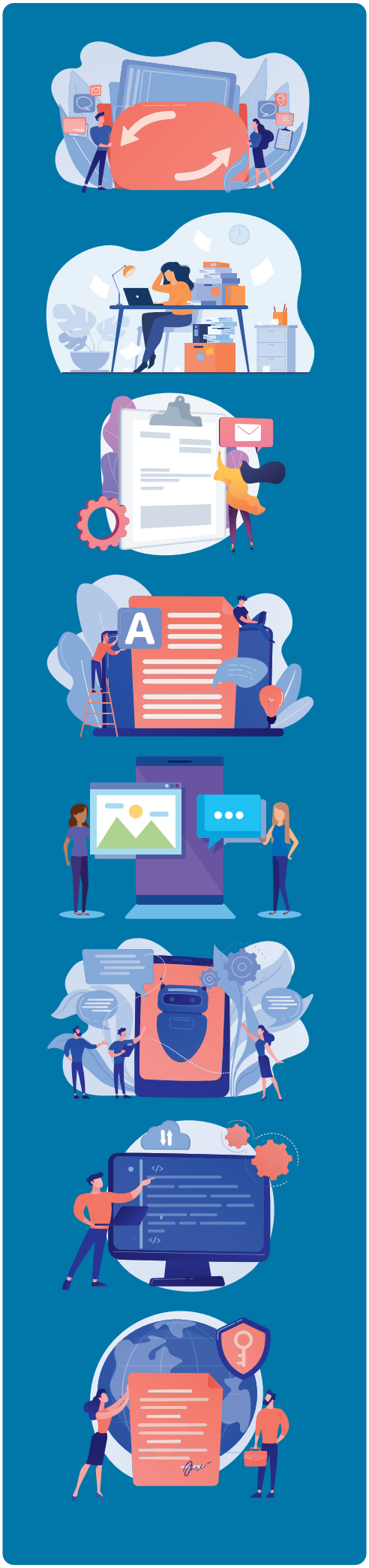
Hessisches Ministerium für
Digitalisierung und Innovation

HESSEN

Generative KI

Eine Einführung mit Blick auf die Landesverwaltung

digitales.hessen


A

Inhalt

1 Vorwort

Prof. Dr. Kristina Sinemus
Hessische Ministerin für
Digitalisierung und Innovation

Liebe Leserinnen und Leser,

in unserem Alltag begegnet sie uns bereits heute: Sie empfiehlt uns Bücher oder Musikstücke, führt uns auf dem schnellsten Weg vom Arbeitsplatz nach Hause und ermöglicht es uns, im Urlaub unkompliziert die Speisekarte in fremden Sprachen zu verstehen. Künstliche Intelligenz (KI) ist eine Schlüsseltechnologie des 21. Jahrhunderts, die Wirtschaft, Wissenschaft und Gesellschaft zunehmend prägt, auch bei uns in Hessen.

Der Hessischen Landesregierung ist es ein wichtiges Anliegen, auch in der Landesverwaltung die Potenziale von KI umfassend zu heben, zum Nutzen der Bürgerinnen und Bürger, der Unternehmen und der Beschäftigten. Dabei können wir auf der im Jahr 2022 veröffentlichten hessischen KI-Zukunftsagenda aufbauen, die bereits einen Schwerpunkt auf die Stärkung des Einsatzes von KI in der öffentlichen Verwaltung legt. Projekte und Maßnahmen wie die Forschungsstelle KI am Finanzamt Kassel, der INNOVATION HUB 110 der hessischen Polizei, das Richterassistenztool FraUKe am Amtsgericht Frankfurt am Main oder die Nutzung von Webscraping im Hessischen Statistischen Landesamt zeigen, wie KI in der Landesverwaltung zielführend eingesetzt werden kann.

In den vergangenen Monaten haben sich die potenziellen Anwendungsfelder durch die rasanten Entwicklungen im Bereich der generativen KI noch deutlich erweitert. Beispiele reichen von der Nutzung für das Wissensmanagement und ein erleichtertes

Einarbeiten neuer Kolleginnen und Kollegen über die Auswertung großer Textmengen bis hin zur Unterstützung bei Recherchen und bei der Erstellung von Vermerken oder Terminvorbereitungen.

Generative KI kann also ein wichtiges Unterstützungswerkzeug für die öffentliche Verwaltung sein. Zugleich ist sie kein Allheilmittel: Teils existieren überzogene Erwartungen daran, was generative KI aktuell leisten kann. Mit realistischen Einordnungen zur Rolle dieser Technologie in der Verwaltung, aber auch mit grundlegenden Ängsten vor generativer KI sowie mit den zahlreichen ethischen und rechtlichen Herausforderungen, die dieses Thema aufwirft, müssen wir uns auseinandersetzen.

Das vorliegende Einführungsdokument greift den großen Informationsbedarf innerhalb der Landesverwaltung zum Themenkomplex generative KI gezielt auf. Es ermöglicht einen niedrigschwelligen Zugang zum Thema generative KI, versucht, dieses zu entmystifizieren und die Fakten in den Mittelpunkt zu rücken, und zeigt anhand von konkreten Anwendungsbeispielen, wie generative KI in der Landesverwaltung sinnvoll genutzt werden kann. Es würde mich freuen, wenn das Einführungsdokument Denkanstöße gibt und als Inspirationsquelle für weitere KI-Projekte in der Landesverwaltung dient.

Zur Erarbeitung des Einführungsdokuments hat sich aus dem Arbeitskreis KI-Agenda heraus eine ressortübergreifende Projektgruppe gebildet und sich in mehreren Workshops mit dem Thema generative KI befasst. Den beteiligten Mitarbeiterinnen und Mitarbeitern danke ich herzlich für ihre Offenheit, an diesem ungewöhnlichen Projekt mitzuwirken und ihre ganz unterschiedlichen Perspektiven auf das Thema in die Projektgruppe und in das Einführungsdokument einzubringen. Mein Dank gilt zugleich den Ressorts, die bereit waren, Vertreterinnen und Vertreter in die Projektgruppe zu entsenden, sowie der Hessischen Zentrale für Datenverarbeitung, die das Projekt engagiert unterstützt hat.

Ihnen allen wünsche ich eine angenehme Lektüre und viel Spaß beim Eintauchen in die Welt der generativen KI!

Prof. Dr. Kristina Sinemus

2 Einleitung

Sie könnte beim Schreiben von Vermerken und Terminvorbereitungen unterstützen, helfen, wenn uns Ideen für die Strukturierung des nächsten Redeentwurfs fehlen, und neuen Kolleginnen und Kollegen in kürzester Zeit die Einarbeitung in ihre Themenfelder ermöglichen. Zugleich „halluziniert" sie, ist eine „Blackbox" und wirft zahlreiche rechtliche und ethische Fragen auf. Generative KI, eine für Nutzerinnen und Nutzer besonders intuitive Ausprägung der Künstlichen Intelligenz (KI), ist seit einiger Zeit in den Medien sehr präsent, wird bestaunt und kritisiert und hat weltweit das Interesse zahlreicher Menschen am Thema KI geweckt.

Auslöser für diese Entwicklungen war die Veröffentlichung von ChatGPT im November 2022. Bereits im Januar 2023 wurden weltweit über 100 Millionen Nutzerinnen und Nutzer gezählt. Seither sind rasante Entwicklungen erfolgt und die Debatten rund um generative KI sind vielfältiger geworden. Längst gibt es eine Vielzahl von Anwendungen neben ChatGPT, die auf generativer KI basieren, und zusätzlich zur Erzeugung von Texten sind auch Bilder, Videos oder Audios in den Blickpunkt gerückt.

Diese Entwicklungen haben auch Hessen und die hessische Landesverwaltung erreicht. Hier existiert mit der im Jahr 2022 veröffentlichten und in einem ressortübergreifenden Prozess entwickelten hessischen KI-Zukunftsagenda bereits ein strategischer Rahmen, der auch die großen Potenziale der Nutzung von KI in der Landesverwaltung hervorhebt. Zugleich betont die KI-Agenda, dass stets der Nutzen für die Bürgerinnen und Bürger, die Unternehmen und die Beschäftigten sowie ein verantwortungsbewusster Umgang mit dieser Technologie im Mittelpunkt stehen müssen. Dies gilt auch für den speziellen Bereich der generativen KI.

Der Umgang mit generativer KI bringt zahlreiche Fragen für die Landesverwaltung mit sich. Neben ressortspezifischen Herausforderungen – z. B. der Nutzung von generativer KI in den Schulen und Hochschulen des Landes – existieren ressortübergreifende Fragen und Ansatzpunkte. Hinzu kommt, dass sich der Bereich der generativen KI rasant weiterentwickelt. Darüber hinaus variiert der Wissensstand der Beschäftigten innerhalb der Landesverwaltung sehr stark: Teils werden bereits erste Projekte aufgegleist, teils existiert durch die private Nutzung solcher Tools großes Interesse an einer Nutzung im beruflichen Kontext, teils ist aber auch noch keinerlei Erfahrung mit solchen Anwendungen vorhanden.

Vor diesem Hintergrund wurde aus dem Arbeitskreis KI-Agenda – einem ressortübergreifenden Arbeitskreis, der vom Hessischen Ministerium für Digitalisierung und Innovation geleitet wird – heraus eine Projektgruppe Generative KI gegründet. Diese hat sich von Oktober bis Dezember 2023 intensiv mit dem Einsatz von generativer KI in der Landesverwaltung befasst. Unter gemeinsamer Federführung von HMD und HZD sowie unter Beteiligung des HMdJ, des HMWVW, des HMdI und der Staatskanzlei wurde das vorliegende Einführungsdokument erarbeitet. Der HBDI wurde zu Fragen des Datenschutzes beratend einbezogen.

Das Einführungsdokument bündelt den aktuellen Wissensstand zu generativer KI für Beschäftigte der Landesverwaltung, sowohl in theoretischer Sicht als auch mit Blick auf erste konkrete Projekte. Es richtet sich an Führungskräfte, die sich über den potenziellen Einsatz von generativer KI informieren wollen, aber auch an Mitarbeiterinnen und Mitarbeiter, die einen Einblick in das Thema erhalten möchten. Zugleich bietet es eine Grundlage für die notwendigen nächsten Schritte, d. h. die Entwicklung rechtlich verbindlicher Regularien zum Umgang mit KI in der Landesverwaltung und das Aufsetzen von Pilotprojekten. Die Entwicklung eines Muster- und Regelungsentwurfs zur rechtssicheren Nutzung von generativer KI in der hessischen Landesverwaltung ist bereits in Planung.

Um diesen Anforderungen gerecht zu werden, stellt das Einführungsdokument zunächst Grundwissen zu generativer KI vor. Hierauf aufbauend werden Einsatzbereiche in der öffentlichen Verwaltung geschildert, bevor ein Blick auf die organisatorischen Fragestellungen und technologischen Grundlagen und Entwicklungen geworfen wird. Im abschließenden Teil werden Handlungsempfehlungen formuliert, die dabei helfen sollen, über das weitere Vorgehen zu entscheiden, und um letztlich im nächsten Schritt einen Rahmen für die Nutzung von generativer KI in der Landesverwaltung zu definieren. Dieser sollte es unter Beachtung der aufgeworfenen ethischen und rechtlichen Fragen ermöglichen, dass die Bürgerinnen und Bürger und die Beschäftigten des Landes bestmöglich von den Potenzialen dieser Anwendungen profitieren können.

3 Grundwissen

Der Begriff generative KI umfasst jede Art von KI, die neue Texte, Bilder, Video- oder Audioinhalte erzeugen kann. Technisch gesehen greift diese Form der KI auf Trainingsdaten zurück und erzeugt aus zugrunde liegenden Mustern neuartige Ergebnisse.

3.1 Einordnung und Unterscheidung zu anderen Technologien

Eine **Abgrenzung von generativer KI zu anderen, bereits etablierten Systemen bzw. Technologien** wird im Folgenden vereinfacht und überblicksartig dargestellt:

Generative KI (→Glossar) **und Large Language Models (LLMs)** (→Glossar)**:** Diese Modelle zeigen, dass sie eine Vielzahl von Aufgaben mit einem einzigen, vielseitigen Modell bewältigen können. Sie nutzen die menschliche Sprache, um Aufgaben durch sogenannte Prompts (→Glossar) zu erhalten. Diese Modelle sind in der Lage, neuartige Inhalte zu generieren, Fragen zu beantworten, Texte zu verfassen und viele andere Aufgaben durchzuführen, die zuvor jeweils spezialisierte Modelle erforderten.

Hintergrundwissen: Foundation Models

LLMs gehören zur Kategorie der *Foundation Models* (→Glossar). Für diese Modelle gibt es fünf Hauptmerkmale:

- **Vortrainiert** – unter Verwendung großer Datenmengen und massiver Rechenleistung, sodass es ohne zusätzliches Training einsatzbereit ist.
- **Verallgemeinert** – ein Modell für viele Aufgaben (im Gegensatz zur traditionellen KI, die spezifisch für eine Aufgabe wie z. B. die Bilderkennung ist).
- **Anpassungsfähig** – durch Eingabeaufforderungen.
- **Groß** – in Bezug auf die Größe des Modells und die Datenmenge. Zum Beispiel wurden einige Modelle mit mehr als 500 Milliarden Wörtern trainiert. Um auf diese Anzahl zu kommen, müssten mehr als zehn Menschen ihr Leben lang ununterbrochen lesen!
- **Selbstüberwacht** – es werden keine spezifischen Bezeichnungen (*Annotationen* (→Glossar)) vorgegeben und das Modell muss aus den Mustern in den bereitgestellten Daten lernen.

Spezialisierte KI-Modelle: Diese Modelle sind in der Lage, komplexe Aufgaben wie *Sentiment-Analyse* (→Glossar), *Intent-Erkennung* (→Glossar), *Entity-Extraktion* (→Glossar) und *Klassifikation* (→Glossar) durchzuführen. Jedes dieser Modelle ist für eine spezifische Funktion optimiert und zeigt in seinem Bereich außergewöhnliche Leistung. Trotz ihrer Effizienz sind diese Modelle jedoch in ihrer Anwendung begrenzt und können nicht über ihre spezifischen Funktionen hinaus eingesetzt werden.

Expertensysteme: Diese Systeme stützen sich auf vorprogrammierte Algorithmen und Regeln. Expertensysteme sind damit in der Lage, spezifische Aufgaben in eng definierten Domänen zu lösen. Sie basieren auf einem festen Satz von Regeln, die von Expertinnen und Experten auf ihrem jeweiligen Gebiet erstellt werden. Diese Systeme sind effektiv für Aufgaben mit klaren, logischen Schritten, aber sie können nicht mit neuen, unbekannten Situationen umgehen oder von Erfahrungen lernen.

Alle oben angeführten Technologien bzw. Modelle haben ihre Stärken und Schwächen. Welches das jeweils passende ist, ist immer abhängig vom Anwendungsfall und den konkreten Rahmenbedingungen.

3.2 Wie „denkt" generative KI?

Generative KI „denkt" nicht im menschlichen Sinne. KI stellt keine digitale Reproduktion des menschlichen Denkvermögens dar, sondern sie verwendet künstliche neuronale Netze (KNN) mit einer Vielzahl von Parametern, um Muster in großen Datenmengen zu erkennen (Maschinelles Lernen / *Machine Learning* (→ Glossar)). Das generative KI-Modell generiert eine Antwort basierend auf Wahrscheinlichkeiten und Verbindungen, die während des Lernprozesses antrainiert wurden. Die generative KI reflektiert nicht Bewusstsein oder menschliches Denken, sondern reproduziert auf intelligente Weise Muster aus den Trainingsdaten und gibt dabei lediglich Wahrscheinlichkeiten aus. Die KI versteht auch Sprache nicht im eigentlichen Sinne, sondern setzt ihre Antworten ausgehend von den Trainingsdaten und daraus ermittelten Wahrscheinlichkeiten zusammen. Ein einfaches Beispiel kann wie folgt lauten: Der Satz „Das Auto biegt nicht rechts ab, sondern […]." wird in einer entsprechenden Anfrage durch die generative KI mit „links" vervollständigt, weil diese Antwort (trotz mehrerer denkbarer Alternativen) die wahrscheinlichste ist. Es liegt auf der Hand, dass diese Arbeitsweise umso präziser ist, je mehr Daten verfügbar sind und je umfassender die KI durch menschliche Rückmeldung darauf trainiert wird, dass die gelieferten Antworten auch den menschlichen Erwartungen entsprechen.

