Transkript Video 1 „Gameshow und Tiefpunkte – Fakten auf dem Prüfstand“

Hallo und herzlich willkommen zum Promptlabor!

Dieses Video ist das erste echte inhaltliche Video im Kurs. Und damit wir optimal auf die kommenden Videos eingestimmt sind, beginnen wir mit einem ab-so-lu-ten… Tiefpunkt. Wobei, für uns als Menschen ist dieser Tiefpunkt eigentlich ein echter Triumph. In diesem Video lernen wir den Schwachpunkt kennen, den alle generativen KI-Tools heute haben: die Faktentreue. Wir werden erleben, wie sich dieser Schwachpunkt in verschiedenen KI-Tools zeigt. Und wir machen erste Erfahrungen damit, in welchen Einsatzszenarien wir den Aspekt der Faktentreue ganz besonders im Auge haben müssen. Lasst uns loslegen!

Allen von uns fallen sofort eine Vielzahl von Szenarien ein, die wir an Hochschulen mit KI-Tools bespielen können. Dazu gehören natürlich die Generierung von Texten, von Plänen, von Illustrationen, aber natürlich auch von Prüfungsaufgaben. Wir können Zusammenfassungen und Übersetzungen anfertigen, aber auch einfach nur Ideen generieren. Wir können uns bei Gutachten oder Bewertungen unterstützen lassen, wir können Alltagsaufgaben automatisieren – und viele Hochschulen experimentieren damit, Chatbots auch in der Beratung und Betreuung einzusetzen, zum Beispiel in der Studienberatung. All das werden wir uns im Rahmen dieses Kurses anschauen. Das wird richtig gut und wir werden in ganz vielen Fällen sehen, wie wir generative KI-Tools effektiv, nutzbringend und mit großem Mehrwert einsetzen können.

Nun ist aber all diesen Szenarien gemein, dass sie mit Wissenschaft zu tun haben. Und in der Wissenschaft geht es um: Fakten. Damit sind wir direkt schon mittendrin im versprochenen Tiefpunkt. Um den zur Gänze würdigen können, müssen wir zuerst drei Basics über textgenerative KI-Tools besprechen. Viele von euch werden die schon kennen, aber damit wir auf einem Stand sind, schauen wir ganz schnell auf drei Begriffe. Das, was wir an die KI schicken, bezeichnen wir als Prompt, zu deutsch die Eingabe an die KI. Der Prompt kann ausschließlich aus Text bestehen, aber auch multimediale Elemente enthalten, wie etwa Bilder oder ganze PDF-Dateien. Die Antwort der KI wird englisch in der Regel als Response bezeichnet. Diese Response der KI besteht für uns menschliche Nutzende aus Worten, für die KI aber zerfallen diese Worte noch einmal in die sogenannten Tokens. Tokens können ganze Worte sein, aber auch Wortteile oder Satzzeichen. Ich habe hier mal ein Beispiel angezeigt. Das Wort aktuelle besteht für die KI aus zwei Bestandteilen. Zum einen der Wortteil “aktu” und zum zweiten der Wortteil “elle”. Beides sind Tokens, die dann zusammengesetzt ein vollständiges Wort ergeben. Der konkrete Umrechnungsfaktor von Tokens in Worte hängt dabei von der Sprache ab, in der gechattet wird. Im englischen braucht man für das durchschnittliche Wort etwas weniger Tokens als im deutschen – einfach weil deutsche Worte länger sind. Ganz grob können wir damit rechnen, dass ein Token je nach Sprache etwa einem halben bis einem dreiviertel Wort entspricht. Mit diesem Wissen können wir in unser erstes Kräftemessen mit der KI einsteigen. Ich habe ChatGPT in der Version Vier-o folgenden Prompt gegeben: “Zähle die Wörter in diesem Prompt und nenne mir das Ergebnis als Zahl”. So, und jetzt können wir alle kurz innehalten und diese Aufgabe für uns selbst erledigen. Also, ihr Lieben, zählt bitte die Wörter in diesem Prompt und behaltet die Zahl im Kopf. ChatGPT jedenfalls antwortet, dass der Prompt dreizehn Wörter enthält. Und wer von euch ebenfalls auf die Zahl dreizehn gekommen ist… herzlichen Glückwunsch – das ist die richtige Antwort.

Also alles gut? Nee, nicht ganz. Wir werden nämlich noch sehr oft in unserem Kurs die Erfahrung machen, dass wir dieselbe Sache nicht nur ein einziges Mal an eine KI prompten sollten, sondern denselben Prompt mehrfach ausführen, um zu schauen, wie unterschiedlich die Antworten der KI sind.

Das stellt sicher, dass wir nicht zufällig einmal eine besonders gute oder auch im Gegenteil eine ungewöhnlich schlechte Antwort der KI bekommen haben. Heißt: wir machen das Ganze nochmal! Dazu öffnen wir einen neuen Chat und geben denselben Prompt noch einmal ein. Der neue Chat ist deshalb wichtig, weil die KI im alten Chatfenster ja nun weiß, was sie vorher geantwortet hat. Entsprechend wird sich ihre Antwort vermutlich nicht ändern. Also, noch einmal: "Zähle die Wörter in diesem Prompt und nenne mir das Ergebnis als Zahl". Jetzt antwortet ChatGPT, dass der Satz achtzehn Wörter enthält. Das stimmt nicht mehr. Also nochmal versuchen. Beim dritten Mal zählt ChatGPT zwölf Wörter, beim vierten Mal vierzehn und beim fünften Mal elf. Von fünf Versuchen hat ChatGPT also genau einmal – und das vermutlich rein zufällig – die richtige Anzahl von Wörtern genannt. Das ist seltsam. Vermutlich haben die allermeisten von uns richtigerweise dreizehn Wörter gezählt und diejenigen von euch, die auf eine andere Zahl gekommen sind, waren vermutlich nicht so weit vorbei wie die achtzehn oder die elf.

Woran das liegt, sehen wir im nächsten Video. Vorher aber probieren wir das Ganze zur Sicherheit in verschiedenen KI-Tools aus – vielleicht ist ChatGPT ja besonders schlecht im Zählen. Also, schnell Claude geöffnet und dort dasselbe eingegeben. Claude haut direkt beim ersten Versuch vollkommen daneben und zählt achtundzwanzig Wörter. Beim zweiten Versuch sind es fünfundzwanzig, beim dritten achtzehn, beim vierten fünfzehn und beim letzten Versuch sind es dann dreißig Wörter. Wir waren also nur einmal relativ nah dran, in allen anderen Fällen aber so weit weg, dass sich vermutlich niemand von uns auch nur annähernd so heftig verzählt hat. Rüber zu Google: was sagt Gemini dazu? Google Gemini zählt direkt im ersten Versuch einundsiebzig Wörter. Da muss man nicht mal zählen, um zu sehen, dass das keine einundsiebzig Wörter sein können. Nächster Versuch vierzehn. Im dritten Versuch: Korrekt! Dreizehn Wörter. Direkt im nächsten Versuch sind es dann aber wieder siebzehn und im letzten Versuch achtzehn Wörter. Wir sehen also: keine unserer KIs kann zählen – und das liegt am Funktionsprinzip von KI, dass wir uns im nächsten Video näher anschauen werden.

Im Augenblick würde ich also sagen: Eins zu Null für die Menschheit. Das war aber nur der erste Wettbewerb. Einen hab' ich noch. Für diesen werden wir ein bisschen textbasierter, denn vielleicht war das Zählen ja einfach eine Aufgabe, mit der die KI grundsätzlich nicht so viel anfangen kann. Lasst uns also etwas fairer werden und tatsächlich Text generieren lassen. Dazu stellen wir der KI zwei verschiedene Fragen. Die Struktur der beiden Fragen ist dieselbe, nur beziehen sie sich auf unterschiedliche Textkörper. Und um dieses Mal wirklich fair zu sein, chatten wir in Englisch. Das ist für viele KI-Tools so etwas wie die "First language", weil der Anteil englischsprachiger Trainingsdaten deutlich höher ist als in anderen Sprachen. Entsprechend gab es Berichte, dass Prompting in englischer Sprache zu besseren Ergebnissen führt. Selbst wenn das einmal gestimmt haben sollte, dürfte das ein vorübergehendes Phänomen sein, dessen Beobachtung ohnehin eher auf anekdotischer Evidenz als auf verlässlichen Forschungsdaten beruht hat.

Also: Wenn Ihr auf Deutsch promptet – keine Sorge… das läuft schon. Unser erster Prompt jedenfalls lautet: "Cite the second sentence of Chapter four of the novel "Ulysses" by James Joyce.". Auch bei der zweiten Frage geht es um den zweiten Satz eines vierten Abschnitts, nur dieses Mal in der Verfassung der Vereinigten Staaten. Der Prompt ist: "Cite the second sentence of Article Four Section One of the Constitution of the United States." Macht gerne wieder mit – haltet das Video kurz an und recherchiert selbst – ohne KI-Hilfe! Beide Aufgaben brauchen nicht mehr als eine Internet-Suche mit drei Worten. Meine Suche nach "Joyce Ulysses PDF" hat mich zu Projekt Gutenberg geführt. Da findet man definitiv die Antwort… der zweite Satz in Kapitel vier ist: "He liked thick giblet soup, nutty gizzards…" und so weiter. Üh-ähh. In derselben Weise führt die Suche nach "Verfassung USA PDF" ebenfalls zum Ziel, nämlich zur Website des US Constitution Center. Der zweite Satz in Artikel Vier Abschnitt Eins ist ganz offensichtlich "And the Congress may by general Laws prescribe the Manner in which… undsoweiter." Und jetzt schauen wir, wie unsere KI-Tools abschneiden. Aber auch hier: probiert gerne erst mal selbst aus, was eure KI zu dieser Frage sagt. Bei mir jedenfalls ist ChatGPT-Vier-o der Meinung, der erste Satz in James Choice Roman Ulysses sei: „Millie Bloom, fair-haired, fresh cheeked, looked at the photograph and then looked at her daughter-in-law." Das ist kompletter Unsinn. Dieser Satz kommt im

ganzen Roman nicht ein einziges Mal vor. Das Ergebnis wird auch nicht besser, wenn man es nochmal versucht… und nochmal… und nochmal… und nochmal… ChatGPT denkt sich jeweils wunderbare Texte aus – von einem korrekten Zitat sind wir allerdings weit entfernt.

Google Gemini macht das etwas besser. Es sagt nämlich, dass es aufgrund seiner Eigenschaft als generative KI keine Zitate produzieren kann. Das ist zwar unbefriedigend, aber aus meiner Sicht deutlich besser als eine komplette Erfindung wie bei ChatGPT. Claude verhält sich genauso wie Gemini. Es führt aus, dass es als KI Modell nicht zitieren kann, gibt aber wie auch Gemini ein paar Hinweise, wie wir als Nutzende weiter vorgehen könnten. Fairerweise muss ich sagen, dass bei unseren vielen Versuchen mit ChatGPT genau einmal eine ähnliche Antwort wie bei den anderen beiden KI-Tools dabei war und ChatGPT kein Zitat gerade mal frei erfunden hat. Richtig gut ist das alles trotzdem nicht.

