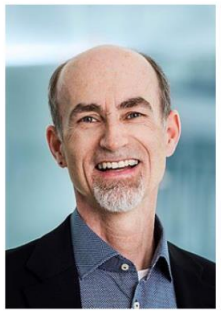
Leitfaden digitale Verwaltung: KI, Ethik und Recht

Praxisleitfaden für die Verwaltung V2.0

Dieser Foliensatz basiert zum Teil auf Vorarbeiten des AIT AI Ethics Lab

AIT AI Ethics Lab






Peter Biegelbauer
Sven Schlarb
Pia Weinlinger
Caroline Lackinger
Rodrigo Conde
Alexander Schindler

https://oeffentlicherdienst.gv.at/wp-content/uploads/2023/11/Leitfaden-Digitale-Verwaltung-Ethik.pdf

Leitfaden digitale Verwaltung 2.0

- Zusammenarbeit AIT AI Ethics Lab + Research Institute


Madeleine Müller


Heidi Scheichenbauer


David M. Schneeberger

  - Einbeziehung von neuen **technischen Entwicklungen** (LLMs, RAG, Quantencomputing)
  - stärkere Beschäftigung mit **rechtlichen Themen** (AI Act, Vergaberecht, Legalitätsprinzip, Folgenabschätzungen)

  - **Fertigstellung**: (geplant) Dezember 2024

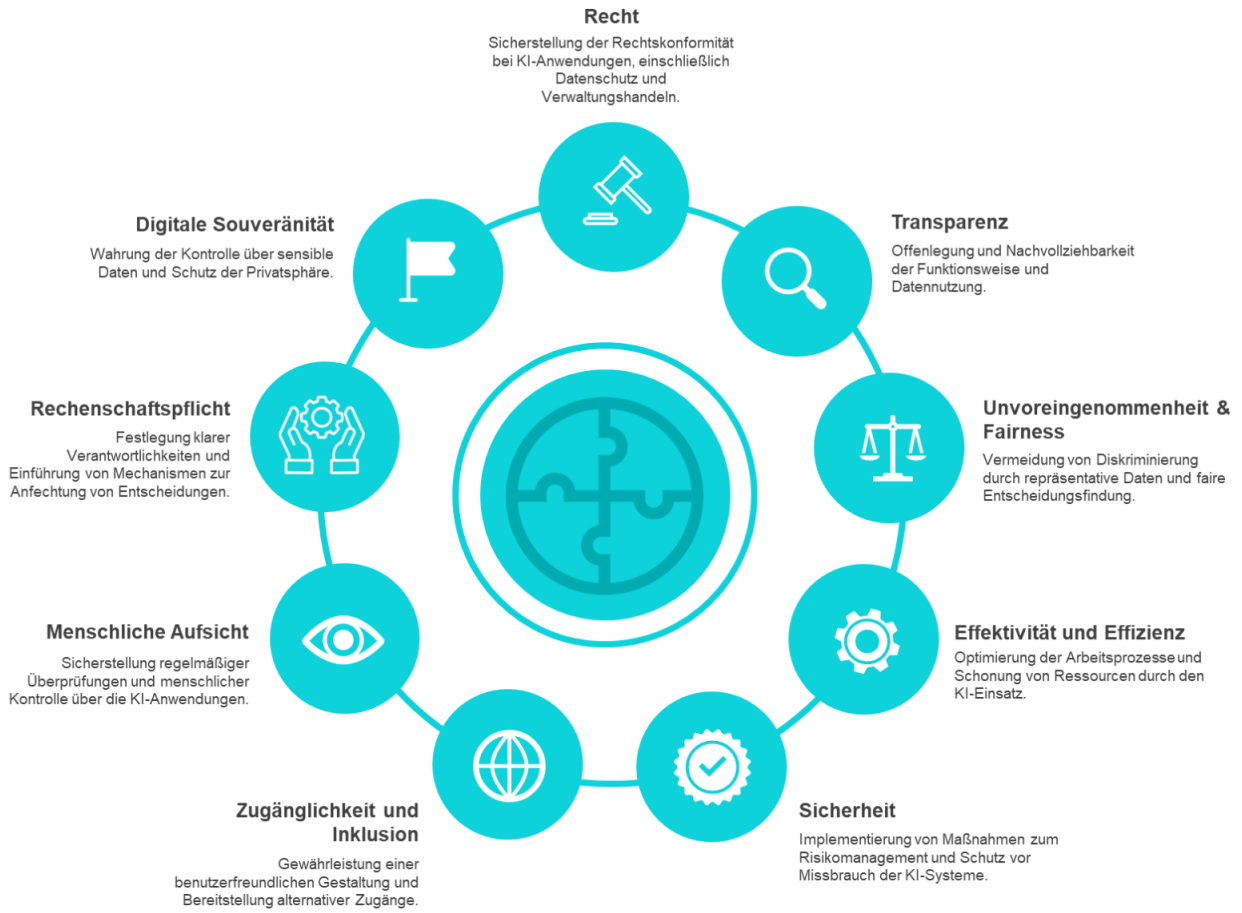
Ziele des Leitfadens

- Unterstützung von **Verwaltungsbediensteten** in
  - Planung,
  - Design,
  - Erstellung bzw. Vergabe,
  - Einsatz und
  - Evaluierung
- digitaler, insbesondere aber Künstlicher Intelligenz (KI)-basierter Anwendungen

- Definition von **grundlegenden Rahmenbedingungen** (für Förderung, Einführung, Anwendung)
- welche **Auswirkungen** haben digital unterstützte Prozesse und Entscheidungen auf Individuen und die Gesellschaft?
- Schaffung von **Vertrauen** in öffentliche Institutionen
- **Menschenzentrierter Ansatz**

Checkliste für ethische KI

- KI-Einsatz: **ethisch** und **verantwortungsbewusst**
  - technische und organisatorische Maßnahmen
  - menschenzentrierter Ansatz
