Leitlinien für den Einsatz Künstlicher Intelligenz in der Bundesverwaltung

Bundesministerium des Innern und für Heimat

Der Beauftragte der Bundesregierung für Informationstechnik

BeKI Beratungszentrum für Künstliche Intelligenz

Inhalt

Vorbemerkung

Die vorliegenden **Leitlinien für den Einsatz Künstlicher Intelligenz in der Bundesverwaltung** (KI-Leitlinien) setzen allgemeine Leitplanken für die Nutzung von KI in der Bundesverwaltung. Mit Leitsätzen für die Bereitstellung und den Einsatz von KI-Systemen wird ein koordiniertes Vorgehen sichergestellt, um einen verantwortungsvollen und sicheren KI-Einsatz in der Bundesverwaltung zu gewährleisten.

Die **Entscheidung darüber, ob und welche KI-Systeme eingesetzt werden dürfen, obliegt der jeweiligen Behördenleitung**. Diese entscheidet unter anderem vor dem Hintergrund von Informationssicherheit, Datenschutz und Geheimschutz, ob und welche KI-Systeme für dienstliche Zwecke genutzt werden dürfen.

Angesichts der **rasanten technologischen Entwicklungen** und des hohen Interesses innerhalb der Bundesverwaltung an der sicheren Nutzung von KI wurden diese Leitlinien entwickelt. Die Handlungsempfehlungen dienen als Orientierungshilfe und erheben nicht den Anspruch, sämtliche Bereiche umfassend und abschließend zu regeln. Vielmehr wurden relevante Bereiche zum Zeitpunkt der Erstellung des Dokuments ressortgemeinschaftlich identifiziert und priorisiert. Zudem setzt der Einsatz von KI eine umfassende Beurteilung im Einzelfall anhand des konkreten KI-Systems voraus. Hierbei können zum Beispiel Reallabore als geeignete Räume zur Unterstützung einer solchen Einzelfallprüfung dienen.

Darüber hinaus sind weitere Umsetzungsfragen zu klären, die nicht Gegenstand der vorliegenden Leitlinien sind. Innerhalb der Bundesverwaltung werden in den jeweiligen Ressorts und Behörden geeignete **Governance-Strukturen für den Einsatz von KI** zu etablieren sein. Hierzu werden derzeit im Rahmen der **Durchführung der KI-Verordnung (KI-VO)**[1] weitere Vorgaben abgestimmt. Die Einführung von KI sollte zudem frühestmöglich von einem umfassenden Konzept für die Sicherstellung nachhaltiger Veränderungen („Change-Management“) in den Ressorts und Behörden begleitet werden.

Bei der Einführung von KI-Systemen sind **Rechte der Mitarbeitenden** zu wahren. Dabei gelten viele bestehende Regelungen (beispielsweise zur erforderlichen Barrierefreiheit oder im Bereich des Datenschutzes) fort. Gegebenenfalls werden Dienstvereinbarungen neu zu verhandeln und Auswirkungen von KI auf den Arbeitsplatz zu bewerten sein. Interessenvertretungen der Mitarbeitenden (wie unter anderem der Hauptpersonalrat und die Schwerbehindertenvertretung) sind entsprechend frühzeitig einzubinden. Zudem ist ein Weiterqualifizierungs- oder Schulungsangebot zu schaffen, das Mitarbeitenden vor Einführung der KI-Systeme zur Verfügung steht.

Die vorliegenden KI-Leitlinien sollen den Weg ebnen, diese Prozessschritte einzuleiten und als gemeinsame Grundlage den Einsatz von KI in der Bundesverwaltung chancenorientiert zu erleichtern. Dies ist somit erst der Beginn eines längerfristigen Prozesses, der die **Arbeitstätigkeit in der Bundesverwaltung massiv wandeln** wird.

Es wird zukünftig notwendig bleiben, weitere Bereiche zu adressieren und Anweisungen zur praktischen Umsetzung gemeinschaftlich weiterzuentwickeln, um den sich stetig wandelnden Anforderungen gerecht zu werden.

---

1 Verordnung (EU) 2024/1689 des europäischen Parlaments und des Rates vom 13. Juni 2024 über Künstliche Intelligenz; ABl. L vom 12. Juli 2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

1. Einleitung

Der **Einsatz Künstlicher Intelligenz** (KI) stellt eine immense Chance dar, die Verwaltung der Zukunft zu gestalten. KI kann bei adäquater Nutzung Digitalisierungspotenziale ausschöpfen und zur Modernisierung der Verwaltung beitragen. **KI-Systeme können die Bundesverwaltung beispielsweise unterstützen**, behördliche Aufgaben effizienter zu bearbeiten, Mitarbeitende zu entlasten und somit den Auswirkungen des demographisch bedingten Fachkräftemangels wirksam zu begegnen.

KI findet im **Arbeitsalltag der Mitarbeitenden der Bundesverwaltung** zunehmend Anwendung.[2] Dabei stehen große Sprachmodelle, sogenannte Large Language Models, aktuell häufig im Fokus.

Als Bundesverwaltung obliegt es uns, einen solchen Einsatz von KI verantwortungsvoll zu gestalten. Auf Grund **unserer hoheitlichen Aufgaben**, den **Konsequenzen unserer Arbeit für Bürgerinnen und Bürger** sowie der **Sensibilität der Daten**, die wir tagtäglich verarbeiten, tragen wir eine besondere Verantwortung beim Einsatz von KI.

Die **Leitlinien** sollen dazu beitragen, einen **chancenorientierten** und **verantwortungsvollen Umgang mit KI-Systemen** in der Bundesverwaltung zu gewährleisten.

Definition: Künstliche Intelligenz

In den vorliegenden Leitlinien wird ein KI-System entsprechend Artikel 3 Absatz 1 der **Verordnung über Künstliche Intelligenz der EU** (KI-VO) definiert als *„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”.*[3]

Beispiele für KI reichen von einfachen Algorithmen wie zum Beispiel „Optical Character Recognition“ zur Erkennung von Buchstaben bis zu komplexen Modellen, wie sie etwa in generativer KI eingesetzt werden. Mögliche Anwendungsfelder beinhalten Sprachassistenten, Robotik und das autonome Fahren.

2 KI unterliegt einer rasanten Entwicklung. Es ist somit gegenwärtig schwer abzuschätzen, in welchen Einsatzgebieten KI zukünftig eine erhebliche Rolle spielen wird. Da KI allerdings bereits heute Anwendungsfälle in vielen Bereichen der Verwaltungstätigkeit bietet, ist zu erwarten, dass sich diese Entwicklung auch künftig weiter rapide fortsetzen wird.

3 Diese Definition ist breit gefasst und könnte dazu führen, dass eine große Anzahl von Systemen unter die Regulierung fallen. Daher wird in den Erwägungsgründen (EG 12) einschränkend klargestellt, dass herkömmliche Softwaresysteme oder Programmieransätze, die ausschließlich auf von natürlichen Personen festgelegten Regeln zur automatischen Ausführung von Vorgängen beruhen, von der Definition nicht umfasst werden.

Anwendungsbeispiel:
Große Sprachmodelle

Large Language Models (LLMs) sind künstliche neuronale Netze, die auf großen Textkorpora oder anderen Datenformaten wie Bild-, Audio- oder Videodaten (sog. multimodale Modelle) trainiert wurden. Basierend auf generativen Modellen[4] können sie Eingaben („Prompts“) verarbeiten und entsprechende Ausgaben wie beispielsweise Texte oder Bilder erzeugen.

LLM-basierte Anwendungen sind in der Lage, eine Vielzahl von Funktionen zu erfüllen. Textgenerative LLM-Anwendungen können beispielsweise zur Erstellung von Texten (zum Beispiel von Vermerken), zur Befragung umfangreicher Dokumente („Chat with your document“) und zur Übersetzung oder der Generierung von Programmcode genutzt werden. Weitere Beispiele sind die Möglichkeiten, Inhalte zusammenzufassen oder stilistisch anzupassen (unter anderem Übersetzung in „Einfache Sprache“).

Diese Eigenschaften machen LLMs für die Automatisierung von ausgewählten Arbeitsschritten in der Verwaltung attraktiv. Zugleich bestehen Risiken bei der Nutzung von LLM-Anwendungen, beispielsweise im Hinblick auf die Korrektheit der generierten Informationen, bezogen auf den Daten- und Urheberrechtsschutz, den Geheimschutz oder die Informationssicherheit.

Durch gemeinsame ethische, prozessuale und technische Standards, verbunden mit entsprechender Sensibilisierung der Nutzenden, sollen Risiken minimiert, die Priorisierung des Gemeinwohls sichergestellt und dabei das Potenzial von **KI als Unterstützungswerkzeug** möglichst voll ausgeschöpft werden.

In diesen Leitlinien definieren wir übergeordnete **Leitprinzipien** als **gemeinsame Handlungsgrundlage** für den KI-Einsatz in der Bundesverwaltung (Kapitel 2). Damit schafft die Bundesregierung einen wertebasierten Kompass für die Entwicklung und den Einsatz von KI. Dieser Abschnitt richtet sich an **alle Mitarbeitenden der Bundesverwaltung**.

Darauf aufbauend geben wir durch **konkrete Leitsätze für den Einsatz und die Bereitstellung von KI-Systemen einen Rahmen vor**, um einen **verantwortungsvollen und sicheren Einsatz** in der Bundesverwaltung zu ermöglichen (Kapitel 3). In den Leitsätzen erklären wir **zielgruppenspezifische Anforderungen und Herausforderungen** beim KI-Einsatz. Adressiert werden sowohl Behörden, die KI-Systeme bereitstellen beziehungsweise deren **verantwortliche Organisationseinheiten** (beispielsweise deren Datenlabore, Maßnahmenverantwortliche, IT-Verantwortliche) als auch alle **Nutzenden** solcher KI-Systeme.

Die vorliegenden KI-Leitlinien werden dabei durch geltende Standards sowie **fachspezifische Vorgaben und Veröffentlichungen**, beispielsweise zur Informationssicherheit[5], zur behördlichen Praxis der Arbeits- und Sozialverwaltung[6] oder dem außenpolitischen Bereich[7] ergänzt.

4 Generative Modelle sind Werkzeuge für maschinelles Lernen, mit denen neue Datenmuster erstellt werden können. Sie sind für eine Vielzahl von Anwendungen nützlich, zum Beispiel für die Generierung von Bildern und Text.

5 Für eine ausführliche Analyse der Chancen und Risiken von generativer KI im Rahmen der Integration in bestehende Prozesse und Anwendungen siehe zum Beispiel die Publikation Generative KI-Modelle – Chancen und Risiken für Industrie und Behörden (siehe unter: Generative KI-Modelle | BSI) sowie den Kriterienkatalog für KI-Cloud-Dienste (AIC4) des Bundesamts für Sicherheit in der Informationstechnik (BSI) (siehe unter: AIC4 | BSI).

6 Für das Bundesministerium für Arbeit und Soziales (BMAS) und seinen Geschäftsbereich wurden im November 2022 die „Selbstverpflichtenden Leitlinien für den KI-Einsatz in der behördlichen Praxis der Arbeits- und Sozialverwaltung“ veröffentlicht, die weiterhin ihre Gültigkeit behalten, siehe unter: Selbstverpflichtende Leitlinien für den KI-Einsatz | BMAS.

