

Verwendung von generativer KI im BMBF

Die Sicherheitsrichtlinie für die Verwendung von generativer KI im BMBF definiert verbindlich die Rahmenbedingungen für den Einsatz KI-basierter Dienstleistungen. Sie deckt selbstgehostete, behördenspezifische sowie externe Dienste ab, berücksichtigt den gesamten Nutzungszyklus von der Kontoeinrichtung bei einem Anbieter, über die Verwendung der KI bis hin zur Kontolöschung ab und gibt insbesondere vor, welche Informationen damit verarbeitet werden dürfen.

1. VORWORT

Seit Dezember 2022 sind große generative KI-Sprachmodelle, insbesondere ChatGPT des Unternehmens OpenAI, in Zeitungen, Sozialen Medien und anderen Informationsquellen omnipräsent. Vor allem die Ankündigung und Veröffentlichung von Modellen, die teilweise frei verfügbar sind, haben zu einem rasanten Anstieg hinsichtlich der Popularität und der Nutzung von großen KI-Sprachmodellen geführt. So meldeten sich innerhalb von fünf Tagen beispielsweise eine Million Nutzerinnen und Nutzer bei ChatGPT mit einem enorm wachsenden Trend an. Im Januar 2023 wurde daher bereits die Grenze von 100 Millionen Nutzerinnen und Nutzern überschritten. ChatGPT beeindruckt besonders durch die hohe Qualität der von ihr generierten Texte selbst Fachleute.

Der Einsatz von generativer KI kann auch die Effizienz und die Genauigkeit von Geschäftsprozessen innerhalb der Bundesverwaltung verbessern und zu Arbeitsentlastungen oder Vereinfachungen führen. Vor diesem Hintergrund befinden sich derzeit sowohl eigene behördeninterne, als auch behördenübergreifende KI-Projekte im Aufbau.

Gleichzeitig werden intensive Diskussionen über Anwendungsgebiete der neuen Technologie sowie die aus ihr resultierenden Gefahren geführt.

Unter generativer KI wird im Sinne dieser Sicherheitsrichtlinie eine Technologie verstanden, die in der Lage ist, mithilfe generativer Modelle Texte, Bilder oder andere Medien synthetisch zu erzeugen. Dabei greift sie auf vorhandene Trainingsdaten zurück und nutzt den Ansatz des maschinellen Lernens, d. h. der Mustererkennung.

Als für Forschung und Wissenschaft zuständiges Ministerium auf Bundesebene gehört es einerseits zum Aufgabenbereich des BMBF, Entwicklungen im Bereich der Künstlichen Intelligenz genau zu beobachten und zu fördern. Andererseits ist das BMBF als Teil der Bundesverwaltung nach Art. 20 Abs. 3 GG an Recht und Gesetz gebunden und unterliegt daher besonders strengen Anforderungen an die Informationssicherheit und an die Gewährleistung der Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit.

2. ZIELSETZUNG

Ziel der Sicherheitsrichtlinie für die Verwendung von generativer KI im BMBF ist es daher, Mitarbeiterinnen und Mitarbeitern

- den produktiven Einsatz generativer KI,
- die Integration generativer KI in Geschäftsprozesse sowie
- das Sammeln von Erfahrungen mit Sprachmodellen

zu ermöglichen, ohne dabei das Sicherheitsniveau auf ein inakzeptables Maß zu senken. Hierfür werden mit diesem Dokument verbindliche Regelungen aufgestellt. Darüber hinausgehende Vorgaben und Regelungen in anderen Bereichen (z. B. Geheimschutz, Datenschutz) bleiben hiervon unberührt.

3. GELTUNGSBEREICH

Die Sicherheitsrichtlinie für die Verwendung von generativer KI im BMBF gilt für alle Mitarbeiterinnen und Mitarbeiter des BMBF sowie für externes Personal, das vom BMBF zur Erbringung von Dienstleistungen oder zur Erfüllung von Aufträgen beauftragt wurde und weisungsgebunden ist.

