österreich.
- Schneeberger, David M.** 2024b. Large Language Models in der Verwaltung: (Gen)er(r)are humanum est? In *Digitalisierung und Recht Jahrbuch 2024*. Hrsg. Hoffberger-Pippan, Elisabeth, Ladeck, Ruth, und Ivankovics, Peter. Wien: Verlag Österreich. 223–258.
- Shneiderman, Ben.** 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36 (6), 495–504.
- Schwartzmann, Rolf, Keber, Tobias O., und Zenner, Kai Hrsg.** 2024. KI-Verordnung. Leitfaden für die Praxis. Heidelberg: C.F. Müller.
- Selbst, Andrew D.** 2021. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology* 35 (1): 117–191.
- Sigfrids, Anton, Nieminen Mika, Leikas Jaana, und Pikkuaho 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>.
- Stadt Wien.** 2024. KI-Kompass für Bedienstete der Stadt Wien. <https://digitales.wien.gv.at/ki-kompass-fuer-bedienstete-der-stadt-wien/>. Zugegriffen: 11.November.2024.
- 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.
- The White House.** 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. Zugegriffen: 13. September 2024
- 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 of 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.

- UK Parliament, House of Lords.** 2024. Artificial Intelligence (Regulation) Bill. <https://researchbriefings.files.parliament.uk/documents/LLN-2024-0016/LLN-2024-0016.pdf>. Zugegriffen: 13. September 2024.
- UNESCO.** 2022. Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Zugegriffen: 19. Mai 2023.
- UNESCO.** 2023. Ethical Impact Assessment. A Tool of the Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000386276_eng. Zugegriffen: 29. August 2024.
- 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>.
- Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Łukasz und Polosukhin, Illia.** 2017. Attention Is All You Need. *arXiv*, 1–15.
- Wachter, Sandra.** 2024. Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. *Yale Journal of Law & Technology* 26 (3): 671–718.
- Wendt, Janine und Wendt, Domenik H.** 2024. Das neue Recht der Künstlichen Intelligenz. Artificial Intelligence Act (AI Act). Baden-Baden: Nomos.
- 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.
- Yampolskiy, Roman V.** 2024. On monitorability of AI. *AI and Ethics*. <https://doi.org/10.1007/s43681-024-00420-x>.
- Zetzsche, Dirk A., Ross P. Buckley, Janos N. Barberis, und Douglas W. Arner.** 2017. Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. *Fordham Journal of Corporate and Financial Law* 23:31.

Glossar für Fachbegriffe

AI Act: auch KI-Verordnung genannt. Eine EU-Verordnung, die somit in jedem Mitgliedsstaat unmittelbar anwendbar ist, die KI-Systeme und General-Purpose-AI-Modelle regelt und einem risikobasierten 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 maschinellen lernenden Algorithmen, in anderen Kontexten hat Bias andere Bedeutungen (siehe Abschnitt 7.1)

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 und Bürgerinnen im Chatformat automatisiert zu beantworten versucht (siehe Abschnitt 4, 5, 6)

„Ethics by Design“: methodischer, oft interdisziplinärer Zugang bei der Entwicklung von KI, bei dem ethische Reflexion in allen Stadien berücksichtigt wird (siehe Abschnitt 7.5, 12)

Foundation Model: Großes, maschinell auf der Basis von großen Mengen von Daten erstelltes Modell. Nach diesem Vortraining können diese spezifische Aufgaben, wie z. B. Chat, Übersetzung, Textklassifikation feinjustiert werden (Fine-Tuning). Ein bekanntes Beispiel für Foundation Models sind große Sprachmodelle oder Large Language Models (LLMs), Neuronale Netze mit teilweise mehreren Milliarden von Parametern, die unterschiedlichste NLP-Aufgaben wie Textklassifikation, Textgenerierung, Sprachübersetzung, Sentimentanalyse und Frage-Antwort-Systeme übernehmen können. Außer Sprache gibt es auch visuelle und multimodale Foundation Models, die z. B. auf der Basis von Eingabetext Bilder erzeugen können.