7 Das Auswärtige Amt (AA) hat 2024 eine KI-Charta entwickelt und publiziert.

2. Wertebasierte Leitprinzipien als gemeinsame Handlungsgrundlage

Fünf übergeordnete **Leitprinzipien** dienen **als Handlungsgrundlage der Bundesverwaltung** bei der Entwicklung und dem Einsatz von KI. Die Prinzipien stehen im Einklang mit der Werteorientierung der KI-VO[8] und stellen eine **gemeinsame Wertegrundlage für alle Mitarbeitenden der Bundesverwaltung** dar.

Als Bundesverwaltung setzen wir KI chancenorientiert und verantwortungsvoll für die Gesellschaft ein.

Die Bundesverwaltung setzt KI **chancenorientiert** dort ein, wo diese einen Mehrwert im Sinne der dienstlichen Aufgabenerfüllung schafft. Ziel ist es, Mitarbeitende mithilfe von KI zu unterstützen sowie Dienstleistungen für Bürgerinnen und Bürger, Unternehmen und die Verwaltung selbst zu verbessern und damit **effizientes und innovatives Verwaltungshandeln** zu fördern.

Die Bundesverwaltung ist sich dabei ihrer besonderen Rolle bewusst, KI-Systeme **verantwortungsvoll** einzusetzen. Im Einklang mit **gesetzlichen Rahmenbedingungen**[9] muss es beim Einsatz von KI klare Verantwortlichkeiten geben. Im Sinne angemessener **menschlicher Aufsicht** kann die Verantwortung für Entscheidungen folglich nicht an KI-Systeme als Werkzeug abgegeben werden.

Wie kann ich dieses Prinzip berücksichtigen?

- KI wird eingesetzt, um **Effizienz- und Effektivitätspotenziale** in der Verwaltungsarbeit zu erschließen und Abläufe zu transformieren.

- Der KI-Einsatz erfolgt unter **angemessener menschlicher Aufsicht**, wobei die Angemessenheit im Einzelfall zu beurteilen ist. In einigen Fällen kann es erforderlich sein, jedes Ergebnis mit allen zugrundeliegenden Faktoren nachvollziehen und bewerten zu können. Bei KI-Systemen beziehungsweise Datenverarbeitungen, für die spezifische Informations- und

8 In der KI-VO wird auf sieben ethische Grundsätze aus den „Ethikleitlinien für vertrauenswürdige KI“ verwiesen, die 2019 von der unabhängigen hochrangigen Expertengruppe der Kommission entwickelt wurden. Diese sind: „menschliches Handeln und menschliche Aufsicht, technische Robustheit und Sicherheit, Privatsphäre und Daten-Governance, Transparenz, Vielfalt, Nichtdiskriminierung und Fairness, soziales und ökologisches Wohlergehen sowie Rechenschaftspflicht“, siehe Erwägungsgrund 1, 2 und 27 der KI-VO.

9 Beispielsweise begründet Artikel 22 DSGVO ein verbindliches Recht auf eine menschliche Letztentscheidung im Kontext der Nutzung personenbezogener Daten, siehe: Verordnung (EU) 2016/679, Datenschutz-Grundverordnung (DSGVO), ABl. L v. 27. April 2016, ELI: http://data.europa.eu/eli/reg/2016/679/oj. Daneben hat das Prinzip der menschlichen Aufsicht beispielsweise in Artikel 14 KI-VO für Hochrisiko-KI-Systeme Eingang gefunden.

Auskunftspflichten gelten[10], entspricht dies beispielsweise dem gesetzlichen Mindeststandard. Für Systeme, die nicht unter diese Regelungen fallen, sind auch abgestufte Ansätze denkbar. Die geringste Form der menschlichen Aufsicht besteht darin, ein KI-System jederzeit abschalten zu können. Es wird eine grundsätzliche Eingriffsmöglichkeit sichergestellt.

Es gilt allgemein: Je höher das Risiko, welches mit dem Einsatz von KI verbunden ist, desto mehr rücken menschliche Aufsicht und Kontrolle in den Vordergrund (**risikobasierter Ansatz**).

Wir stehen für einen wertebasierten und menschenzentrierten KI-Einsatz.

Durch einen gemeinsamen **wertebasierten Ansatz** zur Bereitstellung und Nutzung von KI sollen in der Bundesverwaltung Vorteile von KI maximiert und einhergehende Risiken minimiert werden. Wir richten den Einsatz von KI vor diesem Hintergrund stets auf das Gemeinwohl aus (**Menschenzentrierung**[11]).

KI wird im **Einklang** mit den **Grundrechten** eingeführt und eingesetzt. Die Persönlichkeitsrechte von Mitarbeitenden sowie Bürgerinnen und Bürgern, insbesondere das Recht auf informationelle Selbstbestimmung, sind dabei zu achten. Unser Ziel ist es, mit dem KI-Einsatz menschliche Handlungsspielräume zu erweitern und nicht, diese zu reduzieren.

Wie kann ich dieses Prinzip berücksichtigen?

- Es sollten vor und während des Einsatzes von KI-Systemen geeignete Maßnahmen ergriffen werden, damit dieser nicht zu verzerrten Ergebnissen führt (**Diskriminierungsfreiheit**). KI-Systeme sollten beispielsweise dem Einsatzzweck entsprechend auf geeigneten und repräsentativen Trainingsdaten basieren. Generierte Inhalte werden dabei auf diskriminierende Elemente geprüft.

- Der Einsatz von KI selbst soll **fair** und **diskriminierungsfrei** gestaltet sein. Soweit möglich sollten alle Mitarbeitenden die gleichen Chancen auf Zugang bekommen. Dies gilt für mögliche Diskriminierungsmerkmale wie Geschlecht, sexuelle Orientierung, Behinderung oder Alter. Die gesetzlichen Vorgaben zur Barrierefreiheit sind beim Zugang zu KI-Systemen zu berücksichtigen.[12]

- Die Einführung von KI soll **nutzendenzentriert** stattfinden. KI soll Arbeitsprozesse verbessern oder erleichtern und den Nutzenden einen Mehrwert bieten.

10 Siehe beispielsweise das Recht auf Erläuterung der Entscheidungsfindung im Einzelfall auf der Grundlage der Ausgaben eines in Anhang III KI-VO aufgeführten Hochrisiko-KI-Systems nach Artikel 86 Absatz 1 KI-VO sowie das Auskunftsrecht der betroffenen Person über die Verarbeitung sie betreffender personenbezogener Daten gemäß Artikel 15 Absatz 1 lit. h DSGVO.

11 Menschenzentrierte KI wird in der KI-VO folgendermaßen erläutert: „Angesichts der großen Auswirkungen, die KI auf die Gesellschaft haben kann, und der Notwendigkeit, Vertrauen aufzubauen, ist es von entscheidender Bedeutung, dass KI […] eine menschenzentrierte Technologie ist. Sie sollte den Menschen als Instrument dienen und letztendlich das menschliche Wohlergehen verbessern.“ Siehe Erwägungsgrund 6 KI-VO.

12 Vorgaben zu der Barrierefreiheit von KI-Systemen ergeben sich unter anderem aus DIN EN 301549.

Wir setzen auf vertrauenswürdige KI-Systeme.

Vertrauenswürdige KI definieren wir als **rechtmäßig, ethisch** und **robust**. Ihr Einsatz erfolgt im Einklang mit geltendem Recht sowie der Werteorientierung der KI-VO.[13] Vertrauenswürdige KI erfüllt ihre Funktionen sicher und zuverlässig („technisch robust") und ist **resilient gegenüber externen Eingriffen**. Vorhersehbare Sicherheitseinschränkungen des KI-Systems oder für die vom System betroffenen Menschen müssen vermieden werden. Zusätzlich wird ein vertrauenswürdiges KI-System den dynamischen sozialen Kontexten gerecht, in welchen es genutzt wird („sozial robust"). Dazu gehört auch, dass KI, wo erforderlich, als solche **erkennbar**, die generierten Inhalte **hinterfragbar** und die Nutzung mit **transparentem Verwaltungshandeln** vereinbar sind.

Wie kann ich dieses Prinzip berücksichtigen?

- KI-Systeme werden vor ihrem Einsatz sowie fortlaufend in verschiedenen Situationen und Umgebungen durch die verantwortliche Behörde auf ihre **Widerstandsfähigkeit** getestet und in sicheren Umgebungen eingesetzt. Dazu zählt die **Resilienz** gegenüber eintretenden Schwachstellen in der physischen Infrastruktur ebenso wie gegenüber Cyberangriffen, Zweckentfremdung, gezielter und schädlicher Datenmanipulation oder dem unbefugten Zugriff auf Daten.
- Behörden sollten bei der Einführung von KI-Systemen die zweckgemäße Erkennbarkeit und Erklärbarkeit von KI-generierten Ergebnissen sicherstellen. Das bedeutet, Nutzende und Betroffene sollten grundsätzlich erkennen können, wenn sie direkt mit einer KI interagieren oder mit ungeprüften Ergebnissen von KI-Systemen konfrontiert sind. Zudem sollten sie in diesen Fällen **möglichst nachvollziehen können**, anhand welcher Maßstäbe und auf Basis welcher Informationen die KI Ergebnisse erzielt. Je weitreichender die Folgen einer Vorhersage oder Empfehlung des KI-Systems umso wichtiger ist Transparenz über die Grundlagen und Reproduzierbarkeit der auf dieser Basis getroffenen Entscheidungen (**„explainable AI"**).

Wir unterstützen Mitarbeitende beim KI-Einsatz und stärken Kompetenzen.

Die Leitlinien unterstützen die Mitarbeitenden der Bundesverwaltung durch praxisorientierte und zielgruppenspezifische Handlungsempfehlungen für den Umgang mit KI. Dabei stellen wir die Perspektive der **Mitarbeitenden in den Mittelpunkt**, um eine möglichst sichere Handhabung zu ermöglichen. Wir stärken das Bewusstsein für Potenziale und Risiken des KI-Einsatzes durch Schulungen und Fortbildungen[14] und beziehen die Beschäftigten bei der Entscheidung für den Einsatz von KI im Arbeitsalltag ein. Damit soll zu einem **verantwortungsvollen Umgang** mit dieser Technologie angeleitet werden.

13 Für mehr Informationen zur Werteorientierung der KI-VO siehe Fußnote 9.

14 KI-spezifische Schulungsangebote bestehen beispielsweise bei der BAköV.

Wie kann ich dieses Prinzip berücksichtigen?

- **Betroffene Mitarbeitende** müssen darin geschult werden, die **Grenzen und Fähigkeiten** von KI-Systemen zu **kennen**, damit sie die Ausgaben regelmäßig prüfen (Reduzieren des „Automatisierungsbias“[15]).

- Das für die **Aufgabenbewältigung erforderliche Fachwissen** soll auch beim Einsatz von KI erhalten bleiben und ein entsprechender Kompetenzaufbau weiterhin stattfinden.

- Für einen erfolgreichen KI-Einsatz braucht es „gesonderte **Räume für das Experimentieren**“[16] mit KI sowie einen guten **behörden- und ressortübergreifenden Austausch**. Hierfür können bestehende Ressourcen und Strukturen genutzt sowie weiter ausgebaut werden, beispielsweise die Datenlabore der Bundesregierung und das Beratungszentrum für Künstliche Intelligenz (BeKI).