Die Sicherheitsrichtlinie findet Anwendung auf

- selbstgehostete Dienste (generative KI, die vom BMBF selbst betrieben und ausschließlich für Mitarbeiterinnen und Mitarbeiter des BMBF bereitgestellt wird),
- Dienste, die von Institutionen innerhalb der Bundesverwaltung für die Bundesverwaltung angeboten werden (generative KI, die von einer Bundesbehörde betrieben und auch anderen Bundesbehörden bereitgestellt wird) sowie
- auf Dienste externer Anbieter (generative KI, die von Institutionen außerhalb der Bundesverwaltung oder Unternehmen betrieben und einem unbestimmten Personenkreis kostenlos oder zahlungspflichtig zur Verfügung gestellt wird).

4. ANFORDERUNGEN

Für die oben genannten Dienste-Kategorien gelten unterschiedliche Anforderungen. Während selbstgehostete Dienste den größten Handlungsspielraum erlauben und Dienste, die von Institutionen innerhalb der Bundesverwaltung für die Bundesverwaltung angeboten werden ein mittleres Schutzniveau erfordern, müssen am anderen Ende für Dienste externer Anbieter strenge Anforderungen gelten, damit die Informationen des BMBF geschützt sind. Vor diesem

Hintergrund werden die Anforderungen in einen allgemeinen Teil, der für alle Dienst-Kategorien gilt sowie Dienste spezifische Teile untergliedert.

4.1. Anforderungen für alle Dienste

KI.GEN.A1 Ethisch und moralisch korrekte Eingaben

Der generativen KI DÜRFEN KEINE ethisch oder moralisch fragwürdigen Fragen oder Aufgabenstellungen übermittelt werden, die diese gezielt insbesondere zu menschenverachtenden Aussagen, Hassreden, Verunglimpfungen, Beleidigungen, Verhetzungen oder Verbreitung von Desinformation veranlasst.

Über Taktiken wie bspw. „Prompt Injection“ oder „Training Data Poisoning“ können auch Nutzende mit nicht-privilegierten Rechten einer generativer KI ein bestimmtes Antwortverhalten antrainieren oder aufzwingen, das eventuell gegen Straftatbestände verstößt oder zumindest eine negative Außenwirkung nach sich zieht. Mit der Vorgabe wird sichergestellt, dass derartige Verhaltensweisen nicht auf das BMBF zurückfallen.

KI.GEN.A2 Plausibilitäts- und Qualitätskontrolle

Antworten und Ergebnisse der generativen KI DÜRFEN NICHT ungeprüft übernommen, sondern MÜSSEN vor einer (weiteren) Verwertung einer Plausibilitäts- und Qualitätskontrolle unterzogen werden. Sofern ein spezifischer genutzter Dienst Anforderungen zu Verwertung oder Plausibilitäts- und Qualitätskontrollen konkretisiert oder diesbezüglich weitergehende Anforderungen stellt, MÜSSEN auch diese beachtet werden.

Werden Informationen aus der Hand Dritter oder sonstiger Externer im BMBF verwertet, MUSS bei diesen vorher erfragt werden, ob diese unter Zuhilfenahme generativer KI erstellt und ebenfalls einer Plausibilitäts- und Qualitätskontrolle unterzogen wurden.

Auftragnehmer/-innen sowie Dienstleister/-innen des BMBF SOLLTEN im Vorfeld darauf hingewiesen werden, mit generativer KI produzierte Arbeitsergebnisse als solche zu kennzeichnen.

Sprachmodelle wie bspw. das von ChatGPT basieren auf einer Musterkennung. Diese funktioniert aber nicht in jedem Fall zuverlässig. Treten beispielsweise Muster oder besondere Fragekonstellationen zum ersten Mal auf oder wurden sie bislang noch nicht trainiert, steigt das Risiko von fehlerhaften Antworten.