- Checkliste als **Leitfaden zur Umsetzung** dieser (abstrakten) Werte
  - strukturierte Unterstützung



Checkliste für ethische KI

Checkliste	Ja	Nein
Recht		
Wird das KI-System im Rahmen des (schlicht) hoheitlichen Verwaltungshandelns eingesetzt? (Siehe 8.1 Die Grundlage des Verwaltungshandelns)	☐	☐
Sofern (schlicht) hoheitliches Verwaltungshandeln vorliegt, wurde abgeklärt, ob eine ausreichende Rechtsgrundlage für den Einsatz der KI besteht? (Siehe Anwendungsfall Entscheidungsbaum)	☐	☐
Verarbeitet die KI-Anwendung Daten im Einklang mit den Anforderungen der Rechtsnormen und -prinzipien, die im nationalen und EU-Rechtsrahmen festgelegt ist? (Siehe 8.2 Datenschutzgrundverordnung)	☐	☐
Gewährleistet der Einsatz des KI-Systems, dass die Grundrechte von Bürger:innen in keiner Weise beeinträchtigt werden? (Siehe 5 Anwendungsfälle Australien und Niederlande)	☐	☐
Wurde die Risikoeinstufung der KI-Anwendung gemäß dem AI Act ermittelt? (Siehe 8.3 Die EU regelt KI: der AI Act)	☐	☐
Falls zutreffend: Ist ein Conformity Assessment laut dem AI Act bereits erfolgt? (High-risk) (Siehe 8.3 Die EU regelt KI: der AI Act, Konformitätsbewertungsverfahren)	☐	☐
Fals zutreffend: Wurden die Transparenzpflichten nach dem AI Act erfüllt? (Medium-risk) (Siehe 8.3 Die EU regelt KI: der AI Act, Konformitätsbewertungsverfahren, Transparenzanforderungen)	☐	☐
Wurde geprüft, ob eine Grundrechte-Folgenabschätzung (AI Act) durchgeführt werden muss? (Siehe 10.1 Grundrechte-Folgenabschätzung im AI Act)	☐	☐
Wurde geprüft, ob eine Datenschutz-Folgeabschätzung durchgeführt werden muss? (Siehe 8.2 Datenschutz-Grundverordnung)	☐	☐