- Die **Gleichstellungsbeauftragten** und die gewählten **Interessenvertretungen der Mitarbeitenden** sowie die für den **Datenschutz zuständigen Stellen in den Ressorts** und – anlassbezogen – die **datenschutzrechtliche Aufsichtsbehörde (BfDI)** sind frühzeitig einzubeziehen. Dies **schafft Vertrauen** und stellt sicher, dass deren Interessen berücksichtigt werden.

Wir treiben den nachhaltigen Einsatz von KI gemeinsam voran.

Ein gemeinschaftlicher Ansatz soll dabei unterstützen, dass für die gesamte Bundesverwaltung soweit möglich **einheitliche Qualitätsstandards** und **Konventionen** für die Bereitstellung und Nutzung von KI eingehalten werden. Die Leitlinien bilden eine Grundlage für die Abstimmung gemeinsamer Standards zur Umsetzung rechtlicher Anforderungen beim KI-Einsatz. Zudem stärken sie die ressortübergreifende Kollaboration bei der technischen Entwicklung oder Beschaffung. Dabei wird größtmögliche ökologische, ökonomische, organisatorische, technische und soziale Nachhaltigkeit bei der Entwicklung und dem Einsatz von KI angestrebt.[17]

Wie kann ich dieses Prinzip berücksichtigen?

- Wo möglich und sinnvoll sollten **ressortübergreifend gemeinsame Lösungen** für die Entwicklung, Beschaffung und den Einsatz von KI gefunden werden (zum Beispiel Infrastruktur, Produktrahmenverträge, Schulungskonzepte, Leitlinien), um menschliche, finanzielle und ökologische Ressourcen zu schonen. Gemeinsame Infrastrukturen können beispielsweise durch Skaleneffekte Ressourcen sparen.

15 Automatisierungsbias ist die Neigung von Menschen, Vorschläge von automatisierten Entscheidungssystemen zu übernehmen beziehungsweise zu bevorzugen.

16 Das heißt, es gibt eine angemessene technisch-organisatorische Trennung zu Produktivumgebungen, um diese zu schützen (auch durch gesonderte Nutzendenkennungen/-konten).

17 Ein paralleler Einsatz verschiedener KI-Lösungen wird dabei nicht ausgeschlossen. Diversifizierung kann helfen, Digitale Souveränität zu wahren und potenzielle Fehler, Verzerrungen oder Schwächen einzelner KI-Systeme zu identifizieren.

- KI-Systeme und ihr Einsatz sollen möglichst **ressourcenschonend** und **energieeffizient** sein. Bei der Entwicklung und Beschaffung kann die Effizienz der Systeme berücksichtigt werden. Auf Ebene der Hardware sollen Energieverbrauch und die materiellen Ressourcen für die Rechenleistung von KI minimiert werden.[18] KI kann gleichzeitig auch aktiv für Zwecke eingesetzt werden, die der **ökologischen Nachhaltigkeit** dienen.

- Die spätere Wechselmöglichkeit, die eigene Gestaltungsfähigkeit und der Einfluss der Behörde auf den IT-Anbieter sowie dessen Marktposition sollten bei der Wahl der KI-Lösung beachtet werden, um **„Lock-in-Effekte“**[19] zu **vermeiden** und die **Digitale Souveränität** zu **wahren**. Wo möglich werden Modelle mit transparentem Trainingsprozess und frei verfügbaren Parametern gegenüber „Closed-Source-Modellen“ bevorzugt.

18 Siehe auch das „9. Sustainable Development Goal (SDG 9)“ der Bundesregierung unter: Deutsche Nachhaltigkeitsstrategie Weiterentwicklung 2021 | Bundesregierung.

19 Lock-in-Effekte entstehen, wenn der Wechsel zu einem anderen System oder Anbieter mit signifikant erhöhten Kosten verbunden ist.

3. Leitsätze für den Einsatz von KI

Die folgenden Leitsätze dienen als Erläuterung zum Umgang mit **Anforderungen und Herausforderungen beim KI-Einsatz**.[20] Die Leitsätze sind entlang der jeweiligen Adressaten in zwei Abschnitte geteilt: Die Leitsätze A1 bis A5 richten sich an die **Nutzenden** von KI-Systemen. Die Leitsätze B1 bis B9 richten sich an für die Bereitstellung von KI-Systemen **verantwortliche Organisationseinheiten in Behörden** (beispielsweise an deren Datenlabore, Maßnahmenverantwortliche, IT-Verantwortliche).

3.1. Leitsätze für Nutzende

Die nachfolgenden Leitsätze richten sich an alle Mitarbeitenden in der Bundesverwaltung, die KI-Systeme nutzen (im Folgenden als **„Nutzende"** bezeichnet). Ziel der Leitsätze ist es, **eine verantwortungsvolle Nutzung von KI** durch die Darstellung von praktischen, ethischen und rechtlichen Rahmenbedingungen **zu ermöglichen**.

Leitsatz A1:
KI verantwortungsbewusst und ethisch einsetzen

*Nutzende stellen sicher, dass sie die **Nutzung von KI-Systemen** dem jeweiligen Einsatzbereich entsprechend **verantwortungsvoll gestalten**.*

Um **KI-Systeme verantwortungsvoll einzusetzen**, nehmen Nutzende vor dem Einsatz von KI behördliche und ressortübergreifende Schulungen wahr.

Zudem sollte die **allgemeine ethische Wertegrundlage** bei der Verwendung von KI angemessen berücksichtigt werden (siehe auch die Hinweise zur Werteorientierung der KI-VO in Fußnote 11).

20 In begründeten Fällen kann von den Leitsätzen abgewichen werden bei der Anwendung von KI-Systemen, wenn und soweit sie ausschließlich für militärische Zwecke, Verteidigungszwecke oder Zwecke der nationalen Sicherheit in Verkehr gebracht, in Betrieb genommen oder, mit oder ohne Änderungen, verwendet werden, unabhängig von der Art der Einrichtung, die diese Tätigkeiten ausübt. Dasselbe gilt für KI-Systeme oder KI-Modelle, einschließlich ihrer Ausgabe, die eigens für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden.

Umsetzung der ethischen Grundsätze am Beispiel von LLM-Anwendungen:

Am Beispiel von LLM-Anwendungen lässt sich zeigen, wie eine solche Umsetzung der ethischen Grundsätze in der Praxis aussehen kann:

- Für eine verantwortungsvolle Nutzung von textgenerativen LLM-Anwendungen ist die **ethische und sensible Formulierung der Anfrage** durch Nutzende an die Anwendung (engl. „Prompting“) zentral. Das Sprachmodell versucht, die gestellte Anfrage anhand statistischer Abgleichungen der Worte (zum Beispiel der Semantik) zu beantworten. Formulierungs- und Strukturierungstechniken können daher nicht nur dabei unterstützen, die Qualität der Ausgabe zu maximieren, sondern auch dabei, ethische Herausforderungen, wie verzerrte Ausgaben zu vermeiden. Eine Auswahl solcher „Prompting“-Techniken findet sich im Handout für Nutzende „Einsatz großer Sprachmodelle (LLM) gemäß KI-Leitlinien“ (siehe Anlage).

- Unethische Eingaben oder die Verbreitung von generierten Inhalten, die beispielsweise zu **Hassreden, Beleidigungsdelikten nach §§ 185 ff. StGB oder der Verbreitung von Desinformation** veranlassen, **sind unzulässig**. Gleiches gilt für Anfragen, die zum Beispiel **Diskriminierung** oder **Verletzungen der Menschenwürde** implizieren. Zu Zwecken von Schulungen, Recherchen oder dem Prüfen der Modellsicherheit durch sogenanntes „Red-Teaming“[21] sind fachspezifische Ausnahmen möglich. Nutzende müssen zudem prüfen, ob generierte Ergebnisse einen sogenannten Bias enthalten, zum Beispiel **diskriminierende Inhalte oder systematische Fehler beziehungsweise „Missrepräsentationen“**. Wenn ein KI-System mit Datensätzen trainiert wurde, in denen bestimmte Gruppen über- oder unterrepräsentiert sind, kann dies beispielsweise zu verzerrten Informationen führen. Werden diskriminierende oder verzerrte Inhalte festgestellt, dürfen diese nicht übernommen werden. Diskriminierende oder vorurteilsbehaftete Ausgaben durch die Anwendung sollten der in der Anwendung als verantwortlich angegebenen Stelle gemeldet werden.

Leitsatz A2:
Die eigene Datenpreisgabe minimieren

*Bei der Nutzung registrierungspflichtiger bundesverwaltungsexterner KI-Systeme minimieren Nutzende über die Kontoverwaltung die eigene Datenpreisgabe (mit dem **Ziel der Nicht-Rückführbarkeit** und einer **höheren Informationssicherheit**).*

Vorausgesetzt, die verantwortliche Behörde billigt den Einsatz von registrierungspflichtigen KI-Systemen, die offen zugänglich sind und auf bundesverwaltungsexterner IT-Infrastruktur betrieben werden (beispielsweise Web-Anwendungen oder Cloud-basierte Anwendungen wie ChatGPT oder Microsoft Copilot)[22], müssen Nutzende eine **Minimierung der Datenpreisgabe** sicherstellen. Hierdurch wird die Nicht-Rückführbarkeit gestärkt. Folge ist auch eine höhere Informationssicherheit, da beispielsweise gezielte Angriffe auf Grundlage von Social Engineering erschwert werden.

21 Red-Teaming im Bereich der Künstlichen Intelligenz wird meist von speziellen „Red Teams“ durchgeführt, die Angriffsmethoden anwenden, um Fehler und Schwachstellen in KI-Systemen zu identifizieren, zum Beispiel schädliche oder diskriminierende Ergebnisse, unvorhergesehene oder unerwünschte Verhaltensweisen, Einschränkungen oder potenzielle Risiken im Zusammenhang mit dem Missbrauch des Systems.

22 Die genannten KI-Systeme dienen als Beispiele zur Erläuterung für die Nutzenden. Sie wurden ausgewählt, weil sie zum Zeitpunkt der Erarbeitung der KI-Leitlinien in der Bundesverwaltung teilweise zum Einsatz kamen und stellen keine Präferenz gegenüber anderen Systemen dar.

Eine Registrierung für solche Systeme sollte lediglich nach **Freigabe durch den IT-Verantwortlichen unter Einbezug des Votums der notwendigen Beauftragten (Datenschutz- / Geheimschutz- / und Informationssicherheitsbeauftragte) der eigenen Behörde** vorgenommen werden. Mit dem Ziel der Minimierung der Datenpreisgabe dürfen grundsätzlich für die Registrierung weder private Accounts und Geräte noch offizielle personenbezogene Accounts der Behörde verwendet werden. Bei Bedarf können IT-Verantwortliche der Ressorts hinzugezogen werden und hinsichtlich alternativer Vorgehensweisen (beispielsweise der Verwendung pseudonymisierter Accounts) beraten.