Übrigens haben einige KI-Tools schon seit geraumer Zeit die Fähigkeit, für ihre Antworten im Internet zu recherchieren. Das tun sie nicht immer von sich aus, sondern man muss zuweilen explizit darauf hinweisen. Lasst uns das ausprobieren! Wir beginnen wieder mit ChatGPT. Hier ergänze ich unseren Ulysses Prompt um den Zusatz: „Search the internet if you are unsure.“ Das Ergebnis überzeugt: offenbar hat ChatGPT drei Webseiten durchsucht und auf einer davon das richtige Zitat gefunden. Netterweise gibt es uns auch noch die Quellen seiner Suche an. Wunderbar, genauso will man es doch haben… könnte man denken, aber weit gefehlt! Denn schon beim nächsten Versuch haut die KI wieder eine fehlerhafte Antwort raus. Das Zitat ist zwar weniger falsch als ohne die Internetsuche – das, was da steht, ist der erste Satz im vierten Kapitel, nicht der zweite – aber falsch ist das Ergebnis eben trotzdem. Und das, obwohl die KI für Ihre Antwort im Internet gesucht und dort tatsächlich auch die richtige Quelle gefunden hat. Bei Google Gemini bekommen wir direkt beim ersten Versuch ein falsches Zitat, das im ganzen Roman nicht ein einziges Mal auftaucht. Interessanterweise hat Google Gemini die Fähigkeit, für einen Prompt mehrere Antwortvorschläge zu formulieren. Wenn wir auf den entsprechenden Button klicken, sehen wir, dass der zweite Vorschlag die korrekte Antwort enthält. Also ebenfalls besser als ohne Internetrecherche, aber immer noch extrem unbefriedigend. Okay, wie sieht es bei unserem Zitat aus der US amerikanischen Verfassung aus? Hier ist das Ergebnis deutlich besser. Sowohl ChatGPT als auch Google Gemini und auch Claude liefern bis auf das letzte i-Tüpfelchen ein korrektes Zitat. Das ist erstmal ziemlich verblüffend. Zwei weltbekannte Texte, zwei sinn-gleiche Prompts, aber vollkommen unterschiedliche Antwortqualitäten. Im einen Fall versagen alle KI-Tools auf die eine oder andere Weise, im anderen Fall liegen alle drei Tools komplett richtig. Vermutlich war das bei Euren Versuchen ähnlich. Für uns im Wissenschaftssystem ist das ein ziemlich dramatisches Ergebnis. Wenn wir uns bei den Antworten eines KI-Tools nicht darauf verlassen können, ob wir tatsächlich Fakten geliefert bekommen, oder reine Fiktion – dann haben wir ganz offensichtlich ein gewaltiges Problem. Oder vielleicht doch nicht?

Im Video zum Thema „Faktentreue bei generativen KI-Tools werden wir aber sehen, dass der Schlüssel zum Umgang mit diesem Problem ein solides Verständnis davon ist, wie generative KI-Tools genau arbeiten. Sobald wir das haben, werden wir selbst sehr gut bewerten können, welche Anfragen an ein generatives KI-Tool sinnvoll sind und welche eben auch nicht. Und das war's auch schon für dieses Video. Ciao und… bis gleich.

Transkript Video 2 „Funktionsprinzipien generativer KI-Tools“

Hallo und herzlich willkommen zum Promptlabor!

In diesem Video lernen wir das grundlegende Funktionsprinzip generativer KI-Tools kennen. Vermutlich habt ihr das alle schon auf dem Schirm und deswegen könnte es gut sein, dass euch wenig von dem, was wir in diesem Video erarbeiten, wirklich neu ist. Wir erfahren nämlich unter anderem, warum sich alle KI-Tools, die auf dem Prinzip des Large-Language-Models basieren, einfach an rein gar nichts erinnern. Wir lernen auch, was das für uns bedeutet. Aber das Wichtigste kommt am Schluss. Wir analysieren da nämlich ein Prompt-Beispiel, um zu prüfen, ob wir das Funktionsprinzip von Large-Language-Modellen tatsächlich zu 100% durchschaut haben. Selbst wenn ihr also den Inhalt dieses Videos schon weitgehend kennt, spult vor bis zu diesem Timecode und macht unsere kleine Denkaufgabe mit. Lasst uns loslegen! In unserem Gameshow-Video – den Link findet ihr in der Videobeschreibung – haben wir direkt einen Tiefpunkt bei der Leistungsfähigkeit generativer KI-Modelle erlebt. Zum einen konnten die Large-Language-Modelle nicht rechnen und zum zweiten haben sie entweder fantasiert oder sie haben uns gesagt, dass sie die Anfrage nicht beantworten können. Ausprobiert haben wir das mit einer Passage aus dem Roman Ulysses von James Joyce. ChatGPT hat munter darauflos fantasiert und uns Zitate gebracht, die im ganzen Roman nicht ein einziges Mal vorkommen, während Google Gemini und Claude zugegeben haben, dass sie die Frage prinzipiell nicht beantworten können, weil sie keinen Zugriff auf den Originaltext haben. Bei der Verfassung der Vereinigten Staaten sah das gänzlich anders aus. Alle Modelle konnten diese Passage perfekt zitieren. Woher kommt diese Diskrepanz zwischen zwei eigentlich sinngleichen Aufgaben? Naja, erstmal muss zugeben, dass schon das Wort “Zitieren” eine komplett falsche Wortwahl von mir war. Large-Language-Modelle – LLMs… und das sind alle textgenerativen KI-Tools, mit denen wir aktuell arbeiten – haben nämlich keinen Zugriff mehr auf die Trainingsdaten, mit denen sie ursprünglich gefüttert worden sind. Ihr alle wisst, dass LLMs während ihres Trainings unfassbar viele Textdokumente verarbeitet haben. Nach dem Training aber haben sie auf diese Textdokumente keinen Zugriff mehr. Sie können also bei einer Anfrage nicht mal eben die Quelle nachschlagen und dann “abschreiben” oder eben “zitieren”. Wenn wir sie also nach einem Zitat fragen, dann ist das eine Aufgabe, die sie grundsätzlich nicht erfüllen können. Ihr seht aber die Fußnote: Bei aktuellen LLMs ist das womöglich ein bisschen anders. Einige davon können nämlich zum Beispiel im Internet suchen und hätten so Zugriff auf alle Wissensdatenbanken, die wir als Menschen auch benutzen könnten. Darunter natürlich auch Wikipedia oder Projekt Gutenberg im Falle von freien Werken aus der Literatur. Auch das haben wir mit ChatGPT probiert und es war tatsächlich in der Lage, die Originalquelle von Ulysses zu finden und auszuwerten. Nicht immer fehlerfrei, aber immerhin. Begründet sind solche Schwachpunkte im Funktionsprinzip von Large-Language-Modellen. Lasst uns dieses Funktionsprinzip kurz anschauen. Und auch wenn ihr jetzt sagt: ach komm, kenne ich doch schon. Es dauert nicht lange und vielleicht nehmt ihr noch ein bisschen was mit, um den Kenntnis-Check am Ende des Videos problemlos abzuhaken. Also, wie funktionieren LLMs? Sobald eine generative KI unsere Anfrage verarbeitet hat, sucht die KI nach einem Wort – oder richtiger einem Token –, das als erstes Token ihrer Antwort am besten passen könnte. Dazu hat sie auf ihr riesengroßes Sprachnetzwerk zurückgegriffen, das sie während des Trainings aufgebaut hat. Sie stellt also eine Liste von Tokens zusammen, die jetzt mit der höchsten Wahrscheinlichkeit kommen könnten. Für Ihre tatsächliche Antwort nimmt sie aber nicht immer direkt das erste Wort aus dieser Liste, sondern sie würfelt, um zu entscheiden, welches der Worte auf der Liste denn tatsächlich eingesetzt werden soll. Wir simulieren mal diesen Würfelwurf und sehen: die Reihenfolge hat sich geändert. Nun steht das Wort

“The” an erster Stelle und genau das wird jetzt auch als erstes Wort in die Antwort der KI eingefügt. Und so geht das weiter. Die KI schaut sich nun unsere Frage plus ihr erstes Wort an und sucht nach Tokens, die jetzt folgen könnten. Das könnten wieder ganz naheliegende Dinge sein, wie second oder auch “2-nd” mit einer Zahl vorne, aber auch weniger naheliegende Tokens. Für diese neue Liste wird erneut gewürfelt. Diesmal ändert sich an der Reihenfolge nichts, sodass das nächste Wort das Wort second wird. Und genau so geht das jetzt weiter. Die KI listet und würfelt und listet und würfelt und es entsteht so Schritt für Schritt die Antwort. Hier ist es ganz wichtig, dass genau das, was Ihr dort auf der linken Seite seht, auch bei uns im Browser passiert. Die KI schreibt nicht zuerst ihre komplette Antwort, sondern sie streamt die von ihr ausgewürfelten Worte direkt an unseren Browser, sodass wir der KI sozusagen beim Schreiben zuschauen können. Jetzt stellt sich natürlich sofort die Frage, warum sich alle von uns getesteten KI-Tools bei James Joyce so schwer getan haben, während sie die amerikanische Verfassung perfekt zitieren konnten. Vermutlich liegt das daran, dass während des Trainings die amerikanische Verfassung in wahnsinnig vielen Kontexten in den Trainingsdaten der KI aufgetaucht ist – und zwar jeder einzelne Satz und jedes einzelne Wort davon – während vermutlich James Joyce weniger häufig diskutiert worden ist und schon gar nicht der zweite Satz im vierten Kapitel. Deshalb wird die Verfassung der Vereinigten Staaten korrekt wiedergegeben. Aber eben nicht zitiert, sondern es entsteht rein zufällig die richtige Wortreihenfolge und zwar deshalb, weil diese Wortreihenfolge im Kontext des Promptes die mit Abstand wahrscheinlichste Antwort auf unsere Frage ist. Übrigens, Anthropic Claude ist das einzige KI-Tool, das auf diesen Umstand auch bei seiner richtigen Antwort hingewiesen hat. Das hat einen zweiten Haken verdient! Okay, wenn man über Langzeitgedächtnis spricht, sollte man auch über Kurzzeitgedächtnis sprechen und auch das haben Large-Language-Modelle nicht. Sie erinnern sich nämlich an nichts aus einem laufenden Chat, und das kommt uns natürlich erst mal total komisch vor, weil es unserer Alltagserfahrung im Umgang mit KI-Tools vollkommen widerspricht. Ich habe euch hier mal ein Beispiel eingeblendet: Das ist eine Pressemitteilung von der Homepage der RWTH und ich frage ChatGPT nach einer Zusammenfassung in fünf Stichpunkten. Die Pressemitteilung ist lang, da entstand gerade die Zusammenfassung, dann frage ich noch etwas, das ist aber gar nicht so wichtig. Wichtig wird es hier. Als dritten Prompt bitte ich die KI, den zweiten Satz des von mir oben einkopierten Originaltextes wörtlich zu zitieren und genau das tut die KI auch – und zwar richtig. Nun können wir uns denken, Mensch prima, die KI hat sich an das erinnert, was wir oben geschrieben haben. Das ist aber nicht der Fall. Die Tatsache, dass die KI einen früheren Prompt korrekt zitieren kann, liegt schlicht und einfach daran, dass alles, was zuvor im gleichen Chat geschrieben worden ist, also sowohl unsere Anfragen als auch die Antworten der KI, für die nächste Antwort der KI immer wieder mitgeschickt wird. Die KI liest also in längeren Konversationen wieder und wieder und wieder dieselben Inhalte und generiert ihre Antworten jeweils genau passend zu dem, was vorher gesprochen worden ist. Wobei wir hier eine Einschränkung machen müssen. Das Erinnerungsfenster aller Large-Language-Modelle ist natürlich nicht unendlich groß. Wenn unsere Konversation zu lang wird, dann fallen die zuerst innerhalb der Konversation geschriebenen Inhalte aus diesem Erinnerungsfenster heraus. Sie werden dann einfach nicht mehr mitgeschickt. Und die Größe der Erinnerungsfenster der verschiedenen KI-Tools ist massiv unterschiedlich. So hat etwa das Llama- Modell der Firma Meta Erinnerungsfenster zwischen zweitausend und achttausend Tokens, also so etwa eintausend-fünfhundert bis sechstausend Worte. Claude hingegen punktet mit zweihunderttausend Tokens. ChatGPT kann bis zu einhundert-achtundzwanzigtausend Tokens verarbeiten, genauso wie Mistral. Googles Gemini hat kürzlich einen gigantischen Sprung auf Erinnerungsfenster von bis zu einer Million Tokens gemacht. Was passiert nun, wenn unsere Konversation zu lang wird und damit die anfangs geschriebenen Inhalte nicht mehr an die KI geschickt werden? Bei manchen KI-Tools wird man darüber nicht informiert. ChatGPT hingegen tut das netterweise und schreibt uns dann: “Diese Konversation ist zu lang, bitte starte eine neue Konversation.” Das ist zwar lästig, aber immerhin nett, weil es dann nicht passieren kann, dass wir uns plötzlich fragen, warum die KI Dinge nicht mehr weiß, die wir ihr am Anfang mitgeteilt haben. So, mit all diesem Wissen ausgestattet kommen wir jetzt endlich zu unserem Erkenntnischeck. Vor einiger Zeit

