Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

1. Dezember 2023 V1.0[1]

Merkblatt zur Verwendung von generativen KI-Werkzeugen in der Bundesverwaltung

Aktenzeichen: 822.1-1/8/5/1

Was sind generative KI-Werkzeuge?

Im Internet verfügbare Werkzeuge mit generativer künstlicher Intelligenz (KI)[2] – zum Beispiel ChatGPT von OpenAI, Copilot von Microsoft, Bard von Google, Grok von X und zahlreiche mehr – vereinfachen eine Reihe von Aufgaben, die auch in der Verwaltung zum Arbeitsinhalt vieler Mitarbeitenden gehören. Sie ermöglichen es den Nutzenden, beispielsweise Text[3] einzugeben und um eine Stellungnahme zum Inhalt zu bitten oder das Werkzeug aufzufordern, eine Textausgabe zu einem bestimmten Thema zu erstellen. Lange Textbeiträge können so zusammengefasst, eine Antwort von bestimmter Länge auf eine Frage erzeugt oder Code für eine beschriebene Funktion generiert werden.

Diese Werkzeuge sind nicht «intelligent»; sie berechnen z.B. bei der Textgenerierung lediglich die statistische Wahrscheinlichkeit der Wortteilfolge – sie sind also *next token prediction systems* – liefern aber dennoch oft erstaunliche Ergebnisse. Sie werden mit grossen Datenmengen gefüttert, deren Quellen meistens nicht offengelegt sind. Die darauf berechneten Wahrscheinlichkeiten können daher veraltet, irreführend, diskriminierend oder schlicht falsch sein. Ebenso dienen die Eingaben (sog. *Prompts* oder Eingabeaufforderungen) unter Umständen dem weiteren Training des KI-Systems und werden in der Regel auch ausserhalb der Schweiz gespeichert; sie können also in andere Unterhaltungen einfliessen.

Verantwortungsvolles Experimentieren? Ja!

Generative KI-Werkzeuge können Sie bei ihrer täglichen Verwaltungstätigkeit unterstützen. Probieren Sie es aus, lernen Sie dazu! Mit etwas Kreativität tragen Sie so zu einer innovativen Verwaltung bei. Gehen Sie dabei aber vorsichtig vor und beachten Sie die geltenden Vorgaben.

➔ Mögliche Einsatzbereiche: Lassen Sie sich längere öffentlich verfügbare Texte zusammenfassen, holen Sie sich Tipps für die Struktur Ihrer nächsten Präsentation, lassen Sie sich von Programmier-Code-Vorschlägen für Ihre Arbeit inspirieren oder lesen Sie sich schnell und spielerisch in ein neues Thema ein, in dem Sie mit dem Werkzeug in einen Austausch treten – finden Sie heraus, wo es Sie optimal unterstützen kann.

➔ Benutzen Sie zur Registration für berufliche Zwecke auch Ihre berufliche E-Mail-Adresse. Wählen Sie ein starkes Passwort und nutzen Sie dieses nur für diesen Dienst.

Testen Sie die Antworten des generativen KI-Werkzeugs durch unterschiedliche Eingaben und finden Sie so zu einer zielführenden Fragetechnik.

---

[1] Dieses Merkblatt wird regelmässig einer Überprüfung unterzogen, um neue Entwicklungen und ein besseres Verständnis der Anwendungsfälle von generativen KI-Werkzeugen in der Bundesverwaltung zu berücksichtigen.

[2] Zur allgemeinen Terminologie siehe https://cnai.swiss/dienstleistungen/terminologie/ und https://www.bfs.admin.ch/bfs/de/home/dscc/dscc.assetdetail.29325686.html (Kapitel 3.3) - «Generative KI» ist ein weit gefasster Begriff, der sich auf KI-Systeme bezieht, die auf grosse Mengen von Daten aus der realen Welt trainiert werden, um selbst Daten zu generieren (z.B. Texte, Bilder, Tonaufnahmen, Videos, Simulationen, Codes). Sie sind oft multimodal, z.B. mit Eingaben und/oder Ausgaben in einer oder mehreren Modalitäten (z.B. Text, Bild, Video).

[3] Daneben gibt es auch Anwendungen, die Bilder, Tonaufnahmen, Videos, Simulationen oder Codes generieren können.

Bestehende Vorgaben verletzen? Nein!

➔ Vorsicht bei der Eingabe:

Geben Sie niemals persönliche Daten oder sensible Informationen in diese Werkzeuge ein!

- keine Eingabe von als intern, vertraulich oder geheim klassifizierten[4] Informationen;
- keine Eingabe von Texten, die zwar nicht klassifiziert sind, aber sensible Informationen enthalten, etwa weil sie durch eine Geheimhaltungspflicht (Amtsgeheimnisse, besondere (Berufs-) Geheimnisse, vertraglich ausdrücklich geschützte Informationen)[5] geschützt sind;
  Vorsicht auch bei der Eingabe von Bildern, privaten Fotos, Tonaufnahmen, Videos, Simulationen und Code;
- keine Eingabe von Personendaten[6] jeglicher Art. Achten Sie bei anonymisierten oder pseudonymisierten Eingaben darauf, dass nicht aufgrund von zusätzlichen Informationen doch Rückschlüsse auf die Betroffenen gezogen werden können (etwa indem zwar ein Name abgeändert wird, aber aufgrund der Angabe des Geburtsdatums, des Geschlechts und des Wohnquartiers die fragliche Person relativ einfach in Erfahrung gebracht werden kann).

Bei Unklarheit ob der Qualifikation der zu verwendenden Informationen und Daten, verzichten Sie auf deren Eingabe und die Benutzung der generativen KI-Werkzeuge. Die Verwendung bereits öffentlich (im Internet) publizierter Informationen, wie Open Government Data (OGD)[7], ist unproblematisch.