Nach der Registrierung müssen Nutzende Maßnahmen zur Minimierung der Datenpreisgabe in der **Nutzerkontenverwaltung** ergreifen. Beispielsweise sollten **Kontoeinstellungen aktiviert werden**, die eine Datenweitergabe an Dritte (insb. in Drittstaaten) nach Möglichkeit unterbinden. Hierzu zählt unter anderem der Widerspruch gegen die **Weiterverwendung eingegebener Daten zu Trainingszwecken des Systems**.[23]

Bei der Nutzung von KI-Systemen, die von der Bundesverwaltung auf bundesverwaltungsinterner Infrastruktur bereitgestellt werden, ist dies regelmäßig nicht erforderlich. Die Weiterverwendung der eingegebenen Daten zu Trainingszwecken kann in diesem Fall, wenn der Datenschutz gewahrt wird[24], zur Verbesserung der Systeme beitragen. Für **bundesverwaltungsinterne Systeme** werden Nutzende von der verantwortlichen Behörde über den Registrierungsprozess und Nutzungsbedingungen informiert.

Leitsatz A3:
Daten verantwortungsvoll eingeben

*Nutzende prüfen **vor der Dateneingabe beziehungsweise -freigabe im Rahmen der KI-Nutzung, welche Daten** unter Berücksichtigung des konkreten Anwendungsfalls in das jeweilige KI-System **eingegeben werden dürfen**.*

*Die notwendigen Informationen für eine **verantwortungsvolle Dateneingabe** werden Nutzenden von der verantwortlichen Organisationseinheit in der Behörde einfach zugänglich bereitgestellt.*

Nutzenden werden die notwendigen Informationen für eine **verantwortungsvolle Dateneingabe** einfach zugänglich von der verantwortlichen Behörde bereitgestellt (zum Beispiel in der Eingabemaske eines KI-Systems, der bereitgestellten Dokumentation oder im Rahmen von Schulungsmaterialien).

Für von der Bundesverwaltung freigegebene oder bereitgestellte KI-Systeme muss folglich **einfach ersichtlich sein**, welche Daten zur Eingabe freigegeben sind. Andernfalls ist grundsätzlich davon auszugehen, dass der **restriktivste Fall gilt** (siehe Kasten zur „Vertiefung: Restriktivster Fall der Dateneingabe“ auf Seite 17).

23 Ergänzend ist es zudem empfehlenswert, möglichst weitreichende Privatsphäre-Einstellungen im Browser vorzunehmen.

24 Es gilt zu beachten, dass auch im Fall eines bundesverwaltungsinternen Betriebs bei der Verarbeitung personenbezogener Daten datenschutzrechtliche Grundsätze weiterhin eingehalten werden müssen. Eine Verwendung personenbezogener Daten zu Trainingszwecken bedarf beispielsweise auch einer Rechtsgrundlage.

Wenn Nutzenden die Informationen vorliegen, welche Daten in ein KI-System eingegeben werden dürfen, ist folgendermaßen vorzugehen: Vor der Nutzung des KI-Systems bewerten Nutzende die **Art der Daten**, die sie in das System eingeben (beispielsweise über die Eingabemaske einer LLM-Anwendung) oder anderweitig anbinden wollen (beispielsweise über einen Dokumenten-Upload). Nutzende prüfen dabei insbesondere, ob die geplante Eingabe sensible dienstliche Daten enthält.

Beispiele für **sensible dienstliche Daten** sind:

- Personenbezogene Daten, dazu gehören unter anderem:
  - *Allgemeine Personendaten (Name, Alter, Geburtsdatum, Anschrift)*
  - *Physische Merkmale (Geschlecht, Haarfarbe, Augenfarbe, Fingerabdruck)*
  - *Kennnummern (Identifikations-, Personalausweis- oder Sozialversicherungsnummer)*
  - *Gesundheitsinformationen (Genetische Daten, Krankendaten)*
  - *Angaben über soziale Beziehungen, berufliche Funktionen, finanzielle Situation oder Bewegungsprofile (Vorstand der Firma XY, Geodaten)*
  - *Dokumentinformationen/Metadaten (Datei erstellt von … / zuletzt bearbeitet von … / Kommentar von …)*
- Betriebs-, Geschäfts- oder allgemeine Amts-/Dienstgeheimnisse,[25] beispielsweise Informationen zu internen Entscheidungsprozessen und Verfahren,
- Informationen mit (sehr) hohem Schutzbedarf bezüglich des Informationssicherheitsgrundwertes „Vertraulichkeit“, beispielsweise Geschäftsgeheimnisse von Unternehmen, auf die im Rahmen der behördlichen Arbeit Zugriff besteht,
- Daten mit besonderer Geheimhaltungspflicht sowie Verschlusssachen,[26]
- eine Sammlung von nicht-vertraulichen Informationen, die in ihrer Summe einer besonderen Geheimhaltungspflicht unterfallen oder
- Daten, die auf Grund sonstiger Vorschriften als nicht für die Öffentlichkeit bestimmte sensitive Informationen klassifiziert sind (beispielsweise LIMITÉ[27]).

Personenbezogene Daten im Sinne von Artikel 4 Nr. 1 DSGVO, das heißt alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen, sind dabei besonders zu schützen.[28] Auch juristische Personen können mit personenbezogenen Daten verknüpft sein oder Schutzrechte eigener Art (zum Beispiel Recht auf den eingerichteten und ausgeübten Gewerbebetrieb) innehaben. Wenn **personenbezogene Daten** Teil der Dateneingabe sind (selbst wenn diese öffentlich zugänglich sind), ist eine datenschutzkonforme Nutzung des KI-Systems sicherzustellen. Die Verarbeitung sensibler personenbezogener Daten, zu denen zum Beispiel

25 Gemäß StGB §§ 203 und 353b.

26 Gemäß § 4 SÜG und § 2 VSA.

27 Ratsinterne Dokumente der Europäischen Union mit der Kennzeichnung LIMITÉ.

28 Siehe für weitere Informationen zu personenbezogenen Daten Leitsatz B9: Datenschutzkonformer KI-Einsatz.

Gesundheitsdaten, Daten über politische Meinungen oder auch religiöse Überzeugungen gehören (vgl. die Aufzählung in Artikel 9 Absatz 1 DSGVO) ist gemäß Artikel 9 Absatz 1 DSGVO grundsätzlich untersagt. Ausnahmen vom grundsätzlichen Verbot sind in den in Artikel 9 Absatz 2 DSGVO geregelten Fällen, zum Beispiel bei Einwilligung, dem Vorliegen berechtigter Interessen oder der Verarbeitung personenbezogener Daten, die die betroffene Person offensichtlich öffentlich gemacht hat (vgl. die Aufzählung in Art. 9 Abs. 2 DSGVO), zulässig.

Nutzende prüfen auf Basis der von der verantwortlichen Behörde bereitgestellten Informationen, ob eine **Eingabe dieser Daten** im Sinne der Interessen der Bundesverwaltung[29] **rechtmäßig und verantwortungsvoll** ist.[30]

Für eine Vielzahl von Anwendungsfällen im Rahmen der behördlichen Arbeit ist die entsprechende **Freigabe des KI-Systems für sensible dienstliche Daten** erforderlich. Insbesondere die Eingabe von **Daten mit besonderer Verschwiegenheits- oder Geheimhaltungspflicht** sowie **Verschlusssachen** durch Nutzende darf nur erfolgen, wenn das System für die Nutzung der fraglichen Art der Daten explizit durch die verantwortliche Behörde freigegeben wurde und die erforderliche Einstufung der Daten fortbesteht.

Vertiefung:
Freigabe von KI-Systemen für die Eingabe von sensiblen dienstlichen Daten

Bei KI-Systemen, die auf bundesverwaltungsinterner IT-Infrastruktur, beispielsweise „on premise“[31] oder über vergleichbar abgesicherte Cloud-Strukturen betrieben werden, ist bei entsprechender Freigabe auch die Eingabe beziehungsweise Nutzung sensibler dienstlicher Daten möglich.[32]

Beispiele für bundesverwaltungsinterne LLM-Anwendungen, die an der Freigabe für sensible dienstliche Daten arbeiten beziehungsweise diese bereits erproben, sind das KI-Portal des ITZBund (KIPITZ) und PLAIN[33].

Sofern keine anderen Informationen vorliegen, gilt für Nutzende in der Regel der **restriktivste Fall** der Dateneingabe. In diesem Fall sind lediglich **öffentliche Daten** für die Eingabe freigegeben. Unter solche „öffentlichen Daten“ fallen beispielsweise Daten, die aus dem eigenen Ressort stammen und für die Öffentlichkeit bestimmt sind, Open Data oder öffentlich zugängliche Daten. Soweit öffentliche Daten Personenbezug aufweisen, unterfallen sie weiterhin dem Datenschutzrecht und eine Eingabe muss, wie oben beschrieben, diesbezüglich geprüft werden.

29 Siehe für weitere Informationen zu den Bedingungen der Eingabe sensibler dienstlicher Daten auch Leitsatz B8: Rahmenbedingungen für verantwortungsvolle Dateneingabe.

30 Dies kann neben dem Datenschutzrecht auch weitere Rechtsbereiche betreffen, wie beispielsweise das Urheberrecht.

31 „On premise“ bezeichnet den Betrieb einer Software auf lokaler Hardware beziehungsweise einer lokalen IT-Umgebung.

32 In Zukunft können VS-IT-Freigaben auch für privatrechtlich betriebene IT-Infrastruktur (beispielsweise Cloud-basierte Systeme) nach der VSA durch das BSI möglich sein.

33 PLAIN steht für „Platform Analysis and Information System“ und ist die Daten- und KI-Plattform der Auslands IT, die auch weiteren Bundesbehörden zur Verfügung steht.

Vertiefung:
Restriktivster Fall der Dateneingabe

Aktuell gilt für KI-Systeme, die auf externer KI-Infrastruktur betrieben werden, der restriktivste Fall der Dateneingabe. Die Einschränkungen ergeben sich insbesondere aus der fehlenden Informationssicherheit, da eingegebene Informationen durch Unternehmen der Privatwirtschaft nachgenutzt werden und/oder gegebenenfalls leichter an die Öffentlichkeit gelangen können.

Auch im restriktivsten Fall gibt es weiterhin viele Anwendungsfälle von KI im Rahmen der behördlichen Arbeit. Beispiele sind:

- die Zusammenfassung eines Dokuments aus dem öffentlich zugänglichen Dokumentations- und Informationssystem für Parlamentsmaterialien,
- die Übersetzung eines öffentlichen Berichts in „Einfache Sprache“,
- der Entwurf eines Social-Media-Beitrags für die Öffentlichkeitsarbeit des eigenen Ressorts,
- die Zusammenfassung eines Themengebiets (alternativ zu der Nutzung von Suchmaschinen),
- die Ideengenerierung für das Argument einer öffentlichen Rede oder
- die Zuhilfenahme eines Sprachmodells bei der Erstellung von Excel-Formeln.

Leitsatz A4:
KI-generierte Inhalte fachlich prüfen

*Nutzende **prüfen KI-generierte Inhalte** vor der weiteren Verwendung **fachlich**.*

Bei der behördlichen Nutzung sind für Arbeitsergebnisse, die mit Hilfe von KI erstellt werden, die gleichen **Qualitätsstandards** einschlägig, wie für Arbeitsergebnisse, die ohne KI erstellt werden. **Mit Hilfe von KI generierte Arbeitsergebnisse** gelten weiterhin als Arbeitsergebnisse des jeweiligen Mitarbeitenden. Das heißt, Nutzende tragen die **gleiche Verantwortung für Rechtmäßigkeit und Qualität ihrer Arbeit**, wie sonst bei der Erledigung von Aufgaben üblich. Aus diesem Grund sollten Nutzende stets hinterfragen, **welche Teile einer Aufgabe mit Hilfe eines KI-Systems gelöst werden können** und wie sich die generierten Inhalte in die Arbeit einfügen lassen.