hat eine Teilnehmerin in einem Workshop von mir Folgendes gemacht. Es war die Aufgabe zu lösen, ein wissenschaftliches Exposé durch die KI schreiben zu lassen. Die Teilnehmerin hat ihren Prompt angefangen mit der Information an die KI, dass sie eine gewissenhaft und hochintelligente wissenschaftliche Mitarbeiterin an einer Uni sei und das Schreiben eines wissenschaftlichen Exposés unterstützen soll. Sie hat dann die fünf Abschnitte eines Exposés genannt und weil sie mit früheren Antworten der KI unzufrieden war, hat sie sich gedacht, dass sie der KI einfach mal denselben Hinweis gibt, den wir unseren wissenschaftlichen Mitarbeiterinnen oder Mitarbeitern geben würden, die Probleme beim Formulieren von Texten haben. Und das ist: "Du, schreib doch am besten nicht direkt den gesamten Text, sondern mach erstmal Stichwortlisten für die einzelnen Abschnitte. Wenn du die hast, dann formulierst du einen Fließtext drumherum. So besiegst du die Angst vorm leeren Blatt." Alles gut soweit. Jetzt aber kommt die entscheidende Anweisung an die KI. Die Teilnehmerin hat nämlich geschrieben: "Du, liebe KI, gibst mir nicht diese Stichwortlisten aus, die interessieren mich gar nicht, sondern Du gibst mir nur den endgültigen Fließtext aus." Jetzt die Frage an euch. Wenn die KI den Prompt befolgt, existieren zu irgendeinem Zeitpunkt während der Textgeneration der KI die gewünschten Stichwortlisten? Kurz das Video anhalten und nachdenken. Bis vor kurzem lautete die Antwort: nein, die Stichwortlisten existieren nicht. Und das hat mit dem gerade vorgestellten Funktionsprinzip der KI zu tun. Denn die KI streamt ja jedes von ihr erzeugte Wort direkt an unseren Browser. Da **kann** es keine Stichwortlisten geben, die irgendwo auf einem unsichtbaren Notizzettel der KI stehen. LLMs können also aufgrund ihres grundlegenden Funktionsprinzips die Anweisung im Prompt nicht umsetzen. Aber… diese Antwort ist inzwischen nicht mehr korrekt, seit die Firma Anthropic ihr KI-Modell Claude 1.5 Sonnet an den Markt gebracht hat: Claude führt sogenannte Artefakte ein. Das sind Dinge, die während der Text-Generation entstehen, aber nicht sofort Teil der Antwort werden müssen, sondern auf die die KI dann später zurückgreifen kann, wenn sie der Meinung ist, dass sie sie benötigt. Auch hier sehen wir mal wieder: Ki entwickelt sich so schnell, dass das, was heute noch korrekt ist, morgen bereits überholt sein kann. Übrigens, ganz anders wäre die Lage natürlich gewesen, wenn unsere Workshop-Teilnehmerin in ihrem Prompt die Ausgabe der Stichwort explizit angegeben hätte. Dann wären die Stichwortlisten Teil der Antwort der KI geworden und hätten dann die späteren Formulierung des Fließtextes beeinflusst. Die eigentliche Frage, die wir uns auf Basis dieses allgemeinen Funktionsprinzips generativer KI-Tools aber stellen müssen, lautet: wie können wir LLMs als "Wahrscheinlichkeitsautomaten" dazu bringen, faktentreuer zu sein, als wir es etwa am Beispiel von James Joyce Ulysses erlebt haben. Genau auf diesen Punkt der Erhöhung der Faktentreue gehen wir in einem eigenen Video ein. Den Link dazu findet ihr in der Videobeschreibung. Und das war's auch schon für dieses Video. Ciao und… bis gleich.

Transkript Video 3 „Faktentreue von KI-Modellen“

Hallo und herzlich willkommen zum Promptlabor! In diesem Video kümmern wir uns um die Faktentreue von KI-Modellen. Wenn wir dieses Video hinter uns haben, können wir die Faktentreue von KI-Tools selbst beeinflussen. Wir lernen Retrieval Augmented Generation kennen, abgekürzt RAG, und wir werden leider erfahren, dass auch RAG nicht in der Lage ist, die Faktentreue von KI-Modellen zu garantieren, sondern nur in gewissem Umfang zu erhöhen. Lasst uns loslegen! Alle textgenerativen KI-Modelle erlauben es, den Phantasiegehalt der Textgeneration über verschiedene Konfigurationsparameter zu beeinflussen. Bei ChatGPT zum Beispiel heißt ein solcher Parameter Temperature. Ihr Wert kann von Null bis Eins reichen und je höher der Wert ist, desto zufälliger werden die Antworten des KI-Modells. Der Hersteller von ChatGPT, die Firma OpenAI, bietet in ihrem sogenannten Playground die Möglichkeit, die Wirkung solcher Parameter selbst auszuprobieren und genau das wollen wir mal zusammen tun. Ihr seht hier genau diesen Playground vor euch und ich habe dort zwei KI-Chats gleichzeitig gestartet. Die beiden Chats werden hier zwar nebeneinander angezeigt, aber sie sind komplett unabhängig voneinander. In beiden Chats stelle ich dieselbe Frage, nämlich: “Was versteht man unter dem Begriff digitale Ethik?” Auf der linken Seite können wir nun verschiedene der Konfigurationsparameter manipulieren, die sich auf den Phantasiegehalt auswirken. Die Temperature da oben kennen wir schon, das ist tatsächlich die Zufälligkeit der KI-Antworten. Der Parameter tot p hingegen beeinflusst die Variabilität der Antworten der KI. Ein sehr niedriges totP führt dazu, dass die generierten Texte sich sehr ähnlich bis hin zu identisch sind, während höhere Werte die Variabilität der Wortwahl erhöhen. Erinnert Ihr Euch an das Video zum Funktionsprinzip textgenerativer KI-Tools? Da haben wir gezeigt, dass die KI für jedes nächste Token ihrer Antwort eine Liste von potentiellen Kandidaten anlegt und dann auswürfelt, welcher Kandidat es dann tatsächlich in die Antwort schafft. Ein totP von Null bedeutet, dass immer nur das erste Wort aus der Liste genommen wird und ein totP von Eins sagt, dass aus der gesamten Liste gewürfelt wird. Mit einer Temperature von Null, also bitte nicht fantasieren, und einem totP von auch Null, also der kleinstmöglichen Variabilität der Textgeneration, produzieren beide Chats nahezu dieselbe Antwort. Die Ergebnisse unterscheiden sich nur an zwei Stellen. Links schreibt die KI “Fragen” rechts “Aspekte” und auf der linken Seite ist eine Satzkonstruktion etwas anders als auf der rechten Seite. Ansonsten sind die Texte wortwörtlich identisch. Was passiert nun, wenn wir sowohl die Temperature als auch die totP aufdrehen, und zwar jeweils auf den Wert von 1? Man sieht, dass die beiden Texte zwar immer noch die digitale Ethik zum Thema haben, dass aber die erzeugten Wortreihenfolgen deutlich unterschiedlich sind. Auch zentrale konkrete Inhalte sind unterschiedlich. Das ist ja super, könnte man jetzt denken. Aber natürlich muss direkt jetzt der Downer kommen. Denn obwohl alle textgenerativen KI-Systeme solche Konfigurationsparameter haben, können wir als Nutzende diese Parameter in aller Regel nicht ändern. Denn die meisten der generativen KI-Tools erlauben nicht deren direkte Änderung über die Web-Oberfläche, sondern dazu müssten wir die Programmierschnittstelle nutzen. Und das tun oder können natürlich eher wenige von uns. Deshalb helfen uns diese Konfigurationsparameter rein gar nichts. Und das hat natürlich dazu geführt, dass findige Menschen versucht haben, dieses Manko zu umgehen. Eine Idee war, direkt im eigenen Prompt die Einstellungen für die jeweiligen Parameter zu setzen. In aller Regel funktioniert das aber nicht. Wir sehen ein Beispiel hier, wo ich in den Prompt aufgenommen habe, dass das KI-Tool eine Temperature von Null und ein totP von ebenfalls Null benutzen soll. Also der kleinste Zufälligkeits- und Variabilitätsgrad, der überhaupt möglich ist. Wenn wir uns dann mal den Text anschauen, der dabei entsteht, sieht man, dass dieser Text von derjenigen Variante, die wir vorhin im Playground mit Temperature gleich Null und totP gleich Null gesehen haben, massiv verschieden ist. Klappt also nicht. Es gibt aber eine Idee, die ganz gut zu funktionieren scheint. Wenn man nämlich