Checkliste	Ja	Nein
Werden unabhängig von extern durchgeführten Prüfungen interne Konformitätsbewertungen (etwa im Sinne eines Plausibilitätsschecks) durchgeführt, um zu überprüfen, ob das KI-System vor der Einführung alle gesetzlichen Anforderungen erfüllt?	☐	☐
Transparenz		
Sind die spezifischen Ziele und Zwecke des Einsatzes der KI-Anwendung identifiziert und dokumentiert?	☐	☐
Gibt es eine Dokumentation, die die technische Entwicklung des Modells erläutert?	☐	☐
Ist die Funktionsweise der KI-Anwendung nachvollziehbar?	☐	☐
Sind die Datensätze, die mit dem KI-System verbunden sind, bekannt? (Siehe 8.2 Datenschutzgrundverordnung)	☐	☐
Wird den Nutzer:innen, wann immer möglich, erklärt, wie das KI-System zu seinen Ausgaben, Inhalten, Empfehlungen oder Ergebnissen kommt und welche Logik dahintersteckt?	☐	☐
Werden Personen informiert, wann und auf welche Weise sie mit einer KI-Anwendung interagieren?	☐	☐
Unvoreingenommenheit und Fairness		
Sind die Daten, die zum Training des KI-Systems verwendet werden, vielfältig und repräsentativ für den jeweiligen Kontext? (Siehe Wissen: Bias; Anwendung: Geschlechterbias)	☐	☐
Gibt es einen Prozess, um verwendete Datenquellen auf mögliche Verzerrungen und Ungenauigkeiten zu prüfen? (Siehe Entscheidungsbaum)	☐	☐
Ist die KI-Anwendung so konzipiert, dass es die Entmenschlichung, Diskriminierung, Stereotypisierung oder Manipulation von Menschen vermeidet? (Siehe Anwendungsfall: Chatbot)	☐	☐
Gibt es ein Verfahren, mit dem Personen gegen Entscheidungen des KI-Systems Einspruch erheben oder diese anfechten können?	☐	☐
Effektivität und Effizienz (verlinken)		
Ergeben sich konkrete Vorteile für die breite Öffentlichkeit durch den Einsatz dieses KI-Systems? (z.B. Zeitersparnis bei der Beantragung einer staatlichen Leistung)	☐	☐
Hat die KI-Anwendung das Potenzial die Arbeitssituation, der im öffentlichen Dienst tätigen Personen zu verbessern oder zumindest nicht zu verschlechtern? (Siehe 5.1 KI und Auswirkungen auf die Arbeitswelt in der öffentlichen Verwaltung)	☐	☐
Gibt es eine Richtlinie zu Qualitäts- und Leistungszielen für das KI-System?	☐	☐

Technische Grundlagen

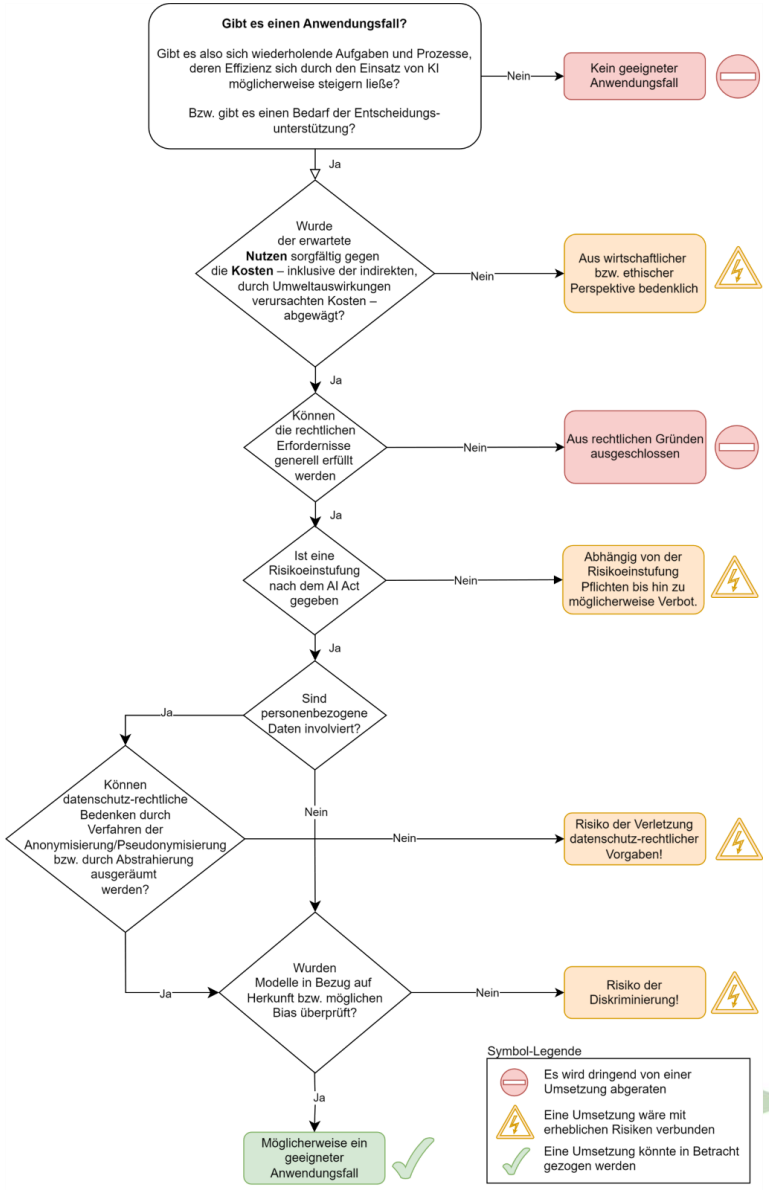
- **Was ist KI?**
  - Entwicklung des KI-Begriffs
  - Starke vs schwache KI
  - Symbolische vs sub-symbolische Systeme
  - Verfahren des maschinellen Lernens
  - Modelle, insb Künstliche Neuronale Netze
  - **Generative KI** (prompts, prompt engineering, Multimodalität, foundation models, **RAG**, etc)
  - **Quantum AI**

