KI-Governance im Kanton St.Gallen

Bern, 3. Mai 2024

16. Sitzung der Fachgruppe von Juristinnen und Juristen im E-Government

Marlène Schürch, Leiterin IT-Recht und Datenschutz

Agenda

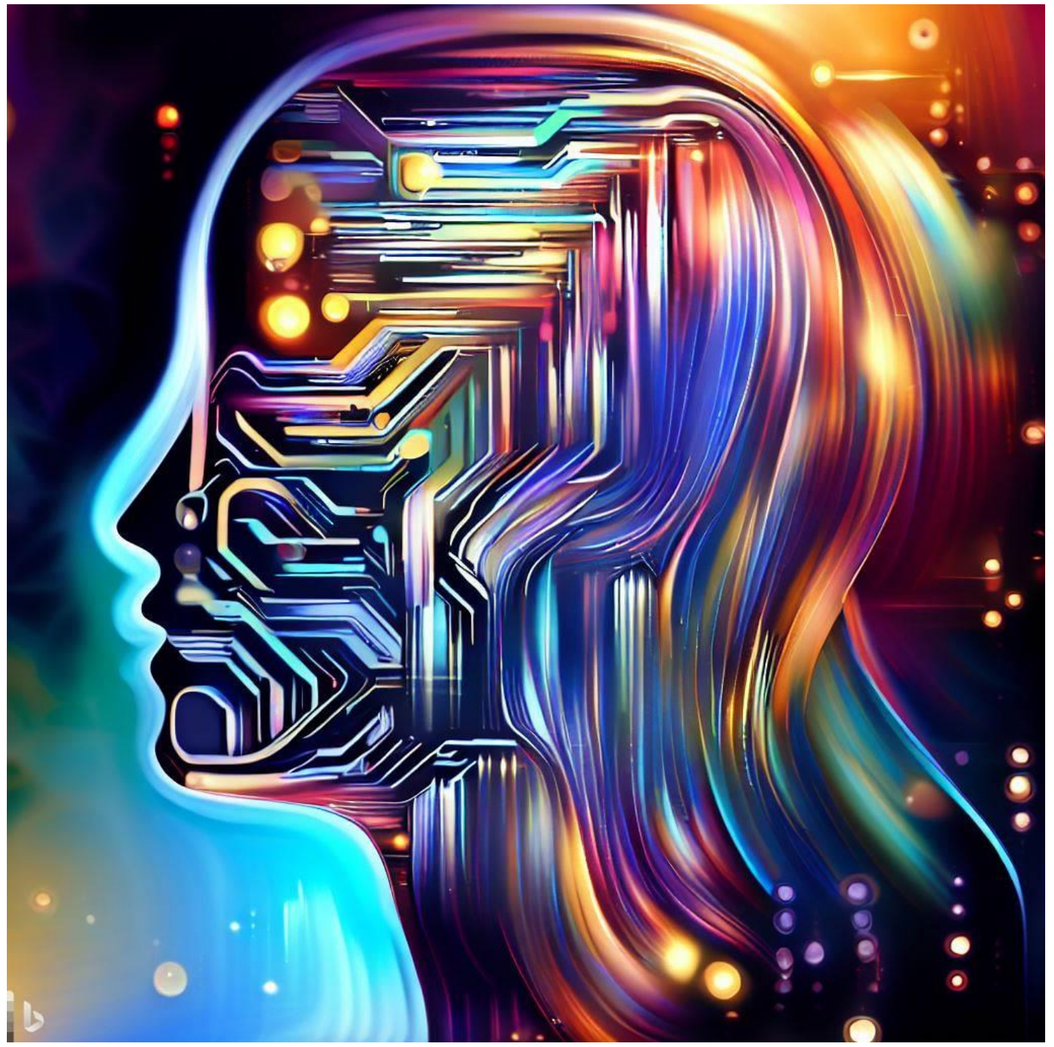
1. Was ist KI?
2. Stand der Regulierung
3. Was macht der Kanton St.Gallen im Bereich von KI?
    1. Auftrag KI-Strategie
    2. Leitlinien zur Verwendung von ChatGPT und ähnlichen Systemen in der Verwaltung
4. Fragen und Diskussion

Was ist Künstliche Intelligenz (KI)?