in die eigenen Prompts folgenden Satz aufnimmt: “Wenn du dir unsicher bist, erfinde keine Fakten, sondern sag mir, dass du unsicher bist.”, scheint dies tatsächlich dazu zu führen, dass KI-Tools weniger erfundene Inhalte produzieren. Inzwischen gibt es aber eine Technologie, die zuverlässig und wirksam in der Lage ist, die Faktentreue von KI-Tools zu erhöhen. Und diese Technologie nennt sich Retrieval Augmented Generation, abgekürzt RAG. Zum ersten Mal beschrieben wurde dieser Ansatz Zweitausend-zwanzig von Patrick Lewis und Kolleg:innen, aber so richtig in den Mainstream gekommen ist das Ganze erst im Jahr Zweitausend-dreiundzwanzig. Da überschlugen sich in den verschiedenen Fachorganen die Berichte, Erklärungen und Analysen eben genau zu Retrieval Augmented Generation. Und wir schauen uns jetzt an, wie dieses Verfahren funktioniert. Das hier ist ein typisches Flussdiagramm für RAG. Wir gehen es mal Schritt für Schritt durch. Alles beginnt mit einer großen Sammlung an Dokumenten, dem sogenannten Textkorpus. Diese Sammlung wird zusammengestellt und dient uns als Grundlage für unsere Fakten. Nennen wir das Ganze mal die Faktenbasis. So ein Textkorpus, das könnten zum Beispiel alle Dokumente aus dem Intranet einer Hochschule sein oder die Inhalte aus den Lernräumen im Lernmanagementsystem oder auch die eigene Literaturdatenbank auf dem Notebook. Die in der Wissensbasis enthaltenen Dokumente werden dann zunächst in ein Large Language Model hineingekippt. Dieses Large Language Model ist nicht eines, wie wir es kennen, denn es ist nicht darauf trainiert, mit Menschen eine Konversation zu betreiben. Stattdessen ist es daraufhin optimiert, die Dokumente auszuwerten, zusammenzufassen, zu verschlagworten und auf die enthaltenen Wissenselemente herunterzubrechen. Dieses in Einzelteile zerlegte Wissen wird dann in eine Vektordatenbank eingefügt. Die Datenbank enthält nun eine komprimierte Repräsentation all des Wissens, das in unserer Faktenbasis vorhanden war. Und damit ist unser RAG-System fast einsatzbereit. Wir müssen nur noch ein echtes Konversations-KI-Modell hinzufügen. Das ist das zweite Large Language Model in diesem Konstrukt und das brauchen wir, damit wir uns wie gewohnt mit der KI unterhalten können. Wenn wir als User nun einen Prompt an die KI schicken, wird sie – anders als es ohne RAG der Fall wäre – nicht sofort an unseren Chatbot geschickt. Statt dessen wird unser Prompt zunächst an das Embedding-Model gesendet. Das Embedding-Model analysiert unseren Prompt und formuliert daraus eine Datenbankabfrage. Das könnte auch ein Programmierer machen, der den Prompt liest und dann daraus eine so genannte Datenbank Query schreibt. Diese Query wird dann an die Datenbank geschickt. Die Datenbank liefert die gewünschten Inhalte und diese Inhalte werden dann gemeinsam mit dem Prompt an das für die Konversation mit uns verantwortliche Large Language Model gesendet. Dieses Large Language Model produziert nun seine Antwort auf unseren Prompt, aber es berücksichtigt dabei eben die Fakten, die es aus der Datenbank erhalten hat. So können wir gewährleisten, dass die Antworten der KI Wissensbestandteile enthalten, die die KI selbst nicht produzieren könnte, sondern die wir ihr nachträglich als Faktenbasis geliefert haben. Das, was wir dann im Browser sehen, ist ein hoffentlich faktentreueres Textgenerat. Das Problem daran ist, dass auch RAG ebenso wie alle anderen Technologien zur Erhöhung der Faktentreue bei generativen KI-Modellen fehlerbehaftet ist. Und speziell im Fall von RAG gibt es nicht nur eine, nicht zwei, sondern gleich drei Fehlerquellen. Fehlerquelle Eins ist natürlich, dass bereits die Eingangsdaten für die Vektordatenbank fehlerhaft sein können. Unsere Faktenbasis enthält also Fehler, die dann natürlich auch Eingang in die Vektordatenbank finden. Das ist das alte Prinzip “Garbage in – Garbage out”. Zusätzlich kann auch das sogenannte Embedding-Model beim Aufbau der Datenbank Fehler machen. Es ist sofort denkbar, dass das Embedding-Model die Dokumente nicht faktengetreu analysiert oder nicht die wesentlichen Informationen herauszieht oder enthaltene Informationen in einer Weise komprimiert, die den ursprünglichen Sinngehalt verschleiern. Die Datenbank enthält also deshalb fehlerhaftes Wissen, weil das Embedding-Model eben nicht korrekt gearbeitet hat. Und schließlich kann es auch sein, dass der Chatbot beim Einbau der aus der Datenbank gelieferten Informationen Fehler macht, weil es die Fakten falsch zusammensetzt oder fehlerhafte Schlüsse formuliert. Und ganz zentral ist, dass auch mit RAG das KI-Modell immer noch nicht zwingend zitieren kann. Das liegt schlichtweg daran, dass in unserer Vektordatenbank die Originaltexte ja nicht mehr enthalten sind. Ein Zugriff auf die auf die exakten Textquellen unserer Faktenbasis ist deshalb auch mit RAG nicht möglich, sondern

erfordert weitergehende Technologien. Diese drei Fehlerquellen sorgen dafür, dass wir uns weiterhin auf die Antworten von generativen KI-Tools nicht verlassen können und dass auf eine Qualitätssicherung durch uns Menschen weiterhin nicht verzichtet werden darf. Techniken wie RAG und vergleichbare Technologien erhöhen die Faktentreue, sie garantieren sie aber nicht. Und das ist die wichtigste Erkenntnis im gesamten Kurs. Auch wenn ihr alles, was wir mit euch in diesen Videos besprechen, wieder vergesst. Die Notwendigkeit der Prüfung der Textgenerate durch uns Menschen ist _die_ eine Erkenntnis, die wir uns irgendwo ganz tief eingravieren müssen. Und das war's auch schon für dieses Video. Ciao und… bis gleich.

Transkript Video 4 „Verhalten generativer KI-Tools beeinflussen“

Hallo und herzlich willkommen zum Promptlabor! In diesem Video kümmern wir uns darum, wie wir das Verhalten von KI-Tools beeinflussen können. Wir lernen drei Stellschrauben kennen, mit denen wir KI tatsächlich für unsere Bedarfe individualisieren können. Wir klären die Bedeutung, die der Upload von Dokumenten in ein KI-System für uns hat. Wir bringen KI das Rechnen bei. Und wir werfen einen kurzen Blick auf System Prompts bzw. Custom Instructions. Lasst uns loslegen. Also, das Verhalten der KI beeinflussen. Die zwei Fragen, die sich da natürlich sofort stellen, ist erstens, wie geht das? Aber zweitens, warum sollte man das überhaupt tun? Um das zu verstehen, lohnt es sich, noch einmal ganz kurz auf den Herstellungsprozess von generativen KI-Tools schauen. Also, wir starten bei jedem KI-System mit dem Programmcode. Das ist die Software, die am Ende die Grundlage des KI-Tools bildet. Wir kippen dann eine hoffentlich gewaltige Menge von Trainingsdaten in dieses KI-Modell hinein und trainieren es mit diesen Daten. Wir haben dann eine prä-trainierte KI. Diese KI kann bereits Texte produzieren, aber sie ist noch nicht das, was wir am Ende als Chatbot benutzen können. Wenn wir uns mit einem solchen prä-trainierten KI-Modell unterhalten würden, dann würden wir sehr schnell feststellen, dass da etwas nicht stimmt. Es kann sich zwar mit uns unterhalten, aber das ist noch keine normale Konversation, die wir mit einer anderen Person haben würden. Um das zu erreichen, gibt es die nächste Phase des Trainings, das Fine-Tuning. Da wird die KI noch einmal mit ganz spezifischen Trainingsdaten gefüttert, um sie auf bestimmte Konversationseigenschaften hin zu optimieren. Im Falle der Chatbots, die wir benutzen können – ChatGPT, Google Gemini, Claude, Mistral, Llama – führt dieser Schritt dazu, dass die KI zu einem sprachlichen Chamäleon wird. Sie kann sich danach unterhalten wie ein vierjähriges Kind, aber auch eine zweiundvierzig-jährige Chemikerin oder Ernest Hemingway. Dieser Chatbot kann jetzt ganz exzellent Text produzieren, aber er hat weiterhin mehrere Defizite. Das erste dieser Defizite ist, dass dieser Chatbot nichts weiß – sondern wenn er korrekte Inhalte produziert, dann tut er dies nur rein zufällig. Das haben wir bereits im Video zum Funktionsprinzip generativer KI-Tools gesehen. Für dieses Defizit haben wir eine zumindest recht weitreichende Lösung, nämlich die Wissensbasis. Wir haben dazu die Technologie der Retrieval Augmented Generation kennengelernt, kurz RAG. Damit garantieren wir zwar noch keine Faktentreue, aber wir sind wesentlich näher dran. Also, Haken an die Faktentreue. Zusätzlich ist der Chatbot nicht auf den ganz spezifischen Sprachstil in bestimmten Anwendungssituationen spezialisiert. Er kann zwar sprechen wie eine Wissenschaftlerin, aber wenn es um die ganz spezifischen sprachlichen Konstrukte geht, die man zum Beispiel im Rahmen eines wissenschaftlichen Artikels zur physikalischen Chemie benutzen müsste, dann kann der Chatbot das vielleicht noch nicht hundertprozentig. Damit er diese Rolle perfekt ausfüllen könnte, benötigt er nicht nur ein Fine-Tuning, sondern ein Fine-Tuning mit Domänenspezifität. Hier wird die KI auf einen ganz bestimmten Anwendungskontext hin optimiert, sodass er die ihm zugewiesene Rolle danach möglichst perfekt übernehmen kann. Andere Rollen gelingen dann vielleicht nicht mehr so gut, aber diese eine kann er dann nahezu perfekt. Das dritte Defizit ist, dass der Chatbot immer noch nicht auf spezifische Konversationsprotokolle vorbereitet ist. Er weiß zum Beispiel noch nicht, ob er uns siezen oder duzen soll, in welcher Sprache zu antworten ist, und er nimmt, wie wir am Anfang des Videos gesehen haben, erst einmal die Rolle eines normalen Gesprächspartners ein. Er verhält sich also noch nicht wie eine Tutorin oder ein Mentor oder eine zentrale Studienberaterin oder ein wissenschaftlicher Mitarbeiter. Wie das geht, müssen wir ihm erst sagen. Und das tun wir entweder im Prompt selber oder wir nutzen dafür spezialisierte Werkzeuge wie die System Prompts oder die Custom Instructions. Da kommen wir gleich noch drauf. Lasst uns den bis hierhin wichtigsten Punkt aber nochmal zusammenfassen: Die typischen Chatbots, mit denen wir uns üblicherweise unterhalten,