Handlungsbedarf

- **Vertrauen in die Verwaltung**
  - Verwaltung oft als „erster Kontaktpunkt“ mit dem Staat
  - Versagen von KI-Systemen: Fälle bspw im Bereich Kindergeld, Steuerbetrug in Australien, Niederlande
- **Automation bias**
  - AI literacy: Ausbildung, Sensibilisierung
  - human-in-the-loop
- **Effizienzgewinn als Motivator**
  - bspw Behördenchatbot
- Vermeidung von **Bias/algorithmischer Diskriminierung**
  - Ungewissheit zB bei Trainingsdaten von LLMs, bei fine-tuning

Entscheidungsbaum

- Liegt ein **geeigneter Anwendungsfall** vor?
- Hilfestellung zur **grundlegenden Prüfung**, ob der Einsatz von KI sinnvoll und gerechtfertigt ist
  - Zweckmäßigkeit
  - Wirtschaftlichkeit
  - Anwendungsbedingungen
  - rechtliche Erfordernisse



Chancen und Herausforderungen beim Einsatz von KI

- **Auswirkungen auf die Arbeitswelt** (zB Recruiting, Kontrolle, Online-Plattformen)
  - Effizienzsteigerung (zB Wissensmanagement)
  - Risiken: Akzeptanz, Verlust von Autonomie, Benachteiligung, Privatsphäre
  - **Umgang mit generativer KI am Arbeitsplatz**
    - Halluzinationen
    - kein Zugriff auf Echtzeitdaten
    - Verzerrungen
    - Datenschutz
      - keinen personenbezogenen Daten als Eingabedaten
      - Trainingsdaten (Extraktion von personenbezogenen Daten)
  - Erkennung von KI-generierten Inhalten

	Empfohlen	Nicht empfohlen
Kreative Inspiration und visuelle Inhalte generieren	• Erstellung nicht-sensibler visueller Inhalte für Präsentationen, Marketingmaterialien • Sicherstellen, dass KI-generierte die zugrunde liegenden Daten oder Informationen korrekt darstellen	• 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	• Erstellung schneller Analysen oder Zusammenfassungen von nicht sensiblen, öffentlich zugänglichen Texten/Berichten/Publikationen • Die Ausgabe vor der Verwendung auf ihre Richtigkeit überprüfen	• Vertrauliche oder geschützte Dokumente zur Analyse/Zusammenfassung durch KI verwenden • Sich ausschließlich auf KI-Analysen/Zusammenfassungen verlassen
Erstellen von Inhalten für Meetings/Projekte	• Erstellung von allgemeinen/generischen Vorlagen für z.B. Projektpläne oder wiederkehrende Meetings, ohne interne und nicht veröffentlichte Informationen preiszugeben	• Angabe von sensiblen Informationen wie Projektnamen, Mitarbeiterdetails oder internen Inhalten