Künstliches neuronales Netz (KNN) (→ Glossar)

Im vorherigen Abschnitt wurde erwähnt, dass KNNs verwendet werden, um Inhalte zu generieren. Aber was ist überhaupt ein KNN? Die wohl bemerkenswerteste Eigenschaft eines KNN ist seine Fähigkeit, im Prinzip jede beliebige Funktion zu approximieren. Nehmen Sie z. B. an, dass Sie den Preis eines Hauses vorhersagen wollen. In einer einfachen linearen Gleichung könnten Sie ihn mit der Wohnfläche korrelieren. Stellen Sie sich nun vor, dass Sie für eine bessere Vorhersage auch viele andere Variablen, z. B. die Nähe zur Stadt, die Infrastruktur, die Nachbarschaft, potenzielle Arbeitgeber usw. in Ihre Berechnung einbeziehen wollen, Sie aber gar nicht wissen, in welcher Weise diese Faktoren den Preis beeinflussen. Hier kommt das KNN ins Spiel, welches in der Lage ist, mit existierenden Daten zu Häusern und ihren Preisen die zugrunde liegende Funktion zu erlernen und anzunähern und dabei komplexe, nicht lineare Zusammenhänge zu erfassen. Dank dieser Anpassungsfähigkeit können KNN komplizierte Beziehungen innerhalb der Daten lernen und ihr Wissen verallgemeinern, um Vorhersagen zu treffen, wenn sie mit neuen, ungesehenen Informationen konfrontiert werden.

Technisch gesehen sind KNN Modelle, die von der Struktur und Funktionsweise des menschlichen Gehirns inspiriert sind. Sie bestehen aus künstlichen Neuronen, die in verschiedenen Schichten organisiert sind. Das Netzwerk nimmt Eingabedaten entgegen, die als Vektoren repräsentiert werden. Jede Verbindung zwischen den Neuronen hat ein zugehöriges Gewicht, das die Stärke der Verbindung repräsentiert. Die Eingabewerte werden mit diesen Gewichten multipliziert. Die gewichteten Eingabewerte werden sodann summiert und es wird eine Aktivierungsfunktion auf das Ergebnis angewendet. Es folgt die Ausgabe des Ergebnisses dieses Verarbeitungsvorgangs. Im Hintergrund laufende Algorithmen passen die Gewichtungen anhand gegebener Rückmeldungen an, um die Vorhersagegenauigkeit zu optimieren.

Weil KNN aus mehreren Ebenen bestehen, werden diese Netzwerke auch als „tief" bezeichnet. Durch die vielen Schichten können die Netzwerke automatisch Merkmale und Hierarchien von Merkmalen lernen (*Deep Learning* (→Glossar)).

Daten als Input für KI

Die Qualität der Ergebnisse generativer KI ist stark abhängig von der verfügbaren Datengrundlage – von deren Qualität wie auch Quantität. Je sauberer und besser annotiert die Daten sind, desto besser kann das Modell sinnvolle Muster lernen. Je größer die Datengrundlage ist, desto breiter können die zu erlernenden Muster gestreut sein und desto zuverlässiger wird das Modell auf unterschiedliche Eingaben reagieren. Idealerweise sind die verwendeten Daten zudem repräsentativ, möglichst vielfältig, aktuell und spiegeln ein ausgewogenes Verhältnis von Klassen und Merkmalen wider, um Verzerrungen oder Vorurteile zu vermeiden. Modelle, die auf unzureichenden oder verzerrten Daten trainiert wurden, neigen dazu, suboptimale oder möglicherweise auch problematische Ergebnisse zu erzeugen (z. B. Fortschreibung einer Diskriminierung durch die KI, die in der Datengrundlage angelegt ist, beispielsweise bei der Kreditvergabe).

3.3 Was kann ich von KI-generierten Ergebnissen erwarten?

Grundsätzlich kann generative KI zu einer gesteigerten Effizienz beitragen und insbesondere auch die öffentliche Verwaltung entlasten. Dies kann vornehmlich dadurch geschehen, dass repetitive, einfache Aufgaben ohne (größeren) Entscheidungsspielraum auf eine KI übertragen werden. Dies kann beispielsweise die Erfassung und Eingabe von Metadaten in verschiedenste Dokumente betreffen. Es ist zudem vorstellbar, vorhandene Datensätze zu erfassen und so komplexere Entscheidungsfindungen mit fundierten Datengrundlagen und Analysen zu ermöglichen bzw. zu unterstützen. Perspektivisch kann auch bei der Erstellung von Dokumenten wie Vermerken, Berichten etc. eine unterstützende KI zum Einsatz kommen. Zusätzlich bietet sich die ergänzende Kommunikation mit Bürgerinnen und Bürgern an, etwa durch KI-gestützte Chatbots. Dadurch können perspektivisch Effektivität und Effizienz der öffentlichen Verwaltung gesteigert werden und dem sich abzeichnenden Fachkräftemangel begegnet werden.

Je geringer die Auswirkungen sind, die potenzielle Fehler in einem spezifischen Bereich haben, und je einfacher sich diese korrigieren lassen, desto eher eignet sich dieser Bereich für den Einsatz von generativer KI. Der Fokus sollte dabei auf einem unterstützenden und begleiteten Einsatz liegen und nicht auf einer vollständigen Automatisierung (Vermeidung einer *Dunkelverarbeitung* (→Glossar)).

Weiterführende Informationen finden Sie im Kapitel *5.2 Erwartungsmanagement.*

3.4 Kann ich den Ergebnissen vertrauen?

Wer generative KI nutzt, muss sich stets darüber im Klaren sein, dass ihm bzw. ihr die Letztverantwortlichkeit für sämtliche mit Unterstützung von generativer KI durchgeführte Tätigkeiten obliegt. Angesichts der beschriebenen Funktionsweise der generativen KI, insbesondere dem wahrscheinlichkeitsbasierten Ansatz, muss stets berücksichtigt werden, dass die gefundenen Ergebnisse keinen Wahrheitsanspruch erfüllen. Vielmehr bilden die Ergebnisse lediglich eine Prognose, beruhend auf einer (mehr oder weniger großen) Datengrundlage über die erwartete Antwort.

Dies kann insbesondere in folgenden Bereichen Auswirkungen haben: Generative KI kann Schwierigkeiten haben, den Kontext in komplexen oder mehrdeutigen Situationen zu verstehen. Sie kann aufgrund vorurteilhafter und verzerrter Datengrundlagen ihrerseits zu unethischen oder diskriminierenden Ergebnissen führen. Generative KI ist zudem nicht in der Lage, Kausalitäten festzustellen oder auszuschließen. Gefundene Muster können irreführend in eine entsprechende Richtung interpretiert werden. Auch in Hinblick auf Nachvollziehbarkeit und Erklärbarkeit der gefundenen Ergebnisse bestehen deutliche Defizite.

Dementsprechend muss (generative) KI verantwortungsbewusst eingesetzt werden, insbesondere in sensiblen oder kritischen Anwendungsbereichen. Ein komplexes und wichtiges Thema bei KI-Modellen sind Verzerrungen (sog. *Bias* (→Glossar)), die in KI-Modellen auf verschiedene Weise auftreten können. Diskriminierung oder Verstärkung bestehender Vorurteile können die Folgen sein wie auch falsche oder ungerechte Ergebnisse.

Es muss ferner ein Bewusstsein dafür geschaffen werden, dass auch der menschliche Umgang mit den durch die KI gelieferten Ergebnissen problematisch sein kann. Menschen neigen dazu, sich stark von den ersten zu einer Thematik gesammelten Informationen beeinflussen zu lassen (sog. *Ankereffekt* (→Glossar)). Im menschlichen Umgang mit generativer KI kann sich dieser Effekt zeigen und verstetigen. Auch in Zeiten großer Arbeitsbelastung dürfen gelieferte Informationen bzw. vorgeschlagene Entscheidungen nicht ungeprüft übernommen werden, sondern können nur als Grundlage der notwendigen menschlichen Bearbeitung dienen und müssen kritisch beleuchtet und hinterfragt werden. Je geringer die potenziellen Fehlerfolgen sind und je größer der nachgelagerte Bereich menschlicher Bearbeitung der gelieferten Ergebnisse ist, desto erfolgversprechender und sicherer stellt sich der Einsatz von generativer KI dar.

Es liegt auf der Hand, dass vor diesem Hintergrund das Verwaltungshandeln an sich nicht vollständig an generative KI übertragen werden kann und darf. Menschliche Entscheidungsträgerinnen und -träger können und sollen in keiner Weise ersetzt werden. Der Fokus liegt auf einem unterstützenden Einsatz zugunsten der Bediensteten und der Bürgerinnen und Bürger.

Auf den Umgang mit KI-generierten Inhalten in der öffentlichen Verwaltung wird im ➜ *Kapitel 5 Organisatorische Fragestellungen* näher eingegangen.

Hintergrundwissen: Bias

Der Begriff Bias bezeichnet in Bezug auf generative KI und LLM eine systematische Verzerrung oder eine unausgewogene Darstellung bestimmter Informationen oder Perspektiven. Dies kann durch unausgewogene Trainingsdaten, ungleiche Gewichtung verschiedener Datenquellen oder inhärente Vorurteile in der Modellarchitektur entstehen.

Bezogen auf den Umgang mit KI-generierten Inhalten ergeben sich außerdem kognitive Bias. Als *Automation Bias* bezeichnet man eine unzureichende Überprüfung von automatisierten Inhalten und dabei eine Zuschreibung einer zu hohen Verlässlichkeit auf das (z. B. KI-)System. Im Gegensatz dazu kann ein *Algorithm Aversion Bias* dazu verleiten, algorithmenbasierte Systeme selbst bei Wissen um eine hohe Ergebnisqualität abzulehnen. Eine weitere kognitive Verzerrung ist der *Ankereffekt* (englisch *„Anchoring Bias"*), bei dem der KI-generierte Inhalt als fester Bezugspunkt für anschließende Entscheidungen herangezogen wird. Dies kann zu fehlerhafter Entscheidungsfindung und zur Verbreitung falscher oder voreingenommener Annahmen führen, die auf unvollständigen oder ungenauen Ausgangsinformationen beruhen.

3.5 Was sind typische Vorurteile gegenüber generativer KI?

Die Nutzung von generativer KI stößt häufig auf Skepsis. Im Folgenden werden einige der typischen Vorurteile kurz kommentiert.

Die Antworten einer generativen KI sind intransparent und nicht nachvollziehbar.

Dies trifft nicht pauschal zu. Im Moment werden Mechanismen in diesem Bereich entwickelt, die das Thema Nachvollziehbarkeit adressieren – Stichwort *Explainable AI (XAI)* [→Glossar]. Es gibt bereits erste Anbieter von Modellen, die diese Mechanismen nutzen.

Zudem gibt es Anwendungsfälle und Lösungsansätze, die einen hohen Grad an Nachvollziehbarkeit erreichen, z. B. bei der Einbindung von eigenen Wissensdatenbanken in Verbindung mit Quellangaben. Mehr dazu in ➜ *Kapitel 6.3 Wie kann ich mein internes Organisationswissen mit generativer KI nutzen?*

Alle generativen KI-Systeme halluzinieren und sind daher nicht seriös nutzbar.

Dies trifft nicht pauschal zu. Bei der Anwendung von generativer KI besteht grundsätzlich die Möglichkeit von *„Halluzination"* [→Glossar], d. h. das Generieren von plausibel wirkenden, aber nicht auf Daten und Fakten basierenden Inhalten. Während Halluzinationen beim Erstellen kreativer

Inhalte ein interessanter und nützlicher Aspekt von generativer KI sind, werfen sie beispielsweise bei deren Anwendung als Rechercheassistenz Bedenken und Herausforderungen in Bezug auf die Zuverlässigkeit und Vertrauenswürdigkeit der generierten Inhalte auf. Allerdings gibt es mittlerweile zahlreiche Maßnahmen, dies zu erkennen und mehr oder weniger stark einzuschränken. Zusammenfassend kann man sagen, dass die konkrete Problematik einer möglichen Halluzination stark vom verwendeten Kontext bzw. dem Anwendungsfall abhängt. Eine Pauschalisierung ist daher nicht zielführend.

Generative KI verwendet meine Daten zum Trainieren des Modells.

Dies trifft nicht pauschal zu. Es gibt mittlerweile Anbieter, die das Trainieren bzw. Optimieren des Modells auf Basis von Nutzereingaben entweder ganz oder unter bestimmten Bedingungen (z. B. Art des Zugriffs, Nutzungsvariante usw.) ausschließen. Werden die Lösung und das Modell *On-Premises* (→Glossar) gehostet, so liegt die Handhabung ohnehin in der eigenen Hand. Mehr dazu in ➔ *Kapitel 6.2 Laufzeitumgebungen von generativer KI*.

Lösungen auf Basis generativer KI sind nicht datenschutzkonform nutzbar.

Dies trifft nicht pauschal zu. Wie bei jeder anderen Softwarelösung gibt es Strategien, um datenschutzrechtliche Vorkehrungen und Rahmenbedingungen zu schaffen, die z. B. einen DSGVO-konformen Betrieb der Lösung ermöglichen. Mehr dazu im ➔ *Kapitel 5 Organisatorische Fragestellungen*.

Damit ich generative KI mit meinen eigenen Daten nutzen kann, muss ich sie mit viel Aufwand antrainieren.

Dies trifft nicht zu. Um eigene Organisationsdaten in Verbindung mit generativer KI nutzen zu können, gibt es mittlerweile verschiedene Lösungsansätze und Szenarien, bei denen ein Trainieren eines Basismodells nicht nötig ist. Mehr dazu in ➔ *Kapitel 6.3 Wie kann ich mein internes Organisationswissen mit generativer KI nutzen?*

In jeder generativen KI steckt eine Art von Verzerrung (Bias), und deshalb kann sie nicht zuverlässig genutzt werden.

Ja, man muss davon ausgehen, dass grundsätzlich in jedem Modell im Bereich generativer KI Bias in irgendeiner Form enthalten ist. Sowohl die zugrunde liegenden Daten als auch das nachgelagerte Training wurden in großen Teilen mithilfe von Menschen erstellt. Es kommt also immer zu einer Form von Verzerrung bzw. Bias, da der Mensch nicht wertfrei agieren kann.

Allerdings stellt sich die Frage, ob dieser Umstand dazu führen sollte, pauschal auf generative KI zu verzichten. Es gibt auch Anwendungsfälle, in denen Bias entweder keine oder eine untergeordnete Rolle spielt, z. B. im technischen Kontext oder in der Softwareentwicklung.

Generative KI ist nicht nutzbar im Bereich der Entscheidungsfindung.

Dies trifft nicht pauschal zu. Automatisierte Entscheidungssysteme (engl. *Automated Descision-Making System, ADMS*) sind technische Systeme, die unterstützend oder auch vollständig

autonom komplexe Entscheidungen treffen. Aus verschiedenen Gründen wird generative KI im Bereich **autonome** Entscheidungsfindung zu Recht nicht eingesetzt.

Es gibt allerdings im Bereich **Assistenzsysteme** (hier ist der Mensch bei Entscheidungen einbezogen) sinnvolle Anwendungsfelder mit dem Ziel, die menschliche Arbeit zu unterstützen. Mögliche Maßnahmen zur Stützung dieses Ansatzes sind u. a. Risikoeinschätzung (Identifizierung und Minimierung potenzieller Risiken, z. B. in Hinblick auf Datenschutz, Datensicherheit, Zuverlässigkeit, Genauigkeit, Bias usw.) und Selbstdeklaration (z. B. Kennzeichnung).

3.6 Funktionsbereiche generativer KI

Nachfolgend werden beispielhaft einige Funktionsbereiche aufgeführt, in denen Anwendungen unter Verwendung generativer KI aktuell genutzt werden.

3.6.1 Text

- **Wissensmanagement**: Unternehmens- bzw. Organisationsdaten können analysiert und wichtige Informationen extrahiert werden, um beispielsweise Nutzerinnen und Nutzer bei der Entscheidungsfindung zu unterstützen.
- **Suche**: Generative KI kann in Suchmaschinen integriert werden, um die Suchanfragen der Benutzerinnen und Benutzer besser zu verstehen und präzisere Ergebnisse zu liefern.
- **Texterzeugung**: KI kann automatisch Artikel oder Berichte zu vorgegebenen Themen erstellen. Beispiele sind das Analysieren langer Dokumente oder Artikel und die Erstellung einer prägnanten Zusammenfassung.
- **Personalisierung**: Erstellung personalisierter Texte.
- **Textzusammenfassung**: Lange Dokumente oder Artikel können von KI analysiert und prägnant zusammengefasst werden.
- **Textbearbeitung**: Überprüfung von Texten auf Grammatik, Stil und Klarheit und Erstellung von Vorschlägen zur Verbesserung.
- **Extrahierung von Informationen**: Extrahieren spezifischer Informationen aus großen Textmengen.
- **Textklassifizierung**: Einteilung von Texten in Kategorien.
- **Übersetzung**: KI-basierte Übersetzungstools ermöglichen das schnelle und genaue Übersetzen von Texten in verschiedene Sprachen.
- **Sentiment-Analyse**: KI kann Meinungen und Emotionen in Texten analysieren, um beispielsweise die öffentliche Meinung zu bestimmten Themen oder die Kundenzufriedenheit zu erfassen.