sind von Hause aus auf das Führen durchschnittlicher Gespräche mit normalen Personen trainiert. Wenn wir wollen, dass sie sich anders verhalten oder schlauer sind, müssen wir Lösungen implementieren. Zum Glück haben wir gerade gesehen, dass wir für alle drei Herausforderungen ziemlich gute Lösungen haben. Allerdings sind nur zwei davon für uns als End-User tatsächlich realistische Optionen. Und damit sind wir bei den aktuellen Entwicklungen. Für die Wissensbasis können wir als Nutzende solcher KI-Tools in aller Regel kein ausgewachsenes Retrieval Augmented Generation durchführen – dafür braucht es Spezialist:innen. Aber wir können in fast allen Systemen inzwischen Dokumente hochladen. Und genau das ist ein Weg für uns als User, dem KI-System eine Wissensbasis zumindest in begrenztem Umfang bereitzustellen. Zusätzlich haben nicht alle Anbieter die sogenannten System Prompts oder Custom Instructions. Was aber insofern kein großes Problem ist, als das nahezu alles, was wir damit machen könnten, auch mit ganz normalen Prompts an die KI umzusetzen ist. Problematisch wird es beim Feintunen. Lange war das tatsächlich den KI-Expert:innen vorbehalten. Denn für das Feintuning sind zum einen größere Textkorpora notwendig, mit denen die KI dann nachtrainiert werden kann. Und die Technologie selbst war lange so komplex zu bedienen, dass für uns als Nutzende eigentlich keine Chance bestand, ein Finetuning selbst umzusetzen. Das ändert sich gerade, denn es gibt erste KI-Anbieter wie zum Beispiel OpenAI, die ein solches Feintuning auch für uns als User geöffnet haben. Allerdings sind die Einstiegshürden immer noch sehr hoch und nach meiner Einschätzung eben noch nichts für uns als Otto-Normaluser. Damit sollten wir noch einmal einen näheren Blick auf Dokumente werfen, die wir User der KI als Wissensbasis unterschieben können. Die meisten KI-Tools beherrschen das inzwischen und zum Beispiel in Claude sieht das aus wie hier dargestellt. Wir hängen einfach eines oder mehrere Dokumente an und beziehen uns bei unseren Prompts auf diese Dokumente. Die KI wertet die Dokumente dann aus und integriert die Inhalte dieser Dokumente in ihre Antworten. Wir sollten kurz klären, was eigentlich nach dem Upload der Dokumente in der KI selbst passiert. Denn da gibt es im Wesentlichen drei Varianten, wie die verschiedenen KI-Anbieter damit umgehen. Variante Eins ist die komplexeste. Da wird tatsächlich ein RAG-artiges Verfahren benutzt, um aus den hochgeladenen Dokumenten zunächst eine Vektordatenbank zu generieren, sodass bei allen folgenden Anfragen von uns nicht mehr das komplette Dokument selbst ausgewertet werden muss, sondern nur noch die zuvor extrahierten Wissensinhalte aus der Datenbank herausgesucht werden müssen. Das spart massiv Rechenkapazität auf Seiten der Herstellenden. Variante Zwei ist sozusagen die weniger schlaue Variante. Hier werden bei jeder Anfrage von uns, die sich auf die hochgeladenen Dokumente bezieht, diese erneut komplett gelesen. Das dabei entstehende Verarbeitungsvolumen ist entsprechend riesig. Lasst uns den Unterschied zwischen diesen beiden Varianten nochmal ganz klar grafisch fassen. Wenn ich zum Beispiel bei Google Gemini eine Literaturquelle hochlade und dann nach der ersten Zitation im Literaturverzeichnis dieser Quelle im APA-Format frage, dann passiert mit der Retrieval Augmented Generation Variante folgendes. Das Dokument ist bereits beim Upload verarbeitet worden und es ist eine Datenbank daraus entstanden. Diese Datenbank liefert dann dem KI-Modell Informationen zu unserer Anfrage, wie etwa das Jahr der Zitation, den Titel, die Autorinnen und Autoren und so weiter. Und diesen Datenbankeintrag wandelt die KI dann auf Basis seines internen Wissens über das APA-Format in die gewünschte Antwort um. Ohne RAG sieht das Ganze völlig anders aus. Hier wird der komplette Artikel gelesen. Die KI findet hoffentlich den ersten Eintrag im Literaturverzeichnis, liest diesen Eintrag und erzeugt auf Basis dieses Eintrags dann eine APA-formatierte Variante davon. Das Ergebnis kann in beiden Fällen dasselbe sein. Der Weg dorthin ist aber jeweils komplett unterschiedlich. Es gibt noch eine dritte Variante, denn die Mischung aus den beiden vorgenannten ist eigentlich der Optimalfall. Wir haben ja gesehen, dass RAG nicht garantiert, dass aus Artikeln zitiert werden kann. Wenn also Variante Eins durchgeführt wird, geht eine Frage wie: “Zitiere den dritten Satz im Artikel” ins Leere, weil die aufgebaute Vektordatenbank nicht mehr den Text im Original enthält. In Variante Drei entscheidet das KI-Modell selbst auf Basis des Inhalts unserer Anfrage, ob sie auf die Datenbank zugreift oder zur korrekten Beantwortung unserer Frage den Text der Originalquelle noch einmal lesen muss. Die gute Nachricht ist, dass alle kommerziellen KI-Tools inzwischen Datei-Uploads erlauben, aber noch nicht alle Open-Source-Benutzeroberflächen dies auch implementieren. Es kann also sein,

dass ihr bei den Systemen, die eure Hochschule euch anbietet, noch nicht die Möglichkeit habt, Dokumente hochzuladen. Zum Abschluss dieses Videos werfen wir noch einen schnellen Blick auf System Prompts und Custom Instructions. Was ist das? Wir beginnen mit einem ganz einfachen Beispiel. Wenn wir ChatGPT in der Version 4o-Mini fragen, was das Ergebnis von viertausend-vierhundert-dreiunddreißig geteilt durch siebzehn ist, dann antwortet ChatGPT, dass dieses Ergebnis zweihundersechzig sei. Und jetzt frage ich euch, stimmt das oder hat die KI hier gerundet? Der Antwort können wir das nicht ansehen und ich wäre nicht in der Lage zu entscheiden, ob viertausend-vierhundert-dreiunddreißig ganzzahlig durch siebzehn zu teilen ist. Der Taschenrechner sagt uns, dass genau das nicht geht, sondern dass das Ergebnis ungefähr zweihundertsechzig-komma-sieben-sechs-vier ist. Wir können ChatGPT 4o-Mini nun aber explizit im Prompt bitten, Python zu verwenden, um unsere Anfrage zu bearbeiten. Dann wird zunächst der Python-Code ausgegeben, die Berechnung tatsächlich intern mit Python durchgeführt und das Ergebnis an uns zurückgemeldet. In ähnlicher Weise hätten wir auch im Rahmen unserer Gameshow die KI-Tools anweisen können, Python zu benutzen, um die Worte in diesem Prompt zählen zu lassen. Es ist aber mühselig, diese Anweisung immer als Teil unseres Prompts aufzuschreiben. Deshalb gibt es in verschiedenen KI-Tools die Möglichkeit der sogenannten Custom Instructions oder der System Prompts. Intern unterscheiden sich diese beiden Konzepte leicht, aber für uns sind sie im Wesentlichen dasselbe. In ChatGPT sieht das Ganze folgendermaßen aus: Wir können über ein separates Menü eine individuelle Konfiguration vornehmen, indem wir dort Anweisungen hineinschreiben, die das KI-Modell ab dann bei jedem neuen Chat immer wieder berücksichtigt. Man kann sich diese Custom Instructions oder System Prompts wie einen unsichtbaren Prompt vorstellen, der jeder Konversation vorangestellt wird, die wir mit der KI beginnen. In einem solchen System Prompt können wir der KI alles mitgeben, was wir von ihr grundsätzlich über alle Konversationen hinweg erwarten. Ich habe in diesem System Prompt mal aufgenommen, dass die KI Python benutzen soll, wann immer möglich, und dass sie im Internet recherchieren soll, wenn sie sich unsicher ist oder eine Frage ohne eine Recherche im Internet nicht beantworten kann. Dieser Prompt wird ab jetzt Teil jeder Konversation, die wir beginnen. Was damit passiert, seht ihr hier. Die KI liefert plötzlich das richtige Ergebnis und rechts neben diesem Ergebnis taucht ein kleines Icon auf. Wenn wir da draufklicken, dann zeigt uns die KI den Python-Code an, den sie zur Beantwortung unserer Anfrage geschrieben und ausgeführt hat. Das Ergebnis hat sie dann in ihren Prompt übernommen. ChatGPT ist also in der Lage, nicht nur Python-Code zu schreiben, sondern diesen Python-Code auch direkt auszuführen, um unsere Anfragen damit zu ergänzen. Google Gemini kann das auch. Wenn wir Google Gemini anweisen, dass es Python nutzen soll, um die Aufgabe des Wortezählens zu lösen, dann wird wie bei ChatGPT dieser Python-Code ausgegeben. Allerdings ist die Antwort der KI unter diesem Code immer noch inkorrekt. Da steht die Zahl Zehn und nicht, wie es korrekt wäre, Dreizehn. Wenn wir dann aber auf den Playbutton klicken, führt auch Google den Python-Code aus und liefert natürlich das richtige Ergebnis. Die falsche Antwort im Chat selber ändert sich dadurch aber noch nicht. Das ist noch ein Fehler von Google Gemini, den Google wahrscheinlich in der nächsten Zeit beheben wird. Claude hingegen kann das noch nicht. Mit System Prompts und Custom Instructions müssen wir diese Bitte um die Verwendung von Python nicht wieder und wieder in jeden Prompt aufnehmen, sondern sie werden immer unsichtbar als erster Prompt in eine Konversation eingefügt. Sie zeichnen sich vor allem auch dadurch aus, dass – selbst wenn wir einmal eine so lange Konversation schreiben sollten, dass sie das Token-Window überschreitet – diese System Prompts oder Custom Instructions niemals aus diesem Token-Window herausfallen, sondern immer mitgeschickt werden. Und sie haben je nach KI-Tool auch ein höheres Gewicht als andere Prompts, sodass es nicht so leicht ist, Vorgaben aus den System Prompts oder Custom Instructions in den nachfolgenden Prompts, die wir als User an die KI schicken, zu überschreiben. Das Problem daran ist: Nicht alle KI-Tools erlauben solche System Prompts bzw. Custom Instructions. Wenn Euer KI-Tool das aber anbietet: nutzt sie! Und das war es auch schon für dieses Video. Ciao… und bis gleich.

Transkript Video 5 „Urheberrecht und KI“

Scene 1:

Herzlich willkommen zu Modul 2: Recht, Ethik und Verantwortung.

Scene 2:

In diesem Video zeigen wir dir, wo es beim Thema Urheberrecht mit KI Probleme geben kann, wann die Hochschule haftet und wie du KI-Ergebnisse sicher im Arbeitsalltag nutzt.

Scene 3:

Im nächsten Abschnitt erfährst du welche Ziele dich erwarten.

Scene 4:

Du lernst, warum das Urheberrecht in Deutschland meist nur für Werke von Menschen gilt. Außerdem erfährst du, dass reine KI-Texte oft keinen Schutz genießen. Wir zeigen dir, worauf du bei Lizenzen und Haftung achten solltest. Am Ende erhältst du eine kurze Checkliste, mit der du KI-Texte, Bilder oder Code sicher an der Hochschule einsetzen kannst.

Scene 5:

Jetzt zeigen wir dir, warum das Thema KI und Urheberrecht für deinen Hochschul-Alltag wichtig ist.

Scene 6:

Immer mehr Hochschulverwaltungen nutzen heute Programme mit künstlicher Intelligenz – kurz KI. Das kann zum Beispiel beim Auswerten von Formularen sein oder mit Chatbots, die Anfragen beantworten. Viele fragen sich aber: Wem gehören eigentlich die Texte, die so entstehen? Wir geben dir einen Überblick, damit du im Arbeitsalltag auf der sicheren Seite bist.

Scene 7:

Im nächsten Abschnitt erfährst du die wichtigsten rechtlichen Grundlagen zum Urheberrecht und KI.

Scene 8:

Das Urheberrecht schützt in Deutschland Werke nur dann, wenn sie von Menschen kreativ geschaffen wurden. Das heißt: Ein Computerprogramm, das Texte oder Bilder erzeugt, ist nicht kreativ im menschlichen Sinne. Erst wenn du als Mensch etwas dazutust – zum Beispiel, indem du den Text umschreibst oder ergänzt – kann daraus ein geschütztes Werk werden. Dann gilt der Schutz für dich oder für die Hochschule.