Definition CNAI - Kompetenznetzwerk für künstliche Intelligenz (Bund)

Künstliche Intelligenz (KI) wird definiert als:
«einen Computer so bauen oder programmieren, um Dinge zu tun, die normalerweise menschliche oder biologische Fähigkeiten («Intelligenz») erfordern»,
z.B. visuelle Wahrnehmung (Bilderkennung), Spracherkennung, Sprachübersetzung, visuelle Übersetzung und Spiele spielen (mit konkreten Regeln).

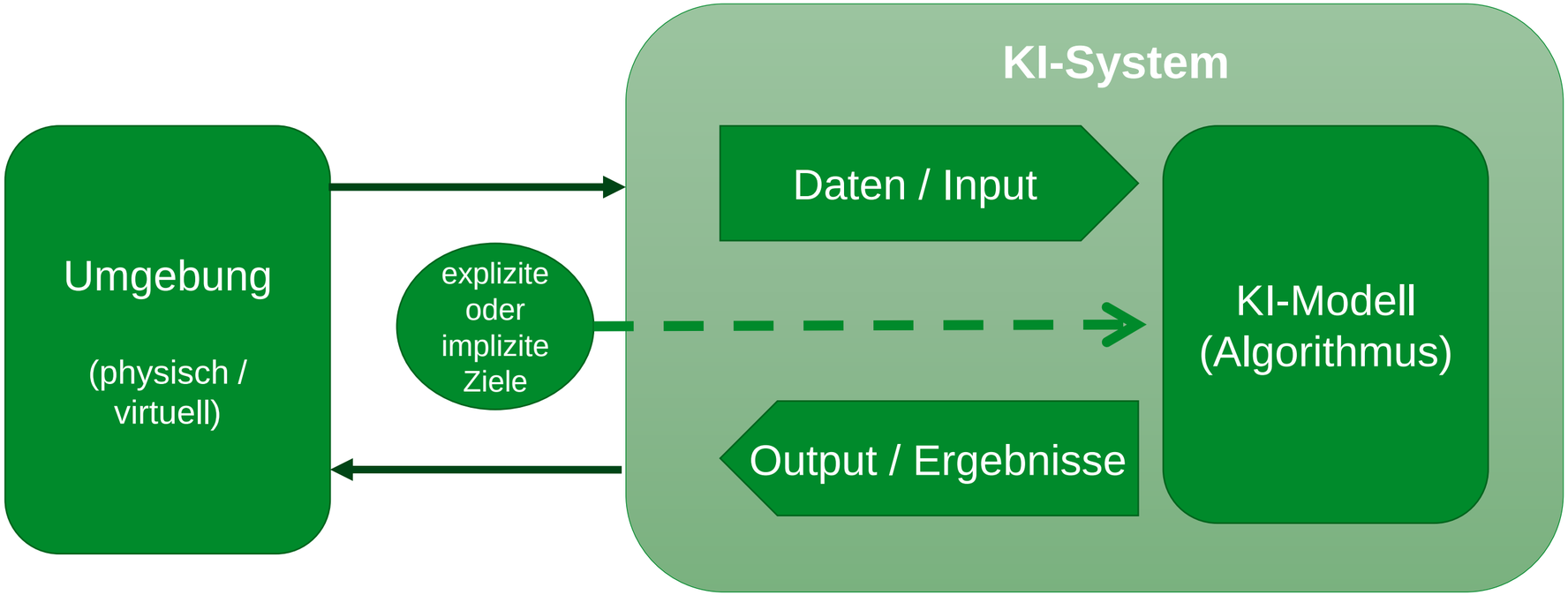
Bei KI geht es um «intelligente» Maschinen, die Aufgaben ausführen können, die normalerweise von Menschen ausgeführt werden, d.h. Maschinen «intelligent» machen.



Quelle: Bundesamt für Statistik, Terminologie Kompetenznetzwerk CNAI, Version 2.0, 21.12.2023, https://cnai.swiss/dienstleistungen/terminologie-2/.