- **Dialogsystem – Chatbot**: Generative KI-Modelle sind darauf trainiert, menschliche Sprache zu verstehen und zu generieren. Dies ermöglicht eine natürlichere und fließende Interaktion mit Benutzerinnen und Benutzern.
- **Programmiercode**: Schreiben, Interpretieren und Ausführen von Code in unterschiedlichsten Programmiersprachen wie beispielsweise Python, JavaScript, HTML und SQL.

3.6.2 Bild und Video

- **Automatisierte Bildbeschriftung und -annotation**: KI kann Bilder und Videos analysieren und automatisch mit relevanten Beschriftungen oder *Annotationen* versehen.
- **Bildgenerierung**: Generative KI kann neue Bilder basierend auf bestimmten Vorgaben wie Stil, Thema oder Farbschema erzeugen.
- **Videogenerierung**: Ähnlich der Bildgenerierung kann KI auch kurze Videos oder Animationen erstellen, die auf spezifischen Anweisungen oder Skripten basieren.
- **Bild- und Videobearbeitung**: KI kann automatisiert Fotos und Videos bearbeiten, um beispielsweise die Bildqualität zu verbessern, Objekte zu entfernen oder Farbkorrekturen vorzunehmen.
- **Gesichts- und Objekterkennung in Bildern und Videos**: KI kann Gesichter oder Objekte in Bild- und Videomaterial erkennen und klassifizieren.
- **3D-Modellierung**: Generative KI kann dazu verwendet werden, realistische 3D-Modelle von Objekten oder Landschaften zu erstellen.

3.6.3 Audio

- **Automatisierte Transkription**: KI-basierte Systeme können gesprochene Worte in Text umwandeln, was bei der Untertitelung oder bei der Erstellung von Sitzungsprotokollen hilfreich ist.
- **Sound Design**: Soundeffekte und Klanglandschaften für Filme, Spiele und virtuelle Realitäten können mittels KI erstellt werden.
- **Audio-Restaurierung und -Verbesserung**: Mittels KI können beschädigte oder schlecht aufgenommene Audiodateien verbessert, Hintergrundgeräusche reduziert und die Klangqualität erhöht werden.
- **Spracherkennung und -verständnis**: Generative KI kann nicht nur Sprache erkennen, sondern auch den Kontext und die Absichten hinter den Worten „interpretieren".
- **Personalisierung von Audioinhalten**: KI kann Audioinhalte basierend auf den Vorlieben und dem Hörverhalten der Nutzerin bzw. des Nutzers personalisieren.

3.7 Aktuelle Beispiele generativer KI

Um die Potenziale und Grenzen von generativer KI zu verstehen, ist es hilfreich, sich konkrete Beispiele anzusehen. Exemplarisch sollen drei unterschiedliche Modelle betrachtet werden, von denen ChatGPT die vermutlich größte Bekanntheit weltweit besitzt. Auch in Deutschland werden leistungsstarke Werkzeuge entwickelt, wie die beiden anderen Beispiele *Luminous* [→Glossar] und *LeoLM* [→Glossar] zeigen.

3.7.1 Was ist ChatGPT?

Noch eine Erklärung?

Obwohl ChatGPT bereits in aller Munde ist, soll kurz darauf eingegangen werden. Nicht zuletzt aufgrund der aktuellen Marktführerschaft im Bereich generativer KI-Werkzeuge und des nahezu monatlich wachsenden Funktionsumfangs darf ein Überblick und eine grobe Einordnung nicht fehlen. Zudem wird nicht selten in Gesprächen das Wort „ChatGPT" als Synonym für das Thema generative KI verwendet. Diese sprachliche Vereinfachung ist vergleichbar mit analogen Themen in der Vergangenheit: So wurde „Facebook" lange Zeit als Synonym für Social Media Tools verwendet, und noch heute wird „googeln" als Synonym für suchen oder recherchieren im Internet genutzt.

Ganz kurz

ChatGPT, das für *„Chat Generative Pre-trained Transformer"* [→Glossar] steht, ist ein auf einem großen KI-Sprachmodell bzw. LLM basierender Chatbot, der von dem US-amerikanischen Unternehmen OpenAI entwickelt und am 30. November 2022 eingeführt wurde. Das zugrunde liegende KI-Sprachmodell wurde auf riesigen Datenmengen, z. B. Büchern und Webseiten, vortrainiert, um relevante und semantisch kohärente natürliche Sprache zu erzeugen.

Das Tool wird in der Regel als Webanwendung oder als mobile App genutzt. Technisch betrachtet ist ChatGPT eine Cloudanwendung (Software as a Service – SaaS), die auf einem amerikanischen Server gehostet wird. Mit ChatGPT können Nutzerinnen und Nutzer in einen Dialog zu nahezu jedem erdenklichen Thema treten. Es beantwortet Fragen oder schreibt Texte, die als Grundlage für weitere Anfragen dienen können.

Der Kenntnisstand (z. B. Aktualität und Umfang der Wissensbasis) und die Leistungsfähigkeit von ChatGPT hängen sehr stark vom dahinter genutzten KI-Modell ab. So ist das Modell *GPT-4* [→Glossar] Turbo im Vergleich zu GPT-3 aktueller (Wissenstand Anfang 2023 versus 2021) und um Längen leistungsfähiger. Hinzu kommt z. B. die Möglichkeit zur Spracheingabe und Sprachausgabe.

Einige Beispiele für die Verwendung von ChatGPT

ChatGPT ist vielseitig und kann für mehr als nur Unterhaltungen verwendet werden. Hier werden einige Beispiele aufgeführt:

- E-Mails entwerfen
- Computerprogramme programmieren
- Zusammenfassen von Artikeln, Podcasts oder Präsentationen
- Komplexe Themen einfacher beschreiben
- Übersetzungen vornehmen

Multimodalität (→ Glossar)

Während zunächst nur Text verarbeitet wurde, können Nutzerinnen und Nutzer Bilder oder Audiodateien als Eingabe verwenden.

3.7.2 Was ist Luminous?

Das Luminous-Modell ist ein LLM, ähnlich wie die GPT-Modelle von OpenAI. Es stammt von Aleph Alpha, einem deutschen Unternehmen, das sich auf die Entwicklung von KI-Technologien spezialisiert hat.

Ein wichtiger Aspekt von Luminous ist, dass es speziell für den europäischen Markt entwickelt wurde, mit einem Fokus auf Datenschutz und Konformität mit den restriktiven Datenschutzgesetzen der EU.

Alle Modellvarianten wurden in den fünf am häufigsten gesprochenen europäischen Sprachen trainiert: Englisch, Deutsch, Französisch, Italienisch und Spanisch. Wie ChatGPT und im Gegensatz zu vielen anderen großen Sprachmodellen ist die Luminous-Familie multimodal, das heißt, sie kann auch mit Bildern arbeiten.

3.7.3 Was ist LeoLM?

LeoLM ist ein LLM, entwickelt von der deutschen Organisation LAION, die sich auf die Forschung und Entwicklung im Bereich KI spezialisiert hat, und dem Hessischen Zentrum für Künstliche Intelligenz hessian.AI. Dieses vereint die KI-Kompetenzen von dreizehn hessischen Hochschulen und bietet Spitzenforschung sowie anwendungsorientierte Forschung und Transfer.

LeoLM ist ein kommerziell nutzbares, quelloffenes LLM mit Fokus auf die deutsche Sprache. Es basiert auf dem LLM Llama 2 des Unternehmens Meta und wurde mit einem umfangreichen Korpus deutscher und landesspezifischer qualitativ hochwertiger Texte trainiert.

3.8 Wissenswertes

3.8.1 Die Grenzen aktueller Sprachmodelle

Viele große Sprachmodelle bringen trotz ihrer beeindruckenden Leistung Nachteile mit sich.

- **Begrenzte Aktualität**: LLMs sind auf den Stand ihrer letzten Trainingssession beschränkt und können daher nicht auf Informationen zugreifen, die nach diesem Zeitpunkt veröffentlicht wurden.
- **Mangel an domänenspezifischem Wissen**: Obwohl sie für eine breite Palette von Aufgaben trainiert wurden, fehlt LLMs oft das spezifische Wissen, das für die besonderen Anforderungen eines bestimmten Unternehmens bzw. einer bestimmten Organisation relevant ist.
- **Undurchsichtigkeit in der Informationsquelle**: Es ist oft nicht nachvollziehbar, auf welche Quellen sich ein LLM stützt, um seine Antworten zu generieren, was zu einer Blackbox-Problematik führt.
- **Hoher Ressourcenbedarf**: Die Entwicklung und Implementierung dieser Modelle ist mit erheblichem Aufwand und hohen Kosten verbunden, was sie ineffizient machen kann.

3.8.2 Was ist Prompt Engineering? (→ Glossar)

Im Zusammenhang mit generativer KI bzw. LLMs wird nicht selten der Begriff Prompt Engineering verwendet. Prompt Engineering ist das geschickte Formulieren von Anweisungen oder Fragen, um aus einer generativen KI die gewünschten Antworten oder Ergebnisse zu erhalten. Es wird benötigt, weil die Qualität der Eingabe (der *Prompt* (→Glossar)) oft die Qualität der Ausgabe (die Antwort oder das generierte Ergebnis) bestimmt. Das Ziel ist es, die Anfrage so zu gestalten, dass die KI möglichst präzise versteht, was der Nutzer bzw. die Nutzerin möchte.

Unterschiedliche KI-Modelle gehen unterschiedlich mit ein und demselben Prompt um. Zudem entwickeln sich die KI-Modelle immer weiter. Das bedeutet, dass einmal zurechtgelegte Prompts immer wieder neu getestet werden sollten.

Prompt-Beispiele

Ein einfacher Prompt: „Erkläre die Relativitätstheorie."

Ein etwas erweiterter Prompt: „Erkläre die Grundprinzipien der Relativitätstheorie und ihre Auswirkungen auf die moderne Physik. Stelle das Ergebnis in Form einer Pressemitteilung dar."

Tipps für den Umgang mit Prompts

Im Internet finden sich unzählige Hilfestellungen zum Thema Prompt Engineering und der effektiven Nutzung von Prompts. Sie werden nicht selten als „Cheat Sheets" veröffentlicht.

Exemplarisch seien hier einige grundsätzlichen Hilfestellungen genannt:

- **Geben Sie einen Kontext.**
  - Schildern Sie die aktuelle Situation, um die Problemstellung möglichst eindeutig und umfassend zu beschreiben.
- **Nutzen Sie Anführungszeichen, um wichtige Teile des Prompts zu betonen.**
- **Weisen Sie dem System eine Rolle zu.**
  - Das System sollte wissen, welche Rolle oder Perspektive eingenommen werden soll.
- **Geben Sie klare Anweisungen.**
  - Zum Beispiel: „Analysiere …“, „Schreibe …“, „Nenne mir …“
  - Vermeiden Sie irrelevante Informationen im Prompt.
- **Führen Sie Beispiele an.**
  - Hilfreiche Ergänzungen für präzisere Antworten sind positive und negative Beispiele.
- **Begrenzen Sie die Antwortlänge.**
  - Zum Beispiel: „Fasse mir den folgenden Artikel in nicht mehr als 500 Wörtern zusammen.“
- **Verfeinern Sie die Prompts via Konversation.**
  - Iteratives Ausbessern – z. B. wenn Sie mit der Antwort nicht zufrieden sind.
  - Unterstützt das System eine Konversation (z. B. bei ChatGPT), so kann die Antwort mit weiteren Anweisungen konkretisiert oder erweitert werden.

In Zusammenhang mit Prompts und einem möglichen Missbrauch stehen auch die Begriffe *Prompt Injection* (→ Glossar) und *Indirect Prompt Injection* (→ Glossar), die im Glossar näher erläutert werden.

4 Einsatzbereiche in der öffentlichen Verwaltung

Während die öffentliche Verwaltung in den vergangenen Jahren bereits zahlreiche KI-Projekte durchgeführt hat und bereits Erfahrungen hinsichtlich sinnvoller Anwendungsbereiche sammeln konnte, ist die generative KI ein neues Werkzeug. Daher werden aktuell in verschiedenen Behörden Anwendungsfelder für generative KI identifiziert.

4.1 Anwendungsfelder

Bei der Definition von Einsatzbereichen von KI in der öffentlichen Verwaltung stellt sich stets die Frage, wo die Verwendung von KI einen Mehrwert bietet, denn der Einsatz von KI ist kein Selbstzweck. Die „klassische" Einschätzung hierzu lautet, dass der Einsatz von KI in der Verwaltung vor allem dort sinnvoll ist,

- wo es stark strukturierte Prozesse gibt,
- wo große Datenmengen ausgewertet werden müssen,
- wo Verwaltungen mit einer großen Zahl ähnlich gelagerter Fälle umgehen müssen.

Durch die Nutzung von generativer KI können diese klassischen Anwendungsfelder deutlich erweitert werden.

Typische Anwendungsfelder von generativer KI in der öffentlichen Verwaltung bestehen im Bereich Texterzeugung. Hier finden sich auch zahlreiche der in Kapitel 4.1 genannten Anwendungsfälle wieder.

So kann generative KI im Wissensmanagement von Behörden eingesetzt werden. Sie kann die Suche nach Informationen erleichtern und eine unkomplizierte Einarbeitungsphase für neue Mitarbeiterinnen und Mitarbeiter ermöglichen. Auch in Bereichen, in denen große Textmengen ausgewertet und analysiert werden müssen, wie der Justiz, kann generative KI unterstützend eingesetzt werden. Generative KI kann bei Recherchen und bei der Erstellung von Vermerken, Grußworten oder Terminvorbereitungen unterstützen. Mit Anwendungen, die auf generativer KI basieren, können zudem vorhandene Texte in leichte Sprache überführt oder in andere Sprachen übersetzt werden.

Behörden können aber nicht nur im Bereich Text, sondern auch in den Feldern Bild, Video und Audio vom Einsatz generativer KI profitieren. So kann die automatisierte Bildbeschriftung und *-annotation* beispielsweise die Arbeit von Archiven unterstützen, und Pressestellen können von der KI-gestützten Bild- und Videobearbeitung profitieren. Schließlich entstehen in Behörden täglich zahlreiche Protokolle von Sitzungen und Veranstaltungen. Hier kann die automatisierte Transkription und automatisierte Erstellung von Protokollen einen erheblichen Mehrwert darstellen.

4.2 Katalog möglicher Anwendungsfälle

Eine Zusammenstellung von möglichen Anwendungsfällen (*Use Cases*) beschreibt und veranschaulicht, wie generative KI in der Landesverwaltung in der Praxis eingesetzt werden könnte. Hierzu finden Sie im **Anhang 1 – Katalog möglicher Anwendungsfälle** eine erste Sammlung aus dem Bereich der hessischen Landesverwaltung. Diese möglichen Anwendungsfälle werden gegenwärtig, auch im Rahmen von Machbarkeitsprüfungen, evaluiert.

Übersicht der Anwendungsfälle

1. **Leichte/einfache Sprache – Teilhabe ermöglichen**
   (am Beispiel Pressemitteilung)
2. **Textzusammenfassung – komplexe/lange Dokumente schnell überblicken**
   (am Beispiel Stellungnahme zu Gesetzesentwurf)
3. **Texterstellung – Anfragen effizienter beantworten**
   (am Beispiel Antwort auf Bürgeranschreiben)
4. **Texterstellung – Standarddokumente effizient verfassen**
   (interne Verwendung)
5. **Bildgenerierung – passende Illustrationen effizient erzeugen**
6. **Chatbot – Beantwortung wiederkehrender Fragestellungen**
   (Ergänzung zum internen Mitarbeiterportal MAP)
7. **Programmcodeentwicklung bzw. -analyse – Optimierung des Softwareentwicklungsprozesses**
8. **KI-Unterstützung zur Erhöhung der IT-Sicherheit**

Der Anwendungsfälle-Katalog für generative KI bietet einen systematischen Überblick über denkbare erste Einstiege in diese Technologie und trägt dazu bei, deren Leistungsfähigkeit zu verdeutlichen und das Verständnis dafür zu vertiefen. Er verhilft gleichzeitig dazu, das Anwendungspotenzial für die hessische Landesverwaltung zu ermitteln und zu evaluieren.