KI.GEN.A3 Datenschutzkonforme Verarbeitung

Enthalten die Antworten oder sonstigen Ergebnisse der generativen KI personenbezogene Daten, DÜRFEN diese NUR unter der Beachtung der gesetzlichen und regulatorischen Vorgaben des Datenschutzes (u. a. Erforderlichkeit einer Rechtsgrundlage, Informationspflicht) durch das BMBF verarbeitet werden.

KI.GEN.A4 Dokumentation von Art und Umfang der Nutzung

Um Rechtssicherheit zu schaffen, Informationspflichten nachzukommen und bei Sicherheitsvorfällen das Ausmaß genau bestimmen zu können, MÜSSEN im BMBF Art und Umfang der Nutzung von generativer KI im damit verknüpften Verwaltungsvorgang dokumentiert werden.

KI.GEN.A5 Entscheidungen mit Rechtswirkung

Die generative KI DARF NICHT dazu verwendet werden, Entscheidungen mit Rechtswirkung zu treffen.

KI.GEN.A6 Exklusive dienstliche Nutzung

Im dienstlichen Kontext angelegte Accounts DÜRFEN NICHT zu privaten Zwecken genutzt werden.

4.2. Anforderungen an selbstgehostete Dienste

KI.INT.A1 Erlaubte Informationen und Inhalte bei selbstgehosteten Diensten

Informationen und Inhalte, die an selbstgehostete KI-Dienste (z. B. via API) übertragen oder auf der Anwendungsoberfläche (z. B. über den sog. „Prompt“) eingegeben werden,

- DÜRFEN NUR einen Personenbezug im Sinne des Art. 4 Nr. 1 DSGVO aufweisen, wenn die gesetzlichen und regulatorischen Vorgaben des Datenschutzes (u. a. Erforderlichkeit einer Rechtsgrundlage sowie Informationspflicht der betroffenen Person) eingehalten werden,
- DÜRFEN NICHT als Verschlusssachen (§ 2 VSA) des Geheimhaltungsgrads VS-VERTRAULICH (VS-Vertr.), GEHEIM oder STRENG GEHEIM eingestuft sein und
- MÜSSEN ihren Ursprung direkt im BMBF haben oder es MUSS eine Einwilligung des Urhebers vorliegen. Auf diese Weise werden die Gefahren einer Urheberrechts- oder Persönlichkeitsrechtsverletzung reduziert.

Betriebs-, Geschäfts- und Amtsgeheimnisse sowie Verschlusssachen des Geheimhaltungsgrads VS-NUR FÜR DEN DIENSTGEBRAUCH (VS-NfD) dürfen hingegen mit selbstgehosteten KI-Diensten im Rahmen der bestehenden gesetzlichen und regulativen Vorgaben verarbeitet werden.

4.3. Anforderungen an Dienste der Bundesverwaltung

KI.BVW.A1 Anzeigepflicht beim Informationssicherheitsmanagement

Organisationseinheiten oder Mitarbeiterinnen und Mitarbeiter, die ein Konto zur Nutzung eines KI-Dienstes von der Bundesverwaltung für die Bundesverwaltung erstellen möchten, MÜSSEN dieses sowie den verwendeten Dienst dem Informationssicherheitsmanagement im BMBF (vorzugsweise [per E-Mail](#)) anzeigen.

Dies ermöglicht dem Informationssicherheitsmanagement ein schnelles Reagieren im Falle eines Datenlecks und eine genauere Risikoeinschätzung.

Darüber hinaus MUSS die dienstliche Erforderlichkeit, insbesondere bei der Erstellung eines personenbezogenen Kontos, zur Nutzung gegeben sein.