Konkreter: Was ist ein KI-System?
Definition OECD

Ein **KI-System** ist ein maschinenbasiertes System, das für explizite oder implizite Ziele aus den empfangenen **Inputs** schlussfolgert, wie es **Outputs** wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erzeugen kann, welche die physische oder virtuelle Umgebung beeinflussen können. KI-Systeme können mit unterschiedlichem Ausmass an **Autonomie** ausgestattet werden.



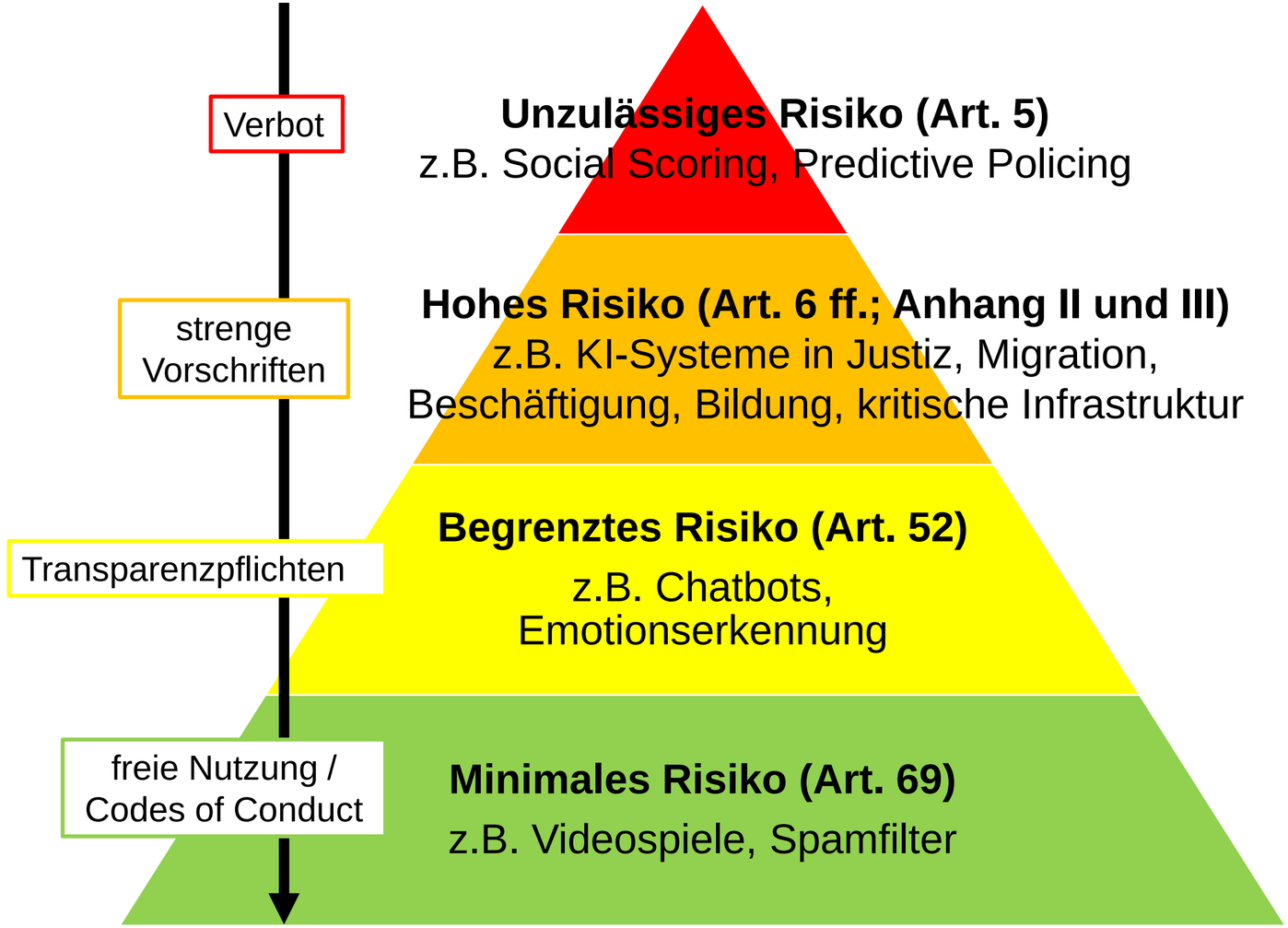
Quelle: OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2024, https://oecd.ai/en/ai-principles.