5 Organisatorische Fragestellungen

In der öffentlichen Verwaltung existieren bisher nur wenige Anwendungsbeispiele für die Nutzung von generativer KI, die bereits in der Praxis umgesetzt wurden. Zudem fehlt es aktuell noch an Leitlinien, an denen sich die Landesbeschäftigten bei der Nutzung von generativer KI orientieren können. Im Folgenden wird daher zunächst auf den bestehenden Regulierungsbedarf eingegangen, bevor organisatorische Fragestellungen aufgegriffen werden, denen sich Behörden stellen sollten, wenn sie die Nutzung von generativer KI in Betracht ziehen.

5.1 Regulierungsbedarf

Bei der Nutzung von Anwendungen, die auf generativer KI basieren, können sich verschiedene juristische Problemfelder ergeben. Die genauen Fragestellungen hängen u. a. davon ab, ob es sich um frei zugängliche Anwendungen handelt oder um Software, die speziell für Behörden angeboten wird. Typische Herausforderungen sind u. a. die folgenden:

- **Urheberrecht**: Anwendungen, die auf generativer KI basieren, ziehen oftmals eine Vielzahl an urheberrechtlich geschützten Inhalten als Datenbasis heran.
- **Haftungsrecht**: Es stellt sich die Frage, wer für fehlerhafte Tipps, Handlungsanweisungen und Informationen haftet.
- **Strafrecht**: Es ist fraglich, wer verantwortlich ist, wenn strafrechtlich relevante Inhalte veröffentlicht werden.
- **Arbeitsrecht / Dienstrecht**: Hier besteht die Herausforderung, dass sich die Qualität eigener Leistung bei der Nutzung von generativer KI kaum feststellen lässt.
- **Deutsches und europäisches Datenschutzrecht**: Hier steht infrage, inwieweit die Einhaltung dieser Regelungen sichergestellt ist und wie mit personenbezogenen Daten umgegangen wird.

Die Nutzung von generativer KI in der hessischen Landesverwaltung ist bislang nicht geregelt. Es besteht ein Bedarf an Regelungen, die den Beschäftigten einen Rahmen für die generelle Zulässigkeit und einen rechtssicheren Gebrauch von Anwendungen aus dem Bereich der generativen KI geben. Die Entwicklung eines Muster- und Regelungsentwurfs zur rechtssicheren Nutzung von generativer KI in der hessischen Landesverwaltung ist in Planung.

Hintergrundwissen: Die europäische KI-Verordnung

Die Entwicklung der europäischen KI-Verordnung stellt den weltweit ersten Versuch dar, ein umfassendes Gesetz zur Regulierung von KI auf den Weg zu bringen. Bereits im Jahr 2021 hat die Europäische Kommission den Entwurf der Verordnung lanciert. Nach der Abstimmung im Europäischen Parlament am 14. Juni 2023 begann der sogenannte Trilog zwischen Vertreterinnen und Vertretern des Europäischen Parlaments, des Rats und der Kommission. Abgeschlossen wurden die Trilog-Verhandlungen am 8. Dezember 2023. Seitdem hat der Entwurf weitere Hürden passiert. Es steht nur noch die finale Zustimmung des Rates der Europäischen Union aus, die noch vor der Europawahl Anfang Juni 2024 erwartet wird. Nach Verabschiedung beginnen gestaffelte Übergangsfristen: nach sechs Monaten greift das Verbot bestimmter KI-Systeme (manipulative, diskriminierende Systeme, z. B. social scoring), nach zwölf Monaten gelten die Regeln für Modelle mit allgemeinem Verwendungszweck. Nach 24 Monaten findet bis auf letzte Ausnahmen die ganze Verordnung Anwendung.

Mit der Verordnung sollen Möglichkeiten geschaffen werden, die Potenziale von KI auszuschöpfen und zugleich sicherzustellen, dass die Nutzung von KI im Einklang mit den ethischen Prinzipien und Grundwerten der EU erfolgt. Die Verordnung sieht vor, KI nach den Risiken ihrer Anwendungszwecke zu klassifizieren. Je nach Risikoklasse müssen unterschiedliche Anforderungen erfüllt werden. Beispielsweise sind für Anwendungen mit geringem Risiko Transparenzpflichten vorgesehen. Für KI-Systeme, die als Hochrisikosysteme eingestuft werden, gelten demgegenüber umfangreiche Regeln z. B. hinsichtlich Sicherheit, Dokumentation und menschlicher Aufsicht.

Das Thema der generativen KI wurde zu Beginn des Entstehungsprozesses der KI-Verordnung noch nicht in den Entwurf der Kommission eingebunden, fand aber nach der Veröffentlichung von ChatGPT zunehmend Aufmerksamkeit in Brüssel. Die Frage, wie Basismodelle (Foundation Models) reguliert werden sollen, war bis zur Einigung im Zuge des Trilogs eine der umstrittensten Fragen der Verhandlungen. Geregelt wurde das Thema schlussendlich unter dem Begriff der Modelle mit allgemeinem Verwendungszweck (general purpose AI), für welche besondere Regeln gelten.

5.2 Erwartungsmanagement

Die mediale Präsenz des Themas generative KI ist im Augenblick enorm. Praktisch jeden Tag erscheinen Neuerungen von Sprachmodellen und Erfolgsmeldungen in der Presse, was in vielen Fällen zu überzogenen Erwartungen an die Technologie führt. Generative KI ist derzeit noch eine vergleichsweise neue Technologie, und man muss sich deren Fähigkeiten und Limitationen bewusst sein, um sie zielführend in der Praxis einsetzen zu können. Dazu müssen Organisationen sowie Nutzerinnen und Nutzer Expertise erarbeiten und sich mit generativer KI vertraut machen.

Auch für die öffentliche Verwaltung ist es wichtig, Fachwissen in diesem Bereich aufzubauen und die Praxistauglichkeit durch entsprechende Pilotprojekte zu testen.

Das Potenzial, dass KI innerhalb der hessischen Landesverwaltung dazu eingesetzt werden kann, um Mitarbeiterinnen und Mitarbeiter bei ihrer täglichen Arbeit zu unterstützen, ist groß.

In Kapitel 4 wurden dazu bereits verschiedene Anwendungsbereiche von generativer KI in der öffentlichen Verwaltung vorgestellt. Beispielsweise könnte KI in der Verwaltung dazu genutzt werden, Dokumente auf gewisse Themenstellungen und Schlagworte hin zu durchsuchen, sie zusammenzufassen oder zu klassifizieren. Auch für die Erstellung von Texten im Sinne von Entwürfen ist die Verwendung generativer KI-Modelle denkbar.

Der Einsatz von generativer KI ist hierbei rein als nützliche Unterstützungsleistung zu betrachten. Generative KI wird in keinem Fall eingesetzt, um Entscheidungen zu treffen oder menschliche Bearbeiterinnen und Bearbeiter zu ersetzen. Der Mensch steht beim Einsatz KI-basierter Unterstützungsfunktionen jederzeit im Fokus.

Die Qualität KI-generierter Inhalte ist dabei nicht zu unter-, aber vor allem auch nicht zu überschätzen. So hat die Nutzung beispielsweise in dem Bewusstsein zu erfolgen, dass im Hinblick auf die Nutzung von Automatisierungstechnologien und generative KI im Speziellen gewisse Tendenzen (Bias) entstehen können, die einer optimalen Nutzung von KI entgegenstehen.

5.3 Vertraulichkeit der eingegebenen Daten

Alle Eingaben bei der Nutzung externer Schnittstellen von LLMs fließen zunächst an den Betreiber des Modells ab. Eine mögliche nachgelagerte Nutzung dieser Daten durch den Betreiber, sei es zur Speicherung oder zum Training, ist sehr unterschiedlich geregelt. Betrachtet man die Ausgaben des Modells, so hat der Betreiber oftmals uneingeschränkten Zugriff darauf. Zur Erweiterung des Funktionsumfangs bieten zudem einige Hersteller Erweiterungen bzw. Plugins von Drittanbietern an. In diesem Fall werden die eingegebenen Daten gegebenenfalls an weitere Anbieter weitergereicht. Dies ist beim Aufsetzen der Architektur etwaiger KI-Systeme zu beachten und vor einer Implementierung entsprechend mit dem jeweiligen IT-Dienstleister bzw. den Dienstleistern abzustimmen, sodass eine datenschutzkonforme Anwendbarkeit sichergestellt werden kann.

Aus diesen Gründen ist die Eingabe von sensiblen und vertraulichen Informationen im Rahmen der Nutzung von LLMs über externe Schnittstellen kritisch zu betrachten.[1]

5.4 Umgang mit KI-generierten Inhalten

Die Verantwortung für jegliche innerhalb der hessischen Landesverwaltung entstehenden Dokumente verbleibt in jedem Fall bei den Bearbeiterinnen und Bearbeitern – mit oder ohne Einsatz von Unterstützungsleistungen, die auf generativer KI basieren. Inhalte, im Verwaltungskontext insbesondere Texte, die mittels KI generiert wurden, sind daher stets als Entwürfe zu betrachten, die grundsätzlich immer von menschlicher Hand zu prüfen und im Bedarfsfall zu überarbeiten und zu finalisieren sind.

1 Bundesamt für Sicherheit in der Informationstechnik (2023): Große KI-Sprachmodelle – Chancen und Risiken für die Industrie und Behörden. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf?__blob=publicationFile&v=2

Die Qualität von KI-generierten Inhalten muss kontinuierlich überprüft werden, nicht nur durch einfache Plausibilitätskontrollen. Insbesondere die Richtigkeit der beinhalteten Informationen ist bei jeder Nutzung durch menschliche Bearbeiterinnen oder Bearbeiter zu kontrollieren und Unrichtiges zu identifizieren sowie zu korrigieren. Das KI-basierte Unterstützungssystem sollte gemäß den technischen Möglichkeiten bei der Identifizierung von Fehlern helfen, z. B. durch die Ausgabe von Trefferwahrscheinlichkeiten bei gegebenen Antworten und insbesondere bei generierten Textpassagen. Jeglichen – meist unterbewusst auftretenden – kognitiven Bias ist bewusst entgegenzuwirken, damit der Einsatz KI-basierter Unterstützungsleistungen ein Erfolg werden kann.

5.5 Vorgehensweise bei identifizierten Anwendungsfällen

Wurde ein spezifischer Anwendungsfall identifiziert, so ist zu validieren, ob der Einsatz von generativer KI für eine Unterstützung des jeweiligen Arbeitsschrittes relevant ist. Eignet sich der Anwendungsfall für den Einsatz von generativer KI, z. B. aufgrund von Fallzahlen, Nutzungsfrequenzen sowie gesehener Zeit- und Kosteneinsparungspotenziale, kann eine Analyse der Ist- und Soll-Prozesse erfolgen. Wie sieht der jetzige Prozess aus und wie ist er zukünftig, bei Einsatz von generativer KI, erwünscht? Stehen diese Anforderungen fest, so kann ein Projekt geplant werden. Hierbei sind u. a. folgende beispielhaft aufgeführte Aspekte zu bedenken:

(a) Umfang des Projektes (u. a. Ziel, Anforderungen, grobe Zeitplanung)

(b) Einbezug der betroffenen Personen und Bereiche (u. a. Kolleginnen und Kollegen des Referats, IKT-Referat des Hauses, weitere Fachreferate des Hauses, HMD, IT-Dienstleister: erstens zur Entwicklung und zweitens zum Betrieb)

(c) Kostenschätzung (u. a. für die Entwicklung, die benötigte Infrastruktur, die Einführung im Arbeitsprozess und den Betrieb)

(d) Modus der Umsetzung (u. a. Feinplanung des Zeitablaufs, Festlegung der Art der Zusammenarbeit)

(e) Berücksichtigung von bestehenden IT-Infrastrukturen und Vorgaben zu IT-Standards in der Landesverwaltung

(f) Auswahl und Vergabe IT-Dienstleister zur Entwicklung

(g) Umsetzung des Projekts

(h) Schulung bzw. Fortbildung der betroffenen Mitarbeiterinnen und Mitarbeiter

Diese Aufzählung erhebt keinen Anspruch auf Vollständigkeit, sondern dient der ersten Anregung bei Umsetzungsinteresse eines Technologieprojektes, das auf generativer KI basiert.

Nicht zuletzt aufgrund der oft nicht vorhersagbaren Qualität und Passgenauigkeit der generativen KI auf einen neu identifizierten Anwendungsfall ist es ratsam, im Vorfeld eines finalen Umsetzungsprojektes eine Machbarkeitsprüfung (*Proof of Concept*) durchzuführen. Damit wird es möglich, eine Aussage über die technische Machbarkeit der geplanten Lösung zu treffen und unvorhersehbare Herausforderungen frühzeitig zu identifizieren. Da bei LLMs je nach

Anwendungsfall zum Teil enorme Qualitätsunterschiede auftreten, lohnt es sich dabei, die Ergebnisse verschiedener Modelle, Modellvarianten bzw. Hersteller zu vergleichen.

5.6 Akzeptanzmanagement

Generative KI ist ein Themenfeld, auf das Personen außerhalb von Fachkreisen in der Regel erst durch die Veröffentlichung von ChatGPT im November 2022 aufmerksam geworden sind. Auch aktuell ist nicht davon auszugehen, dass die Nutzung von generativer KI und Kenntnisse zu dieser Thematik in breiten Bevölkerungsschichten bereits der Normalfall sind. Vielmehr hat mit Blick auf die Landesverwaltung bisher nur ein Teil der Beschäftigten im privaten Bereich, noch weniger Kolleginnen und Kollegen im beruflichen Kontext bereits Erfahrungen mit generativer KI gesammelt. Vor diesem Hintergrund und aufgrund der zahlreichen Berichte über generative KI in den Medien besteht innerhalb der Landesverwaltung ein hoher Informationsbedarf zum Thema generative KI. Es bestehen viele Unsicherheiten, von den oben genannten rechtlichen Herausforderungen über Ängste z. B. vor massiven Veränderungen des Aufgabenprofils oder vor Arbeitsplatzverlust, bis hin zu Bedenken, ob Umstellungen in der Arbeitsweise gemeistert werden können. Um generative KI in der Landesverwaltung einzuführen, ist daher ein durchdachtes Akzeptanzmanagement eine wichtige Voraussetzung.

Am Anfang eines gelungenen Akzeptanzmanagements steht stets die Erkenntnis, dass KI nicht als Ersatz von menschlichen Beschäftigten gedacht ist, sondern die bestehenden Handlungsabläufe in Form einer sinnvollen, rechtskonformen und durchdachten Automatisierung effektivieren kann.

Entscheidend sind ein möglichst transparenter Umgang mit dem Einsatz von KI und den gelieferten Ergebnissen sowie frühzeitige Schulungs- und Mitbestimmungsmöglichkeiten der Bediensteten, etwa in Hinblick auf die Ermittlung möglicher Anwendungsfälle. Fähigkeiten und Grenzen der KI müssen klar thematisiert und dem konkreten Einsatz zugrunde gelegt werden. Angesichts der beschriebenen Defizite muss auch die menschliche Letztverantwortlichkeit betont und in Form geeigneter Kontrollmechanismen implementiert werden.

Durch einen auf den jeweiligen Anwendungsfall bezogenen, schrittweisen Einsatz von generativer KI mit ausreichender Testphase und einer Erprobungs- und Mitwirkungsmöglichkeit für die Bediensteten wird eine Überforderung vermieden. Der tatsächliche Mehrwert der KI und die Effektivität der geschaffenen Arbeitsabläufe sind kontinuierlich zu evaluieren.

5.7 Datenschutz

Sollen mittels KI personenbezogene Daten verarbeitet werden, sind die Anforderungen der DSGVO, des HDSIG (→Glossar) sowie gegebenenfalls spezialgesetzliche Anforderungen des Datenschutzes zu erfüllen. Zu beachten ist, dass für eine datenschutzkonforme Verwendung alle Systeme und Stellen zu erfassen und zu bewerten sind, die in der Verarbeitungskette der Gesamtanwendung beteiligt sind.

Bei der Beantwortung der Frage, ob personenbezogene Daten verarbeitet werden, müssen alle Aspekte berücksichtigt werden. Hierzu zählen z. B.

- personenbezogene Daten, die zu Trainings-, Test- und Evaluationszwecken genutzt werden,
- personenbezogene Daten, die im Rahmen der Nutzung der Systeme eingegeben werden, sowie
- Daten der Anwenderinnen und Anwender, die bei der Nutzung der KI-Systeme anfallen und
- Daten, die das KI-System ausgibt.

Neben den allgemein zu erfüllenden datenschutzrechtlichen Anforderungen[2] stellen sich bei Einsatz von KI zusätzliche, KI-spezifische Anforderungen. Hierzu hat sich die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder bereits im Jahr 2019 geäußert.[3] Die dort gemachten Ausführungen haben weiterhin Gültigkeit. Es ist jedoch absehbar, dass aufgrund der Entwicklungen auf diesem Gebiet kontinuierlich Aktualisierungs- und Anpassungsbedarf entstehen wird. Aktuell stehen der Hessische Beauftragte für Datenschutz und Informationsfreiheit (HBDI) und andere europäische Datenschutzaufsichtsbehörden im Zusammenhang mit ChatGPT zur Klärung datenschutzrechtlicher Fragen mit dem Anbieter OpenAI im Kontakt.[4] Hinzu kommen legislative Initiativen auf europäischer Ebene, die erheblichen Einfluss auf den Einsatz von KI haben werden. Besonders hervorzuheben ist hier die europäische KI-Verordnung.

