

Kompass für den dienstlichen Umgang mit generativer Künstlicher Intelligenz (KI)

Version 2.0, 10. April 2024

Erstellt vom „AI-Kompetenznetzwerk“ (Leitung: MD-OS)



1. Kompass für den dienstlichen Umgang mit generativer künstlicher Intelligenz (KI)

Generative KI ist eine Technologie, die über eine Eingabe (engl. Prompt) aus vielen Daten verschiedene Arten von Inhalten wie Texte, Bilder, Videos und Audios erstellen kann. Obwohl KI-Technologien große Möglichkeiten bieten, bringen sie auch Herausforderungen in der Anwendung und im Umgang mit den Ergebnissen mit sich.

Im Sinne des Digitalen Humanismus ist es unser Ziel, Wien zur „Digitalisierungs-Hauptstadt“ zu machen, wobei der Mensch im Mittelpunkt steht. Wir wollen bewusst mit generativer KI umgehen und die Chancen, die sie bietet, aktiv nutzen. Diese Unterlage soll uns als Mitarbeiter*innen der Wiener Stadtverwaltung dabei helfen, generative KI-Werkzeuge¹ verantwortungsvoll einzusetzen.

Dabei liegt besonderes Augenmerk darauf, dass diese neuen Arbeitswerkzeuge Mitarbeiter*innen bzw. Teams unterstützen und ihre Arbeitsabläufe optimieren. Durch die gezielte Implementierung von KI-Technologien streben wir danach, Innovation zu fördern, die menschliche Expertise zu ergänzen und die Mitarbeiter*innen in die Lage zu versetzen, sich auf anspruchsvollere Aufgaben zu konzentrieren, die kreatives Denken und zwischenmenschliche Fähigkeiten erfordern.

Wir verzichten vorerst darauf, eine Richtlinie mit detaillierten Vorgaben zu erstellen, da sich die Entwicklung von KI dynamisch gestaltet und in den kommenden Monaten und Jahren weitere Innovationen zu erwarten sind. Außerdem stehen noch rechtliche Regelungen aus, die eine weitere Anpassung notwendig machen könnten. Dieser Kompass für den dienstlichen Umgang mit generativer KI wird regelmäßig überprüft und bei Bedarf aktualisiert.

1.1. Dienstlicher Umgang mit generativer KI

Die Nutzung von im Internet verfügbaren generativen KI-Werkzeugen für den dienstlichen Gebrauch ist erlaubt und erscheint im Sinne einer modernen und effizienten Stadtverwaltung unumgänglich. Die bereits geltenden Rahmenbedingungen dürfen dabei nicht außer Acht gelassen werden.²

Es ist wichtig, dass wir als Mitarbeiter*innen eigenverantwortlich handeln und diese Technologien mit Erfahrung sowie Haus- und Sachverstand einsetzen. Dies sollte durch Abstimmung im Team

¹ Wie z. B. der auf einem Sprachmodell basierte „ChatGPT“ oder der Bildgenerator „Stable Diffusion“ etc.

² Erlässe wie z. B. [MD-OS-338293/2019 „Internet und elektronische Kommunikation: offizielle Dienststellen-Postfächer“](#) oder [MDK – 2008102-2022-1 „Datenschutz im Magistrat der Stadt Wien“](#) oder [MD-OS 1416182-2023 „Sicherheit in der Informations- und Kommunikationstechnologie“](#) werden in Erinnerung gerufen. Zudem wird auf das [Handbuch „Das Erscheinungsbild der Stadt Wien“](#) und den [„Leitfaden für geschlechtergerechtes Formulieren und diskriminierungsfreie Bildsprache“](#) hingewiesen.

und mit den Führungskräften erfolgen, um gemeinsam geeignete Anwendungsbereiche zu identifizieren, denn nicht in jedem Arbeitsbereich eignet sich der Einsatz von KI-Werkzeuge gleichermaßen.

Dabei müssen wir – wie immer in unserem Arbeitsleben – folgende Grundsätze beachten:

- a. Rechtmäßigkeit (z. B. Datenschutz, Urheberrecht),
- b. Sparsamkeit, Wirtschaftlichkeit (Kosten-Nutzen-Verhältnis), Zweckmäßigkeit (Förderung von Innovation, Verbesserung der Arbeitsleistung),
- c. Berücksichtigung weiterer rechtlicher und ethischer Standards.

Auf diese Weise können wir einen sinnvollen Einsatz dieser Technologien zur wirkungsvollen Unterstützung in unserem Arbeitsbereich finden.

1.2. Was bedeutet dieses Grundprinzip, wenn ich selbst KI-Werkzeuge dienstlich anwende?

1.2.1. Der Mensch trägt die Verantwortung bei der Verwendung von KI-generierten Inhalten.

Beim Einsatz von generativen KI-Werkzeugen ist es unerlässlich, die Ergebnisse kritisch zu hinterfragen und auf inhaltliche Richtigkeit sowie ethische Aspekte zu prüfen (z. B. Gleichbehandlung, Antidiskriminierung). Dies ist besonders wichtig, da textbasierte KI-Werkzeuge hauptsächlich Sprachmodelle (und nicht „Wissensbibliotheken“) sind und sich auch oft Inhalte ausdenken („halluzinieren“). Die Verantwortung für die weiterverwendeten Inhalte liegt stets bei der*dem menschlichen Anwender*in.

