



Freiburg, 18. November 2024

generativer KI für administrative Aufgaben

—

Richtlinie über den Einsatz

Da der Einsatz von generativer künstlicher Intelligenz in der kantonalen Verwaltung noch in Anfängen steckt, muss diese Richtlinie im Laufe der Zeit regelmässig überprüft und an die Praxis angepasst werden.

Bei der Verwendung Tools der generativen KI für administrative Aufgaben ist Folgendes zu beachten:

I. Definition

Der Begriff Künstliche Intelligenz (KI) bezeichnet die breite Disziplin der Entwicklung von Software, die menschliche Intelligenz simulieren kann. Genauer gesagt bezieht sich KI auf Technologien und Algorithmen, die eingesetzt werden, um Systemen beizubringen, intellektuelle Aufgaben auszuführen.

Generative künstliche Intelligenz (GenKI) ist ein Teilbereich der künstlichen Intelligenz (KI), der auf die Erstellung neuer Inhalte ausgerichtet ist, z. B. Bilder, Texte, Videos, Musik usw. Der Begriff «generativ» bezieht sich auf die Fähigkeit dieser Systeme, neue Daten und Informationen zu erzeugen, die vorhandenen Daten und Informationen ähneln.

II. Ziel und Geltungsbereich

Diese Richtlinie bietet den Mitarbeitenden der kantonalen Verwaltung eine Orientierung für den sicheren und mit dem Schweizer Recht konformen Einsatz der KI. Sie gilt für die Mitarbeitenden aller Direktionen und Anstalten des Staates Freiburg gemäss dem Geltungsbereich der Verordnung des Staatsrats über die Governance der Digitalisierung und der Informationssysteme des Staates, die im Rahmen ihrer beruflichen Tätigkeit generative KI-Tools verwenden und einsetzen.

III. Welche Tools der generativen KI sind betroffen?

Die von dieser Richtlinie betroffenen Tools der generativen KI umfassen alle Anwendungen, die administrative Such- und Schreibaufgaben unterstützen. Dazu gehören z. B. Textgeneratoren wie Copilot, Übersetzungsprogramme wie DeepL oder Microsoft Translator oder Bildgeneratoren wie DALL-E. Diese Tools analysieren die von den Benutzerinnen und Benutzern gemachten Eingaben und generieren auf dieser Grundlage Ausgaben wie Texte, Bilder oder Programmcode.

IV. Nutzung von Tools der generativen KI

Der Fantasie sind bei den Anwendungen der generativen KI kaum Grenzen gesetzt. Generative KI-Tools können auf vielfältige Weise als Unterstützung bei der täglichen Arbeit eingesetzt werden, z. B.:

- > für das Verfassen von öffentlich zugänglichen Informationszusammenfassungen;
- > für das Brainstorming zu einem Thema;
- > zur Redaktionshilfe für bereits bestehende Texte, E-Mails; ohne persönliche Daten oder vertrauliche Inhalte;
- > für Übersetzungen;

- > für die Zusammenstellung von Folienserien oder Zusammenfassungen von Videos, die im Internet verfügbar sind (z. B.: Medienkonferenzen auf Youtube).

Sie sollten nicht verwendet werden:

- > im Zusammenhang mit der Bearbeitung eines persönlichen Dossiers einer Bürgerin oder eines Bürgers;
- > für die Vorbereitung rechtsverbindlicher Entscheide;
- > bei der Erstellung von Dokumenten auf der Grundlage von Informationen, deren Offenlegung dem Staat schaden könnte.

Generell ist davon auszugehen, dass der Output aus den entsprechenden Tools je nach Anwendungsfall von der Benutzerin oder dem Benutzer kritisch bewertet und überarbeitet werden muss.

V. Einschränkungen und Datenschutz

Bei der Verwendung von KI-Tools dürfen keine heiklen, vertraulichen oder geheimen Informationen oder persönliche Daten, einschliesslich pseudonymisierter Daten eingegeben werden. Selbst anonymisierte Daten können unter bestimmten Umständen Rückschlüsse auf Personen zulassen und sollten daher sehr restriktiv analysiert und verwendet werden. Der Einsatz von KI darf unter keinen Umständen dem Ansehen des Staates schaden.