Scene 9:

Jetzt schauen wir uns an, welchen Schutz KI-generierte Inhalte tatsächlich genießen.

Scene 10:

Ein Beispiel: Du lässt dir von ChatGPT einen kurzen Info-Text erstellen, etwa für das Studierendenportal. Solange du diesen Text direkt übernimmst, kann jeder andere ihn auch nutzen – es besteht kein besonderer Schutz. Das ist praktisch, aber auch ein Nachteil: Der Text ist nicht einzigartig. Wenn du den Text aber änderst, ergänzt oder mit eigenen Ideen verbindest, kann daraus ein geschütztes Werk werden. Wichtig ist also: Dein eigener kreativer Beitrag zählt!

Scene 11:

Im nächsten Schritt klären wir, warum Lizenzbedingungen und Haftung beim Einsatz von KI wichtig sind.

Scene 12:

Die meisten Anbieter:innen von KI-Programmen schreiben in ihren Geschäftsbedingungen, was mit den Daten und Texten passieren darf. Wer gegen diese Regeln verstößt, kann abgemahnt werden. Das kann teuer werden. Die Hochschule haftet für Fehler, die ihre Mitarbeitenden machen. Darum ist es wichtig: Lies die Regeln (AGB) der Anbieter:innen. Frag im Zweifel die Rechtsabteilung und halte fest, was du geprüft hast.

Scene 13:

Jetzt kommt unsere praktische Checkliste für den sicheren Umgang mit KI-Inhalten.

Scene 14:

Vor der Veröffentlichung von KI-Inhalten gibt es fünf einfache Punkte zu prüfen: Erstens – Darfst du laut Lizenz und AGB den Text wirklich nutzen, auch öffentlich? Zweitens – Gib keine sensiblen oder persönlichen Daten in die KI ein. Drittens – Frag bei Unsicherheiten früh die Rechtsabteilung. Viertens – Gib immer an, woher der Text oder das Bild stammt. Fünftens – Schau alle paar Monate nach, ob sich Regeln geändert haben. So bleibst du immer rechtlich auf dem neuesten Stand!

Transkript Video 6 „Datenschutz-Grundlagen zum Einsatz von Chatbots“

Scene 1:

In diesem Video zeige ich dir was du zum Thema Datenschutz beim Einsatz von Künstlicher Intelligenz wissen musst.

Scene 2:

In dieser Einheit zeigen wir dir, welche Pflichten beim Einsatz von Chatbots besonders wichtig sind.

Scene 3:

Am Ende weißt du, welche sechs Grundpflichten die DSGVO für den Umgang mit Chatbots vorschreibt.

Scene 4:

In diesem Abschnitt erfährst du, warum Datenschutz bei Chatbots an der Hochschule so wichtig ist.

Scene 5:

Studierende nutzen einen Chatbot viel lieber, wenn sie sicher sein können, dass ihre Daten geschützt werden. Schon ein kleiner Vorfall kann Vertrauen zerstören. Mit einigen einfachen Regeln zeigst du dass die Hochschule verantwortungsvoll mit persönlichen Informationen umgeht. So wird Datenschutz verständlich und ist kein reines Juristenthema mehr.

Scene 6:

In diesem Abschnitt lernst du die wichtigsten Gesetze kennen, die für den Einsatz von Chatbots gelten.

Scene 7:

Für alle Chatbots in Deutschland und der EU gilt die DSGVO, das Datenschutz-Grundgesetz. Das deutsche Bundesdatenschutzgesetz regelt einige Details, zum Beispiel, welche Behörde kontrolliert. Für Chatbots mit künstlicher Intelligenz gibt es keine besonderen Ausnahmen. Sobald persönliche Daten wie Name oder IP-Adresse verarbeitet werden, müssen die allgemeinen Datenschutzregeln eingehalten werden.

Scene 8:

Jetzt zeigen wir dir die sechs wichtigsten Datenschutz-Grundpflichten, die du beim Einsatz von Chatbots beachten musst.

Scene 9:

Merk dir diese sechs wichtigen Punkte des Artikel 5 der Datenschutzgrundverordnung: Erstens: Rechtmäßigkeit und Transparenz. Du musst Einwilligung oder Vertrag klar erklären. Zweitens: Zweckbindung. Daten nur für das vorher angekündigte Ziel verwenden. Drittens: Datenminimierung. Wirklich nur das abfragen, was gebraucht wird. Viertens: Richtigkeit. Die Daten aktuell halten, zum Beispiel wenn jemand den Kurs wechselt. Fünftens: Speicherbegrenzung. Chats nach einer festgelegten Zeit löschen. Sechstens: Integrität und Vertraulichkeit. Daten mit technischen Maßnahmen schützen, etwa durch Verschlüsselung oder Zugang nur für bestimmte Rollen. Und für alle Punkte gilt: Du musst nachweisen können dass du dich daran hältst.

Transkript Video 7 „Die KI-Verordnung der EU: Der AI Act“

Scene 1:

Seit August 2024 gibt es ein neues, einheitliches KI-Gesetz für ganz Europa den sogenannten AI Act. In diesem Video erfährst du für wen dieses Gesetz gilt warum es sowohl technische Neuerungen als auch den Schutz von Grundrechten fördert und was das für die tägliche Arbeit bedeutet zum Beispiel auch welche Schulungen gemacht werden müssen.

Scene 2:

Jetzt geht es um die zentralen Anforderungen und Ziele des neuen AI Act.

Scene 3:

Du lernst die drei wichtigsten Ziele des AI Act kennen, erfährst, wer sich an das neue Gesetz halten muss und bekommst einen Überblick über die Schulungspflicht nach Artikel 4.

Scene 4:

Wir zeigen dir jetzt, warum der AI Act für alle EU-Länder eine so zentrale Rolle spielt.

Scene 5:

Der AI Act ist das erste große KI-Gesetz, das in allen EU-Staaten gilt. Es will zwei Dinge erreichen: Einerseits fördert es Innovation, also neue Technologien. Andererseits schützt es Grundrechte, Sicherheit und Gesundheit. Praktiken wie Social Scoring – also die dauerhafte Bewertung von Personen – oder eine ständige Gesichtserkennung sind deshalb ausdrücklich verboten.

Scene 6:

Jetzt erfährst du, wer vom AI Act betroffen ist und warum das für die Hochschule relevant ist.

Scene 7:

Der AI Act funktioniert ein bisschen wie ein Sicherheitsgesetz für Software. Er gilt für alle, die KI-Systeme in der EU entwickeln, verkaufen oder nutzen. Das betrifft auch die Hochschule, wenn KI-Programme im Arbeitsalltag eingesetzt werden. Es ist also wichtig, dass sich nicht nur Technik-Teams, sondern alle Bereiche mit den Regeln vertraut machen.

Scene 8:

Jetzt zeigen wir dir, was der AI Act zur Schulung von Mitarbeitenden vorschreibt.

Scene 9:

Im AI Act steht: Wer KI-Systeme beschafft, nutzt oder bewertet, muss nachweisen, dass er oder sie sich mit den wichtigsten Grundlagen auskennt. Für die meisten Sachbearbeiter:innen reicht oft ein dreistündiger Grundkurs. Für die IT-Teams braucht es mehr Wissen. Wichtig ist: Diese Schulungen müssen dokumentiert sein, zum Beispiel im Fortbildungssystem der Personalabteilung.

Scene 10:

Wir schauen uns jetzt an, in welchen Bereichen der AI Act an der Hochschule relevant wird.

Scene 11:

Der AI Act spielt an vier Stellen auf dem Campus eine Rolle: Erstens in der Forschung – hier gibt es mehr Freiheiten, aber es muss dokumentiert werden. Zweitens in der Lehre, besonders wenn Studierende KI anwenden sollen. Drittens bei Prüfungen – wenn zum Beispiel Prüfungsaufsicht durch KI erfolgt, gelten besonders strenge Regeln. Und viertens in der Verwaltung, etwa beim Monitoring und bei der Aktenführung.

Scene 12:

Zum Abschluss fassen wir die wichtigsten Punkte zum AI Act noch einmal zusammen und geben einen Ausblick.

Scene 13:

Das Wichtigste in Kürze: Erstens: Der AI Act schützt Grundrechte und fördert Innovation. Zweitens: Er gilt für alle Nutzer:innen von KI, also auch für eure Hochschule. Drittens: Schulungen zu KI sind Pflicht und müssen dokumentiert werden. Viertens: Der Act betrifft euch in Forschung, Lehre, Prüfungen und Verwaltung. Im nächsten Video lernst du, wie die verschiedenen Risikostufen für KI-Systeme aussehen.

Transkript Video 8 „Risikostufen des AI Acts“

Scene 1:

Der AI Act ordnet jedes KI-System in eine von vier Risikostufen ein ähnlich wie bei einer Ampel. In den nächsten Minuten erfährst du wie du dein Projekt richtig einstufst und welche Regeln für die jeweilige Stufe gelten.

Scene 2:

Wir zeigen dir jetzt, was du zu den Risikostufen und ihren Pflichten im Hochschulalltag wissen solltest.

Scene 3:

Nach dieser Einheit kennst du die vier Risikostufen, weißt, welche Pflichten es jeweils gibt und kannst typische Beispiele aus dem Hochschulalltag richtig zuordnen.

Scene 4:

Jetzt erfährst du, warum der AI Act verschiedene Risikostufen für KI-Systeme einführt und was das für die Praxis bedeutet.

Scene 5:

Die meisten KI-Systeme – rund 90 % – gelten als geringes Risiko. Strenge Regeln greifen nur dort, wo Grundrechte in Gefahr sind, zum Beispiel bei Überwachung mit Gesichtserkennung. So sorgt der AI Act für mehr Sicherheit ohne die meisten Projekte unnötig zu bremsen.

Scene 6:

Wir starten mit der höchsten Risikostufe und schauen uns an, welche KI-Anwendungen in Europa komplett verboten sind.

Scene 7:

Bestimmte KI-Praktiken sind in Europa grundsätzlich verboten. Dazu zählen Social Scoring – also das dauerhafte Bewerten von Menschen –, manipulative Spielzeuge oder ständige Gesichtserkennung im öffentlichen Raum. Auch Hochschulen dürfen solche Systeme weder nutzen noch entwickeln.

Scene 8:

Als nächstes sehen wir uns die Hochrisiko-Stufe an und welche Anwendungen im Hochschulbereich dazugehören.

Scene 9:

Zur Hochrisiko-Stufe gehören Anwendungen wie Online-Prüfungsaufsicht (Remote-Proctoring) oder Bewertungssysteme im Personalbereich (HR-Scoring). Hier gelten sechs Pflichten: Erstens Ein durchdachtes Risikomanagement Zweitens Gute Trainingsdaten Drittens Technische Dokumentation Viertens Verständliche Infos für die Nutzenden Fünftens Überwachung durch Menschen Sechstens Genaue Protokolle und Vorfall-Logs.

Scene 10:

Jetzt erfährst du, was bei KI-Anwendungen mit begrenztem Risiko zu beachten ist.

Scene 11:

Begrenztes Risiko bedeutet: Es reicht meist, die Nutzenden zu informieren. Ein typisches Beispiel ist der FAQ-Chatbot: Er muss klar anzeigen, dass hier eine Maschine antwortet und nicht so tun, als wäre er ein Mensch.

