

Leitlinien über die Verwendung von ChatGPT und ähnlichen Systemen in der Verwaltung¹

1 Was sind ChatGPT und ähnliche Systeme?

ChatGPT ist ein computerbasiertes Programm, das menschenähnliche Texte erzeugen kann. Es nutzt künstliche Intelligenz (KI)², um aufgrund von Mustern und Daten Antworten auf Texteingaben zu generieren. Neben ChatGPT gibt es auch weitere textverarbeitende KI-Systeme. Bekannte Anwendungen sind z.B. das Übersetzungsprogramm DeepL oder DALL-E, das aus Textbeschreibungen Bilder erstellen kann.

Diese Systeme sind für die Öffentlichkeit meist kostenfrei zugänglich und können für verschiedene Aufgaben eingesetzt werden, z.B. für die Erstellung eines Berichts oder einer Zusammenfassung. Neben der Textgenerierung können gewisse KI-Anwendungen auch Codeblöcke ausgeben oder Bilder erzeugen. Der Fokus der vorliegenden Leitlinien liegt indessen auf textverarbeitenden KI-Systemen.

Eine vertiefte Auseinandersetzung zum Umgang mit KI in der Staatsverwaltung erfolgt im Rahmen der Bearbeitung des Auftrags zur Erarbeitung einer KI-Strategie, den der Kantonsrat in der Sommersession 2023 mit dem Kantonsratsbeschluss über die Rechnung 2022 ([33.23.01](#)) erteilt hat.

2 Welche Chancen und Risiken stellen sich?

Chancen: ChatGPT und andere textverarbeitende KI-Systeme bergen hohes Potenzial, da sie Menschen Arbeit abnehmen können, die sonst sehr viel Zeit in Anspruch nimmt. Informationen können schnell und effizient erlangt werden, was die Entscheidungsfindung erleichtern kann. Zudem können Texte leicht in verschiedene Sprachen übersetzt werden, strukturiert oder verständlicher formuliert werden.

Risiken: Durch die Nutzung solcher Systeme stellen sich indessen auch Risiken:

- Die Nachvollziehbarkeit und Transparenz der Textausgaben ist nicht gewährleistet.
- Antworten können falsch, irrelevant oder nicht werteneutral sein.
- Geheime und vertrauliche Daten könnten entgegen der Vorgaben der Informationssicherheit und des Datenschutzes³ an Dritte weitergegeben werden. Die eingegebenen Daten werden teilweise für die Weiterentwicklung dieser Systeme verwendet.
- Die Verwendung KI-basierter Textgeneratoren kann auch urheberrechtlich problematisch sein, wenn nicht die nötigen Rechte erworben wurden.

3 Wie dürfen ChatGPT und ähnliche Systeme bei der Arbeit verwendet werden?

Die Mitarbeitenden der Verwaltung können ChatGPT und andere textverarbeitende KI-Systeme grundsätzlich unter Beachtung der geltenden rechtlichen Vorgaben im Arbeitsalltag verwenden. Es ist wichtig, solche Systeme mit der notwendigen Umsicht und Vorsicht nutzen. Betreffend die Texteingabe und die anschliessende Verwendung der Textausgaben gelten folgende Leitlinien:

¹ Verabschiedet vom Vorstand der Vereinigung St.Galler Gemeindepräsidentinnen und Gemeindepräsidenten am 31. August 2023 und von der Regierung des Kantons St.Gallen am 5. September 2023.
² Vgl. die Definition des Kompetenznetzwerks für künstliche Intelligenz des Bundes: <https://cnaai.swiss/>.
³ Datenschutzgesetz ([sGS 142.1](#); abgekürzt DSG) und Verordnung über die Informatiksicherheit ([sGS 142.21](#)).

3.1 Leitlinien für die Texteingabe

3.1.1 Schutz von geheimen und vertraulichen Daten

Geheime und vertrauliche Daten dürfen nicht an ChatGPT und ähnliche Systeme weitergegeben werden. Dies umfasst insbesondere:

1. durch das **Datenschutzgesetz** ([sGS 142.1](#); abgekürzt *DSG*) geschützte Daten:
 - *Personendaten* (Art. 1 Bst. a DSG), d.h. Angaben, die sich auf eine bestimmte oder bestimmbare natürliche Person beziehen. Ausnahme: Entsprechende Eingaben können erfolgen, wenn diese *vorab* anonymisiert bzw. pseudonymisiert werden und damit keine Rückschlüsse auf natürliche Personen möglich sind.⁴
 - *besonders schützenswerte Personendaten* (Art. 1 Bst. b DSG), d.h. Angaben über:
 - religiöse, weltanschauliche sowie politische Ansichten und Tätigkeiten. Ausgenommen sind Angaben über die Mitgliedschaft bei einer Religionsgemeinschaft, einer Organisation oder einer politischen Partei, wenn die betroffene Person diese selbst bekannt gegeben hat oder für ein öffentliches Amt kandidiert;
 - Gesundheit, Intimsphäre und ethnische Zugehörigkeit;
 - genetische Daten;
 - biometrische Daten, die eine natürliche Person eindeutig identifizieren;
 - Leistungen und Massnahmen der sozialen Hilfe;
 - strafrechtliche sowie disziplinarische Verfahren und Sanktionen.
 - *Persönlichkeitsprofile* (Art. 1 Bst. d DSG), d.h. eine Zusammenstellung von Personendaten, welche die Beurteilung der Persönlichkeit einer natürlichen Person erlaubt.
 - *Profiling* (Art. 1 d^{bis} DSG), d.h. jede Art der automatisierten Bearbeitung von Personendaten, die darin besteht, dass diese Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten.⁵
2. durch eine **Geheimhaltungspflicht** geschützte Daten:
 - dem (*allgemeinen*) *Amtsgeheimnis* unterstehende Daten.⁶ Das Amtsgeheimnis umfasst nur die Informationen, die gemäss dem Öffentlichkeitsgesetz ([sGS 140.2](#)) nicht zugänglich sind. Es handelt sich dabei um Daten, die ihrer Natur nach geheim sind, d.h. sie sind nur einem beschränkten Personenkreis bekannt, sie sind nicht allgemein zugänglich, es bestehen berechnigte (öffentliche oder private) Geheimhaltungsinteressen sowie ein Wille des Arbeitgebers an ihrer Geheimhaltung.
 - *besonderen (Berufs-)Geheimnissen* unterstehende Daten (z.B. Arztgeheimnis, Anwaltsgeheimnis, Steuergeheimnis).
 - *vertraglich* ausdrücklich geschützte Daten.