VI. Gesetzliche Anforderungen

Der Einsatz von KI-Tools unterliegt der Bundes- und Kantonsgesetzgebung, insbesondere was Vertraulichkeit, Informationssicherheit und Datenschutz angeht. Diese umfassen unter anderem (in nicht erschöpfender Form):

- > den Schutz personenbezogener Daten gemäss Datenschutzgesetz (Bundesgesetz über den Datenschutz, DSG, SR 235.1 und kantonales Datenschutzgesetz, DSchG, SGF 17.1);
- > die Einhaltung des Amts- oder Berufsgeheimnisses (Gesetz über das Staatspersonal, SGF 122.70.1);
- > die Einhaltung der fachspezifischen Vertraulichkeitsregelungen;
- > das Urheberrecht und das Recht auf geistiges Eigentum.

VII. Haftung und Überprüfung der Ergebnisse

Die Mitarbeitenden selbst sind für die Eingabe und Nutzung von KI-Tools und deren Ergebnisse sowie für alle rechtlichen Konsequenzen, die sich daraus ergeben, verantwortlich. Die Genauigkeit und Vollständigkeit aller erzeugten Ergebnisse und Quellen müssen vor ihrer Verwendung oder Verbreitung kritisch überprüft werden, etwa durch das 4-Augen-Prinzip. Ethisch fragwürdige oder voreingenommene Ergebnisse sollten identifiziert und ihre Wiederverwendung vermieden werden. Würde, Ethik und das öffentliche Interesse müssen immer gewährleistet sein.

Darüber hinaus ist zu beachten, dass in Tools der generativen KI zwei Phänomene auftreten können, welche die Überprüfung durch die Benutzenden so wichtig machen, insbesondere:

Bias: Der vom Modell verwendete Datensatz ist in einen räumlichen und historischen Kontext eingebettet und repräsentiert nicht immer die gesamte erwartete Vielfalt und Neutralität. Eine BIAS (Verzerrung) entsteht durch eine Überrepräsentation von bestimmten Trainingsdaten (westliche oder östliche Daten, Geschlecht, Ethnie usw.). Da die KI die Daten widerspiegelt, kann dies die mithilfe der KI ermittelten Ergebnisse verfälschen, die dann von der Benutzerin oder dem Benutzer korrigiert werden müssen. Jede Person muss vor Diskriminierung oder Stigmatisierung geschützt werden. Der Einsatz von KI-Tools darf diesem Grundrecht nicht zuwiderlaufen, und bei der Überprüfung der Ergebnisse muss dies berücksichtigt werden.

Halluzinationen: Es handelt sich um ein von einer generativen KI überzeugend formuliertes Ergebnis, das durch die Trainingsdaten des Modells nicht gerechtfertigt ist und objektiv falsch ist. Solche Ergebnisse müssen vernichtet werden und man muss auf eine andere Weise zu einem Ergebnis zu kommen versuchen.

VIII. Kennzeichnung und Transparenz

Der Einsatz von generativen KI-Tools muss transparent erfolgen. Ergebnisse, die ohne wesentliche Überarbeitung übernommen werden, sind entsprechend zu kennzeichnen. In einem Hinweis ist auf transparente Weise der verwendete generative KI-Generator, die Version desselben und den Umfang von dessen Verwendung anzugeben.

Die 7 Regeln in Kürze:

1. Geben Sie unter dem Aspekt des Berufsgeheimnisses keine heiklen, vertraulichen oder geheimen Daten in eine generatives KI-Tool ein.
2. Geben Sie keine persönlichen Daten in KI-Tools ein.
3. Sind Sie sich der Verantwortung für die Nutzung von generativer KI bewusst: Wer sie nutzt, ist allein für ihren Inhalt und die Einhaltung der gesetzlichen Normen verantwortlich.
4. Überarbeiten Sie die von der KI erzeugten Ergebnisse kritisch und überarbeiten Sie sie, um Verzerrungen und unvollständige oder falsche Informationen zu vermeiden.
5. Überprüfen Sie die Quellen der Informationen.
6. Machen Sie den Einsatz von KI transparent (auch wenn er nur teilweise erfolgte).
7. Schaden Sie mit dem Einsatz von KI dem Image des Staates Freiburg nicht.

Verabschiedet in der Sitzung des Staatsrates vom 18. November 2024