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 9.2)

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 11, 12)

Grundlagenmodell: siehe „Foundation Model“

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 8, 9.1)

Anhang

Fallbeispiel für die „Checkliste ethische, rechtlicher KI in der öffentlichen Verwaltung“

Anwendungsfall: Einführung einer KI zur Verbesserung des internen Wissensmanagements in der öffentlichen Verwaltung

In einer Stadtverwaltung sind wertvolle Informationen und Erfahrungsberichte über verschiedene Projekte und Prozesse in unterschiedlichen Abteilungen verteilt und oft schwer auffindbar. Wissen zu Abläufen, Anträgen, Anfragen und Anträgen von Bürgerinnen und Bürgern, Gesetzen, Projekten oder Problemlösungen wird häufig in E-Mails, Dokumenten oder einzelnen Ordnern gespeichert, was zu einem ineffizienten Wissensmanagement führt. Dadurch kommt es öfters zu einer doppelten Bearbeitung von Aufgaben, einer erschwerten Einarbeitung neuer Mitarbeiterinnen und Mitarbeiter, längeren Suchzeiten für relevante Informationen und einer Verzögerung bei der Erstellung von verwaltungsrechtlichen Entscheidungsentwürfen.

Die Stadtverwaltung möchte daher eine KI-basierte Wissensmanagementlösung implementieren, die alle internen Informationen in einem zentralen System organisiert und zugänglich macht. Die KI soll Inhalte automatisch klassifizieren, relevante Informationen aus Dokumenten extrahieren und diese intelligent verknüpfen. Mitarbeitende sollen so leichter auf relevantes Wissen zugreifen können und Suchzeiten werden deutlich verkürzt. Die KI wird außerdem in der Lage sein, auf Basis früherer Suchanfragen ähnliche Informationen vorzuschlagen, um den Zugang zu Ressourcen zu verbessern. Darüber hinaus kann das System basierend auf Erfahrungswerten Empfehlungen für Entscheidungsprozesse abgeben. Beispielsweise kann eine Priorisierung von Anträgen unterstützt werden, indem auf besonders zeitkritische oder komplexe Fälle hingewiesen wird.

Anhand des dargestellten Beispiels wurde die Checkliste im Folgenden ausgefüllt und die Antworten erläutert. Damit wird veranschaulicht, wie die Entscheidung für ein „Ja“- oder „Nein“-Häkchen in der Checkliste zustande kommen kann.

Checkliste Anhang

Recht

Wird das KI-System im Rahmen des (schlicht) hoheitlichen Verwaltungshandelns eingesetzt?

Ja

Nein

Ja. Die Einführung des KI-Systems unterstützt die Stadtverwaltung bei der Erfüllung ihrer hoheitlichen Aufgaben, indem es das Wissensmanagement optimiert und die Zusammenarbeit zwischen den Abteilungen fördert. Außerdem unterstützt das System die Verwaltungsbediensteten mit Vorschlägen und Empfehlungen im Hinblick auf Entscheidungsprozesse.

Sofern (schlicht) hoheitliches Verwaltungshandeln vorliegt, wurde abgeklärt, ob eine ausreichende Rechtsgrundlage für den Einsatz der KI besteht?

Ja

Nein

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

Ja

Nein

Ja. Die Verarbeitung personenbezogener Daten erfolgt unter Einhaltung der Datenschutzgrundverordnung (DSGVO) und die Verwaltungsdokumente erfüllen die jeweils einschlägigen rechtlichen Anforderungen, wie beispielsweise Geheimhaltungspflichten, Datenschutzstandards und die Einhaltung des nationalen und EU-Rechtsrahmens.

Gewährleistet der Einsatz des KI-Systems, dass die Grundrechte von Bürgerinnen und Bürger in keiner Weise beeinträchtigt werden?

Ja

Nein

Ja. Da die Ergebnisse der KI, insbesondere die Empfehlungen zu Entscheidungen werden von fachkundigen Mitarbeiterinnen und Mitarbeiter vor der weiteren Verwendung überprüft. Der Zugang zu Informationen ist auf Mitarbeitende beschränkt und sensible Daten werden gemäß interner Datenschutzrichtlinien gesichert. Somit bestehen insgesamt geringe Risiken für eine Beeinträchtigung der Grundrechte von Bürgerinnen und Bürger.