Stand der Regulierung

EU Artificial Intelligence Act («AI Act»)

- **2018:** Europäische KI-Strategie
- **13.03.2024:** Annahme EU Artificial Intelligence Act («AI Act») durch EU-Parlament
  - Inkrafttreten: vermutlich Mai/Juni 2024
  - Umsetzungsfrist: 2 Jahre

- **Inhalt AI Act:**
  - Anwendungsbereich: alle Anbieter von KI-Systemen im EU-Markt
  - Kategorisierung von KI-Systemen: risikobasierter Ansatz
  - Eigenkontrolle der Anbieter mit Beschwerdemöglichkeit der Bevölkerung
  - Sanktionen (bis zu 35 Mio. Euro)
  - allgemeine KI-Grundsätze (menschliche Kontrolle, Sicherheit, Datenschutz, Transparenz, Nichtdiskriminierung, Fairness, Soziales und ökologisches Wohlergehen)



Quellen: Europäisches Parlament, https://www.europarl.europa.eu/portal/de; EUR-Lex, 21.02.2021, https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52021PC0206.

Stand der Regulierung
KI-Konvention des Europarates

- «Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law» (KI-Konvention)
- völkerrechtlich verbindliches Abkommen
- Dezember 2023: Veröffentlichung des zweiten Entwurfs
- März 2024: abschliessende Verhandlungen in Strassburg
- ausstehend: formelle Verabschiedung durch Ministerkomitee, Ratifikation durch Staaten und nationale Umsetzung

Quelle: Council of Europe, DRAFT FRAMEWORK CONVENTION ON ARTIFICIAL INTELLIGENCE, HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW, 18.12.2023, https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043.

Stand der Regulierung
KI-Regulierung in der Schweiz (Bund)

- bis November 2024: Auslegeordnung zur Regulierung von künstlicher Intelligenz (Auftrag Bundesrat)
  - Stossrichtung:
    - Aufbau auf bestehendem Schweizer Rechtsrahmen
    - Komptabilität mit AI Act (EU) und KI-Konvention (Europarat)
  - Fokus:
    - Grundrechte
    - technische Standards
    - finanzielle und institutionelle Auswirkungen
  - interdisziplinär: Recht, Wirtschaft, Europapolitik
  - Organisation:
    - Interdepartementale Koordinationsgruppe EU-Digitalpolitik mit Federführung durch BAKOM (UVEK) und Abteilung Europa (EDA)
    - Einbezug:
      - Plateforme Tripartite (Austauschplattform zu digitaler Gouvernanz und KI)
      - Kompetenznetzwerk KI des Bundes – CNAI
      - Arbeitsgruppe KI in der Bundesverwaltung
- ab 2025: konkreter Auftrag für KI-Regulierungsvorlage

Quelle: Medienmitteilungen des Bundesrates, 22.11.2023, https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-98791.html.

Was macht der Kanton St.Gallen im Bereich KI?

Auftrag KI-Strategie

Klassische Strategieentwicklung
(Analysen, Handlungsfelder, Umsetzungsplanung)

Erweiterung um spezifische Vorgaben
Rechtsrahmen, Risiken, Förderung Diskurs, Zusammenspiel mit anderen Staatsebenen und weiteren Akteuren, Befähigung

Antrag der Finanzkommission vom 25. Mai 2023

Ziff. 3 (neu):