1.2.2. Generative KI kann in der öffentlichen Verwaltung dabei helfen, Arbeitsprozesse zu unterstützen, zu vereinfachen und zu beschleunigen.

Die Qualität der Ergebnisse von generativen KI-Werkzeugen hängt von der Qualität der Eingaben der Anwender*innen (sogenannte „Prompts“) ab. Sie verbessert sich oft durch weiterführende Frage-Antwort-Schleifen zwischen Anwender*innen und der KI. Die Anwendungshinweise in den jeweiligen Nutzungsinformationen generativer KI-Werkzeuge sind zu beachten.

1.2.3. Es ist wichtig, rechtskonform mit personenbezogenen Daten und dem Amtsgeheimnis umzugehen.

Wenn Inhalte in das generative KI-Werkzeug eingegeben werden, ist wie bisher im Internet oder in sozialen Medien darauf zu achten, dass mit personenbezogenen Daten (auch Bilddaten), internen sensiblen Informationen und Amts-, Geschäfts- oder Betriebsgeheimnissen rechtskonform umgegangen wird.

1.2.4. Es ist wichtig, die Verwendung von Inhalten, die mit Hilfe von KI erstellt wurden, transparent zu kennzeichnen:

„Der Inhalt dieses Texts/dieses Bilds wurde unter Zuhilfenahme eines Künstlichen Intelligenz-Werkzeugs erstellt und von einem Menschen überprüft.“

Wenn generative KI für die Erstellung von Texten, Bildern usw. verwendet wird, ist dies jedenfalls nach außen transparent zu machen, zum Beispiel durch einen Hinweis in einer Fußzeile, Bildunterschrift oder im Signaturbereich einer E-Mail. Dieser Hinweis entbindet die Person, die das Dokument erstellt oder unterzeichnet hat, jedoch nicht von ihrer Verantwortung für die Inhalte. Eine Kennzeichnung ist nicht erforderlich, wenn bestehende Inhalte lediglich modifiziert und nicht erweitert wurden, z.B. bei Rechtschreibprüfungen, Übersetzungsprogrammen, Textzusammenfassungen, E-Mail-Ausformulierungen oder bei Fotos, die mithilfe von KI freigestellt wurden.

1.3. Was bedeutet dieses Grundprinzip, wenn ich dienstlich mit Ergebnissen konfrontiert bin, die von KI-Werkzeugen sind/sein können?

Wie bisher ist es wichtig, Informationen aus externen Quellen (wie Social Media, Internet usw.) kritisch zu hinterfragen und auf ihre Nachvollziehbarkeit zu prüfen, bevor sie in eigene Entscheidungsprozesse oder dienstliche Erledigungen einfließen.

Besonders KI-generierte Deepfakes, also täuschend echt wirkende, manipulierte Bild-, Audio- oder Videoaufnahmen, und ihre schnelle Verbreitung in sozialen Medien stellen uns vor große Herausforderungen in Bezug auf Desinformation und deren mögliche Folgen, auch in Bezug auf die Glaubwürdigkeit der Stadtverwaltung.³

Es ist daher entscheidend, Faktenchecks durchzuführen, um echte von gefälschten Inhalten zu unterscheiden, insbesondere im Zusammenhang mit potenziell von KI-Werkzeugen generierten gefälschten Bildern.

Zudem können eine Reihe von Kriterien⁴ Hinweise auf die Authentizität geben, z. B.:

- a. Erscheint der Kontext des Bildes schlüssig?
- b. Sind alle Komponenten des Bildes stimmig?
- c. Sind augenscheinliche Fehler und Ungereimtheiten bei Körperteilen u. ä. erkennbar?
- d. Sind verschwommene oder verschmolzene Übergänge (v. a. zwischen Personen und Hintergründen) erkennbar?
- e. Fallen un schlüssig abgebildete Reaktionen umstehender Personen oder unübliche Erscheinungsformen ins Auge?

³ <https://www.wien.gv.at/medien/fake-news/>

⁴ vgl. <https://apa.at/faktencheck/ki-generierte-bilder-haeufig-geteilt/>.

1.4. KITT – Künstliche Intelligenz Tipps und Tricks

Im Magistrat gibt es für Interessierte mit KITT die Möglichkeit, sich zum Thema KI auszutauschen und weiterzubilden. Es ist kein (technisches) Vorwissen dafür notwendig und steht allen Mitarbeiter*innen offen.

KITT-Online-Terminserien:

- a) Einführung, Inspiration und Herausforderung (1x pro Monat für Neueinsteigende)
- b) Erfahrungsaustausch und aktuelle Entwicklungen (1x pro Monat, um den Überblick zu behalten)
- c) Prompt Together (2x pro Monat zum gemeinsamen Aufbauen von effektiven Prompts)

Per Mail an kitt@ma53.wien.gv.at mit Betreff „Terminverteiler“ lässt man sich in den Terminverteiler aufnehmen. Zudem kann man zur KITT-Intranet-Gruppe oder KITT-Signal- Messenger-Gruppe beitreten.

Titel	KI-Kompass
Version	2.0
Gültig ab	10. April 2024
Ersetzt	Version 1.0
Herausgebende Stelle	Magistrat der Stadt Wien Geschäftsbereich Organisation und Sicherheit Gruppe Organisation (MD-OS/ORG)
Beiträge von	AI-Kompetenznetzwerk
Dokumentenklasse	Frei verfügbar
Beilagen	-