KI.BVW.A2 Pflicht zur Löschung des Benutzerkontos

Wird der Account für einen KI-Dienst von der Bundesverwaltung für die Bundesverwaltung nicht länger benötigt, MUSS er unverzüglich gelöscht werden. Die Löschung MUSS dem Informationssicherheitsmanagement im BMBF (vorzugsweise per [E-Mail](#)) angezeigt werden. KI.BVW.A3 Erlaubte Informationen und Inhalte bei Diensten der Bundesverwaltung

Informationen und Inhalte, die an KI-Dienste der Bundesverwaltung (z. B. via API) übertragen oder auf der Anwendungsoberfläche (z. B. über den sog. „Prompt“) eingegeben werden,

- DÜRFEN NUR einen Personenbezug im Sinne des Art. 4 Nr. 1 DSGVO aufweisen, wenn die gesetzlichen und regulatorischen Vorgaben des Datenschutzes (u. a. Erforderlichkeit einer Rechtsgrundlage sowie Informationspflicht der betroffenen Person) eingehalten werden,
- DÜRFEN NICHT als Verschlusssachen (§ 2 VSA) eingestuft sein und
- MÜSSEN ihren Ursprung direkt im BMBF haben oder es MUSS eine Einwilligung des Urhebers vorliegen.

Betriebs-, Geschäfts- und Amtsgeheimnisse dürfen hingegen mit KI-Diensten der Bundesverwaltung im Rahmen der bestehenden gesetzlichen und regulativen Vorgaben verarbeitet werden.

4.4. Anforderungen an Dienste externer Anbieter

KI.EXT.A1 Anzeigepflicht beim Informationssicherheitsmanagement

Organisationseinheiten, die ein dienstliches Konto zur Nutzung generativer KI erstellen, MÜSSEN dieses sowie den verwendeten Dienst dem Informationssicherheitsmanagement im BMBF (vorzugsweise [per E-Mail](#)) anzeigen.

Dies ermöglicht dem Informationssicherheitsmanagement ein schnelles Reagieren im Falle eines Datenlecks und eine genauere Risikoeinschätzung.

Darüber hinaus MUSS die dienstliche Erforderlichkeit, insbesondere bei der Erstellung eines personenbezogenen Kontos, zur Nutzung gegeben sein.

KI.EXT.A2 Berücksichtigung der Nutzungsbedingungen

Um unerwünschte rechtliche Konsequenzen zu vermeiden und die Sicherheit der mit KI verarbeiteten Daten zu gewährleisten, MÜSSEN vor der Verwendung generativer KI die Nutzungsbedingungen des Dienstes gesichtet und auf Vereinbarkeit mit dem geplanten Einsatzzweck sowie dieser Sicherheitsrichtlinie überprüft werden. Widersprechen die Nutzungsbedingungen dem Einsatzzweck oder dieser Sicherheitsrichtlinie, MUSS das Informationssicherheitsmanagement darüber informiert werden. Bis zur finalen Entscheidung durch das Informationssicherheitsmanagement zum weiteren Vorgehen DARF der betroffene Dienst NICHT genutzt werden.

KI.EXT.A3 Benutzerkonto und verknüpfte E-Mail-Adresse

Um die Preisgabe personenbezogener Daten und den digitalen Fußabdruck möglichst gering zu halten, DÜRFEN für dienstliche Benutzerkonten NUR Organisationspostfächer als E-Mail-Adresse oder allgemein gehaltene Kontokennungen ohne Personenbezug angegeben werden.

Das Organisations-/Funktionskonto SOLLTE nach Möglichkeit von mehreren Personen genutzt werden, um eine Profilbildung zu erschweren.

Werden im Rahmen der Anmeldung Name und Vorname abgefragt, SOLLTE als Name "**BMBF**" und als Vorname die jeweilige Organisationseinheit (z. B. "**ISM**") angegeben werden.