Die Regierung wird eingeladen,[1] dem Kantonsrat gestützt auf eine Potenzial- und Umfeldanalyse zur Nutzung der künstlichen Intelligenz (KI) in der öffentlichen Verwaltung und basierend auf den Leitlinien des Bundes vom November 2020 ihre strategischen Leitplanken zur Nutzung von und zum Umgang mit KI darzulegen. Eine solche «KI-Strategie» sollte namentlich folgende Aspekte umfassen:

a) übergeordnete sowie bereichsspezifische Potenziale der KI-Nutzung innerhalb der kantonalen und kommunalen Verwaltung;
b) strategische Stossrichtung und Handlungsfelder einschliesslich einer Umsetzungsplanung mit konkreten Zielen und Projekten zur Nutzung der KI in der Verwaltung;
c) bestehender Rechtsrahmen für die KI-Nutzung und regulatorischer Handlungsbedarf;
d) Risiken der KI-Nutzung sowie Strategien und Instrumente zur Bewältigung der Risiken;
e) Förderung des öffentlichen Diskurses zu Chancen und Risiken der KI;
f) Zusammenspiel der Staatsebenen, Privatwirtschaft, Wissenschaft sowie Zivilgesellschaft bei der Förderung und kritischen Begleitung von KI-Projekten;
g) Massnahmen im Bereich Personal- und Organisationsentwicklung, um die Mitarbeitenden zur Nutzung von KI zu befähigen und das Zusammenspiel von Mensch und Technologie zu verbessern.

Quelle: Kantonsratsbeschluss über die Rechnung 2022 des Kantons St.Gallen (33.23.01), Link.

Was macht der Kanton im Bereich von KI?
Bereits ergriffene oder umgesetzte Massnahmen

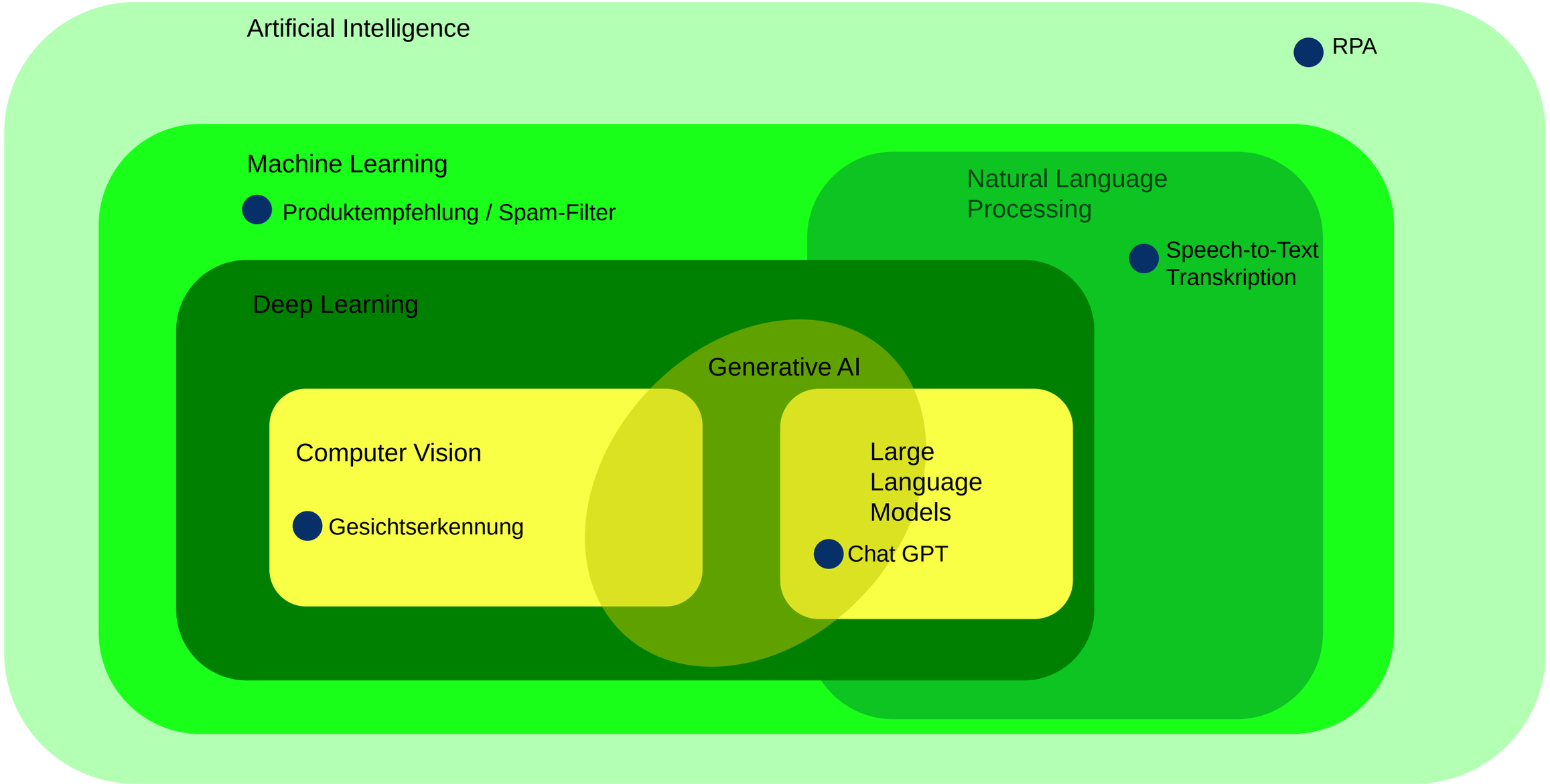
- Leitlinien über die Verwendung von ChatGPT und ähnlichen Systemen in der Verwaltung (Link)
- Beantwortung der Einfachen Anfrage 61.23.53 «Effizienzsteigerung durch künstliche Intelligenz und Prozessautomatisierung» mithilfe von ChatGPT (Link)
- Veranstaltungsreihe digitale Transformation (Link)
- Hack Case am Start Summit zu «Voice Bots»
- interne Weiterbildungen: Kurs «Digital Pioneer» / Cybersicherheit