Für KI-Vorhaben bedeutet dies, dass von Anfang an und dann durchgängig eine umfassende Einbeziehung des operativen Datenschutzes unverzichtbar ist.[5] In allen Projektphasen müssen datenschutzrechtliche Anforderungen als integraler Bestandteil Berücksichtigung finden und umgesetzt werden. Selbiges gilt für den Betrieb, die Nutzung und allgemein für den gesamten Lebenszyklus von KI-Systemen.

2 DSK: Das Standard-Datenschutzmodell – https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methode_V3.pdf

3 DSK: Hambacher Erklärung zur Künstlichen Intelligenz – https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf und DSK: Positionspapier der DSK zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen – https://www.datenschutzkonferenz-online.de/media/en/20191106_positionspapier_kuenstliche_intelligenz.pdf

4 HBDI: Hessischer Datenschutzbeauftragter fordert Antworten zu ChatGPT – https://datenschutz.hessen.de/presse/hessischer-datenschutzbeauftragter-fordert-antworten-zu-chatgpt

HBDI: HBDI veröffentlicht Fragenkatalog zu ChatGPT – https://datenschutz.hessen.de/presse/hbdi-veroeffentlicht-fragenkatalog-zu-chatgpt und HBDI: Hessischer Datenschutzbeauftragter fordert erneut Antworten zu ChatGPT – https://datenschutz.hessen.de/presse/hessischer-datenschutzbeauftragter-fordert-erneut-antworten-zu-chatgpt

5 HBDI: 50. Tätigkeitsbericht Datenschutz, Kapitel 3.2 Digitale Souveränität und erfolgreiche Digitalisierungsprojekte, S. 45ff – https://datenschutz.hessen.de/sites/datenschutz.hessen.de/files/2022-08/50_taetigkeitsbericht_01_0.pdf

5.8 Generative KI und Cybersicherheit

5.8.1 KI ist Software – Software hat Fehler

Zunächst ist KI – und damit auch generative KI – Software. Es gelten alle IT-Sicherheitsanforderungen an Software bei Entwicklung und Betrieb uneingeschränkt auch für generative KI. Software hat Fehler und Schwachstellen, Software benötigt Updates. Software sollte „by design" möglichst sicher programmiert sein. Software muss getestet werden und erfordert einen sicheren Betrieb. Zusammengefasst müssen die Anforderungen des *BSI-Grundschutzes* (→Glossar) erfüllt sein. Dies ist bei einer Software, auf deren Entwicklung und Betrieb das Land Hessen keinen Einfluss hat (wie bei ChatGPT), nicht verlässlich durchsetzbar und muss bei der Nutzung bedacht und beachtet werden.

5.8.2 Generative KI ist angreifbar

Die Modelle, die das „Wissen" der KI repräsentieren, sind angreifbar, sodass die KI nach Angriffen Ausgaben erzeugen kann, die den Anforderungen des Angreifers und nicht denen des legitim Nutzenden entsprechen. Zudem können im Zuge eines Angriffs grundsätzlich durch entsprechende Eingaben (*Prompts*) die Trainingsdaten und damit gegebenenfalls schützenswerte Daten aus dem Modell reproduziert werden. Hier sind durch weitere Forschung Methoden bereitzustellen, die solche Angriffe verhindern, erschweren oder erkennbar machen können.

Innerhalb der G-7-Länder wurde mit den „International Guiding Principles for Advanced AI Systems" ein freiwilliger Verhaltenskodex für KI-Entwicklerinnen und -Entwickler verabschiedet. Dieser umfasst Empfehlungen zur Minderung von Risiken und Missbrauch und zur Ermittlung von Schwachstellen, die Meldung von Sicherheitsvorfällen, Investitionen in die Cybersicherheit sowie ein Kennzeichnungssystem, das es den Nutzerinnen und Nutzern ermöglicht, KI-generierte Inhalte zu erkennen. Dieser Verhaltenskodex sollte bei KI-Entwicklungen angewandt werden.

Hintergrundwissen: Hinweise des Bundesamts für Sicherheit in der Informationstechnik (BSI)

Das BSI stellt regelmäßig aktuelle Informationen zum Thema KI und im Speziellen auch zu generativer KI zur Verfügung, beispielsweise folgende:

Zentrale BSI-Informationsseite zum Thema KI

Die Webseite enthält zahlreiche Links zu weiterführenden Informationen.

➜ https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/kuenstliche-intelligenz_node.html

Große KI Sprachmodelle – Chancen und Risiken für Industrie und Behörden

Das Dokument enthält Hintergrundinformationen zu LLMs inklusive einer Betrachtung der Chancen und Risiken.

➜ https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf?__blob=publicationFile&v=2

6 Technologische Grundlagen und Entwicklungen

Technologische Aspekte und die rasante Entwicklung spielen im Bereich der generativen KI eine entscheidende Rolle, da der Fortschritt und die Effizienz dieser Systeme maßgeblich von den zugrunde liegenden Algorithmen, Rechenleistungen und Daten abhängen.

6.1 Die Rolle von Open Source im Bereich generativer KI

Der überwiegende Teil der aktuellen KI-Anwendungen basiert auf kommerziellen Schnittstellen (*AI-as-a-Service – AIaaS* [→Glossar]). Mögliche Nachteile solcher kommerziellen Lösungen sind u. a. die nicht absehbaren Kosten, mangelnder Datenschutz und unzureichende Berücksichtigung von Cybersicherheit.

Dem steht eine schnell wachsende Open Source Community gegenüber. Das Unternehmen Meta hat 2022 mit *LlaMA* ein eigenes LLM veröffentlicht, das seit Juli 2023 in der Version LlaMA-2 auch kommerziell genutzt werden darf. Dies hat die Entwicklung von leistungsfähigen Open-Source-Modellen erheblich vorangetrieben. Diese und andere Open-Source-Modelle haben sich schnell weiterentwickelt und erreichen in vielen Benchmarks ähnliche oder sogar bessere Ergebnisse als ChatGPT und andere *Closed Source Modelle wie GPT-4*.

Beim **Einsatz von Open-Source-Modellen** kommen verschiedene Aspekte zum Tragen:

Lizenz und Fine-Tuning: Die Leistung von LLMs hängt maßgeblich von der Qualität und Menge der Trainingsdaten ab. Besonders das sogenannte Instruction Fine-Tuning erfordert spezielle Trainingsdatensätze, die reich an Frage-Antwort-Paaren mit Instruktionen sind. Häufig wurden solche Datensätze unter Verwendung von Closed-Source-Anwendungen wie ChatGPT oder GPT-4 generiert. Da OpenAI die Verwendung seiner Produkte zum Training anderer LLMs untersagt, können *Open Source LLMs* [→Glossar], die mit solchen Daten (weiter-)trainiert wurden, nicht kommerziell genutzt werden. Daher ist es empfehlenswert, bei der Wahl eines Open Source LLMs auch die Lizenzen der verwendeten Trainingsdatensätze zu überprüfen.

Inbetriebnahme und Performance: Ein wesentlicher Vorteil aktueller Open Source LLMs gegenüber Closed Source LLMs [→Glossar] liegt in ihrer deutlich geringeren Größe und ihrer offenen Verfügbarkeit. Dies ermöglicht es, solche Systeme eigenständig zu hosten und zu betreiben. Dadurch können Herausforderungen wie fehlende Transparenz (beispielsweise im Kontext der KI-Verordnung), Datenschutz und Sicherheit oder die Notwendigkeit von Anpassungen effektiver bewältigt werden. Allerdings sind auch kleinere Modelle rechenintensiv und benötigen spezialisierte Ressourcen, um die Inferenzzeit, also die Dauer für die Generierung einer Antwort durch ein LLM, möglichst kurz zu halten.

Kosten und Auswahl eines LLMs: Obwohl leistungsfähige generalistische Open-Source-Modelle oft deutlich kleiner sind, fallen Betriebskosten an. Daher ist es entscheidend, den Anwendungsfall zu betrachten und zu überlegen, ob für die spezifische Problemstellung stets das neueste Open Source LLM erforderlich ist. Falls der Einsatzbereich sich z. B. lediglich auf die automatische Klassifizierung von Dokumenten beschränkt, könnte beispielsweise ein angepasstes BERT-Modell (ein Open-Source-*Algorithmus* (→Glossar) für Sprachmodelle, Abkürzung für Bidirectional Encoder Representations from Transformers) mit wenigen hundert Millionen Parametern die effizientere und kostengünstigere Option sein, da es günstiger zu betreiben ist. In der Tat benötigen viele Anwendungsfälle nicht die breit angelegten, aufgabenagnostischen LLMs, sondern lassen sich mit deutlich kleineren, dafür aber spezialisierten Modellen realisieren. Somit sind der Verwendungszweck und die damit verbundenen Kosten immer entscheidende Faktoren bei der Auswahl.

Fazit

In relativ kurzer Zeit hat die Open Source Community Modelle hervorgebracht, die mit Closed Source LLMs konkurrieren können. Diese Modelle erreichen vielleicht nicht in allen Aspekten die Vielseitigkeit und Generalität von Systemen wie GPT-4, bieten jedoch den bedeutenden Vorteil, dass sie mit einem angemessenen Aufwand individuell angepasst und betrieben werden können. Besonders die Möglichkeit, diese Modelle auf spezifische Anwendungsfälle zuzuschneiden, hat einen großen Einfluss und führt oft auch bei älteren Modellen zu ausgezeichneten Ergebnissen.

Es ist daher davon auszugehen, dass Open-Source-Modellen eine zunehmend größere Rolle zukommen wird.

6.2 Laufzeitumgebungen von generativer KI

In der dynamischen Welt der generativen KI spielt die Cloud-Technologie eine wichtige Rolle, vor allem bei der Bereitstellung und Nutzung von LLMs. Nachfolgend werden deshalb die zentralen Gründe für die aktuelle Vormachtstellung der Cloud als Laufzeitumgebung beleuchtet, darunter Ressourcenintensität, Skalierbarkeit, einfache Aktualisierungen und Wartung sowie die Zugänglichkeit und Kosteneffizienz. Zugleich wird auf die Herausforderungen und Potenziale von On-Premises-Lösungen eingegangen, die trotz einiger Beschränkungen zunehmend an Bedeutung gewinnen könnten.

6.2.1 Cloud

Warum spielt die Cloud-Technologie im Bereich der generativen KI eine so große Rolle?

Wesentliche Gründe für eine Laufzeitumgebung in der Cloud ergeben sich aus folgenden Merkmalen dieser Technologie:

- **Ressourcenintensität**: Generische KI-Modelle, insbesondere fortschrittliche Modelle wie GPT-4, erfordern erhebliche Rechenleistung und Speicherplatz. Cloud-Systeme bieten die notwendige Infrastruktur, um diese Ressourcen effizient und kosteneffektiv bereitzustellen.
- **Skalierbarkeit**: Cloud-Plattformen ermöglichen es, dass Dienste je nach Bedarf skaliert werden können. Dies ist besonders wichtig für KI-Anwendungen, da die Nutzerlast stark variieren kann.
- **Aktualisierungen und Wartung**: KI-Modelle benötigen regelmäßige Aktualisierungen und Wartungsarbeiten, um ihre Genauigkeit und Relevanz zu bewahren. Cloud-Systeme erleichtern die Bereitstellung dieser Updates, ohne dass Endbenutzer eingreifen müssen.
- **Zugänglichkeit und Vernetzung**: Cloudbasierte KI-Lösungen sind von überall und auf verschiedenen Geräten zugänglich, was für Benutzer bequem ist. Diese Zugänglichkeit fördert auch die Vernetzung und Integration mit anderen Online-Diensten und Datenquellen.
- **Kosten**: Der Betrieb eigener Server für leistungsstarke KI-Anwendungen kann für die öffentliche Verwaltung unerschwinglich sein. Cloud-Dienste bieten eine kosteneffiziente Alternative, da sie auf einem *Pay-as-you-go-Modell* [→Glossar] basieren, das es ermöglicht, nur für die tatsächlich genutzten Ressourcen zu bezahlen.

Wie sind die KI-Modelle aus der Cloud nutzbar?

Führende kommerzielle Anbieter von KI-Modellen bieten zum Teil ihre KI-Modelle in ihrer eigenen Umgebung (oft als Public Cloud bezeichnet) oder in einer für den Kunden dedizierten Umgebung (Private Cloud) an.

Der Unterschied zwischen Public Cloud und Private Cloud liegt hauptsächlich im Bereich Zugänglichkeit und Besitz. Während eine Public Cloud die Ressourcen über das Internet auf einer *Multi-Tenant* [→Glossar]-Basis anbietet (verschiedene Kunden teilen sich dieselbe Infrastruktur), ist eine Private Cloud exklusiv für einen einzigen Kunden reserviert und kann entweder intern oder von einem Drittanbieter verwaltet werden.

Die Cloud-Anbieter stellen Schnittstellen (APIs) zur Nutzung der Modelle bereit. Diese Art der Bereitstellung wird auch *AI-as-a-Service (AIaaS)* genannt.

6.2.2 On-Premises

Im Unterschied zu As-a-Service-LLMs oder Cloud-Hosting behält eine Organisation bei einem in Eigenregie gehosteten LLM vollständige Kontrolle über die Verwendung und Weitergabe ihrer Daten. Dies bezieht sich sowohl auf die für das Training genutzten Daten als auch auf Logging und Monitoring. Da das Logging auf der eigenen Hardware der Organisation stattfindet, können entsprechende Schutzmaßnahmen angewandt werden. Zudem entscheidet die Organisation eigenständig, welche Daten in den Logs erfasst werden, und hat diese bei Bedarf, beispielsweise

bei einem Audit, sofort zur Verfügung. Somit bietet ein selbst gehostetes LLM deutlich bessere Kontroll- und Überwachungsmöglichkeiten im Vergleich zu externen Anbietern.

Warum gibt es bisher so wenige On-Premises-Lösungen?

Wie im vorangegangenen Kapitel erläutert, sind u. a. die **Hardwareanforderungen** zum Betrieb von großen KI-Sprachmodellen enorm. Die Anschaffungskosten für eine solche Hardware sind sehr hoch. Zudem ist die Verfügbarkeit der Hardware am Markt im Moment eingeschränkt, woraus lange **Lieferzeiten** resultieren.

Darüber hinaus bieten namhafte kommerzielle Hersteller vom LLMs ihre Modelle ausschließlich in der Cloud an. Ein Übertrag des Service in eine On-Premises-Umgebung ist in vielen Fällen bisher nicht vorgesehen.

Ausblick

Die ersten Einsätze zeigen, dass auch mit kleineren Modellen (in der Regel auf Open-Source-Basis) je nach Anwendungsfall gute Ergebnisse erzielt werden können. Das resultiert in einer Senkung der Hardwareanforderungen. Dieser Umstand wird dazu führen, dass in naher Zukunft immer mehr Modelle auch in einer On-Premises-Umgebung zum Einsatz kommen können.

6.3 Wie kann ich mein internes Organisationswissen mit generativer KI nutzen?

Es gibt unterschiedliche Möglichkeiten und Ansätze, um ein KI-Sprachmodell mit eigenem Wissen anzureichern. Einige davon werden nachfolgend aufgeführt. Je nach Anwendungsfall können die unterschiedlichen Methoden auch miteinander kombiniert werden.

6.3.1 Nutzung im Prompt / Kontext

Die wohl bekannteste und gleichzeitig auch einfachste Möglichkeit ist die Übergabe des Fach- und Domänenwissens als Teil des **Prompts** bei der Anfrage an das LLM. Das Wissen wird als Kontext dem LLM mitgegeben und fließt somit in die Antwortfindung ein.

Allerdings ist die Größe der übergebenen Datenmenge durch das sogenannte **Token Limit** (→ Glossar) eingeschränkt. Das Token Limit bei großen Sprachmodellen (LLMs) zeigt die maximale Anzahl von Tokens an, die das Modell in einer einzigen Anfrage verarbeiten kann. Ein Token ist dabei die grundlegende Einheit der Verarbeitung und kann je nach Sprachmodell und Sprache unterschiedlich definiert sein. Oft repräsentiert ein Token ein Wort, einen Teil eines Wortes oder ein Satzzeichen.