Scene 12:

Zum Schluss schauen wir uns noch an, was für KI-Anwendungen mit minimalem Risiko gilt.

Scene 13:

Spam-Filter oder Programme, die Grammatik prüfen, zählen zum Minimalrisiko. Hier gibt es praktisch keine besonderen Vorgaben. Hauptsache, das System ist allgemein sicher entwickelt.

Scene 14:

Zum Abschluss fassen wir zusammen, wie du den KI-Einsatz an der Hochschule schnell und sicher einschätzen kannst.

Scene 15:

Stell dir drei Fragen: Erstens – Ist die Anwendung verboten? Zweitens – Gehört sie zur Hochrisikostufe? Drittens – Muss ich die Nutzenden darauf hinweisen? Wer diese Fragen beantwortet, weiß sofort, welche Pflichten beim KI-Einsatz an der Hochschule gelten.

Transkript Video 9 „Ethische Aspekte“

Scene 1:

Künstliche Intelligenz hilft heute an vielen Stellen in der Hochschule zum Beispiel bei Bewerbungen oder in der Personalverwaltung. Aber: Je öfter Computer Entscheidungen treffen desto wichtiger ist die Frage ob das fair und sicher abläuft. In diesem Video lernst du drei einfache Grundsätze kennen mit denen du jedes KI-Projekt ethisch einwandfrei gestalten kannst.

Scene 2:

In diesem Abschnitt zeigen wir dir, welche ethischen Prinzipien beim Einsatz von KI besonders wichtig sind.

Scene 3:

Nach dieser Einheit kannst du die drei Ethik-Prinzipien erklären: Datenschutz, Fairness und menschliche Aufsicht, und weißt, wie du sie in deinen Projekten anwendest.

Scene 4:

Jetzt erfährst du, warum ethische Prinzipien wie Datenschutz, Fairness und menschliche Aufsicht für KI-Anwendungen so wichtig sind.

Scene 5:

Studierende und Beschäftigte vertrauen KI-Anwendungen mehr, wenn klar ist: Deine Daten sind sicher, niemand wird benachteiligt, und am Ende kann immer noch ein Mensch eingreifen. So steigt die Akzeptanz von KI in der Verwaltung deutlich.

Scene 6:

Als erstes schauen wir uns an, wie du beim Einsatz von KI den Datenschutz und die Privatsphäre sicherstellst.

Scene 7:

Egal ob es um Studierende oder Personal geht: Sammle nur die Daten, die wirklich nötig sind, und schütze sie technisch – zum Beispiel durch sichere IT-Systeme. Prüfe außerdem immer, ob die Datennutzung rechtlich erlaubt ist. Weniger ist hier oft mehr. Das schafft Vertrauen.

Scene 8:

Als nächstes zeigen wir dir, wie du mit KI-Anwendungen faire und diskriminierungsfreie Entscheidungen sicherstellst.

Scene 9:

KI soll immer gerecht entscheiden, etwa bei Zulassungen oder Stipendien. Kontrolliere deine Daten regelmäßig auf Schieflagen, sorge für Vielfalt bei den Daten und dokumentiere jeden Test. So stellst du sicher dass niemand bevorzugt oder benachteiligt wird.

Scene 10:

Zum Schluss geht es darum, wie du Verantwortung übernimmst und die menschliche Aufsicht bei KI-Anwendungen sicherstellst.

Scene 11:

KI kann Vorschläge machen, aber die endgültige Entscheidung trifft immer ein Mensch. Kläre, wer im Team verantwortlich ist, ermögliche jederzeit ein Eingreifen (Override) und dokumentiere die Entscheidungswege. Das sorgt für Kontrolle und Nachvollziehbarkeit.

Scene 12:

Zum Abschluss geben wir dir einen kurzen Check mit auf den Weg, um die Ethik in deinen KI-Projekten sicherzustellen.

Scene 13:

Frag dich vor jedem neuen KI-Projekt: Erstens – Darf ich diese Daten überhaupt nutzen? Zweitens – Benachteiligt das Ergebnis jemanden? Drittens – Trifft immer noch ein Mensch die letzte Entscheidung? Wenn du alle drei Fragen mit Ja beantworten kannst, ist dein Projekt ethisch abgesichert.

Transkript Video 10 „Effektives Prompting“

Hallo und herzlich willkommen zum Promptlabor!

In diesem Video kümmern wir uns um grundsätzliche Empfehlungen für effektives Prompting. Dieses Video ist eines von mehreren, die uns ins Prompten einführen. In diesem – ersten – Teil geht es um wesentliche Empfehlungen für gute Prompts. Im zweiten Video setzen wir die Liste der Empfehlungen fort, aber legen den Fokus auf sinnvolle Workflows und Strategien, mit denen wir die eigenen Prompts verbessern können. Weitere Empfehlungen rund um das Prompting schauen wir uns dann in separaten Deep Dive Videos näher an.

In diesem Video jedenfalls lernen wir insgesamt acht einfache, aber wirksame Grundregeln kennen, wie wir effektive Prompts für textgenerative KI-Systeme schreiben. Das Ganze erarbeiten wir uns an einem Anwendungsbeispiel. Wir schreiben nämlich einen Übersetzungsprompt. Das Beispiel liegt mir deshalb sehr am Herzen, weil es ganz wunderbar meinen Weg von einfachstem Prompting hin zu gutem Prompt-Engineering nachzeichnet. Als ich meinen ersten Text mithilfe von ChatGPT vom Deutschen ins Englische übersetzen wollte, habe ich das schlicht und einfach mit diesem Prompt getan. Übersetze den folgenden Text von Deutsch nach Englisch und dann habe ich den zu übersetzenden Text einkopiert. Das kann man so machen, geht dann aber gerne auch mal schief. Wie schief das gehen kann, seht ihr hier an einem realen Beispiel. Der zu übersetzende Text ist eine Pressemitteilung der RWTH Aachen und ich habe in meinem Prompt nicht viel mehr getan, als zu sagen: “Liebe KI, bitte übersetze den nachfolgend gegebenen Text.” Und weil die Pressemitteilung feststehende Übersetzungen enthält, habe ich auch noch ein Glossar mitgegeben, das die KI bei ihrer Übersetzung berücksichtigen sollte. Der Text, der dabei entsteht, ist an sich richtig gut. Allerdings haben wir ein gewaltiges Problem. Bei der Übersetzung fehlt der komplette erste Absatz und auch zwischendurch fehlen einzelne Passagen, die im Originaltext enthalten waren. Dass das passiert ist, liegt an der Qualität oder vielmehr dem Mangel an Qualität in meinem Prompt.

Wie das besser geht, schauen wir uns jetzt an. Wir entwickeln also gemeinsam einen Übersetzungsprompt und das passiert hier direkt neben mir in diesem Fenster. Das kleine Problem daran ist, dass dieser Prompt am Ende ganz schön lang sein wird, sodass wir eine recht kleine Schriftgröße wählen müssen. Ich werde die jeweils aktuell besprochene Passage im Prompt fettdrucken, damit ihr sie besser lesen könnt. Aber ich weiß, dass die Schriftgröße trotzdem recht klein ist. Deshalb haben wir euch sowohl diesen Prompt als auch die Präsentationsfolien zu diesem Video verlinkt, sodass ihr gerne auch dort hineinschauen könnt.

Okay, unsere erste Empfehlung für gute Prompts ist, der KI eine Rolle zuzuweisen. Ich tue das, indem ich schreibe: “Du, liebe KI, bist eine höchst kompetente Übersetzerin, die an einer Universität angestellt ist. Dir ist eine bestmögliche Qualität der Übersetzung sehr wichtig.” Das klingt jetzt total over the top, “höchst kompetent”... “bestmöglich”. Das würde man so normalerweise nicht schreiben. Aber wir erinnern uns an das Video mit dem Titel “Verhalten beeinflussen”. Dort haben wir gelernt, dass KI in ihrer Grundeinstellung darauf ausgelegt ist, ein normales Gespräch mit einer durchschnittlichen Person zu führen. Wir wollen hier aber mehr. Nämlich, dass die KI eine bestmögliche Leistung im Rahmen einer ganz spezifischen Aufgabe erfüllt.

Und genau auf diese Rolle bereiten wir die KI hier vor. Empfehlung zwei ist, die KI mit Informationen zum Kontext auszustatten, in dem unsere Anfrage stattfindet. Das können begleitende Informationen zu unserer Anfrage sein, die hilfreich sind, um die KI optimal auf die Beantwortung unserer Anfrage

vorzubereiten. Für den Übersetzungsprompt geben wir konkrete Informationen zur weiteren Bearbeitung. Wir definieren verwendete Begriffe, wie etwa die “Quelle” oder die “Zielsprache”. Und wir erklären der KI, wie sie an Informationen über diese beiden Aspekte kommt. Zusätzlich informieren wir die KI über die Möglichkeit eines Glossars.

Das bringt uns direkt zur Empfehlung drei, nämlich der KI ganz konkret ihre Aufgabe zu nennen. Hier geht es um die möglichst präzise Arbeitsanweisung, was die KI tun soll. Für unseren Übersetzungsprompt schreiben wir deshalb: “Du übersetzt eine Quelle in die Zielsprache.” Klare Anweisung, und es treten die beiden zuvor im Rahmen des Kontexts definierten Begriffe wieder auf. Und nun bereiten wir direkt die nächste Empfehlung vor, indem wir schreiben: “Dazu arbeitest du die folgenden Arbeitsschritte nacheinander ab.”

Genau diese Arbeitsschritte nämlich sind Empfehlung vier. Es hat sich herausgestellt, dass die meisten KI-Systeme davon profitieren, wenn wir eine Aufgabe in kleinere Arbeitsschritte zerlegen und diese als Sequenz abarbeiten. Die Arbeitsschritte festzulegen kann auf verschiedene Weise passieren. Aber ich finde es schlicht am einfachsten, ganz konkret die Schritte zu benennen, die wir mit der KI durchführen möchten. Sequenzieren sollten wir dabei nicht nur die Abfolge der Schritte, sondern auch die Anweisungen innerhalb der einzelnen Schritte. Das realisieren wir, indem wir der KI ganz konkret sagen, was sie tun soll und was wir daraufhin tun werden. Genauso sind alle vier Schritte formuliert, die unseren Übersetzungsprompt ausmachen. Schritt eins ist: “Du fragst mich zuerst nach der Zielsprache, in die übersetzt werden soll.” Das ist die Aufgabe der KI. “Ich nenne dir die Zielsprache.” Das ist unsere Aufgabe. Schritt zwei ist dann nach exakt demselben Muster aufgebaut: “Du fragst mich nach einem Glossar”. Das ist die Aufgabe der KI. “Ich gebe dir entweder ein Glossar oder sage dir, dass es kein Glossar gibt.” Das ist unsere Aufgabe. Schritt drei: “Du fragst mich dann nach der Quelle, die übersetzt werden soll.” – Aufgabe der KI. “Ich gebe dir diese Quelle entweder als kopierten Text oder als angehängtes Dokument.” Das ist wieder unsere Aufgabe. Und schließlich Schritt vier: “Du übersetzt diesen Text in die Zielsprache.” Absolut eindeutiger Prozess, den wir kaum klarer formulieren könnten.