Wurde die Risikoeinstufung der KI-Anwendung gemäß dem AI Act ermittelt?

Ja

Nein

Ja. Da das System auch Entscheidungsempfehlungen vorbereiten kann, kann es unter bestimmten Umständen (etwa Entscheidungsunterstützungssysteme bei Justizbehörden, oder wenn es um Sozialleistungen geht) ein High Risk System sein, ansonsten handelt es sich um ein System mit „geringem Risiko“.

Falls zutreffend: Ist ein „Conformity Assessment“ laut dem AI Act bereits erfolgt? (Hohes Risiko) Ja Nein

Ja, ein ex-ante „Conformity Assessment“ wurde durchgeführt, weil ein Einsatz in entsprechenden Hochrisikobereichen nicht unwahrscheinlich ist.

Falls zutreffend: Wurden die Transparenzpflichten nach dem AI Act erfüllt? (Mittleres Risiko) Ja Nein

Die Transparenzpflichten sind in den Maßnahmen, die für die Kategorie „Hohes Risiko“ erforderlich sind, bereits enthalten.

Falls zutreffend: Wurde geprüft, ob eine Grundrechte-Folgenabschätzung (AI Act) durchgeführt werden muss? (Hohes Risiko) Ja Nein

Ja. Die Grundrechte-Folgenabschätzung wurde vor der ersten Verwendung des Systems durchgeführt, da die Anwendung (abhängig vom konkreten Anwendungsbereich in der Stadtverwaltung) ein entsprechendes Hochrisiko-KI-System darstellt. Durch die GRFA wurden potenzielle Schadensrisiken und Grundrechtsbeeinträchtigungen identifiziert und Maßnahmen für den Fall des Eintretens dieser Risiken erarbeitet. Die Ergebnisse der Folgenabschätzung wurden der zuständigen Marktüberwachungsbehörde übermittelt.

Wurde geprüft, ob eine Datenschutz-Folgenabschätzung durchgeführt werden muss? Ja Nein

Ja. Die Datenschutz-Folgenabschätzung (DSFA) wurde durchgeführt, da die Anwendung personenbezogene Daten verarbeitet. Durch die DSFA und die darauf basierenden Maßnahmen wurden Risiken entsprechend minimiert.

Transparenz

Sind die spezifischen Ziele und Zwecke des Einsatzes der KI-Anwendung identifiziert und dokumentiert? Ja Nein

Ja. Die spezifischen Ziele der KI-Anwendung wurden klar definiert und umfassen die Verbesserung des internen Wissensmanagements, eine deutliche Reduzierung der Suchzeiten, die Förderung der Zusammenarbeit über Abteilungen hinweg, die Minimierung redundanter Aufgaben und die Unterstützung bei der Erstellung von Entscheidungsentwürfen. Diese Ziele wurden schriftlich dokumentiert und auch dem Auftragnehmer im Vergabeverfahren bezüglich des KI-Systems zur Kenntnis gebracht.

Gibt es eine Dokumentation, die die technische Entwicklung des Modells erläutert? Ja Nein

Ja. Die technische Dokumentation umfasst die Entwicklungsgrundlagen, Modellarchitektur sowie die eingesetzten Algorithmen und Technologien. Sie steht den Nutzerinnen und Nutzer zur Verfügung.

Ist die Funktionsweise der KI-Anwendung nachvollziehbar? Ja Nein

Ja. Die Funktionsweise des KI-Systems ist klar und für alle betroffenen Mitarbeitenden nachvollziehbar dokumentiert.

Sind die Datensätze, die mit dem KI-System verbunden sind, bekannt? Ja Nein

Ja. Alle für die KI-Anwendung verwendeten Datensätze, wie z.B. interne Protokolle, Projektdokumentationen und Prozessanweisungen, sind erfasst.