Einige Beispiele von Token Limits:

- Llama2 (→ Glossar): 4.096 Token
- GPT 3.5 Turbo: 4.096 Token
- GPT-4: 32.768 Token
- GPT-4 Turbo: 128.000 Token

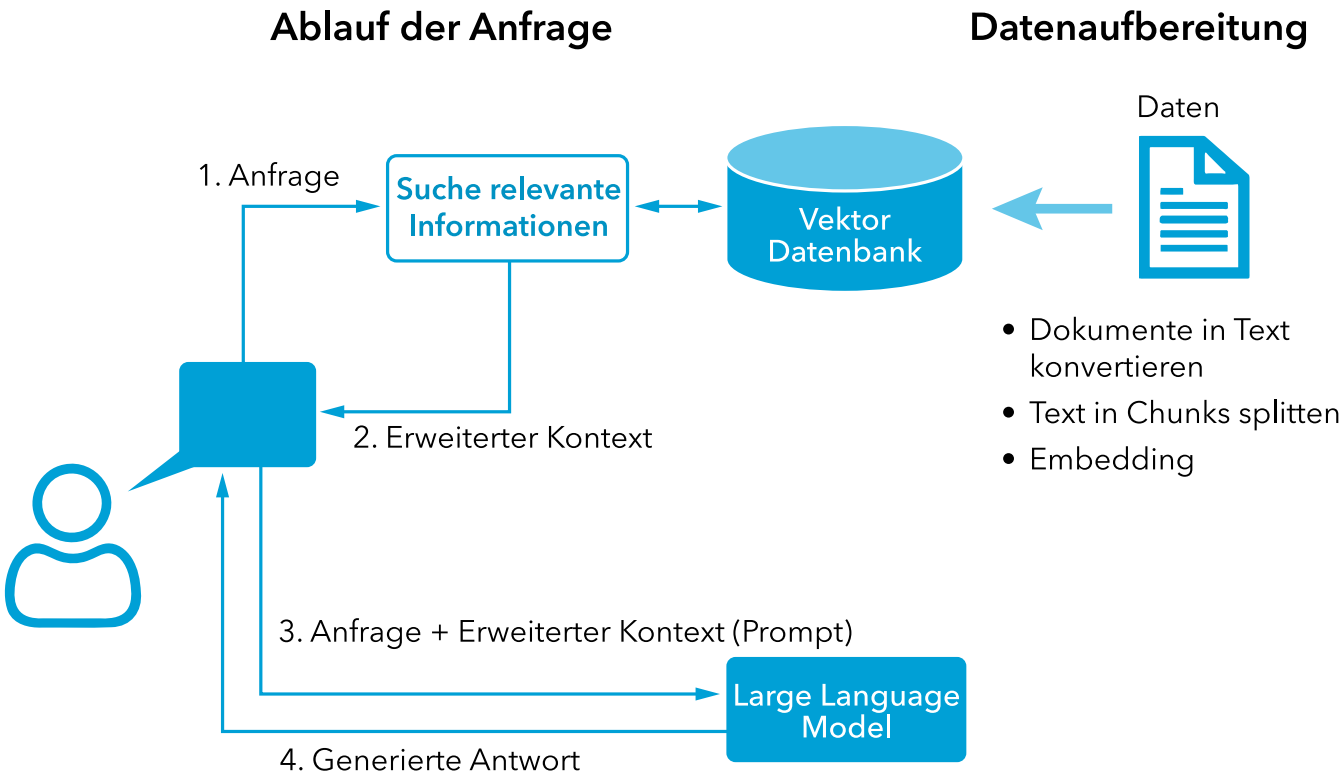
Zur einfachen Einordnung: Im Falle eines Token Limits von 4.096 wäre nach einer groben Schätzung die Anzahl an möglichen Wörtern ca. 2.700, was in etwa fünf bis sechs DIN A4-Seiten entsprechen könnte.

Auch wenn das Token Limit im Zuge der Weiterentwicklung von Sprachmodellen immer wieder erhöht wird, ist dieser Ansatz demnach nicht geeignet, um eine größere Wissensbasis (z. B. ein sehr großes Dokument oder Dokumentsammlungen) zu berücksichtigen.

6.3.2 Retrieval Augmented Generation (→ Glossar)

Retrieval Augmented Generation (RAG) stellt derzeit eine vielversprechende Möglichkeit dar, eigene Daten sicher in ein LLM zu integrieren. Dabei wird die begrenzte Menge an Inhalten pro Anfrage (Token Limit) berücksichtigt.

In einem vorgeschalteten System (z. B. Vektordatenbank) werden jeweils die für die Beantwortung der Frage relevanten eigenen Dokumentteile ermittelt – und diese Fragmente als erweiterter Kontext mit der Anfrage an das LLM geliefert.



Datenaufbereitung

Die benötigten Organisationsdaten werden im Vorfeld in Text konvertiert und in sogenannte Chunks (Textschnipsel) aufgeteilt. Diese werden dann über eine Embedding-Technologie (Transformation von Text in einen hochdimensionalen Vektorraum) konvertiert und in der Vektordatenbank gespeichert. Ein großer Vorteil von Vektordatenbanken ist ihre Fähigkeit, semantische Beziehungen zwischen Begriffen zu erfassen. Sie verwenden Vektorrepräsentationen von Textdaten, um eine effiziente Ähnlichkeitssuche und -analyse zu ermöglichen.

Ablauf der Anfrage durch den Nutzer

1. Die Anfrage wird genutzt, um mithilfe einer Ähnlichkeitssuche in der Vektordatenbank nach relevanten Informationen zu suchen.
2. Die Anfrage wird um relevante Informationen erweitert.
3. Die erweitere Anfrage wird an das LLM übergeben.
4. Das LLM generiert eine spezifische Antwort auf die Anfrage.

Der RAG-Ansatz hat u. a. folgende Vorteile:

Aktualität und domänenspezifisches Wissen

RAG-Modelle kombinieren die Fähigkeiten von LLMs mit der Möglichkeit, auf externe Datenquellen zuzugreifen. Dies ermöglicht es ihnen, aktuellere und spezifischere Informationen in ihre Antworten zu integrieren, als es allein durch das Training auf einem festen Datensatz möglich wäre.

Verbesserte Überprüfbarkeit

RAG ermöglicht generativen KI-Anwendungen, ihre Quellen anzugeben, ähnlich wie in wissenschaftlichen Arbeiten. Dies verbessert die Überprüfbarkeit und macht die inneren Abläufe von generativen KI-Anwendungen transparenter.

Reduzierung von Verzerrungen

Da LLMs auf historischen Daten trainiert werden, können sie Verzerrungen und veraltete Informationen enthalten. RAG kann durch den Zugriff auf aktuelle und vielfältige externe Quellen dazu beitragen, die Auswirkungen solcher Verzerrungen zu verringern.

Kosteneffizienz

Das Fine-Tuning großer LLMs kann sehr rechenintensiv und teuer sein, insbesondere wenn es um umfangreiche Datensätze und komplexe Modelle geht. RAG kann durch die Nutzung vorhandener Datenbanken und die Kombination mit einem generativen Modell eine kostengünstigere Alternative bieten.

Fazit

Die Anwendung von Open-Source-Sprachmodellen birgt großes Potenzial für die Analyse organisationsinterner, unstrukturierter Daten. Um dieses Potenzial jedoch voll auszuschöpfen, ist es entscheidend, dass das eigene Fach- und Branchenwissen effektiv eingesetzt wird. Zudem ist die geschickte Handhabung und Integration dieser Daten z. B. durch den RAG-Ansatz (in Verbindung mit Vektordatenbanken) ausschlaggebend für eine erfolgreiche Nutzung.

6.3.3 Fine-Tuning

Fine-Tuning ist die wohl bekannteste Methode, um die Leistung von LLMs zu verbessern. Es bezeichnet einen Prozess, bei dem ein vorhandenes LLM auf eine spezifische Aufgabe trainiert wird, indem es mit spezifischen Daten angepasst wird. Um effektiv zu sein, erfordert Fine-Tuning jedoch einige Voraussetzungen. Die wichtigsten sind:

- **Leistungsstarke Hardware**: Zum Trainieren und Fine-Tuning von LLMs ist eine leistungsstarke Hardware erforderlich.
- **Zugang zu großen Datenmengen**: Für das Fine-Tuning benötigt man spezifische Datensätze, die relevant für den Anwendungsbereich des Modells sind.
- **Fachwissen in maschinellem Lernen und KI**: Fachkenntnisse in den Bereichen maschinelles Lernen, KI und Natural Language Processing (NLP) (→Glossar) sind entscheidend. Dies umfasst Kenntnisse über Modellarchitekturen, Trainingsalgorithmen und Techniken zur Vermeidung von Overfitting.
- **Software- und Programmierkenntnisse**: Kenntnisse in Programmiersprachen wie Python sowie Erfahrungen mit Machine Learning Frameworks wie TensorFlow oder PyTorch sind erforderlich, um Trainingsprozesse zu gestalten und anzupassen.
- **Budget und Zeit**: Das Fine-Tuning großer Modelle ist kosten- und zeitintensiv.

Fazit

Fine-Tuning ist eine etablierte, aber oft aufwendige und kostspielige Methode, um die Leistung von LLMs für spezifische Aufgaben zu verbessern. Es erfordert erhebliche Ressourcen, einschließlich leistungsstarker Hardware, Zugang zu umfangreichen Datensätzen, Fachwissen in maschinellem Lernen und KI sowie Programmierkenntnisse. Darüber hinaus sind ein beträchtliches Budget und ausreichend Zeit notwendig, um den Prozess effektiv durchzuführen.

6.4 Einfluss von generativer KI auf Chatbot-Plattformen

Werden Chatbot-Plattformen durch generative KI überflüssig?

Die umfangreichen Fähigkeiten von generativer KI im Bereich der automatisierten, dialoggestützten Kommunikation mittels natürlicher Sprache lassen vermuten, dass herkömmliche Chatbot-Plattformen gar nicht mehr nötig sind. Dass dies bei modernen Conversational-AI-Plattformen nicht der Fall ist und welche Ansätze und Synergien zwischen den beiden Technologien *Conversational AI* (→ Glossar) und *Generative AI* (→ Glossar) möglich sind, wird nachfolgend kurz erläutert.

Wie kann ein Zusammenspiel der beiden Technologien aussehen?

Moderne Chatbot-Plattformen sind in der Lage, generative KI, also unterschiedliche LLMs, über vorgefertigte Schnittstellen in die eigene Plattform zu integrieren. Über diesen Weg können die Stärken der Chatbot-Plattform (siehe folgenden Abschnitt) und die Stärken von LLMs (z. B. Formulierung, Zusammenfassung, Personalisierung etc.) miteinander kombiniert werden.

Was sind die wesentlichen Stärken von modernen Chatbot-Plattformen im Vergleich zur generativen KI?

Die konfigurierte Chatbot-Plattform „kennt" die Organisation und die Anwendungsfälle. Die definierten **Prozesse** und **Flows** sind spezifisch und zielorientiert umgesetzt.

Es gibt die Möglichkeit einer vorgefertigten und anpassbaren **Datenintegration** (Backend) und der Integration in das Ökosystem des Kunden.

Die Chatbot-Plattform bietet umfangreiche Integrationsmöglichkeiten in die **Kommunikationskanäle** (z. B. Web, Telefon, Messengerdienste etc.) der Organisation.

Moderne Chatbot-Systeme bieten auch die Möglichkeit, an eine „echte" Person in Form eines Live-Chats zu übergeben (Human Handover). Ist eine Person im Live-Chat-Modul angemeldet, kann diese fachkundige Person versuchen, das Problem zu lösen. Nach der Problemlösung übergibt die Fachkraft gegebenenfalls wieder an den Chat zurück.

Wie kann generative KI eine Chatbot-Plattform besser machen?

- **Erweiterung des Sprachverständnisses und der Antwortmöglichkeiten durch generative KI**

  Moderne Chatbots mit KI-Unterstützung (auf Basis Conversational-AI-Plattformen) waren bereits in der Lage, die natürliche Sprache des Menschen zu verstehen und den vermeintlichen Wunsch bzw. das Anliegen (*Intent*) abzuleiten. Darüber hinaus können auch bestimmte Kategorien und Kontexte (Entitäten) extrahiert werden (z. B. Fachbegriffe, Datum, Zahlen, Länder). Die Basis hierfür bildet NLP. Dieses Sprachverständnis wird durch die neuen Möglichkeiten von LLMs signifikant erweitert. Zudem ergeben sich weitreichende Potenziale bei der Erstellung von Antworttexten.

- **Verbesserte Einbindung von externen Informationsquellen**

  Moderne Chatbot-Plattformen unterstützen die Einbindung von externen Informationsquellen (z. B. Webseiten, Word-Dokumente, PDF-Dokumente usw.) über die im LLM-Kontext bekannten Ansätze wie RAG (siehe Kapitel *6.3.2 Retrieval Augmented Generation*). Dadurch erhält man eine weitere Option zur Einbindung von Organisationsdaten in die Konversation.

- **Reduktion des Pflegeaufwandes**

  Vor allem bei einer großen Anzahl an *Intents* ist der Pflegeaufwand (Erstellung und Anpassung von Beispielfragen und Antworten) für Bot-Redakteure sehr hoch. Über den oben genannten RAG-Ansatz wird der Aufwand erheblich reduziert. Zudem entfällt in vielen Fällen die doppelte Pflege der Informationen.

- **Assistenz im Live-Chat-Modul**

  Die im Zusammenhang mit Chatbot-Plattformen im Einsatz befindlichen Live-Chat-Module profitieren ebenfalls von den sprachlichen Möglichkeiten von LLMs. Sie unterstützen den Agenten in Echtzeit im Dialog mit dem Kunden bei der Beantwortung der Fragen (z. B. durch Vorschläge) und Formulierung der Antworten.

7 Fazit und Ausblick

Vor dem Hintergrund der rasanten Entwicklungen im Bereich der generativen KI seit der Veröffentlichung von ChatGPT im November 2022 und der starken Präsenz des Themas in den Medien sind in der hessischen Landesverwaltung in den letzten Monaten erste Ideen zu Anwendungsmöglichkeiten von generativer KI, aber auch große Unsicherheiten und ein enormer Informationsbedarf rund um das Thema generative KI entstanden.

Diese Herausforderungen und Chancen hat das vorliegende Einführungsdokument in den vorangehenden Kapiteln aufgegriffen und versucht, einen realistischen Blick auf das Thema generative KI aus der spezifischen Perspektive einer Landesverwaltung zu werfen. Es zeigt sich, dass generative KI keine „Wunderwaffe" ist und teils überzogene Erwartungen an die Möglichkeiten bestehen, die generative KI heute bietet. Zugleich sollten die Potenziale, die der Einsatz von KI in Behörden aufweist, keinesfalls unterschätzt oder generative KI gar als Bedrohung aufgefasst werden. Vielmehr existieren in der Landesverwaltung zahlreiche Bereiche, in denen generative KI bereits heute sinnvoll eingesetzt werden kann.

Der Einsatz von generativer KI sollte kein Selbstzweck sein. Stattdessen sollte genau analysiert werden, wo generative KI sinnvoll eingesetzt werden kann. Dabei entstehen vielfältige Herausforderungen, denen sich die öffentliche Verwaltung stellen muss und für die Lösungswege erarbeitet werden müssen, beispielsweise in den Bereichen Datenschutz und Cybersicherheit. Aus Sicht der Projektgruppe Generative KI sind für die erfolgreiche Entwicklung von generativer KI in der hessischen Landesverwaltung vor allem drei Elemente entscheidend:

Praxisprojekte

Das Einführungsdokument stellt mehrere interessante Projektansätze aus dem Bereich der generativen KI aus verschiedenen Ressorts vor. Diese zeigen den ganz konkreten Nutzen auf, den generative KI für die Arbeit der Landesverwaltung haben kann. Hierauf sollte aufgebaut und generative KI in der Landesverwaltung schnell in die praktische Anwendung gebracht werden. Wichtig sind die Identifizierung weiterer Anwendungsfälle und das Aufsetzen konkreter Pilotprojekte. Nur mithilfe solcher konkreten Projekte können Wege gefunden werden, die Potenziale von generativer KI zu heben und die Herausforderungen, die sich durch sie stellen, zu überwinden. Die im Einführungsdokument vorgestellten Anwendungsbeispiele können als Inspirationsquelle für weitere Projekte dienen.

Zu bedenken ist hierbei, dass viele Landesbeschäftigte in unterschiedlichen Dienststellen aktuell ähnliche Überlegungen zur Nutzung von generativer KI anstellen und sich hier große Potenziale für Synergien ergeben können. Ressortübergreifende Ansätze bieten daher einen deutlichen Mehrwert.