⁴ Die Eingabe von besonders schützenswerten Personendaten, Persönlichkeitsprofilen und Profiling ist auch in anonymisierter bzw. pseudonymisierter Form aufgrund des hohen Schutzbedarfs der Daten und der Rückschlussgefahr auf eine natürliche Person nicht zulässig.

⁵ Dies betrifft insbesondere Aspekte bezüglich Arbeitsleistung, wirtschaftlicher Lage, Gesundheit, persönlicher Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder um Ortswechsel dieser natürlichen Person zu analysieren oder vorherzusagen.

⁶ Art. 67 des Personalgesetzes ([sGS 143.1](#)) i.V.m. Art. 3a des Staatsverwaltungsgesetzes ([sGS 140.1](#)) und Art. 99 des Gemeindegesetzes ([sGS 151.2](#)) sowie Art. 320 des Schweizerischen Strafgesetzbuchs ([SR 311.0](#)).

3.1.2 Gezielter Einsatz und korrekte Anwendung

Die Eingabe von *ganzen Dokumenten der Verwaltung oder längeren Textabschnitten daraus* («copy paste») in solche Systeme (z.B. zur Übersetzung) ist ohne eine *vorgängige* Prüfung des Inhalts nicht zulässig. Es muss vor einer Texteingabe überprüft werden, ob ein Text vertrauliche oder geheime Daten enthält und ob die entsprechenden Inhalte bzw. bestimmte Begriffe weggelassen oder umformuliert werden können (vgl. vorstehend Abschnitt 3.1.1).

ChatGPT und ähnliche Systeme eignen sich lediglich für einfache, standardmässige Anfragen. Sie eignen sich nicht für die Bearbeitung von komplexen oder kreativen Aufträgen. Um qualitativ gute Antworten zu erzielen, sollten *klare und spezifische Fragestellungen* gewählt werden, ohne zu viel über den Arbeitgeber preiszugeben.

Beispiel einer Texteingabe:

Nenne mir die drei wichtigsten Punkte, die den Kanton St.Gallen als modernen und zukunftsgerichteten Kanton beschreiben, und gib eine kurze Erklärung.

Antwort:

- *Bildung und Forschung:* Fokus auf erstklassige Bildungseinrichtungen wie Universität St.Gallen (HSG) und Fachhochschulen, die zur Innovationskraft des Kantons beitragen.
- *Wirtschaftliche Diversifikation und Innovation:* Vielfältige Wirtschaftsstruktur von traditionellen bis modernen Branchen, Förderung von Unternehmertum und enge Wissenschaft-Industrie-Kooperation.
- *Nachhaltige Entwicklung und Lebensqualität:* Betonung von Umweltschutz, erneuerbaren Energien, und Förderung eines aktiven Lebensstils für hohe Lebensqualität.

3.2 Leitlinien für die Verwendung von Textausgaben

Bei der Verwendung von generierten Textausgaben durch ChatGPT oder ähnlichen Systemen gilt Folgendes:

- *Kritische Prüfung:* Ein generierter Text und die darin enthaltenen Informationen sind kritisch auf Qualität und Korrektheit zu prüfen.
- *Vergleich mit anderen Quellen:* Um die Genauigkeit einer Textausgabe sicherzustellen, sollte ein Vergleich mit anderen, zuverlässigen Quellen vorgenommen werden.
- *Sachgerechte Anpassung:* Von einer direkten Übernahme von generierten Textpassagen ist abzusehen, insbesondere, wenn eine Quellenangabe erforderlich wäre. Textausgaben sind folglich sachgerecht anzupassen und an die jeweils geltenden Schreibweisungen (insbesondere kantonale [Schreibweisungen](#)) anzugleichen.
- *Objektivität:* Die gemachten Aussagen von ChatGPT und ähnlichen Systemen sind unter Umständen nicht werteneutral. Es muss geprüft werden, ob ein Text diskriminierend wirkt.
- *Urheberrecht:* Das Urheberrecht (v.a. das eidgenössische Urheberrechtsgesetz [\[SR 231.1\]](#)) ist zu beachten. Es kann vorkommen, dass Text aus unterschiedlichen Quellen generiert wurde, für die aber nicht die erforderlichen urheberrechtlichen Rechte erworben wurden.