Wird den Nutzerinnen und Nutzer, wann immer möglich, erklärt, wie das KI-System zu seinen Ausgaben, Inhalten, Empfehlungen oder Ergebnissen kommt und welche Logik dahintersteckt? Ja Nein

Ja. Nutzerinnen und Nutzer erhalten Erläuterungen darüber, wie das KI-System Informationen klassifiziert, verknüpft und Vorschläge generiert. Schulungen und Informationsmaterialien für Mitarbeiterinnen und Mitarbeiter wurden entwickelt, die die Funktionslogik des Systems verständlich machen, darunter eine Übersicht zu den Entscheidungsmechanismen und Suchalgorithmen, die auf den bisherigen Nutzungsmustern basieren.

Werden Personen informiert, wann und auf welche Weise sie mit einer KI-Anwendung interagieren? Ja Nein

Ja. Mitarbeitende werden bei der Nutzung des Systems darüber informiert, dass sie mit einer KI-Anwendung interagieren, die ihre Suchanfragen analysiert und entsprechende Inhalte oder Empfehlungen bereitstellt. Diese Informationen werden sowohl in Schulungen als auch durch Hinweise innerhalb der Benutzeroberfläche vermittelt, sodass Nutzende stets nachvollziehen können, wann sie mit der KI interagieren.

Unvoreingenommenheit und Fairness

Sind die Daten, die zum Training des KI-Systems verwendet werden, vielfältig und repräsentativ für den jeweiligen Kontext? Ja Nein

Ja. Die Datenbasis des KI-Systems wurde gezielt so ausgewählt, dass sie den vielfältigen Kontext und die Anforderungen der verschiedenen Abteilungen innerhalb der Stadtverwaltung widerspiegelt. Es wurde darauf geachtet, dass Dokumente aus allen relevanten Verwaltungsbereichen und für unterschiedliche Anwendungsfälle integriert werden.

Gibt es einen Prozess, um verwendete Datenquellen auf mögliche Verzerrungen und Ungenauigkeiten zu prüfen?

Ja Nein

Ja. Ein Prüfprozess ist etabliert, um die Datenquellen auf Verzerrungen, Ungenauigkeiten und Relevanz für den spezifischen Anwendungsfall zu kontrollieren. Dieser Prozess umfasst eine regelmäßige Analyse der verwendeten Daten auf Anzeichen potenzieller Bias-Faktoren z.B. ein Übergewicht spezifischer Abteilungen oder Projektarten, die eine bestimmte Sichtweise oder Schwerpunktbildung repräsentieren oder die Kontrolle, ob die KI aufgrund vergangener Daten bestimmte Anträge von Bürgerinnen und Bürger unsachgemäß bevorzugt.

Ist die KI-Anwendung so konzipiert, dass sie die Entmenschlichung, Diskriminierung, Stereotypisierung oder Manipulation von Menschen vermeidet?

Ja Nein

Ja. Die KI wurde bewusst so entwickelt und trainiert, dass sie keine endgültigen Entscheidungsbefugnisse oder personalbezogene Bewertungen ausführt, sondern ausschließlich unterstützend agiert. Der Fokus liegt darauf, objektiv Inhalte zu organisieren und bereitgestellte Empfehlungen und Informationen basierend auf den Eingaben der Nutzerinnen und Nutzer sachlich und unverzerrt bereitzustellen. Entmenschlichende oder manipulative Mechanismen können vollständig ausgeschlossen werden.

Gibt es ein Verfahren, mit dem Personen gegen den Einsatz bzw. die Ausgabe des KI-Systems Einspruch oder sonstige Rechtsmittel dagegen erheben können?

Ja Nein

Ja. Ein Einspruchsverfahren ist vorhanden, das Mitarbeitenden ermöglicht, bei Bedenken hinsichtlich der von der KI bereitgestellten Informationen, Entscheidungsempfehlungen oder Dokumentenverknüpfungen Feedback zu geben und Korrekturen anzustoßen. Es wurde eine zentrale Anlaufstelle eingerichtet, an die sich Mitarbeitende wenden können, um Unstimmigkeiten zu melden.