Chancen und Herausforderungen beim Einsatz von KI

- **Auswirkungen auf die Bevölkerung**
  - Interaktion mit der Verwaltung
  - Effizienz und Personalisierung
  - Zugänglichkeit und Chancengleichheit
    - oder digital divide?
  - Vertrauens-/Akzeptanzproblem
    - Transparenz
    - Rechenschaftspflicht
    - Datenschutz
    - Gegenmaßnahmen
      - Co-Kreation, Partizipation
      - Benutzerfreundlichkeit, AI literacy
      - Opt-Out

Chancen und Herausforderungen beim Einsatz von KI

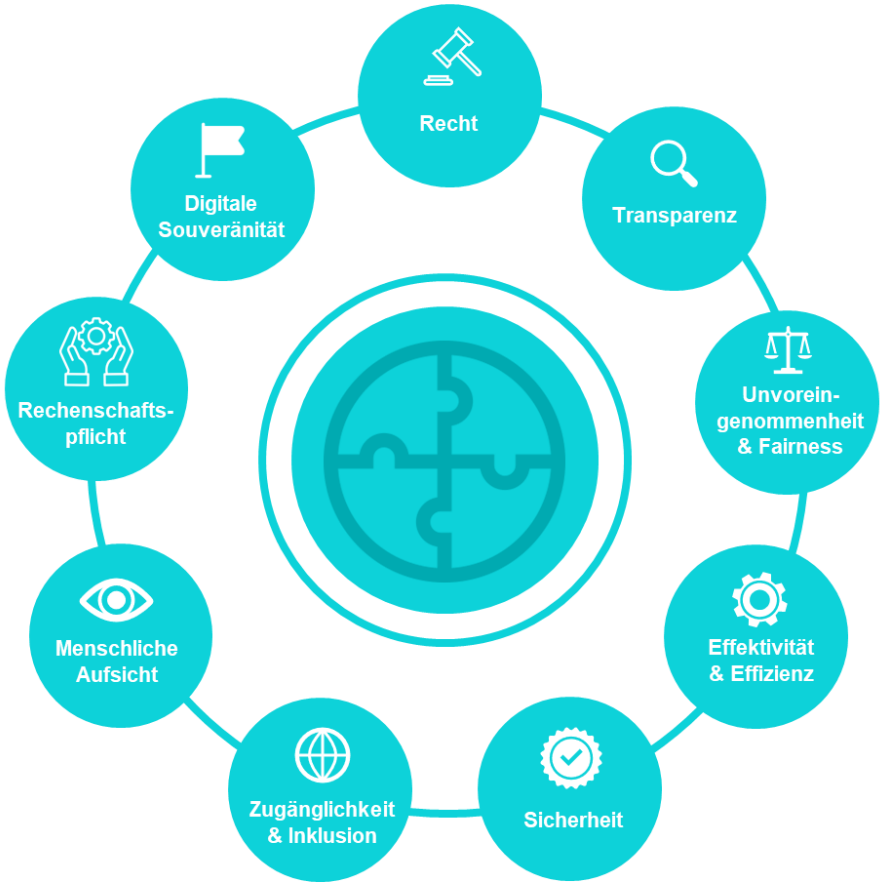
- **KI und Ökologie**
  - Aufwand für finale Optimierung?
  - Basismodelle im open access
  - Zusammenarbeit/gemeinsame Nutzung von Infrastruktur
  - Zertifizierte Rechenzentren
- **Digitale Souveränität**
  - Autonome Entscheidungen über Technologieinfrastruktur oder Abhängigkeit von externen Anbietern?
  - Instrument:
    - öffentliche **Beschaffung**/Vergabe (klare Definition der Anforderungen, ethische Designkriterien zB Transparenz, rechtliche Rahmenbedingen, Data-Governance-Plan)
    - open-source-Software
    - Stärkung europäischer Prozesse (zB Chips, Datensouveränität)