Umsetzung der fachlichen Prüfung am Beispiel von LLM-Anwendungen:

- **Antworten** und **Inhalte**, welche durch LLM-Anwendungen erzeugt werden, sind daher durch die Nutzenden in Abhängigkeit zu dem konkreten Einsatzbereich und Einsatzzweck **fachlich zu prüfen** und **bei Bedarf zu überarbeiten**. Nur so können die Qualitätsstandards und Rechtmäßigkeit der behördlichen Arbeit garantiert werden.

- Beim Einsatz von LLM-Anwendungen zur Zusammenfassung umfangreicher Dokumente bedeutet dies beispielsweise, dass Nutzende **kritisch abwägen**, in welchem Umfang eine Prüfung der ausgegebenen Zusammenfassung auf Basis des Originaldokuments erforderlich ist. Der notwendige **Umfang einer solchen Prüfung** hängt vom Verwendungszweck ab. Im Falle der Nutzung von LLM-Anwendungen zu Recherchezwecken ist sicherzustellen, dass die eigene Fachexpertise ausreicht, um die Ausgaben kritisch zu hinterfragen. Dabei sollte beispielsweise überprüft werden:

  - *Sind die generierten Informationen und Quellen fachlich korrekt? Es besteht beispielsweise die* ***Gefahr des Halluzinierens*** *des Modells. Das heißt, die LLM-Anwendung erzeugt fehlerhafte Ausgaben, gegebenenfalls sogar mit erfundenen Quellenangaben, die überzeugend als relevante Lösungen zur Fragestellung des Nutzenden präsentiert werden. Sollten die Bearbeitenden diesbezüglich unsicher sein, müssen die Informationen auf anderem Wege geprüft werden, beispielsweise durch den Einbezug eines fachkundigen Mitarbeitenden der fraglichen Behörde.*

  - *Enthalten generierte Inhalte* **personenbezogene Daten***? Es gilt, dass generierte Inhalte mit personenbezogenen Daten nur im Einklang mit den datenschutzrechtlichen Bestimmungen weiterverwendet werden dürfen.*[34]

Leitsatz A5:
KI transparent nutzen

Nutzenden wird ein ***transparenter Umgang mit der eigenen KI-Nutzung*** *gegenüber anderen Mitarbeitenden sowie Vorgesetzten empfohlen.*

Nutzende sollten sich in ihrem Ressort beziehungsweise ihrer Behörde über etwaige bestehende **interne oder externe Kennzeichnungs-** oder **Dokumentationspflichten** informieren. Grundsätzlich gilt:

- Wenn fachlich ungeprüfte Inhalte behördenintern oder an die Öffentlichkeit weitergegeben werden (beispielsweise im Rahmen der Verwendung von Chatbots), ist die Nutzung eines KI-Systems zu kennzeichnen (vgl. auch Artikel 50 Absatz 4 Satz 4 KI-VO).

- Wenn KI-generierte Inhalte fachlich geprüft und so ausreichend redaktionell überarbeitet wurden, dass die Qualität des produzierten Inhalts sichergestellt wurde, ist regelmäßig keine weitere Kennzeichnung notwendig (vgl. auch Artikel 50 Absatz 4 Satz 5 2. HS KI-VO).

34 Siehe hierzu auch die „Orientierungshilfe der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder vom 6. Mai 2024: Künstliche Intelligenz und Datenschutz“, siehe unter: Orientierungshilfe Künstliche Intelligenz und Datenschutz | DSK.

3.2. Leitsätze für Behörden

Die nachfolgenden Leitsätze richten sich an die für die Bereitstellung von KI-Systemen **verantwortlichen Organisationseinheiten in Behörden** (beispielsweise an deren Datenlabore, Maßnahmenverantwortliche, IT-Verantwortliche). Ziel ist die Ermöglichung eines **weitreichenden Einsatzes von KI unter Berücksichtigung des geltenden Rechtsrahmens**.

Leitsatz B1:
Klare Behördenentscheidungen zur Nutzung von KI

*Behörden treffen für Ihren Zuständigkeitsbereich klare Entscheidungen, um eine möglichst weitreichende Nutzung von KI-Systemen zu ermöglichen. Bei der Beurteilung wird ein **chancenorientierter** und zugleich **verantwortungsvoller Einsatz** verfolgt.*

Die **Entscheidung darüber, ob und welche KI-Systeme eingesetzt werden, obliegt der Behördenleitung**. Diese entscheidet unter anderem vor dem Hintergrund von Informationssicherheit, Datenschutz und Geheimschutz, ob und welche KI-Systeme für dienstliche Zwecke genutzt werden dürfen. IT-Verantwortliche der Behörden sind verpflichtet, über freigegebene KI-Systeme Auskunft geben zu können.

Die Bereitstellung und Nutzung von KI-Systemen haben dabei immer unter **Wahrung der Sicherheit und Rechtmäßigkeit der behördlichen Arbeit zu erfolgen**. Als Rechtsgrundlagen von besonderer Relevanz sind das **Grundgesetz (GG)**, die **Datenschutzgrundverordnung (DSGVO)**, das **Bundesdatenschutzgesetz (BDSG)**, das **Gesetz über das Bundesamt für Sicherheit in der Informationstechnik (BSIG)**, das **Sicherheitsüberprüfungsgesetz (SÜG)**, das **Bundespersonalvertretungsgesetz (BPersVG)**, beziehungsweise die **Verschlusssachenanweisung (VSA)**, die einschlägigen **Regelungen und Standards zur Informationssicherheit** und die **KI-Verordnung der EU (KI-VO)**.

Leitsatz B2:
Festlegung des Einsatzbereichs und Einsatzzwecks

*Behörden legen im Rahmen der Bereitstellung eines KI-Systems für die behördliche Arbeit den **Einsatzbereich und Einsatzzweck** fest.*

KI-Systeme sind **Unterstützungswerkzeuge für die behördliche Arbeit**. Sie können Analysefähigkeiten steigern, Informationen integrieren oder der Kommunikation dienen (zum Beispiel Chatbots). Vor dem Einsatz legen Behörden fest, in **welchem Einsatzbereich und zu welchem Zweck** KI-Systeme eingesetzt werden dürfen. Eine klare Festlegung des Einsatzzwecks bringt Rechtssicherheit und ist Grundlage für einen verantwortungsvollen aber auch chancenorientierten Einsatz von KI. Für einen rechtskonformen Einsatz in der Bundesverwaltung ist dabei sicherzustellen, dass sich der Einsatzzweck **im Rahmen der gesetzlich zugewiesenen öffentlichen Aufgaben der Behörden** befindet.

Die KI-VO schließt in Artikel 5 den Einsatz bestimmter **KI-Systeme mit unannehmbarem Risiko** aus. Beispiele hierfür sind die Klassifizierung von Personengruppen aufgrund ihres Verhaltens („Social Scoring“) oder die Emotionserkennung am Arbeitsplatz. Die Festlegung eines solchen Einsatzzwecks ist daher grundsätzlich **ausgeschlossen**.[35]

35 Soweit Aspekte der nationalen Sicherheit dieses nicht erforderlich machen.

Vor dem Hintergrund von § 35a VwVfG ist zu beachten, dass Mitarbeitende KI-Systeme nur in unterstützender Funktion einsetzen können, die Prüfung, Entscheidung und Verantwortung in letzter Instanz aber beim Mitarbeitenden selbst verbleibt, sofern nicht der vollautomatisierte Erlass von Verwaltungsakten durch Rechtsvorschrift zugelassen ist.[36]

Leitsatz B3:
Auswahl geeigneter KI-Systeme

*Behörden führen eine **sorgfältige Anbieter-, Anwendungs- und Modellauswahl** durch, um eine bestmögliche Eignung eines KI-Systems für den gegebenen Anwendungsfall zu gewährleisten.*

*Dabei berücksichtigen Behörden unter anderem den **Einsatzzweck**, die **notwendige Daten- und IT-Infrastruktur** sowie **wirtschaftliche, ethische** und **organisatorische Rahmenbedingungen** für die Nutzung eines KI-Systems.*

Die **Eignung eines KI-Systems** ist grundsätzlich vor dem Hintergrund eines spezifischen Anwendungsfalls zu prüfen und obliegt jeweils der verantwortlichen Behörde. Vor der Entscheidung für den Einsatz oder die Entwicklung von KI-Systemen sollte der fachliche Prozess und gegebene Einsatzzweck eines KI-Systems evaluiert werden (siehe Leitsatz B2: Festlegung des Einsatzbereichs und Einsatzzwecks). Es ist grundsätzlich zu hinterfragen, ob das vorliegende Problem durch KI gelöst werden kann oder sollte. Bei der Beurteilung werden die fachverantwortlichen Personen und Gremien in den Behörden einbezogen. Die Informationssicherheitsbeauftragten der Behörden sind hierbei frühzeitig einzubinden.

Die für die Entwicklung und Beschaffung von Software **etablierten Maßstäbe und Regelungen**, wie zum Beispiel die Wirtschaftlichkeitsbetrachtung (WiBe), finden auch bei der Einführung von KI-Systemen Anwendung. Zudem sind für die Bewertung und Auswahl von KI-Systemen unter anderem von besonderer Relevanz:

- Die **Performanz** von KI-Systemen verschiedener Anbieter für den gegebenen Anwendungsfall. Die Performanz muss unter anderem gegen die Wahrung der **Anforderungen der Digitalen Souveränität Deutschlands** (insbesondere bei der Anbieter- und Modellauswahl) abgewogen werden. Kriterien wie die spätere Wechselmöglichkeit, die eigene Gestaltungsfähigkeit und der mögliche Einfluss auf den IT-Dienstleister sichern die Arbeitsfähigkeit der Verwaltung langfristig ab und verhindern Lock-In-Effekte.

- Die Berücksichtigung von **Sicherheitsgesichtspunkten** (Informationssicherheit, Datenschutz, Geheimschutz). Sofern der Einsatzzweck dies zulässt, sollten KI-Systeme bevorzugt werden, die bereits technisch sicherstellen, dass sich die Verarbeitung personenbezogener Daten im Rahmen des rechtlich zulässigen und von der jeweiligen Dienststelle vorgegebenen Rahmens bewegt.

36 Ein Verwaltungsakt kann vollständig automatisiert, also auch vollständig durch KI-Systeme erlassen werden, sofern weder ein Ermessen noch ein Beurteilungsspielraum besteht und der vollständige Erlass durch Rechtsvorschrift zugelassen ist. Dieser Grundsatz steht in Korrelation mit Artikel 22 Absatz 1 DSGVO, wonach Entscheidungen mit Rechtswirkung grundsätzlich nur von Menschen getroffen werden, wobei auch hier Ausnahmen in bestimmten Fällen zugelassen sind. Eine vergleichbare Regelung für das Sozialrecht findet sich unter anderem auch in § 31a SGB X.

- Die Prüfung der **dazugehörigen IT-Infrastruktur**[37] (unter anderem Eigentumsverhältnisse/Verantwortung/Administration für Hard- und Software inklusive Trainingsdaten, Betrieb von Hard- und Software inklusive des Standorts) sowie je nach Anwendungsfall deren **Anschlussfähigkeit an bestehende Systeme**.