Effektivität und Effizienz

Ergeben sich konkrete Vorteile für die breite Öffentlichkeit durch den Einsatz dieses KI-Systems? (z. B. Zeitersparnis bei der Beantragung einer staatlichen Leistung)

Ja Nein

Ja. Auch wenn die KI-Lösung primär intern verwendet wird, bringt sie unmittelbar Vorteile für die breite Öffentlichkeit. Die effiziente Organisation und der schnelle Zugriff auf Wissen trägt langfristig zu einer erhöhten Servicequalität bei, insbesondere bei komplexen Fragestellungen, die eine umfangreichere Recherche erfordern.

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

Ja. Die KI-Lösung verbessert die Arbeitssituation der Verwaltungsbediensteten, indem sie den Zugang zu relevanten Informationen erleichtert, Suchzeiten verkürzt und bei Entscheidungsentwürfen unterstützt. Dies entlastet die Mitarbeitenden und ermöglicht ihnen, ihre Ressourcen effizienter zu nutzen. Auch die Einarbeitung neuer Mitarbeitender und der Wissensaustausch zwischen den Abteilungen wird dadurch erleichtert.

Gibt es eine Richtlinie zu Qualitäts- und Leistungszielen für das KI-System? Ja Nein

Ja. Die Stadtverwaltung hat eine konkrete Richtlinie mit Qualitäts- und Leistungszielen für das KI-System etabliert.

Werden die Verwaltungsbediensteten entsprechend geschult und unterstützt, um die KI-Anwendung wirkungsvoll einzusetzen? Ja Nein

Ja. Es wurden Schulungsprogramme erstellt, die sicherstellen, dass alle Mitarbeitenden die KI-Anwendung optimal nutzen können. Diese Schulungen umfassen sowohl den technischen Umgang mit dem System als auch Hintergrundinformationen zur Funktionsweise und zu den Vorteilen der KI-Lösung.

Gibt es fortlaufende Unterstützung bei Problemen oder Herausforderungen? Ja Nein

Ja. Fortlaufende Unterstützung und Auffrischungsschulungen stehen zur Verfügung, um den Kompetenzaufbau und die Fortbildung der Mitarbeitenden laufend zu fördern. Eine zentrale Anlaufstelle sowie ein technischer Helpdesk stehen zur Verfügung, um bei Problemen oder Fragen zur Anwendung schnell Hilfestellung zu leisten.

Wurden die Umweltauswirkungen der KI-Anwendung berücksichtigt? Ja Nein

Ja. Die Umweltauswirkungen des KI-Systems wurden geprüft. Die Implementierung der Lösung wurde so gestaltet, dass der Energieverbrauch möglichst geringgehalten wird, z.B. durch die Nutzung effizienter Serverstrukturen. Des Weiteren wird geprüft, ob der laufende Betrieb der KI durch weitere Maßnahmen zur Emissionsreduktion ergänzt werden kann, um die Umweltbelastung möglichst gering zu halten.

Sicherheit

Falls zutreffend: Wurde ein Risikomanagementsystem gemäß dem AI Act für die KI-Anwendung geschaffen? (Hohes Risiko) Ja Nein

Anwendungsbereiche wie etwa Justiz sind möglich, wodurch eine Zuordnung zu Hochrisiko-Anwendungen notwendig wird. Dementsprechend wurden die Vorgaben des AI Acts verfolgt und auch ein Risikomanagementsystem implementiert.

Werden Aufzeichnungen über die Betriebsleistung der KI-Anwendung und alle Vorfälle oder Störungen für einen bestimmten Zeitraum archiviert?

Ja Nein

Ja. Es besteht ein Protokollierungssystem, das die Betriebsleistung der KI-Anwendung überwacht und sämtliche Vorfälle oder Störungen dokumentiert. Diese Daten werden über einen festgelegten Zeitraum archiviert und dienen sowohl der Evaluierung der Systemleistung als auch der Identifikation von Optimierungsmöglichkeiten. Die gesammelten Betriebsdaten ermöglichen zudem eine schnelle Ursachenanalyse und bieten eine Grundlage für zukünftige Anpassungen und Verbesserungen der Anwendung.