➔ Vorsicht bei der (Weiter-)Verwendung der Ergebnisse:

Generative KI-Werkzeuge liefern Ergebnisse unterschiedlicher Qualität. Überprüfen Sie die Ergebnisse der Werkzeuge in jedem Fall kritisch auf Richtigkeit und Vollständigkeit und vergleichen Sie diese mit anderen Quellen. **Die Verantwortung für das verwendete Ergebnis bleibt bei Ihnen, sie kann nicht an die generativen KI-Werkzeuge delegiert werden.**

Treffen Sie Entscheidungen, welche auf Ergebnisse von generativen KI-Systemen beruhen, so müssen diese von Ihnen jederzeit begründbar bleiben. Betroffene Personen haben ein Anrecht auf Information über eine Entscheidung, die ausschliesslich auf einer automatisierten Bearbeitung beruht und die für sie mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt (automatisierte Einzelentscheidung)[8].

Weisen Sie gegebenenfalls transparent auf die Nutzung von KI-Systemen hin. Beachten Sie die Nutzungshinweise der gewählten generativen KI, da diese zum Teil auch Hinweise auf die Weiterverwendung von Inhalten (Urheberrechte) beinhalten.

➔ Vorsicht bei der Sicherheit:

Gewisse IKT-Anwendungen (und insbesondere E-Mail-Anmeldedienste) sind aufgrund von Sicherheitsanforderungen innerhalb der Bundesverwaltung gesperrt und die Anzeige vom Internetbrowser abhängig. **Halten Sie die Bestimmungen zur Informatik- und Cybersicherheit jederzeit ein.**

---

4 Siehe Art. 13 ISG (SR 128) und Art. 18 ff. ISV (SR 128.1)

5 Siehe für das Amtsgeheimnis Art. 320 StGB (SR 311.0); Berufsgeheimnisse sind z.B. das Arztgeheimnis oder Steuergeheimnis; vertragliche Non-Disclosure-Agreements sind ebenfalls zu beachten.

6 Alle Angaben, die sich auf eine bestimmte oder bestimmbare natürliche Person beziehen (Art. 5 Bst. a DSG [SR 235.1]); siehe dazu auch die Kurzmeldung des EDÖB.

7 Siehe Art. 10 EMBAG (BBl 2023 787)

8 Art. 21 DSG (SR 235.1)

Beispielübersicht Einsatzkategorien

Generatives KI-Werkzeug erlaubt 👍	Generatives KI-Werkzeug untersagt 🚫
Veröffentlichte Texte (z.B. Berichte) zusammenfassen lassen	Dokumente des Mitberichtsverfahrens zusammenfassen lassen (→ vertraulich / geheim)
Einstiegspforte für ein Thema (analog Google / Wikipedia)	CVs übersetzen lassen (→ Personendaten)
Unterstützung beim Formulieren von Foliensätzen	Konkrete Anfrage von Herr Hans Muster unverändert eingeben (→ Personendaten)
Inspiration für Code-Generierung	Antworten *copy/paste*-artig verwenden (→ Kontrolle)
Bilder für Präsentationen erstellen lassen	bestehenden Software-Code eingeben zum Debugging (→ Urherberrechtsverletzung)

Bei Fragen:

- Zu KI in der Bundesverwaltung: Arbeitsgruppe KI im Kompetenznetzwerk KI (CNAI)
- Zu Informationssicherheit und Datenschutz: die ISBOs und DSBOs Ihrer Verwaltungseinheit
- Für Anwendersupport: IT-Abteilung ihres Vertrauens oder RoBIT (KI-ChatBot des BIT)
- Für konkrete Dienstleistungen im Bereich Datenwissenschaft und KI: DSCC

Weitere Hinweise:

Leitlinien für den Umgang mit KI in der Bundesverwaltung

Die sieben Leitlinien für den Umgang mit KI in der Bundesverwaltung gelten weiterhin fort: Den Menschen in den Mittelpunkt stellen, Rahmenbedingungen für Entwicklung und Anwendung von KI gewährleisten, Transparenz, Nachvollziehbarkeit und Erklärbarkeit einfordern, Verantwortlichkeit klar definieren, Sicherheit gewährleisten, aktive Mitgestaltung der Gouvernanz von KI vorantreiben und dabei alle relevanten nationalen und internationalen Akteuren einbeziehen, bleiben wichtige Handlungsmaxime für die Bundesverwaltung im Umgang mit KI.

Verhaltenskodex des Bundes für menschenzentrierte und vertrauenswürdige Datenwissenschaft (und KI)

Durch den Verhaltenskodex werden die Verwaltungseinheiten des Bundes im Sinne einer Orientierungshilfe mittels praktischer Erläuterungen zum einen für die in der Datenwissenschaftsstrategie des Bundes definierten Grundprinzipien einer menschenzentrierten und vertrauenswürdigen Datenwissenschaft (und KI) sensibilisiert, und zum anderen zu deren Umsetzung im Arbeitsalltag befähigt.
Die Grundprinzipien lauten: Daten- und Informationsschutz, Informationssicherheit, Datensicherheit Datengouvernanz, Nichtdiskriminierung, Erklärbarkeit, Nachvollziehbarkeit, Transparenz, Reproduzierbarkeit, Neutralität, Objektivität und ethischer Umgang mit Daten und Ergebnissen.

Das Merkblatt ist in der Arbeitsgruppe «KI in der Bundesverwaltung» im CNAI unter Mitwirkung von Vertreterinnen und Vertretern aus allen Departementen und der BK entstanden.