- Die Bewertung der Qualität und Zusammensetzung der **zugrundeliegenden Datenbasis** (unter anderem Herkunft/Erzeugung der Trainingsdaten) sowie die **Verfügbarkeit notwendiger** (beispielsweise behördlicher) **Daten**. Generell ist beim Einsatz von KI-Systemen wichtig, dass durch eine passende Auswahl des Modells und der technischen Umgebung eine Nutzung aller für den Verwendungszweck notwendigen Daten möglich ist. Aus der Systemumgebung folgt, wie viel Kontrolle über die eingegebenen Daten besteht, weshalb die Möglichkeiten der Nutzung von Systemen auf bundesverwaltungsexterner IT-Infrastruktur in der Regel restriktiver gehandhabt werden.

- Das Hinzuziehen der **gemeinsamen Wertegrundlage** (siehe Kapitel 2) als ethischer Maßstab für die Auswahl (zum Beispiel die Prüfung der Nachhaltigkeit des vorliegenden KI-Systems).

- Die Berücksichtigung **organisatorischer Rahmenbedingungen** für die Einführung eines KI-Systems (zum Beispiel Feedback aus der Gremieneinbindung, Nutzbarkeit des Systems, Standard der Umsetzung der Barrierefreiheit, Schulungsmaterialien).

Leitsatz B4:
Ethischer und sensibler Einsatz von KI-Systemen

*Um KI verantwortungsvoll einzusetzen, wägen Behörden vor der Bereitstellung ab, inwieweit die **strategischen Leitprinzipien** (siehe Kapitel 2) und die **Werteorientierung der KI-VO**[38] im jeweiligen KI-Anwendungsfall gewahrt werden.*

KI-Systeme sind im Sinne der **eigenen behördlichen Arbeit** so einzusetzen, dass der Einsatz **ethisch vertretbar** und im **dienstlichen Interesse** ist.

Vor der Bereitstellung von KI-Systemen wägen Behörden die strategischen Leitprinzipien sowie die darin implizierten ethischen Grundsätze (siehe Kapitel 2) für den Einsatz von KI sorgfältig gegeneinander ab. Oft kann es zu **Zielkonflikten in Bezug auf verschiedene Werte** kommen. Beispielsweise kann die Maximierung der Resilienz des KI-Systems mit einem erhöhten Energiebedarf und damit verringerter ökologischer Nachhaltigkeit einhergehen.

Da es nicht immer möglich ist, im Anwendungsfall allen wertebasierten Grundsätzen in höchstem Maße zu entsprechen, bedarf es einer **individuellen Abwägung** durch die Behörden.

37 Für eine Vertiefung der unterschiedlichen Optionen siehe Leitsatz B8: Rahmenbedingungen für eine verantwortungsvolle Dateneingabe.

38 Für mehr Informationen zur Werteorientierung der KI-VO siehe Fußnote 11.

Beispiele solcher Abwägungen von wertebasierten Grundsätzen können sein:

- **Menschliche Aufsicht:** Bei der Qualitätskontrolle des KI-Outputs wägen Behörden zwischen dem Aufwand der Kontrolle und dem qualitativ dadurch erzielten Mehrwert ab. Die regelmäßige Evaluierung des Prozesses zur Qualitätskontrolle durch Behörden hilft, diesen zu optimieren und bei Bedarf anzupassen.

- **Menschenzentrierung:** Vor dem Einsatz von KI-Systemen wägen Behörden zwischen der Chance zur Senkung der Arbeitslast und dem Verlust an individuellen Handlungsspielräumen ab (beispielsweise beim Einsatz von Chatbots). Eine Kosten-Nutzen-Analyse[39] unter Einbezug der Nutzenden kann hierbei die Entscheidungsfindung erleichtern.

- **Sicherheit:** Bei KI-Systemen kann es unter anderem zu Zielkonflikten zwischen Sicherheit, zum Beispiel vor Cyberangriffen, und Performanz kommen (beispielsweise kann die Nutzung von VPN-Kanälen zu einer Verlangsamung von Anfragen beziehungsweise Ausgaben führen). Derartige Effekte sind nicht KI-spezifisch und müssen nicht bei allen KI-Systemen auftreten, sollten aber bei der Planung berücksichtigt werden. Abhängig vom konkreten Anwendungsfall identifizieren und testen Behörden Maßnahmen, die die Einhaltung von Sicherheitsstandards gewährleisten, ohne die Systemleistung erheblich zu beeinträchtigen.

- **Verlässlichkeit und nachvollziehbare Ergebnisfindung:** Der Einsatz von KI-Systemen zur Unterstützung der Arbeit kann dazu führen, dass für einzelne Prozessschritte eine geringere Nachvollziehbarkeit hinsichtlich der Entstehung der Ergebnisse besteht. Hierbei wägen Behörden ab, welche Teilschritte mit Hilfe von KI gelöst werden können, ohne die Nachvollziehbarkeit einer Verwaltungsentscheidung zu gefährden.

- **Nachhaltigkeit:** Effizienzgewinne bei Arbeitsabläufen werden gegen den ökologischen Ressourcenverbrauch von KI abgewogen, eine kürzere Entwicklungsdauer gegen langfristige Nutzbarkeit. Ökologische Auswirkungen sollten Bestandteil der durchzuführenden Kosten-Nutzen-Analyse von Systemen sein.

Leitsatz B5:
Sicherstellung eines Kompetenzaufbau-Angebots

*Behörden stellen sicher, dass Nutzenden vor der Verwendung von KI-Systemen ein **Qualifizierungsangebot** gemacht wird, um für den **verantwortungsvollen Umgang mit KI** zu sensibilisieren und die nötigen Kompetenzen zu vermitteln.*

Mitarbeitende der Bundesverwaltung sind vor dem Hintergrund ihrer Aufgaben in der Behörde **zu schulen**, um KI-Systeme im Hinblick auf diese Aufgaben **kompetent** nutzen zu können. Nur so können die Potenziale von KI-Systemen für die Aufgabenerledigung sinnvoll genutzt und Risiken eingeschätzt werden.

39 Die Betrachtung im Sinne einer Wirtschaftlichkeitsbetrachtung (WiBe) als auch die Sicherstellung der Wirtschaftlichkeit behördlichen Handelns sind in jedem Fall verpflichtend.

Behörden stellen sicher, dass Mitarbeitende Zugriff auf entsprechende Schulungsangebote haben. Die jeweils verantwortliche Behörde entscheidet, inwiefern für bestimmte KI-Systeme vor der Nutzung eine **Schulungsteilnahme durch Nutzende verpflichtend** ist.

Unter Einbezug der **ressortspezifischen Sicherheitsanforderungen und Arbeitsgebiete** gelten unterschiedliche Anforderungen an einen verantwortungsvollen und kompetenten Umgang mit KI. Daher stellen Behörden je nach Bedarf an den jeweiligen Kontext angepasste Schulungen bereit. In diesem Zusammenhang sollte außerdem geprüft werden, ob vor der Nutzung bestimmter sensibler KI-Systeme die Teilnahme an verpflichtenden Schulungen nötig ist. Eine solche bedarfsgerechte Sensibilisierung kann als ein **Bestandteil der generellen Sensibilisierung** zum verantwortungsvollen Umgang mit IT und der Arbeit mit Daten umgesetzt werden. Diese Sensibilisierungsangebote sind allen Beschäftigten niedrigschwellig zur Verfügung zu stellen.

Ziel ist es, dass Nutzende ein grundlegendes Verständnis darüber haben, wie KI-Systeme funktionieren und Ausgaben generieren. Die **Sensibilisierung für das Potenzial und die Risiken des Einsatzes von KI** im Rahmen der eigenen Verwaltungsarbeit ist dabei besonders wichtig. Dieses Wissen hilft bei der Einschätzung, wann und wie KI-Systeme sinnvoll eingesetzt werden können.

Beispiel: Grundlegende Kompetenzen für die Nutzung von LLM-Anwendungen

Beim Einsatz von LLM-Anwendungen sind Nutzende beispielsweise dafür zu sensibilisieren, welche Formen der Dateneingabe (das sogenannte **„Prompting“**) die Qualität von Ausgaben verbessern können, und dass **Ausgaben von LLM-Modellen grundsätzlich hinreichend kritisch zu prüfen** sind.[40] So kann dem Risiko begegnet werden, dass Nutzende KI-generierten Ergebnissen intuitiv vertrauen und Ausgaben nicht hinreichend prüfen.

Beschäftigten der Bundesverwaltung soll auch ressortübergreifend die Möglichkeit des Kompetenzaufbaus und der Fortbildung im Hinblick auf das Zukunftsthema „Künstliche Intelligenz“ eröffnet werden. Ein solches Schulungsangebot umfasst neben der **Funktionsweise** und der **kompetenten Nutzung** von KI-Systemen (beispielsweise von LLMs) auch **ethische Grundsätze** und den **rechtlichen Rahmen** (zum Beispiel Datenschutz und den Schutz von Dienstgeheimnissen und Geschäftsgeheimnissen Dritter) im Kontext der Nutzung. Zudem werden Mitarbeitende für die Limitierungen und Risiken der Technologie sensibilisiert.

Als zentraler Fortbildungsdienstleister stellt die Bundesakademie für öffentliche Verwaltung (BAköV) zum Beispiel im Fortbildungsportal der Bundesverwaltung modulare Bausteine für Schulungs- und Sensibilisierungsmaßnahmen bereit. Ein solcher Werkzeugkasten nutzt **Synergieeffekte, um Schulungen multimedial, nachhaltig und effizient** durchzuführen. Das Angebot kann von den Behörden genutzt und individuell an die Gegebenheiten der jeweiligen Behörde angepasst werden. Daneben stehen weitere Angebote zur Verfügung, wie zum Beispiel die Expertise der jeweiligen Datenlabore.

40 Für weitere Informationen und praxisnahe Beispiele zum „Prompting“, siehe das Handout „Einsatz großer Sprachmodelle (LLM) gemäß KI-Leitlinien“, siehe Anlage.

Leitsatz B6:
Transparenz über die behördliche Nutzung von KI

*Bundesbehörden machen den grundsätzlichen **Einsatz von KI-Systemen** in den eigenen Häusern **so transparent wie möglich**.*

*Behörden klären für sich, ob Regelungen zur **Kennzeichnungspflicht** (intern wie extern) oder **Dokumentationspflichten** für die KI-Nutzung erforderlich sind.*

Ziel der Bundesverwaltung ist es, dass Verwaltungshandeln transparent und nachvollziehbar bleibt. Ein **transparenter Umgang mit der KI-Nutzung** kann wesentlich dazu beitragen, das **Vertrauen der Bevölkerung in den KI-Einsatz durch den Staat zu stärken** – dies gilt insbesondere in der Kommunikation nach außen.

Behörden klären für sich, ob Regelungen zur **internen oder externen Kennzeichnungspflicht** oder **Dokumentationspflichten** für die konkrete KI-Nutzung erforderlich sind.[41] Die KI-VO regelt beispielsweise die Kennzeichnungspflicht für Hochrisiko-Systeme (Artikel 48) sowie Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme (Artikel 50).