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

Ja Nein

Ja. Um das KI-System vor Missbrauch und böswilliger Nutzung zu schützen, wurden Sicherheitsvorkehrungen implementiert. Dazu zählen Zugriffskontrollen, die sicherstellen, dass nur autorisierte und entsprechend geschulte Personen mit der jeweils nötigen (Fach) Kompetenz Personen auf das System zugreifen können, sowie Verschlüsselungsmechanismen für den sicheren Datentransfer.

Zugänglichkeit und Inklusion

Ist die KI-Anwendung menschenzentriert konzipiert? Also so, dass sie von verschiedenen Endnutzerinnen und Endnutzer mit unterschiedlichen Kompetenzniveaus verwendet werden kann?

Ja Nein

Ja. Die Benutzeroberfläche ist einfach und benutzerfreundlich gestaltet, sodass auch Mitarbeitende ohne tiefgreifende IT-Kenntnisse mit der Anwendung arbeiten können. Schulungen und Hilfsmaterialien stehen bereit, um den Einstieg zu erleichtern und die optimale Nutzung zu gewährleisten. Regelmäßiges Feedback von Nutzerinnen und Nutzer fließt zudem in laufende Verbesserungen ein. Die Benutzeroberfläche erfüllt außerdem die Anforderungen an die Barrierefreiheit.

Wurde überprüft, ob Alternativen zur KI-Anwendung angeboten werden können, um einen gleichberechtigten nicht-KI-bezogenen Zugang zu gewährleisten?

Ja Nein

Das KI-System wird ausschließlich verwaltungsintern eingesetzt und die zuvor vorhandenen Zugänge zu nicht-KI-basierten Arbeitsabläufen bleiben unverändert bestehen. Dadurch wird gewahrt, dass auch bei einem Ausfall des KI-Systems oder in Fällen, in denen auf die KI-Lösung verzichtet werden soll, alle bisherigen Bearbeitungs- und Recherchemöglichkeiten vollständig zur Verfügung stehen.

Menschliche Aufsicht

Wurde die KI-Anwendung so entwickelt, dass menschliche Aufsicht möglich ist (z. B. human-in-the-loop, human-on-the-loop)? Ja Nein

Ja. Die KI-Anwendung wurde so konzipiert, dass menschliche Aufsicht und Eingriffe jederzeit möglich sind. Das System folgt einem „human-on-the-loop“-Ansatz, bei Mitarbeiterinnen und Mitarbeiter die Möglichkeit haben, Ergebnisse der KI zu überprüfen und (insbesondere im Bereich von Entscheidungsvorschlägen) bei Bedarf einzugreifen oder Korrekturen vorzunehmen.

Wird das KI-System in regelmäßigen Abständen überprüft (zumindest in Bezug auf Leistung/Qualität, Sicherheit, Einhaltung der geltenden Gesetze und Vorschriften)? Ja Nein

Ja. Es wurde ein Überprüfungsprozess eingerichtet, sodass das KI-System in regelmäßigen Abständen hinsichtlich Leistung, Qualität, Sicherheit sowie der Einhaltung gesetzlicher Vorgaben überprüft werden kann. Diese Überprüfungen umfassen eine Leistungskontrolle zur Sicherstellung optimaler Ergebnisse, eine Sicherheitsüberwachung zur Identifikation potenzieller Schwachstellen sowie Compliance-Audits, die sicherstellen, dass das System fortlaufend mit den neuesten rechtlichen Anforderungen konform ist.

Rechenschaftspflicht

Sind klare Verantwortlichkeiten für Entwicklerinnen und Entwickler, Betreiberinnen und Betreiber und Nutzerinnen und Nutzer der KI-Anwendung festgelegt? Ja Nein

Ja. Klare Verantwortlichkeiten wurden festgelegt. Die Entwicklerinnen und Entwickler sind für die technische Implementierung, Wartung und laufende Optimierung des Systems verantwortlich. Die Betreiberinnen und Betreiber überwachen den Betrieb, die Leistung und die Compliance des KI-Systems und stellen sicher, dass das System ordnungsgemäß und nach den vorgeschriebenen Standards funktioniert. Nutzerinnen und Nutzer sind geschult und für die Anwendung im Arbeitsalltag zuständig, wobei ihnen klare Richtlinien zur Nutzung zur Verfügung stehen.