Rechtlicher Rahmen

- **Verfassungs- und Verwaltungsrecht**
  - rechtsstaatliches Prinzip, Legalitätsprinzip
- **Datenschutzrecht**
  - personenbezogenes Datum, Verantwortlicher, Auftragsverarbeiter, Rechtmäßigkeit der Verarbeitung, Datensicherheit, Cloud Services, Betroffenenrechte
- **AI Act**
  - Produktsicherheitsrecht, risikobasiertes System, verbotene Praktiken, Hochrisiko-KI-Systeme (insbesondere eigenständige Hochrisiko-KI-Systeme, primär Zugang zu öffentlichen Diensten und Leistungen), Anforderungen an Hochrisiko-KI-Systeme, Transparenzpflichten, Konformitätsbewertung Pflichten von Anbietern, Betreibern, Änderung der Rolle
- **Haftungsrecht**
  - Produkthaftung
  - Außervertragliche Haftung
- **Data Governance Act**

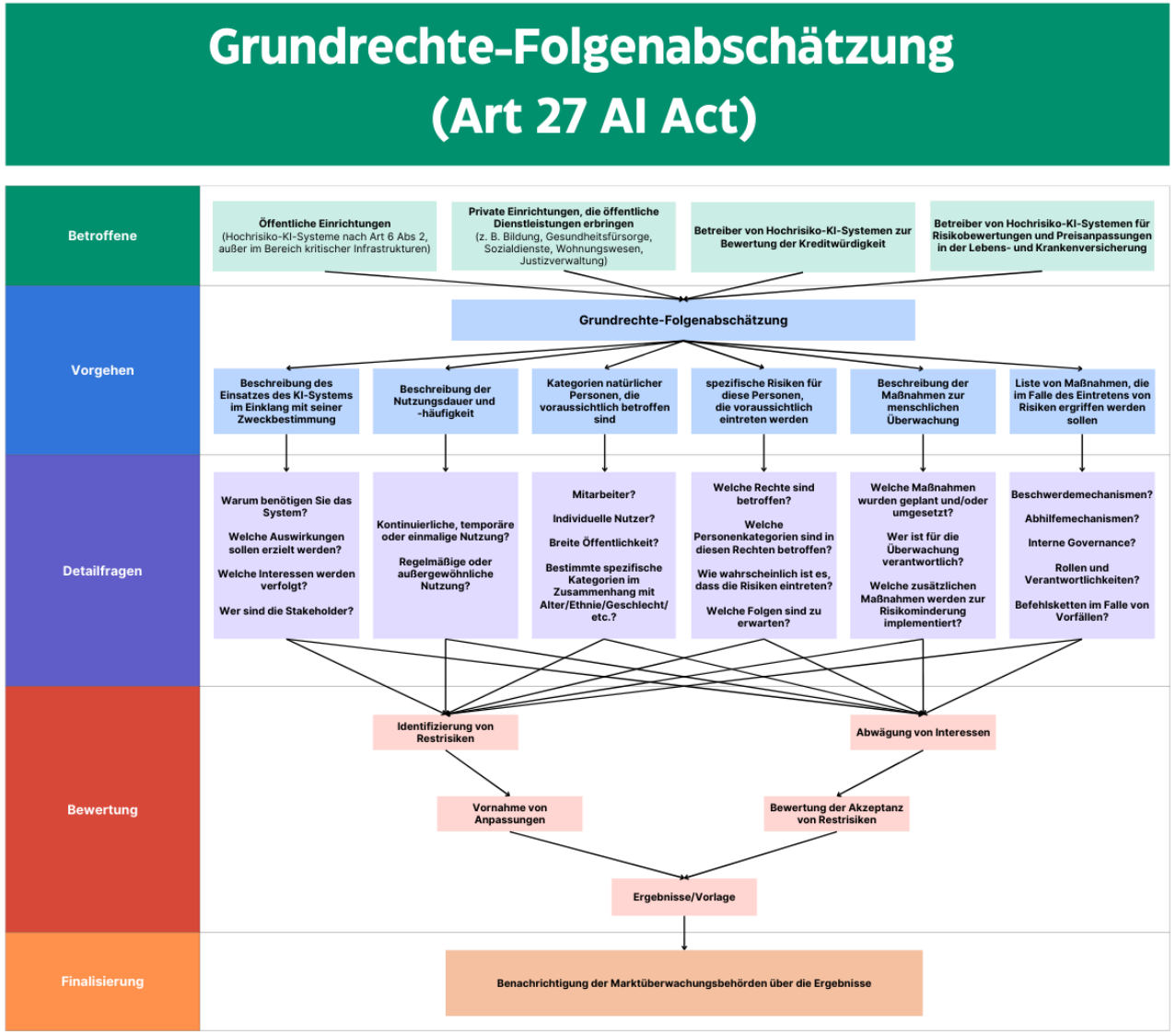
Ethische KI: Prinzipien und Leitlinien

- Auswertung ethischer Leitlinien
  - zugeschnitten auf die Verwaltung
- **Kriterien**
  - Recht
  - Transparenz
  - Unvoreingenommenheit und Fairness
  - Effektivität und Effizienz
  - Sicherheit
  - Barrierefreiheit und Inklusion
  - Menschliche Aufsicht
  - Rechenschaftspflicht
  - Digitale Souveränität



KI-Folgenabschätzung

- **Vorstellung verschiedener Modelle:**
  - Grundrechte-Folgenabschätzung nach Artikel 27 AI Act
  - **DVuE „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)"**
  - Bewertungsliste ALTAI auf EU-Ebene,
  - VCIO-Modell
  - FRAIA und DEDA (Verwaltung Niederlande)



DVuE „Kriterien- und Maßnahmenkatalog für KI in der Verwaltung (EKIV)"

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

- Wenn das System über eine Schnittstelle zur Öffentlichkeit verfügt, 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?

Governance-Struktur

- **EU-Ebene**
  - KI-Büro
  - KI-Gremium
  - Beratungsforum
  - Wissenschaftliches Gremium

- **Nationale Ebene**
  - AI policy Forum
  - KI-Servicestelle
  - KI-Beirat
  - Chief Digital Officer Taskforce
  - AI-Stakeholder-Forum

Empfehlungen für weitere Schritte