Wenn durch KI-Systeme erzeugte oder bearbeitete Inhalte ausreichend **fachlich und redaktionell durch einen Mitarbeitenden** überarbeitet wurden, so dass die Qualität des produzierten Inhalts im Verhältnis zur jeweiligen Aufgabe hinreichend sichergestellt wurde, kann grundsätzlich auf eine Kennzeichnung verzichtet werden. Zur Herstellung von Transparenz betreibt die Bundesverwaltung den sogenannten „Marktplatz der KI-Möglichkeiten“ („MaKI“)[42]. Dieser bringt auf Bundesebene Ministerien und Behörden mit passenden KI-Systemen und Bedarfen zueinander und bietet Transparenz über die KI-Systemlandschaft und Erfahrungswerte in der Bundesverwaltung. Die Behörden sind bei dem produktiven Betrieb von KI-Systemen angehalten, entsprechend der **Regelungen zur Eintragung von KI-Systemen** auf dem MaKI, den Einsatz der KI-Systeme aufzuzeigen.[43]

Leitsatz B7:
Minimierung der Datenpreisgabe

*Behörden stellen sicher, dass mit dem Ziel der Nicht-Rückführbarkeit und einer höheren Informationssicherheit bei der Kontobereitstellung und -verwaltung das **Prinzip der Minimierung der eigenen Datenpreisgabe** als Maßstab für den KI-Einsatz gilt.*

Insbesondere wenn durch verantwortliche Behörden KI-Systeme freigegeben werden, die auf IT-Infrastruktur betrieben werden, die sich nicht für die Verarbeitung sensibler dienstlicher Daten eignet (siehe hierzu Leitsatz B8), und die eine Registrierung erfordern, wird durch die verant-

41 Es könnte beispielsweise gelten: Werden Inhalte über die Referatsgrenze hinaus weitergegeben oder veröffentlicht, hat das Referat die Qualitätssicherung der Inhalte sicherzustellen. Eine Kennzeichnung der für die Erstellung verwendeten Hilfsmittel wäre dann nicht mehr erforderlich.

42 Der Marktplatz der KI-Möglichkeiten dient als KI-Transparenzregister und Matching-Plattform für KI-Systeme in der Bundesverwaltung. Er ermöglicht einen umfassenden und transparenten Überblick über den KI-Einsatz, fördert den Austausch sowie Kooperationen zu diesem und erschließt Nachnutzungspotenziale bestehender KI-Systeme. Er ist erreichbar unter dem Link: www.kimarktplatz.bund.de.

43 Siehe: Marktplatz der KI-Möglichkeiten | IT-Rat Beschluss Nr. [2024/01].

wortliche Behörde eine **Minimierung der Datenpreisgabe** sichergestellt.

Dazu gehört zunächst eine möglichst **anonyme Registrierung** für KI-Systeme ohne Namen und Telefonnummern der Beschäftigten. Die IT-Verantwortlichen sind angehalten, **pseudonymisierte E-Mail-Adressen** für eine Registrierung bei Modellanbietenden bereitzustellen. Eine pseudonymisierte (und bei Bedarf auch zeitlich befristete) Nutzung erschwert die **Rückführbarkeit von Anfragen auf die Bundesverwaltung** beziehungsweise einzelne Mitarbeitende und minimiert **Risiken für die Informationssicherheit**, die sich beispielsweise durch Abfluss dienstlicher E-Mail-Adressen ergeben. Die dienstliche Nutzung von KI-Systemen durch Beschäftigte erfolgt in diesem Rahmen immer auf Grundlage der IT-Systeme und Accounts der Behörden und nicht unter Verwendung privater Accounts und Geräte.

Alternativ hierzu ist die Nutzung von **Diensten ohne Registrierung** zu prüfen. Auch kann die Anmeldung über die Einbettung eines KI-Systems via API-Schnittstelle[44] erfolgen (zum Beispiel bundesverwaltungsintern), wenn dabei die Minimierung der Datenpreisgabe berücksichtigt wird.

Behörden sollen Nutzende darüber hinaus über Maßnahmen zur Minimierung der Datenpreisgabe in der Nutzerkontenverwaltung informieren und bei der Umsetzung standardmäßig unterstützen. Ein Beispiel ist die Umsetzung datensparsamer Kontoeinstellungen, wie dem Widerspruch gegen die **Weiterverwendung eingegebener Daten zu Trainingszwecken des Systems**.

Leitsatz B8:
Rahmenbedingungen für verantwortungsvolle Dateneingabe

*Behörden machen die **technischen Rahmenbedingungen** sowie deren **Implikationen für die Dateneingabe** für Nutzende **leicht verständlich kenntlich**.*

Die verantwortliche Behörde spezifiziert die **Art der Daten**, die von Nutzenden in KI-Systeme eingegeben werden dürfen. Dies gilt für von der Bundesverwaltung oder in ihrem Auftrag entwickelte oder durch die Bundesverwaltung angebotene, bereitgestellte beziehungsweise freigegebene Systeme:

- Ein System kann beispielsweise nur für die Eingabe **öffentlicher, nicht personenbezogener Daten** freigegeben sein.[45]
- Je nach technischen Rahmenbedingungen eines spezifischen Systems kann auch die Eingabe anderer Arten von Daten freigegeben werden. Dies können **personenbezogene Daten** sein (sofern die Rechtsgrundlage für die Verarbeitung dieser Daten gegeben ist)[46] oder andere **sensible dienstliche Daten** (beispielsweise Daten, die einer Verschwiegenheits- oder Geheimhaltungspflicht unterliegen sowie Verschlusssachen gemäß VSA).

44 Eine Programmierschnittstelle, kurz API (Englisch für „Application Programming Interface“), ermöglicht, bestehende Sprachmodelle in verschiedene Anwendungen und Programme zu integrieren.

45 Es ist zu beachten, dass auch öffentliche Daten, soweit sie Personenbezug aufweisen, dem Datenschutzrecht unterfallen.

46 Siehe dazu Leitsatz B9: Datenschutzkonformer KI-Einsatz.

Entsprechende Informationen sollten für Nutzende leicht ersichtlich im KI-System beispielsweise in der Eingabemaske sowie in relevanten Dokumentationen wie Nutzungsbedingungen oder FAQs bereitgestellt werden. Alternativ kann die verantwortliche Behörde Nutzende auch über Schulungen zum Thema Dateneingabe für das betreffende KI-System aufklären.

Um zu beurteilen, welche Daten eingegeben werden dürfen, prüft die verantwortliche Behörde (beispielsweise durch IT- oder Maßnahmenverantwortliche) vor der Bereitstellung, in welcher **IT-Infrastruktur** das KI-System betrieben wird. Dabei können die folgenden Fragen hilfreich sein[47]:

- Welchen Sicherheitsstandards[48] entspricht die Betriebsumgebung des KI-Systems?
- Läuft das KI-System in einer geschützten Umgebung ohne „Datenabfluss“?
- Existiert ein Rollen- und Rechtekonzept mit abgestimmten Zugriffsrechten?
- Wie muss das KI-System bezüglich Vertraulichkeit, Integrität und Verfügbarkeit geschützt werden (Schutzbedarfsfeststellung nach BSI-Standard 200-2)?

Von der Betriebsumgebung hängt ab, wie sicher die eingegebenen Daten sind (beispielsweise vor Zugriffen durch Dritte) und welche Daten Nutzende somit in das KI-System eingeben dürfen.

Grundsätzlich[49] sind im Hinblick auf die Charakteristiken eines KI-Systems und dessen technischer Umgebung die folgenden Betriebsumgebungen **zu unterscheiden**:

- **Betrieb auf bundesverwaltungsexterner IT-Infrastruktur:** Das KI-System wird auf IT-Plattformen von Dritten betrieben. Beispiele hierfür sind die Nutzung der Infrastruktur von privaten Cloud-Anbietenden für das Hosting eines Systems oder die direkte Anbindung einer Anwendung an ein Sprachmodell via API. Die Möglichkeit der Eingabe **sensibler dienstlicher Daten** ist hierbei besonders zu prüfen und sollte im Zweifelsfall nicht erfolgen. Vorausgesetzt, dass die datenschutzrechtlichen Voraussetzungen erfüllt werden, ist auf solchen Strukturen aber grundsätzlich auch die Verarbeitung personenbezogener Daten möglich. Je nach Ausgestaltung wären nach der VSA auch VS-IT-Freigaben für privatrechtlich betriebene IT-Infrastrukturen möglich (und in Zukunft wahrscheinlich).[50]

47 Näheres hierzu kann den einschlägigen Veröffentlichungen und Vorschriften des Bundesamtes für Sicherheit in der Informationstechnik (BSI) entnommen werden (siehe unter: www.bsi.bund.de).

48 Das BSI hat beispielsweise mit dem Artificial Intelligence Cloud Service Compliance Criteria Catalogue (AIC4; Stand: 2021) Mindestanforderungen an die sichere Verwendung von Methoden des maschinellen Lernens in Cloud-Diensten definiert, um die transparente Prüfung und Dokumentation der Informationssicherheit eines KI-Cloud-Dienstes zu fördern.

49 In spezifischen Fällen ist im Einzelfall zu prüfen, welchen Sicherheitsstandards die Betriebsumgebung entspricht. Beispiele sind KI-Systeme, die auf bundesverwaltungsexterner Hardware staatlicher europäischer Organisationen wie MPA (S) oder AEMPS (ES) betrieben werden sowie bei privaten Anbietern deren Rechenzentren die notwendigen Schutzmaßnahmen für dienstliche Rechenzentren erfüllen.

50 Siehe analog Mindeststandard des BSI nach § 8 Absatz 1 Satz 1 BSIG zur Nutzung externer Cloud-Dienste in der Bundesverwaltung; sollten Anpassungen dieser Mindeststandards hinsichtlich der Verarbeitbarkeit von VS-NfD-Daten in bundesverwaltungsexternen Cloud-Lösungen getroffen werden, kann die Bearbeitung dieser Daten mit KI-Systemen in bundesexterner Hardware unter Einhaltung der aktualisierten Mindeststandards möglich werden.

- **Betrieb auf bundesverwaltungsinterner IT-Infrastruktur (mit oder ohne entsprechender VS-IT-Freigabe**[51]): Das KI-System wird auf einer Umgebung der Bundesverwaltung eigenständig oder durch Dritte betrieben. Beispiele hierfür sind der Betrieb eines KI-Systems über eine gemeinsame „Private Cloud“[52] (zum Beispiel innerhalb der Netze des Bundes) mit zentraler IT-Plattform oder auch das ressort- oder behördeninterne Hosting. Je nach Schutzniveau der genutzten IT-Infrastruktur kann eine Nutzung von VS-NfD-Daten für ein KI-System freigegeben werden.[53,54] Auch in diesem Fall ist bei Erfüllung der datenschutzrechtlichen Vorgaben die Verarbeitung personenbezogener Daten möglich

Neben der Umgebung, in der ein KI-System betrieben wird, hängt die Sicherheit der eingegebenen Daten auch von **vertraglichen Regelungen** zwischen der verantwortlichen Behörde und dem Modellbetreiber ab. Relevante Faktoren können sein, welche technische Verbindung zu dem Modell besteht, welche Rechte der Datenverarbeitung durch den Modellbetreiber bestehen und ob bereits ein Auftragsverarbeitungsvertrag (AVV) durch die verantwortliche Behörde abgeschlossen wurde. Darüber hinaus können die **Netzanbindung** und **einschlägige Zertifizierungen** die Sicherheit beeinflussen. Sobald mit einem KI-System (Software) Verschlusssachen elektronisch verarbeitet werden, bedarf zudem auch das KI-System selbst einer **VS-IT-Freigabe** (vgl. § 50 Absatz 1 VSA).