Was macht der Kanton St.Gallen im Bereich von KI?

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

- kurzfristig ergriffene Massnahme
- von der Regierung des Kantons St.Gallen und dem Verband St.Galler Gemeindepräsidien (VSGP) am 31. August 2023 verabschiedet (Link)
- Ziel: Mitarbeitende befähigen, im Internet frei verfügbare, sprachbasierte KI-Systeme sorgfältig im Arbeitsalltag einzusetzen
- Aufbau:
  - Definition ChatGPT und ähnliche Systeme
  - Chancen und Risiken
  - Verwendung bei der Arbeit
    - Texteingabe
    - Textverwendung

Was ist ChatGPT?



Vorteile und Nachteile von ChatGPT und ähnlichen Systemen

Vorteile

- einfacher und schneller Zugang zu Informationen
- Übersetzung in verschiedene Sprachen Formulierungshilfe
- Texte können umgewandelt werden (z.B. von sehr formeller Sprache in einen Marketing-Text und umgekehrt)
- schnelle Erstellung von Zusammenfassungen

Nachteile

- Nachvollziehbarkeit und Transparenz ist nicht gewährleistet.
- Antworten können falsch, irrelevant oder nicht werteneutral sein.
- Die eingegebenen Daten können für die Weiterentwicklung dieser Systeme verwendet werden.
- Geheime und vertrauliche Daten könnten an Dritte weitergegeben werden.
- Urheberrechtlich problematisch / keine Zuordnung an eine Urheberin bzw. einen Urheber möglich

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

Im Arbeitsalltag dürfen ChatGPT, DeepL usw. grundsätzlich eingesetzt werden. Dabei müssen aber die geltenden rechtlichen Vorgaben beachtet werden. Darunter fallen insbesondere:

- Datenschutz und Informationssicherheit
  - kantonales Datenschutzgesetz (sGS 142.1; abgekürzt DSG)
  - Verordnung über die Informatiksicherheit (sGS 142.21)
  - sofern anwendbar:
    - Bundesgesetz über den Datenschutz (SR 235.1)
    - Bundesgesetz über die Informationssicherheit beim Bund (SR 128)
- Geheimhaltungspflichten
  - Berufsgeheimnisse und besondere Amtsgeheimnisse (vgl. Spezialgesetze)
  - (allgemeines) Amtsgeheimnis: 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)
  - Geschäfts- / Fabrikationsgeheimnisse
- Immaterialgüterrechte, vgl. eidg. Urheberrechtsgesetz (SR 231.1)

Texteingabe

Diese Daten dürfen nicht in ChatGPT & Co eingegeben werden:

1. Datenschutz:
   - *(gewöhnliche) Personendaten* (Art. 1 Bst. a DSG), z.B. Name, Alter, Beruf, Wohnadresse, E-Mail, Telefonnummer
     - *Ausnahme: Eingaben können erfolgen, wenn z.B. mittels Pseudonym ein Rückschluss auf die natürliche Person vorab vollständig entfernt werden kann.*
   - *besonders schützenswerte Personendaten* (Art. 1 Bst. b DSG), z.B. religiöse oder politische Ansichten, Gesundheit, strafrechtliche sowie disziplinarische Verfahren und Sanktionen
   - *Persönlichkeitsprofile* (Art. 1 Bst. d DSG), d.h. eine Zusammenstellung von Personendaten zur Zweck einer Persönlichkeitsbeurteilung
   - *Profiling* (Art. 1 Bst. d^{bis} DSG), d.h. automatisierte Bearbeitung von Personendaten zum Zweck der Bewertung (z.B. Arbeitsleistung)
     - *Keine Ausnahme: Eine Pseudonymisierung ist bei besonders schützenswerten Personendaten, Persönlichkeitsprofilen und Profiling aufgrund des hohen Schutzbedarfs nicht zulässig.*

Texteingabe

Diese Daten dürfen nicht in ChatGPT & Co eingegeben werden:

2. Geheimhaltungspflichten:
   - *Berufsgeheimnisse und besondere Amtsgeheimnisse* (z.B. Arzt-, Anwalts-, Steuer-, Opferhilfegeheimnis)
   - (*allgemeines) Amtsgeheimnis:* Informationen, die gemäss dem Öffentlichkeitsgesetz (sGS 140.2) nicht zugänglich sind.
   - *vertraglich ausdrücklich statuiert*, z.B. Geschäfts- / Fabrikationsgeheimnisse von Lieferanten

Wichtig ist ein sorgfältiger Umgang:

- Vor einer Texteingabe muss immer überprüft werden, ob ein Text vertrauliche oder geheime Daten enthält und ob die entsprechenden Inhalte bzw. bestimmte Begriffe weggelassen oder umformuliert werden können.
- Kein «copy paste» von ganzen internen Dokumenten oder längeren Textabschnitten.
- Für komplexe Aufträge, für die internes Fachwissen nötig ist, sind ChatGPT und ähnliche Systeme weniger geeignet.
- Je klarer und spezifischer die Fragestellung, desto besser das Resultat (→ siehe: Tipps und Tricks zum Prompting).

Tipps und Tricks: Prompting

1. Kurze und klare Anweisungen schreiben
2. Verwendung als Formulierungshilfe oder als Rechtschreibekorrektur
3. ChatGPT auffordern für eine gewisse Zielgruppe zu schreiben
   (für Laien / für eine Primarschülerin / …)
4. ChatGPT eine Rolle zuweisen (du bist eine Wissenschaftlerin / du bist ein kritischer Leser / …)
5. ChatGPT anweisen, in einen bestimmten Schreibstil zu formulieren (formell, witzig, in leichter Sprache)
6. Folgefragen stellen (d.h. eine Unterhaltung führen)
7. Den Output formatieren lassen (als Tabelle, Stichwortliste, usw.)

Verwendung von Textausgaben

- Qualität und Korrektheit kritisch prüfen
- Vergleich mit anderen Quellen sicherstellen
- Textanpassung: generierte Textpassagen nicht 1:1 übernehmen, sondern sachgerecht anpassen
- Prüfung Objektivität / allfällige Diskriminierung
- Prüfung Urheberrecht: Text kann aus unterschiedlichen Quellen generiert worden sein, für die aber nicht die erforderlichen Rechte erworben wurden
- Einsatz von KI-Systemen transparent ausweisen

Fragen und Diskussion

- Welche KI-Anwendungen werden in Ihrer kantonalen Verwaltung verwendet?
- Sind weitere Anwendungen konkret geplant?
- Bestehen in Ihrem Kanton zur Nutzung von KI-Anwendungen grundlegende politische oder regulatorische Vorgaben (z.B. Strategien, Rechtsgrundlagen, Richtlinien) oder sind solche in Entwicklung?