Informationsangebote

Generative KI ist für die Landesverwaltung ein neues Thema; der Informationsbedarf ist erheblich. Das vorliegende Einführungsdokument ist hier als erster Schritt zu verstehen, die Landesbeschäftigten zu informieren, dem aber weitere Schritte folgen müssen. So sollten konkrete Projekte aus dem Bereich der generativen KI mit entsprechenden Informationsangeboten für die Mitarbeiterinnen und Mitarbeiter verbunden werden. Denn auch wenn mit dem Einführungsdokument ein relativ niedrigschwelliger Einstieg in das Thema ermöglicht wird, werden im Zuge der konkreten Projekte weitere spezifische Fragen auftreten, die im Rahmen von Informations- und Schulungsangeboten aufgegriffen werden sollten. Letztlich verändert die Nutzung generativer KI in der Landesverwaltung die Aufgaben und Arbeitsplätze der Beschäftigten. Es sollte daher großer Wert auf ein passendes Akzeptanzmanagement gelegt werden; die Mitarbeiterinnen und Mitarbeiter müssen bei der Einführung solcher Anwendungen eingebunden und mitgenommen werden.

Entwicklung eines Muster-Regelungsentwurfs

Im Zuge der Arbeit der Projektgruppe hat sich gezeigt, dass ein großer Bedarf an Leitlinien und Regelungen zum Umgang mit generativer KI in der Landesverwaltung besteht. Die Nutzung von generativer KI wirft zahlreiche rechtliche Fragen auf, insbesondere mit Blick auf das Urheberrecht, das Haftungsrecht, das Strafrecht, das Arbeits- und Dienstrecht sowie das Datenschutzrecht. Die im Einführungsdokument identifizierten Herausforderungen bilden daher auch einen Ausgangspunkt für die Entwicklung eines Muster-Regelungsentwurfs zur rechtssicheren Nutzung von generativer KI in der Landesverwaltung. Ein solcher Muster-Regelungsentwurf soll 2024 entwickelt werden.

Die Entwicklung von rechtlichen Leitlinien, die Bereitstellung von Informationsangeboten sowie das Aufsetzen erster konkreter Projekte, die den praktischen Mehrwert von generativer KI verdeutlichen, können dieses Thema in der Landesverwaltung erheblich voranbringen. Angesichts der rasanten Geschwindigkeit, mit der es sich weiterentwickelt, sind die enormen Potenziale, die in den nächsten Jahren im Bereich der generativen KI für die öffentliche Verwaltung entstehen werden, heute noch gar nicht abzusehen. Die Potenziale, die aktuell bereits sichtbar sind, zeigen aber schon deutlich auf, dass sich Anwendungen, die auf generativer KI basieren, zu bedeutenden Unterstützungsinstrumenten für die Beschäftigten der hessischen Landesverwaltung entwickeln werden.

Anhang

Anhang 1: Katalog möglicher Anwendungsfälle

Jeder Eintrag im Katalog folgt einer einheitlichen Struktur, um Klarheit und Vergleichbarkeit zu gewährleisten. Als festgelegte Elemente für jeden Anwendungsfall sind definiert:

Aussagekräftiger Titel
Ein prägnanter Titel, der den Kern des Anwendungsfalls zusammenfasst.

Beschreibung der Situation
Eine kurze Darstellung des Kontextes, in dem der Anwendungsfall relevant ist.

Das Problem
Klare Definition des Problems oder der Herausforderung, die der Anwendungsfall adressiert.

Benötigte Daten
Angaben zu den für die Lösung erforderlichen Daten.

Potenzielle Nutzergruppen
Identifikation der Zielgruppen, die von der Lösung profitieren.

Lösungsansatz
Beschreibung der vorgeschlagenen KI-basierten Lösung.

Nutzen
Darstellung des Mehrwerts oder der Vorteile, die durch den Einsatz der KI-Lösung entstehen.

Einordnung generative KI
Klassifizierung des Anwendungsfalles gemäß den Einsatzgebieten der generativen KI (z. B. Wissensmanagement, Suche, Texterzeugung, Textbearbeitung, Textzusammenfassung, Extrahierung von Informationen, Übersetzung, Bildung und Training, kreative Texte, Sentiment-Analyse, Textklassifizierung, Chatbot). Dies ermöglicht es gegebenenfalls, die Übertragbarkeit auf andere Anwendungsfälle zu demonstrieren oder Synergien aufzuzeigen.

Datenschutz und Datensicherheit
Grobe Bewertung und Einordnung in Hinblick auf Datenschutz und Datensicherheit.

Anmerkungen
Zusätzliche Bemerkungen, die für das Verständnis relevant sein könnten, oder auch Hinweise in Richtung Übertragbarkeit auf andere Verfahren.

1 Leichte/einfache Sprache – Teilhabe ermöglichen (am Beispiel Pressemitteilung)

Image by vectorjuice / Freepik

Einordnung generative KI
Zusammenfassung, Texterzeugung

Aspekte Datenschutz / Datensicherheit
Öffentlich verfügbare Daten

Anmerkungen
Ansatz auf andere Verfahren gut übertragbar

Situation
Ein neues Förderprogramm für Frauen mit Migrationshintergrund wird mittels Pressemeldung bekanntgegeben.

Problem
Das Sprachniveau der Meldung entspricht B2 und ist für Personen mit geringen Deutschsprachkenntnissen nicht verständlich. Die Zielgruppe kann die Pressemeldung nicht lesen bzw. verstehen. Die Erstellung von Texten in leichter Sprache erfordert spezielle Kenntnisse und Ressourcen, welche in vielen Fällen nicht vorhanden sind.

Benötigte Daten
Die ursprüngliche Pressemeldung auf Sprachniveau B2

Potenzielle Nutzergruppen
Sachbearbeiterin/Sachbearbeiter

Lösungsansatz
Mithilfe von KI-Large-Language-Modellen können Texte stilistisch oder mit einer anderen Komplexität umformuliert werden. Es kann ein Dokument hochgeladen und in leichte (A1/A2) oder einfache (B1) Sprache umformuliert werden. Das Ergebnis kann als alternatives Dokument (hier: Pressemeldung) veröffentlicht werden.

Nutzen
Einhaltung der deutschen und europäischen Vorgaben bzgl. Barrierefreiheit in der IT.
Die automatisierte Erstellung von Pressemitteilungen in leichter Sprache spart Zeit und reduziert die Aufwände. Daraus resutiert auch eine schnellere Bereitstellung.

2 Textzusammenfassung – komplexe / lange Dokumente schnell überblicken (am Beispiel Stellungnahme zu Gesetzesentwurf)


Designed by pch.vector / Freepik

Einordnung generative KI
Zusammenfassung

Aspekte Datenschutz / Datensicherheit
Öffentlich verfügbare Daten

Anmerkungen
Ansatz auf andere Verfahren gut übertragbar

Situation
Es ist sehr zeitaufwendig, lange und komplexe Dokumente schnell zu überblicken. Beispiel: Zu Gesetzesentwürfen müssen regelmäßig und zeitnah Stellungnahmen abgegeben werden.

Problem
Eine erhebliche Anzahl an Gesetzesentwürfen muss in kurzer Zeit analysiert werden. Um sich einen Überblick zu verschaffen und eine Zusammenfassung zu erstellen, wäre ein höheres Zeitbudget nötig.

Benötigte Daten
Bereits veröffentlichte Gesetzesentwürfe

Potenzielle Nutzergruppen
Sachbearbeiterin / Sachbearbeiter

Lösungsansatz
Die KI-gestützte Assistenzlösung ist in der Lage, verschiedene Dateiformate (wie z. B. PDF, Word, Textdateien) zu verarbeiten. Das Ergebnis ist eine Zusammenfassung des Inhaltes. Über benutzerdefinierte Einstellungen können Nutzerinnen und Nutzer Parameter wie z. B. gewünschte Textlänge anwenden. Diese Zusammenfassung dient als Vorschlag bzw. Entwurf.

Nutzen
Entlastung der Nutzerinnen und Nutzer, höhere Effizienz, beschleunigte Bearbeitung von komplexen Dokumenten (hier: Stellungnahme zu Gesetzesentwurf)

3 Texterstellung – Anfragen effizienter beantworten (am Beispiel Antwort auf Bürgeranschreiben)

Image by vectorjuice / Freepik

Einordnung generative KI

Klassifizierung, Zusammenfassung, Texterzeugung

Aspekte Datenschutz / Datensicherheit

Öffentlich verfügbare und interne Daten ⇨ abhängig vom Inhalt des Dokumentes

Anmerkungen

Ansatz auf andere Verfahren gut übertragbar

Situation

In der öffentlichen Verwaltung gehen täglich zahlreiche Anfragen, Beschwerden und Vorschläge von Bürgerinnen und Bürgern ein, die per E-Mail oder als Briefe verfasst sind. Die zuständigen Sachbearbeiterinnen und Sachbearbeiter haben die Aufgabe, diese zu lesen, zu kategorisieren und entsprechend zu beantworten.

Problem

Die hohe Anzahl an Anfragen in Verbindung mit der erforderlichen Bearbeitungszeit lässt oftmals eine zeitnahe Beantwortung nicht zu, was zu einer Unzufriedenheit bei den Bürgerinnen und Bürgern führen kann.

Benötigte Daten

Anschreiben, weitere Quelldateien, Input der Sachbearbeitung

Adressatendaten

Potenzielle Nutzergruppen

Sachbearbeiterin / Sachbearbeiter

Lösungsansatz

Die KI-gestützte Assistenzlösung kategorisiert die eingehende Anfrage automatisch, z. B: Einordnung in Typ (Anfragen zu Dienstleistungen, Beschwerden, Vorschlag etc.) und Themenfeld (zuständiges Ressort etc.). Im Anschluss wird ein erster Vorschlag für eine passende Antwort erstellt.

Nutzen

Beschleunigte Bearbeitung von immer wiederkehrenden Anfragen (hier: Antwort auf Bürgeranschreiben).

4 Texterstellung – Standarddokumente effizient verfassen (interne Verwendung)

Image by vectorjuice / Freepik

Einordnung generative KI
Zusammenfassung, Texterzeugung

Aspekte Datenschutz / Datensicherheit
Interne Daten, Berücksichtigung von Rollen und Rechten, Qualitätssicherung

Anmerkungen
Ansatz auf andere Verfahren gut übertragbar

Situation
In der öffentlichen Verwaltung wird täglich eine Vielzahl von Vermerken, Vorlagen, Zusammenfassungen, Einladungen, Protokollen und weiteren Standardschriftstücken für die interne Verwendung erstellt.

Problem
Die Menge an anzufertigenden Schriftstücken – häufig in Form einer wiederkehrenden Tätigkeit – bindet wertvolle Ressourcen im Arbeitsalltag. Diese Arbeitskraft kann nicht für wichtige Fachtätigkeiten eingesetzt werden.

Benötigte Daten
Verzeichnis der Organisationseinheiten, Inhaltsverzeichnis, Gliederung, Textbausteine

Potenzielle Nutzergruppen
Sachbearbeiterin / Sachbearbeiter

Lösungsansatz
KI-gestützte Assistenzsysteme bieten Funktionen wie Zusammenfassung, Rechercheunterstützung, Fließtextgenerierung und beherrschen verschiedene Formate (z. B. Kabinettsvorlage, Vermerk, Protokoll).

Nutzen
Beschleunigte Erstellung von Standarddokumenten, höhere Effizienz

5 Bildgenerierung – passende Illustrationen effizient erzeugen

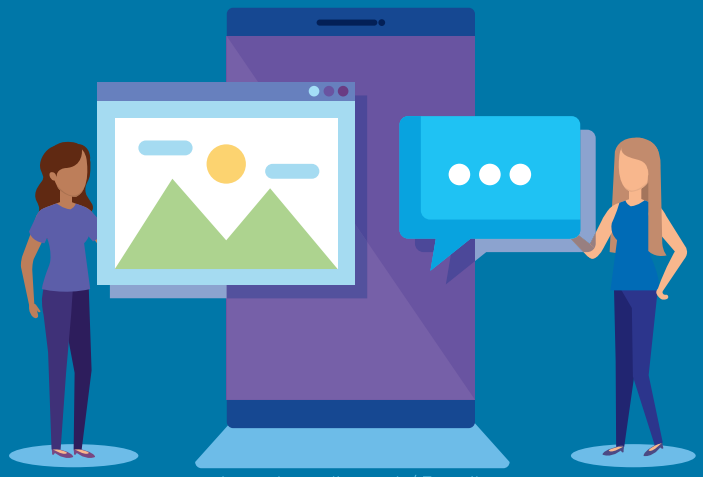


Image by studiogstock / Freepik

Einordnung generative KI
Bilderzeugung

Aspekte Datenschutz / Datensicherheit
Öffentlich verfügbare Daten

Anmerkungen
Ansatz auf andere Verfahren gut übertragbar

Situation

Zur Illustration und Veranschaulichung von Präsentationen, Anleitungen, Webseiten, Berichten, Broschüren etc. werden oft digitale Bilder eingesetzt.

Problem

Erstellung oder Kauf von passendem Bildinhalt sind mit zeitlichem bzw. finanziellem Aufwand verbunden.

Benötigte Daten

Eine genaue textliche Beschreibung des zu erstellenden Bildes

Potenzielle Nutzergruppen

Sachbearbeiterin / Sachbearbeiter

Lösungsansatz

Das KI-Tool ermöglicht den Nutzerinnen und Nutzern, auf Basis von Textbeschreibungen automatisiert digitale Bilder zu erzeugen. Die generierten Bilder sind in vielen Fällen für allgemeine Illustrationszwecke geeignet.

Nutzen

Schnelle und günstige Erstellung von passendem digitalem Bildinhalt durch Nichtgrafikerinnen / Nichtgrafiker.

6 Chatbot – Beantwortung wiederkehrender Fragestellungen (Ergänzung zum internen Mitarbeiterportal MAP)

Image by vectorjuice / Freepik

Einordnung generative KI

Chatbot, Wissensmanagement, Suche, Texterzeugung

Aspekte Datenschutz / Datensicherheit

Interne Daten, Berücksichtigung von Rollen und Rechten, Qualitätssicherung

Anmerkungen

Hohe Komplexität aufgrund geforderter Personalisierung und Systemintegration

Situation

Die öffentliche Verwaltung ist täglich mit wiederkehrenden Anfragen zu gleichen Themen befasst, etwa bei IT-Servicestellen, Personalreferat, internem Dienst etc.; in Zeiten des mobilen Arbeitens auch außerhalb der Servicezeiten.

Besondere Anwendungsfälle sind zum einen die Unterstützung des Onboarding-Prozesses für neue Beschäftigte und zum anderen der Wissenstransfer bei ausscheidenden Beschäftigten (Interviewführung und Dokumentation).

Problem

Die Beantwortung von immer ähnlichen Fragestellungen sowie die Recherche in den Wissensdatenbanken ist zeitaufwendig. Zur Beantwortung der Fragen müssen häufig unterschiedliche Informationen zusammengetragen werden. Die Qualität der übermittelten Antworten variiert in Abhängigkeit von der Bearbeiterin/dem Bearbeiter.

Benötigte Daten

Relevante Wissensdatenbanken bzw. Dokumente, die bereits intern zugänglich sind.

Potenzielle Nutzergruppen

Alle Beschäftigten; neue Mitarbeiterinnen und Mitarbeiter

Lösungsansatz

Die KI gestützte Assistenz in Form eines Chatbots kann z. B. in das Mitarbeiterportal (MAP) integriert werden. Bei Aufruf ist der Chatbot bereits in die passende Rolle (z. B. Onboarding-Prozess) integriert und stellt auch nur Daten passend zur Rolle bereit. Alle relevanten Quellen werden durchsucht und aufbereitet (intelligente Suche und/oder Chatbot).
Ein möglicher technischer Ansatz wäre z. B. die Kombination aus einer KI-gestützten integrativen Chatbot-Plattform (Conversational AI) und einem LLM (Generative AI).

Nutzen

Entlastung der Mitarbeiterinnen und Mitarbeiter, schnellere Beantwortung von individuellen Anfragen, einfache Bedienung, Unterstützung im Onboarding-Prozess, Anreicherung von nachgefragten Informationen (Datenaufbau orientiert am Bedarf)

7 Programmcodeentwicklung bzw. -analyse – Optimierung des Softwareentwicklungsprozesses

Image by vectorjuice / Freepik

Einordnung generative KI
Codeerstellung, -überprüfung und -überarbeitung

Aspekte Datenschutz / Datensicherheit
Mögliche Offenlegung von Quellcode

Situation
Erstellung von Programmcode bzw. Überprüfung und Überarbeitung von bestehendem Code

Problem
Fachkräftemangel in der IT

Benötigte Daten
Anforderungen für einen Programmcode bzw. an eine Software, Spezifikationen

Potenzielle Nutzergruppen
Softwareentwicklerinnen / Softwareentwickler, Softwaretesterinnen / Softwaretester

Lösungsansatz
Mithilfe von KI-Codierungstools, die verschiedene Programmiersprachen beherrschen, lässt sich Softwarecode auf der Grundlage von natürlichsprachlichen Anweisungen generieren bzw. bestehender Code auf Fehler überprüfen.