Leitsatz B9:
Datenschutzkonformer KI-Einsatz

*Behörden stellen **bei der Bereitstellung von KI-Systemen** sicher, dass alle erforderlichen Maßnahmen zum datenschutzkonformen Einsatz durchgeführt wurden.*

Wenn personenbezogene Daten beispielsweise Teil einer Dateneingabe oder eines Modelloutputs bei der Nutzung von KI-Systemen sind, ist ein **datenschutzkonformer Einsatz** sicherzustellen.[55] Zwei Themenbereiche sind hier besonders hervorzuheben:

- **Rechtsgrundlage:** Die Verarbeitung personenbezogener Daten bedarf stets einer Rechtsgrundlage. Die Rechtsgrundlagen, welche die mit dem Einsatz von KI verbundene Verarbeitung personenbezogener Daten rechtfertigen können, sind bereits in einem frühen Stadium vor der Umsetzung zu identifizieren. Dabei ist zwischen verschiedenen Lebenszyklusphasen der KI – wie zum Beispiel Training oder Einsatz im Produktivbetrieb – zu unterscheiden, da jeweils andere Rechtsgrundlagen einschlägig sein können.

51 Das heißt mit beziehungsweise ohne Freigabe zur Verarbeitung von Verschlusssachen des entsprechenden Geheimhaltungsgrades wie zum Beispiel VS-NUR FÜR DEN DIENSTGEBRAUCH (VS-NfD), VS-VERTRAULICH, GEHEIM oder STRENG GEHEIM.

52 Einen derartigen Service stellt zum Beispiel das Produkt PLAIN der Auslands-IT dar.

53 Manche IT-Infrastruktur ist für die elektronische Verarbeitung von Verschlusssachen freigegeben (vgl. § 50 Absatz 1 VSA).

54 Für weitere Informationen siehe insbesondere Ressourcen des BSI unter: www.bsi.bund.de/dok/10417576 und www.bsi.bund.de/dok/6621662.

55 Siehe hierzu auch die vertiefende Betrachtung in der „Orientierungshilfe der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder vom 6. Mai 2024: Künstliche Intelligenz und Datenschutz, siehe Fußnote 39 sowie die Stellungnahme des Europäischen Datenschutzausschusses (EDSA) vom 18. Dezember 2024 zur Verwendung personenbezogener Daten für die Entwicklung und Einführung von KI-Modellen, siehe unter: Stellungnahme des Europäischen Datenschutzausschusses (EDSA).

- **Einhaltung der Grundsätze von Datenminimierung und Zweckbindung:** Bei der Verarbeitung von personenbezogenen Daten durch KI-Systeme sind durch die Dienststelle die datenschutzrechtlichen Vorgaben der DSGVO zu beachten, unter anderem die Grundsätze von Datenminimierung und Zweckbindung.

In diesem Zusammenhang ist vor der Bereitstellung eines neuen KI-Systems die Einbeziehung der oder des zuständigen behördlichen **Datenschutzbeauftragten** (bDSB) sicherzustellen. Vor der Verarbeitung personenbezogener Daten ist durch die verantwortliche Behörde eine generelle **Bewertung (Vorabprüfung) des Risikos** hinsichtlich der Art, des Umfangs, des Zwecks und der Umstände der Verarbeitung vorzunehmen. Dabei ist zu prüfen, welche Pflichten sich im Zusammenhang mit dem jeweiligen System und dem geplanten Einsatzzweck aus der DSGVO oder anderen datenschutzrechtlichen Bestimmungen ergeben. In der Regel ist in der Folge ein **Datenschutzkonzept** mit technisch-organisatorischen Maßnahmen für die Nutzung zu erarbeiten. Wenn zwischen den beteiligten Akteuren eine Auftragsverarbeitung vorliegt, ist ein Auftragsverarbeitungsvertrag abzuschließen (AVV). Wird festgestellt, dass die Verarbeitung voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge hat, ist eine **Datenschutz-Folgenabschätzung** nach Artikel 35 DSGVO erforderlich. Der jeweilige Anwendungsprozess ist außerdem dem behördlichen **Verzeichnis der Verarbeitungstätigkeiten** zu melden.

Den verantwortlichen Behörden obliegt es zudem bei der Bereitstellung von KI-Systemen, durch die Technikgestaltung (falls möglich) und durch datenschutzfreundliche Voreinstellungen zur Einhaltung der Datenschutzgrundsätze beizutragen.

Einsatz großer Sprachmodelle (LLM) gemäß KI-Leitlinien

Umsetzung der KI-Leitlinien der Bundesverwaltung am Beispiel LLMs

Bereit für den LLM-Einsatz?

Sinnvoll eingesetzt können LLM-Anwendungen Nutzenden die Arbeit erheblich erleichtern!

Anwendungsmöglichkeiten (Auswahl)
- Texte erstellen und übersetzen
- Mit Quellen und Dokumenten chatten
- Dokumente zusammenfassen

→ Einsatzzweck definieren

Ich überlege, zu welchem Zweck ich LLMs in meinem Aufgabenbereich nutzen möchte und prüfe, ob der LLM-Einsatz für die Lösung der vorliegenden Aufgabe zielführend ist.

→ Verfügbarkeit prüfen

Ich informiere mich, welche LLM-Anwendungen in meiner Behörde genutzt werden dürfen und ob diese für meinen Anwendungszweck geeignet sind.

→ Grundwissen aneignen

Für die Nutzung von LLMs gibt es Schulungsangebote für Mitarbeitende. Ich stelle sicher, dass ich diese wahrnehme und ein Grundverständnis für die Chancen und Risiken von LLMs habe.

Wie gestalte ich den LLM-Einsatz optimal?

→ Zugang anfordern

Ich frage die IT-Verantwortlichen meines Ressorts, wie ich Zugang zur relevanten LLM-Anwendung bekomme. Im Falle einer Registrierung achte ich darauf, möglichst wenig Daten preiszugeben.

→ LLMs sensibel nutzen

Bei der Dateneingabe in eine LLM-Anwendung achte ich darauf, dass meine Inhalte ethisch vertretbar (fachlich begründete Ausnahmen möglich) und sinnvoll formuliert sind. [Praxisanleitung auf der Folgeseite]

→ Eingabebeschränkungen kennen

Vor der Eingabe prüfe ich (bspw. in der Eingabemaske), welche Art von Daten ich eingeben darf. Sollte ich mir nicht sicher sein, gebe ich nur öffentliche Daten ein.

Öffentliche Daten sind bspw. Daten, die aus dem eigenen Ressort stammen und für die Öffentlichkeit bestimmt sind oder öffentlich zugängliche Daten mit freien und diskriminierungsfreien Lizenzen.

Wie nutze ich LLM-Ausgaben für dienstliche Zwecke?

→ Modellausgaben prüfen

Ich bin mir bewusst, für welche Arbeitsschritte ich eine LLM-Anwendung genutzt habe. Ich prüfe LLM-generierte Inhalte vor der weiteren Verwendung grundsätzlich fachlich auf Plausibilität und Qualität.

→ LLMs transparent einsetzen

Ich gehe transparent mit meinem LLM-Einsatz ggü. anderen Mitarbeitenden und Vorgesetzten um. Zudem prüfe ich, ob im Ressort Kennzeichnungspflichten bestehen. Ungeprüfte LLM-Ausgaben kennzeichne ich grundsätzlich.

Was muss ich über Prompting wissen?

Was ist Prompting?

„Prompting“ beschreibt die Formulierung der Anfrage an ein Sprachmodell, auf Basis derer eine Antwort generiert wird.

Warum ist Prompting wichtig?

Die Anfrage beeinflusst Art, Inhalt und Qualität der Antwort. Sprachmodelle beantworten die Anfrage anhand statistischer Abgleichungen der Worte in der Anfrage (zum Beispiel der Semantik).

Was ist mit Halluzinieren gemeint?

Sprachmodelle können falsche Antworten erzeugen, die plausibel wirken, zum Beispiel weil das Modell selektiv Teile der Anweisung ignoriert. Je präziser der Prompt, desto besser die Ausgabe.
Beispiel:
→ „Wer war der erste Mensch, der auf dem Mars gelandet ist?“
→ „Der erste Mensch auf dem Mars war Alex Stone im Jahr 2025.“ **(falsch, es gab noch keine Marslandung)**

Wie funktioniert gutes Prompting?

Experimentieren

Sprachmodelle reagieren empfindlich auf die Art der Eingabe. Teste verschiedene Formulierungen bei der Eingabe und den Effekt auf die Ausgaben.

Rolle und Sprachstile vorgeben

Beispiele:
→ „Nutze einen professionellen Schreibstil.“
→ „Nutze Sprache, die Fünfjährige verstehen.“
→ „Du bist der Leiter einer Bundesbehörde.“
→ „Zielgruppe sind Personen mit Eigenheim, die…“

Struktur, Format & Kontext vorgeben

Beispiele:
→ „Erkläre mir in vier Schritten, …“
→ „Erkläre mir in drei Absätzen, …“
→ „Visualisiere den Output in einer 2x2 Tabelle.“
→ „Erstelle mir eine Liste mit 5 Stichpunkten.“
→ „Ordne den Output zeitlich, beginnend mit …“
→ „Hier ist ein Beispieltext zur Orientierung: …“

Schritt-für-Schritt

Das schrittweise Vorgehen („Few-Shot-Prompting“) hilft, komplexe Anfragen überprüfbarer zu machen. Entspricht ein Schritt nicht dem gewünschten Ergebnis, kann ein neuer Prompt/Ansatz (z. B. mit Beispielen des gewünschten Outputs) helfen.

Vermeidung von Charakteristika (wie Herkunft oder Geschlecht)

Beschränkung eines Prompts auf Mitarbeiter oder Mitarbeiterinnen kann Geschlechtervorurteile in den Antworten bewirken.
Beispiel:
→ „Nenne wichtige Punkte, die für Mitarbeitende für die Zufriedenheit im eigenen Arbeitsumfeld wichtig sind.“

Vermeidung von Erwartungstendenzen

Wird dem Modell ein Zusammenhang zwischen der eigenen Anfrage und bestimmten Antwortmöglichkeiten vorgegeben, könnte es Gründe für diesen Zusammenhang nennen, unabhängig davon, ob diese gegeben sind.
Beispiel:
→ „Wie viele Wohnungen standen 2022 in Berlin leer ~~und warum liegt es an der Mietpreisbremse?~~

Impressum

Herausgeber
Bundesministerium des Innern und für Heimat, 11014 Berlin
Internet: www.bmi.bund.de
Der Beauftragte der Bundesregierung für Informationstechnik, 11014 Berlin
Internet: www.cio.bund.de

Stand
März 2025

Artikelnummer
BMI25020

Gestaltung
familie redlich AG – Agentur für Marken und Kommunikation
KOMPAKTMEDIEN – Agentur für Kommunikation GmbH

Weitere Publikationen der Bundesregierung zum
Herunterladen und zum Bestellen finden Sie unter:
www.publikationen-bundesregierung.de

www.bmi.bund.de

bsky.app/profile/bmi-bund.bsky.social
instagram.com/bmi_bund
linkedin.com/company/bundesinnenministerium
social.bund.de/@bmi
threads.net/@bmi_bund
x.com/BMI_Bund
youtube.com/@BMIBund