Ziel (Funktion)	Kompetenzaufbau & 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 (NIST), Selbst-verpflichtende Leitlinien (DE BMAS), Risk Assessment Tools und Entwicklungsstandards („Ethics by Design"; IEEE), Entscheidungskriterien für interne / externe Beschaffungsvorgänge (Amsterdam Klauseln)	Diskussionen zu Good Practices und Herausforder-ungen, 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 Herausforder-ungen
Institutiona-lisierung (Struktur)	Curriculum ab Juni 2023, VAB KI-Ethikseminar und Führungskräfte Lehrgang ab WS 2023, themenspezifische interne Kompetenzstellen, Informations- und Diskussionsveranstaltun-gen für und mit Verwaltung, Politik und Öffentlichkeit	Verwaltungs-Ethikrat mit internen und externen Expert:innen (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/ Reallabor, 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, Verarbeitungs-verzeichnis, KI-Behörde, NL Algorithm Register	AI Policy Forum, PIAZZA Format (Algorithm Watch), Interministerielles KI-Ethik Lab

DANKE FÜR IHRE AUFMERKSAMKEIT

Mag. Dr. David M. Schneeberger, BA BA MA
SENIOR RESEARCHER & CONSULTANT
E-Mail: david.schneeberger@researchinstitute.at

Research Institute AG & Co KG
Digital Human Rights Center

Florianigasse 55/10
A-1080 Wien
+43 1 524 3 524 - 0
office@researchinstitute.at
https://researchinstitute.at

EIGENWERBUNG

Schneeberger

Schriften zum österreichischen und europäischen öffentlichen Recht – Band 39

Machine Learning in der Verwaltung

Rechtsfragen der Black-Box-Problematik

Monografie

591 Seiten, broschiert
ISBN 978-3-7046-9377-8 (Print)
ISBN 978-3-7046-9387-7 (eBook)
Erscheinungsdatum: 21. März 2024

- *Schneeberger*, Machine Learning in der Verwaltung (2024).
- *Schneeberger*, Large Language Models in der Verwaltung, Jahrbuch Digitalisierung und Recht 2024 (2024).
- *Schneeberger*, Welcome to the Machine – Machine Learning in der österreichischen Verwaltung, LTZ 2024/4.

Hoffberger-Pippan/Ladeck/Ivankovics (Hrsg)

Digitalisierung und Recht

Jahrbuch 2024

Jahrbuch

369 Seiten, broschiert
ISBN 978-3-7083-4225-2 (Print)
ISBN 978-3-7083-4226-9 (eBook)
Erscheinungsdatum: 9. September 2024