Nutzen
Code-Generierung in verkürzter Zeit
Überprüfung und Aktualisierung von Code
Übernahme von repetitiven Aufgaben
Durchführung von Softwaretests

8 KI-Unterstützung zur Erhöhung der IT-Sicherheit

Image by vectorjuice / Freepik

Einordnung generative KI

KI für operative IT

Aspekte Datenschutz / Datensicherheit

Zugriff auf Netzwerke und Anwendungen

Absicherung des KI-Modells ist wichtig, damit dieses nicht selbst zum Ziel von Cyberattacken wird

Situation

Cyberangriffe treten immer häufiger auf und werden auch immer komplexer.

Problem

Die manuelle Verfolgung von IT-Bedrohungen ist zeitaufwendig und angesichts des Fachkräftemangels in der IT chancenlos gegenüber der Vielzahl an Cyberattacken im Hintertreffen

Benötigte Daten

Extraktion von verdächtigen Mustern aus Netzwerkdaten und aus IT-Systemen

Potenzielle Nutzergruppen

IT-Sicherheitsbeauftragte, IT-Administratorinnen / IT-Administratoren, IT-Betriebsverantwortliche

Lösungsansatz

Anomalie- und Angriffsdetektion in Kommunikationsnetzen und IT-Anwendungen zur Erkennung von verdächtigen Aktivitäten in Netzwerken und Anwendungen durch automatisierte Auswertung großer Datenmengen

Nutzen

Schnelle, weil automatisierte Erkennung von Anomalien im Datenverkehr, was eine schnellere Reaktion und schnellere Abwehr von Cyberattacken ermöglicht

Anhang 2: Glossar

AI-as-a-Service (AIaaS)

Es handelt sich um ein Cloud-Computing-Service-Modell, das KI als Dienstleistung eines externen Anbieters auf seiner Cloud-Plattform bereitstellt. Die KI-Services werden vom Anbieter verantwortet, in seiner Cloud betrieben und verwaltet. Wie üblich für Cloud-Computing-Services erfolgt die Abrechnung in der Regel nutzungsbasiert in Abonnementmodellen. (→zurück zum Text)

Algorithmus

ist ein Rechenvorgang nach einem bestimmten, sich wiederholenden Schema. (→zurück zum Text)

Ankereffekt

bezeichnet eine kognitive Verzerrung, bei der sich die Entscheidungsfindung oder Einschätzung einer Person stark an einem anfänglichen Referenzwert (dem „Anker") orientiert, selbst wenn dieser irrelevant ist. (→zurück zum Text)

Annotieren

Das Annotieren stellt einen Prozess des Hinzufügens von zusätzlichen Informationen oder Markierungen zu Daten dar, um diese für maschinelles Lernen und generative KI-Modelle verständlicher und nutzbarer zu machen. Annotationen können beispielsweise Textpassagen, Bilder oder Audiodaten markieren, um die Modelle beim Verstehen und Verarbeiten dieser Informationen zu unterstützen. (→zurück zum Text)

Bias

In Bezug auf generative KI und LLMs bezeichnet Bias eine systematische Verzerrung oder eine unausgewogene Darstellung bestimmter Informationen oder Perspektiven. Dies kann durch unausgewogene Trainingsdaten, ungleiche Gewichtung verschiedener Datenquellen oder inhärente Vorurteile in der Modellarchitektur entstehen. (→zurück zum Text)

BSI-Grundschutz

ist ein vom Bundesamt für Sicherheit in der Informationstechnik (BSI) entwickeltes Konzept, das standardisierte Verfahren und Maßnahmen zur Risikoanalyse und Absicherung von Informationstechniksystemen bereitstellt. (→zurück zum Text)

Closed Source LLMs

sind Sprachmodelle, deren Quellcode nicht öffentlich zugänglich ist. Sie werden von Organisationen oder Unternehmen entwickelt und gepflegt, die den zugrunde liegenden Code in der Regel proprietär und für die Öffentlichkeit geschlossen halten. Diese Modelle werden oft als kommerzielle Produkte entwickelt und können Lizenzen oder Abonnements für ihre Nutzung erfordern. Die spezifischen Details ihrer Architektur, Trainingsdaten und Algorithmen sind im Allgemeinen nicht öffentlich zugänglich. (→zurück zum Text)

Conversational AI

bezeichnet eine Technologie, die es Maschinen ermöglicht, menschenähnliche Gespräche mit Benutzerinnen und Benutzern zu führen. Conversational-AI-Systeme nutzen natürliche Sprachverarbeitung und maschinelles Lernen, um menschenähnliche Interaktionen in Text- oder Sprachform zu ermöglichen und werden häufig in Chatbots, virtuellen Assistenten und anderen Anwendungen eingesetzt, um menschliche Kommunikation nachzuahmen und Anfragen von Benutzerinnen und Benutzern zu beantworten. (→zurück zum Text)

Deep Learning

ist eine Methode des *Machine Learnings*, bei dem künstliche neuronale Netze mit komplexen inneren Strukturen ausgebildet werden. (→zurück zum Text)

Dunkelverarbeitung

ist ein automatisierter Prozess in Informationssystemen, bei dem Daten ohne menschliche Interaktion verarbeitet und Entscheidungen getroffen werden. (→zurück zum Text)

Entity-Extraktion

bezeichnet einen Prozess in der Datenverarbeitung und im maschinellen Lernen, bei dem spezifische Informationen wie Namen, Orte, Zeitangaben oder andere relevante Daten aus einem Text extrahiert werden. Diese Technik wird häufig in der Informationsgewinnung und im Bereich der natürlichen Sprachverarbeitung eingesetzt, um strukturierte Informationen aus unstrukturierten Textdaten zu erhalten. (→zurück zum Text)

Explainable AI (XAI)

umfasst Methoden und Techniken in der KI, die darauf abzielen, die Entscheidungsprozesse und Ergebnisse von KI-Systemen verständlich und nachvollziehbar für menschliche Nutzerinnen und Nutzer zu machen. (→zurück zum Text)

Foundation Models

sind KI-Modelle, die auf einer breiten Datenbasis trainiert wurden (in der Regel unter Verwendung von Selbstüberwachung in großem Maßstab) und an eine breite Palette von nachgelagerten Aufgaben angepasst werden können, siehe auch LLM. (→zurück zum Text)

Generative KI / generative AI

ist eine Form der KI, die darauf spezialisiert ist, neue Daten, Texte, Bilder oder andere Inhalte zu erstellen, die auf vorhandenen Informationen oder Mustern basieren. Diese Technologie verwendet Modelle des maschinellen Lernens, um kreativ und autonom Inhalte zu generieren, was in verschiedenen Anwendungen wie Textgenerierung, Bildsynthese und Musikkomposition eingesetzt wird. (→zurück zum Text)

Generative Pre-trained Transformer (GPT)

GPT-Modelle sind künstliche neuronale Netzwerke, die auf der Transformer-Architektur basieren, auf großen Datensätzen vorab trainiert werden und in der Lage sind, neuartige, menschenähnliche Inhalte zu generieren. (→zurück zum Text)

GPT-4

ist eine GPT-Technologie in der Version 4, entwickelt von OpenAI, die auf einem großen Sprachmodell basiert und in der Lage ist, komplexe Aufgaben wie Texterstellung, Übersetzung und Problemlösung zu bewältigen. (→zurück zum Text)

Halluzination

Im Kontext generativer KI ist der Begriff „Halluzination" eher metaphorisch und wird verwendet, um Fälle hervorzuheben, in denen das Modell überzeugend formulierte, aber objektiv inkorrekte und damit irreführende Inhalte erzeugt. Der Begriff ist insofern missverständlich, als Halluzination eigentlich eine Abweichung von der Realität, wie sie vom Menschen wahrgenommen wird, impliziert und somit eine Ebene des Bewusstseins bzw. Verständnisses suggeriert, die KI-Modelle nicht besitzen. (→zurück zum Text)

HDSIG

Hessisches Datenschutz- und Informationsfreiheitsgesetz. Das Landesdatenschutzgesetz Hessen regelt vor allem die Verarbeitung personenbezogener Daten durch öffentlichen Stellen des Landes sowie durch die Gemeinden und Landkreise. (→zurück zum Text)

Indirect Prompt Injection

Angriffe können die Daten in Quellen von LLMs manipulieren und dort unerwünschte Anweisungen für LLMs platzieren. Greifen LLMs auf diese Daten zu, werden die unerwünschten Befehle unter Umständen ausgeführt. Dadurch kann das Verhalten der LLMs gezielt manipuliert werden. Die potenziell schadhaften Befehle können kodiert oder versteckt sein und sind für Anwenderinnen und Anwender unter Umständen nicht erkennbar. (→zurück zum Text)

Intent-Erkennung

ist ein Prozess in der KI und im maschinellen Lernen, bei dem die Absicht hinter einer Benutzeranfrage oder einem Textausschnitt identifiziert wird. Diese Technik wird vor allem in Chatbots und Sprachassistenten angewendet, um die Anfragen der Benutzerinnen und Benutzer zu verstehen und entsprechend darauf zu reagieren. (→zurück zum Text)

Klassifikation

ist ein grundlegendes Verfahren im Bereich des maschinellen Lernens und der KI, bei dem Datenobjekte oder -instanzen basierend auf ihren Merkmalen in vordefinierte Kategorien eingeteilt werden. Diese Methode wird häufig verwendet, um Muster in Daten zu erkennen und Entscheidungen zu treffen, beispielsweise in der Bild- oder Texterkennung. (→zurück zum Text)

Künstliches neuronales Netz (KNN)

ist ein Netzwerk mit innerer Struktur, das dem menschlichen Gehirn auf maschineller Ebene nachempfunden ist. (→zurück zum Text)

Large Language Model (LLM)

ist ein Computersystem mit *NLP*, das große und komplexe Mengen an Texten generieren kann. (→zurück zum Text)

LeoLM

ist ein Open Source LLM, bereitgestellt von der deutschen Organisation LAION und dem Hessischen Zentrum für Künstliche Intelligenz hessian.AI. Es basiert auf dem LLM Llama 2 und ist für die deutsche Sprache optimiert. (→zurück zum Text)

Llama 2

ist ein Open Source LLM, bereitgestellt vom Unternehmen Meta. (→zurück zum Text)

Luminous

ist ein kommerzielles LLM des deutschen Unternehmens Aleph Alpha. (→zurück zum Text)

Machine Learning / Maschinelles Lernen (ML)

findet statt, wenn Computersysteme darauf trainiert werden, sich aus Daten weiterzuentwickeln anstatt hierfür explizite Programmierungen vorzunehmen. (→zurück zum Text)

Multimodales LLM

Ein multimodales Large Language Model ist eine Art von KI-System, das fähig ist, neben Text auch visuelle Elemente zu verarbeiten. (→zurück zum Text)

Multi-Tenant

ist eine Software-Architektur, bei der eine einzelne Instanz einer (IT-)Anwendung mehrere Kunden oder Benutzergruppen (sogenannte „Tenants“) gleichzeitig bedient, wobei die Daten und Konfigurationen jedes Tenants isoliert und unabhängig voneinander verwaltet werden. (→zurück zum Text)

Natural Language Processing (NLP)

bezeichnet die Fähigkeit eines Computersystems, menschliche Sprachen zu verstehen, zu interpretieren und zu manipulieren. (→zurück zum Text)

On-Premises

ist ein Ansatz in der Informationstechnologie, bei dem Hard- und Softwareprodukte lokal im physischen Gebäude der Organisation installiert und betrieben werden, im Gegensatz zu Lösungen, die in der Public Cloud gehostet werden. (→zurück zum Text)

Open Source LLMs

sind Sprachmodelle, deren Quellcode öffentlich zugänglich ist und von jedermann frei genutzt, verändert und verbreitet werden kann. (→zurück zum Text)

Pay-as-you-go-Modell

ist ein flexibles Zahlungsmodell, bei dem Kundinnen und Kunden nur für die tatsächlich genutzten Ressourcen oder Dienstleistungen bezahlen, anstatt pauschale oder vorab festgelegte Gebühren zu entrichten. Dieses Modell ist besonders in der Cloud-Computing-Branche verbreitet, wo es Nutzerinnen und Nutzern ermöglicht, ihre Kosten entsprechend der Skalierung und dem tatsächlichen Verbrauch von Rechenleistung, Speicherplatz oder anderen Diensten anzupassen. (→zurück zum Text)

Prompt

ist eine initiale Text- oder Befehlseingabe, die eine Benutzerin oder ein Benutzer einem Large Language Model gibt, um eine spezifische Antwort oder Ausgabe zu generieren. (→zurück zum Text)

Prompt Engineering

meint die Eingabe der Beschreibung einer durch die KI durchzuführenden Aufgabe in ein Bedienfeld. ➜ Da es eines der zentralen Elemente im Bereich generativer KI darstellt und oft missverstanden wird, gibt es ein eigenes Kapitel mit weiteren Erläuterungen. (→zurück zum Text)

Prompt Injection

Dabei manipulieren Angreifer die Eingabeaufforderungen (Prompts) einer generativen KI, um die Ausgabe des Systems zu steuern oder zu verändern, oft mit dem Ziel, unerwünschte Ergebnisse oder Verhaltensweisen zu provozieren. (→zurück zum Text)

Retrieval Augmented Generation (RAG)

ist eine Technik zur Verarbeitung natürlicher Sprache, die die Stärken von abfragebasierten und generativen Modellen der KI kombiniert. (→zurück zum Text)

Sentiment-Analyse

ist ein Verfahren der KI und des maschinellen Lernens, das darauf abzielt, die emotionale Tonlage hinter Textdaten zu identifizieren und zu klassifizieren. (→zurück zum Text)

Token / Token Limit

Das Token Limit bei *LLMs* bezieht sich auf die maximale Anzahl von Tokens, die das Modell in einer einzigen Anfrage verarbeiten kann. Ein Token ist dabei die grundlegende Einheit der Verarbeitung und kann je nach Sprachmodell und Sprache unterschiedlich definiert sein. Oft repräsentiert ein Token ein Wort, einen Teil eines Wortes oder ein Satzzeichen. (→zurück zum Text)

Impressum

Herausgeber

Hessisches Ministerium für Digitalisierung und Innovation
Georg-August-Zinn-Straße 1
65183 Wiesbaden

Pressestelle: 0611 32 11 4222
E-Mail: pressestelle@digitales.hessen.de
Internet: www.digitales.hessen.de

Verantwortlich im Sinne des Presserechts

Markus Büttner, Pressesprecher
Hessisches Ministerium für Digitalisierung und Innovation

Redaktion:

Dr. Tanja Klein, Dr. Tina Klug, Frauke Werner
Hessisches Ministerium für Digitalisierung und Innovation

Dirk Dohn, Dr. Caroline Wagner
Hessisches Ministerium des Innern, für Sicherheit und Heimatschutz

Judith Drebert, Ph.D., Erin Polster
Hessisches Ministerium für Wirtschaft, Energie, Verkehr, Wohnen und ländlichen Raum

Dr. Petra Förg, Dr. Magdalena Kircher, Hessische Zentrale für Datenverarbeitung
Michael Fritz, Hessische Zentrale für Datenverarbeitung (ext.)

Dr. Christian Hermann, Hessisches Ministerium der Justiz und für den Rechtsstaat

Winfried Hock, Hessische Staatskanzlei

Christian Voß, Finanzamt Kassel – Forschungsstelle Künstliche Intelligenz

Abbildung Cover:
stock.adobe.com: Leonid

Grafiken:
freepik.com: vectorjuice, pch.vector, studiogstock

Gestaltung:
Theißen-Design, Lohfelden

Stand: Mai 2024

Ausschluss Wahlwerbung:

Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit der Hessischen Landesregierung herausgegeben. Sie darf weder von Parteien noch von Wahlbewerbern oder Wahlhelfern während eines Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für Landtags-, Bundestags- und Kommunalwahlen sowie Wahlen zum Europaparlament. Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen, an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel.

Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl darf die Druckschrift nicht in einer Weise verwendet werden, die als Parteinahme der Landes-regierung zugunsten einzelner politischer Gruppen verstanden werden könnte.

Die genannten Beschränkungen gelten unabhängig davon, auf welchem Wege und in welcher Anzahl diese Druckschrift dem Empfänger zugegangen ist. Den Parteien ist es jedoch gestattet, die Druckschrift zur Unterrichtung ihrer eigenen Mitglieder zu verwenden.

HESSEN

Hessisches Ministerium für
Digitalisierung und Innovation

digitales.hessen