KI.EXT.A4 Verifizierung über dienstliche Mobilfunknummer

Soweit zur Verifizierung der Account-Informationen eine telefonische Rufnummer erforderlich ist, DARF bei der Anmeldung NUR eine dienstliche Rufnummer angegeben werden. Die verwendete dienstliche Rufnummer DARF NICHT auf eine Person registriert sein, was bei dienstlich bereitgestellten SIM-Karten regelmäßig der Fall ist. KI.EXT.A5 Aktivieren von Funktionen zur Datensparsamkeit und -minimierung

Vor der ersten Nutzung des Accounts MUSS nach Möglichkeit in den Einstellungen der Chatverlauf deaktiviert und der Weiternutzung der eingegebenen Daten zu Trainingszwecken widersprochen werden.

Sollte es zu einem Datenleck beim Anbieter der genutzten generativen KI kommen, ist damit sichergestellt, dass Chatverläufe unberechtigten Dritten nicht angezeigt oder sonst zur Kenntnis gebracht werden können. Darüber hinaus wird damit die dauerhafte Speicherung von Anfragen, die Informationen des BMBF beinhalten, in Trainingsmodellen und als Material zur Beantwortung von Fragen Dritter über das BMBF unterbunden.

KI.EXT.A6 Erlaubte Informationen und Inhalte

Informationen und Inhalte, die zu Anbietern generativer KI (z. B. via API) übertragen oder auf der Anwendungsoberfläche (z. b. über den sog. „Prompt“) eingegeben werden,

- MÜSSEN ihren Ursprung direkt im BMBF haben oder es MUSS eine Einwilligung des Urhebers vorliegen;
- DÜRFEN KEINE personenbezogenen Daten und
- KEINE Betriebs-, Geschäfts- und Amtsgeheimnisse enthalten und
- DÜRFEN NICHT als Verschlusssachen (§ 2 VSA) eingestuft sein.

Da das BMBF als nutzende Institution generativer KI dem betreibenden Dienstleistungsunternehmen keine Vorgaben zum gewünschten Sicherheitsniveau machen kann, sondern von dessen technischen und organisatorischen Maßnahmen abhängig ist und diese nicht umfassend überprüfen kann, dürfen nur unkritische Informationen mit dem Dienst bzw. der generativen KI verarbeitet werden. Dies gilt umso mehr, als dass es beispielsweise bei ChatGPT in der Vergangenheit bereits zu einem Datenleck gekommen ist.

KI.EXT.A7 Pflicht zur Löschung des Benutzerkontos

Wird der Account bei einem Dienstanbieter für generative KI nicht länger benötigt, MUSS er unverzüglich gelöscht werden. Die Löschung MUSS dem Informationssicherheitsmanagement im BMBF (vorzugsweise per [E-Mail](#)) angezeigt werden.

Damit wird dem Grundsatz der Zweckbindung und Datenminimierung Genüge getan.

KI.EXTA8 Übernahme von Texten und sonstigen Ergebnissen

Soweit es die Nutzungsbedingungen nicht ausdrücklich erlauben, DÜRFEN Antworten und sonstige Ergebnisse der eingesetzten generativen KI NICHT im direkten Wortlaut bzw. ohne Anpassungen übernommen, sondern MÜSSEN sprachlich oder anderweitig medial abgeändert werden.

Da die Frage nach der Urheberschaft von KI-generierten Inhalten und Produkten (das betreibende Unternehmen oder die fragestellende Person) noch nicht abschließend geklärt ist, werden damit etwaige Rechtsverletzungen von vornherein vermieden.

Dies bezieht sich ebenso auf die Verwertung von Trainingsmaterial und frei zugänglichen Informationen, auf die von der generativen KI für die Beantwortung von Fragen zurückgegriffen wird.

[Nach oben](#)

ZURÜCK ...

[... zur Kapitelauswahl](#)

KAPITELINHALTE

[Vorwort](#)

[Zielsetzung](#)

[Verwendete Modalverben](#)

[Geltungsbereich](#)

[Anforderungen für alle Dienste](#)

[Anforderungen an selbstgehostete Dienste](#)

[Anforderungen an Dienste der Bundesverwaltung](#)

.....