Wurde festgelegt, wer die letztendliche Verantwortung und Rechenschaftspflicht für den KI-Einsatz sowie die Ausgaben des KI-Systems trägt? Ja Nein

Ja, die Verantwortung und Rechenschaftspflichten sind klar geregelt. Da das KI-System keine Entscheidungen trifft, sondern ausschließlich unterstützend wirkt, liegt die Verantwortung für sämtliche Entscheidungen bei den Verwaltungsbediensteten. Diese tragen die Verantwortung für die korrekte Nutzung des Systems im Einklang mit den Dienstpflichten und Nutzungsrichtlinien. Die Behörde selbst trägt die übergeordnete Verantwortung für die Einführung und den Betrieb des KI-Systems.

Digitale Souveränität

Sind Maßnahmen zur Daten Governance vorhanden, um festzulegen, wie Daten im Zusammenhang mit der Nutzung von KI-Systemen gesammelt, verwendet, gespeichert, gepflegt und verbreitet werden?

Ja Nein

Ja. Ein Data-Governance-Plan wurde implementiert, der Richtlinien und Prozesse für den gesamten Lebenszyklus der Daten festlegt. Diese Maßnahmen umfassen die Erfassung, Nutzung, Speicherung, Pflege und gegebenenfalls die Weitergabe der Daten, die im Zusammenhang mit der KI-Anwendung verwendet werden.

Wird die Datensouveränität der Verwaltung durch den Einsatz der KI gewahrt, insbesondere im Hinblick auf die Privatsphäre der Bürgerinnen und Bürger?

Ja Nein

Ja. Der Einsatz des KI-Systems wurde unter Berücksichtigung der Datensouveränität der Verwaltung konzipiert. Alle personenbezogenen und sensiblen Daten werden innerhalb der Verwaltung gesichert und der Zugang zu diesen Daten ist streng reguliert.

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

Die Entwicklung des KI-Systems wird ausgelagert, wobei für das Training der KI personenbezogene, insbesondere sensible Daten anonymisiert werden. Der Betrieb erfolgt innerhalb der Verwaltung. Es ist sichergestellt, dass ausschließlich berechnete Verwaltungsbedienstete Zugriff auf die personenbezogenen, insbesondere sensiblen Daten haben.

Abbildungsverzeichnis

Abbildung 1: KI-Ethikprinzipien für die österreichische Verwaltung.....	13
Abbildung 2: Generisches KI-System.....	23
Abbildung 3: Beispiel für Human in the Loop	34
Abbildung 4: Repräsentative Umfrage APA/OGM-Vertrauensindex 2024.....	36
Abbildung 5: Entscheidungsbaum zur Verwendung von KI-Technologie.....	41
Abbildung 6: 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.....	59
Abbildung 7: Die Entwicklung des weltweiten Energiebedarfs von Rechenzentren mit einer Prognose für 2026 laut der internationalen Energieagentur (IEA).....	62
Abbildung 8: Risikopyramide AI Act (eigene Darstellung).....	80
Abbildung 9: Zeitplan AI Act © BKA/Digital Austria.....	87
Abbildung 10: Kriterien für ethische KI-Anwendungen in der öffentlichen Verwaltung.....	95
Abbildung 11: Grundrechte-Folgenabschätzung (Artikel 27 AI Act).....	101
Abbildung 12: Kriterien (Außenkreis) und Maßnahmen (Innenkreis) für ethische KI in der Verwaltung.....	103

Tabellenverzeichnis

Tabelle 1: Unterschied starke und schwache KI.....	24
Tabelle 2: Good und Bad Practice Beispiele für die Anwendung generativer KI.....	50
Tabelle 3: Empfehlungen KI-Governance.....	115