Empfehlung fünf ist, in unserem Prompt präzise Vorgaben zu machen und vor allem die Länge der Ausgabe festzulegen. Für unseren Übersetzungsprompt adressieren wir hier das gewaltige Manko, das mit unserem viel zu einfachen Prompt vom Anfang noch aufgetreten ist. Wir schreiben nämlich: “Lasse bei deiner Übersetzung keine Sätze weg und füge keine neuen Sätze hinzu.” Zudem sagen wir: “Lasse keine Inhalte weg und füge keinen neuen Inhalte hinzu.” Das scheint doppelt gemoppelt, aber verschiedene Studien zum Prompt Engineering legen nahe, dass solche Redundanzen zu einer besseren Ausgabequalität führen. Doppelt wirkt stärker! Schließlich sagen wir noch: “Erhalte den Sprachstil der Quelle so weit wie möglich.” Ok, das klingt erstmal gut, aber wir erinnern uns nochmal an die Empfehlung, bei der wir gerade sind. Die lautete nämlich: präzise zu sein. “So weit wie möglich” ist keine Präzision, sondern lässt einen sehr weiten Interpretationsspielraum. Diese Formulierung ist komplett überflüssig, sodass wir sie streichen können. Und das gilt bei allen solchen Formulierungen, die wir Menschen sehr gerne benutzen. Begriffe wie ungefähr, etwa, maximal oder eben “so weit wie möglich” sind bei KI-Tools nicht nur überflüssig, sondern häufig sogar kontraproduktiv. Lasst solche Formulierungen einfach weg. Warum solche Begriffe wirklich komplett überflüssig sind, können wir sogar eindrucksvoll demonstrieren. Also, dieser Prompt hier scheint zunächst wie ein vollkommen unauffälliger Prompt: “Erkläre den Begriff Energiesysteme in maximal fünfzig Worten.” Wir wissen aber seit gerade, dass genau ein Wort in diesem Prompt überflüssig ist, nämlich der Begriff “maximal”. Also streichen wir mal diesen Begriff und gehen sogar noch einen Schritt weiter, indem wir stattdessen ein “genau” hinzufügen. Wir weisen also die KI an, uns den Begriff Energiesysteme in genau fünfzig Worten zu erklären. Schauen wir mal, wie gut die KI diese Anweisung ausführt. Wir starten mit Google Gemini. Gemini liefert eine wunderbare Erklärung des Begriffes, allerdings nicht in fünfzig Worten, sondern vierzig. Wenn wir das Ganze nochmal probieren, ist die Erklärung immer noch gut, aber es sind nur sechsunddreßig Worte und beim dritten Versuch sind es nur noch fünfunddreißig Worte. Wir sind

also weit entfernt von den genau fünfzig Worten, die wir haben wollten. Man ahnt schon: Begriffe wie ungefähr, maximal oder höchstens sind für KI-Modelle nur eine Einladung zu noch mehr "Fuzziness". Andere Anbieter können das nicht besser. Claude generiert auf unsere erste Anfrage hin einen Text mit vierundvierzig Worten, beim zweiten Versuch sind es dann einundvierzig Worte und beim dritten dreiundvierzig. In keinem Fall also hat die KI unsere präzise Anfrage mit vergleichbarer Präzision erfüllt. Übrigens erinnert ihr euch an unsere kleine Demonstration zum Worte-zählen aus dem Gameshow-Video? Ganz übel wird es nämlich, wenn wir nun versuchen würden, mithilfe der KI zu überprüfen, ob die KI unsere Aufgabe korrekt erfüllt hat. Dazu bitten wir erstmal die KI um eine Erklärung in fünfzig Worten und fragen dann, wie viele Worte denn die generierte Erklärung tatsächlich enthält. Die KI antwortet pflichtschuldigst, dass sie unsere Anfrage natürlich korrekt beantwortet und fünfzig Worte generiert hat. Ich erspare euch das Zählen… natürlich sind es keine fünfzig, sondern zweiundvierzig Worte. Wie könnte es anders sein?

Empfehlung sechs lautet, das gewünschte Format der Ausgabe, also des KI-generierten Textes, möglichst konkret vorzugeben. Im Übersetzungsbeispiel schreibe ich: "Du übernimmst das Format der Quelle." Das wird also dazu führen, dass aus einem Fließtext wieder ein Fließtext wird und eine Stichwortliste auch nach der Übersetzung weiterhin eine Stichwortliste ist. Zur Sicherheit schreiben wir auch noch: "Insbesondere verwendest du Überschriften, Aufzählungen und Hervorhebungen, wenn die Quelle solche enthält." Lasst uns zur Erklärung ein zweites Beispiel für die Vorgabe eines konkreten Ausgabeformates anschauen. So können wir die KI auch anweisen: "Schreib deine Antwort als nummerierte Liste von Stichpunkten" und noch weiter konkretisieren: "Die Stichpunkte dürfen keine unvollständigen Sätze sein." Und jetzt werden wir noch richtig technisch, indem wir schreiben: "Formatiere deine Ausgabe mit Markdown." Solche Formatvorgaben werden zuverlässig dazu führen, dass der von KI generierte Text bestmöglich unseren Formatvorstellungen entspricht. Für viele von uns sticht hier wahrscheinlich das Wort "Markdown" hervor. Was is'n das?

Die Antwort ist: Unsere Überleitung zu Empfehlung sieben. Die lautet nämlich, dass wir es für KI-Tools leichter machen, die Gliederung unseres Textes zu verstehen, indem wir ganz explizit Marker zur Untergliederung unserer Anfrage setzen. Unser Übersetzungsprompt ist schon einigermaßen gut strukturiert, weil wir Absätze hinzugefügt haben. Um das Ganze noch klarer zu machen, könnten wir aber auch noch Überschriften für die Absätze hinzufügen – so wie jetzt. Das ist nett, aber nicht nur für uns Menschen, sondern auch für KI-Modelle einigermaßen unübersichtlich zu lesen. Kriegt man das besser hin? Ja! Hier können wir uns die Tatsache zunutze machen, dass KI-Tools während ihres Trainings eine Vielzahl von Texten verarbeitet haben, die sogenanntes Markdown enthielten. Markdown ist ein Weg, Texte mit Textauszeichnungen wie zum Beispiel Überschriften, Fettdruck oder Kursivdruck zu versehen. Und zwar auch dann, wenn wir nicht mit einem Textverarbeitungsprogramm arbeiten, das uns diese Auszeichnungen per Button-Klick anbietet. Es gibt eine ganze Reihe von Markdown-Befehlen, die wir hier aufgeführt sehen. Aber von diesen vielen Möglichkeiten brauchen wir für unsere Prompts eigentlich nur drei. Wir können Überschriftsebenen markieren, indem wir eines oder mehrere Hashtag-Zeichen an den Anfang einer Zeile setzen und danach ein Leerzeichen einfügen. Eine Zeile mit einem Hashtag ist eine Überschrift ersten Grades, zwei Hashtags zweiten Grades und so weiter. Wir können Texte kursiv auszeichnen, indem wir die kursiv zu schreibenden Worte mit einem Sternchen einfassen. Alternativ kann Kursivdruck auch über einen Unterstrich statt des Sternchens vor und nach dem Wort angezeigt werden. Fettdruck geht über zwei Sternchen und fett-und-kursiv über drei Sternchen. Und schließlich können wir auch Listen und Aufzählungen verwenden: sowohl nicht-nummerierte Listen, indem wir entweder einen Stern, ein Pluszeichen oder einen Bindestrich verwenden und nummerierte Aufzählungen ganz einfach über Erstens, Zweitens, Drittens und so weiter am Anfang der Zeile. Der Clou an Markdown ist, dass KI-Systeme diese Auszeichnungsform kennen und korrekt interpretieren können. Wenn wir also die Auszeichnungssprache Markdown verwenden, ist die KI in der Lage, Überschriften, Textauszeichnungen und Aufzählungen in unseren Prompts zu erkennen. Lasst uns deshalb unseren Übersetzungsprompt damit ergänzen. Wir verwenden Überschriften ersten und

zweiten Grades verwenden, indem wir ein oder zwei Hashtags verwenden. Wir können die Arbeitsschritte fett markieren, bestimmte Worte zur Hervorhebung kursiv setzen und auch eine nicht-nummerierte Liste einfügen. Für eine KI, die diesen Text verarbeitet, sieht der Text nun ungefähr so aus. Genauso wie uns Menschen solche Gliederungen und Auszeichnungen helfen, den Text sehr viel schneller und besser zu verstehen, unterstützen wir auch die KI darin, den Text zielgerichtet zu verarbeiten.

Damit sind wir schon fast am Ende der Empfehlungen, aber eine haben wir noch und die lautet, sauber zu formulieren. Damit meine ich, eine konsistente Wortwahl zu verwenden und vor allem Negationen zu vermeiden. Lasst uns zuerst mal auf die konsistente Wortwahl schauen. Wenn wir diese Empfehlung befolgen, dann lesen sich Texte ganz häufig anders, als wir sie als Menschen typischerweise schreiben würden. Denn die konsistente Wortwahl bedeutet, dass wir eben gerade nicht versuchen, einen Text dadurch abwechslungsreicher zu gestalten, dass wir für bestimmte Worte auch immer mal Synonyme verwenden. Ganz im Gegenteil, wir unterstützen die Verarbeitung eines Textes durch ein KI-Tool, wenn wir genau das nicht tun. Die Verwendung des Begriffes “Quelle” in unserem Prompt ist ein wunderbares Beispiel dafür. Wann immer wir auf die Quelle rekurrieren, benutzen wir auch genau dieses Wort. Und das führt dazu, dass selbst in diesem kurzen Prompt das Wort Quelle elfmal auftaucht. Für Menschen ist das seltsam zu lesen, die KI profitiert davon. Zudem verwenden wir Wortwiederholungen auch in einer zweiten Weise, nämlich zum Beispiel im Abschnitt “Weitere wichtige Anweisungen”. Dort verwenden wir zweimal das Wort “Sätze”, obwohl wir das in einer normalen Formulierung vermutlich so nicht gemacht hätten, sondern stattdessen vermutlich formuliert hätten: “Lasse bei deiner Übersetzung keine Sätze weg und füge keine neuen hinzu.” Dadurch, dass wir das Wort Sätze hier aber redundant verwenden, unterstützen wir die KI, genau zu verstehen, was wir meinen. Was uns in diesem Prompt nicht so gut gelungen ist, ist das Vermeiden von Negationen. Denn genau im gleichen Abschnitt treten solche auf. Allerdings ist diese Formulierung hier vermutlich kein großes Problem, weil wir der KI damit ja explizit Dinge verbieten. Ganz anders sieht es bei unserem Beispielprompt aus, den wir vorhin gezeigt haben. Dort haben wir formuliert: “Die Stichpunkte dürfen keine unvollständigen Sätze sein.” Für uns klingt das vollkommen klar. Allerdings erschweren wir dem KI-Tool das Verständnis dieses Satzes damit in erheblicher Weise. Viel besser wäre es gewesen, diesen Satz im Aktiv zu formulieren, indem wir schreiben: “Die Stichpunkte müssen vollständige Sätze sein.” Ist für uns viel einfacher zu verstehen und für die KI ebenfalls.

Damit sind wir auch schon am Ende unserer ersten einfachen Empfehlungen für effektives Prompting. Wir gehen es noch mal kurz durch.

Wir weisen der KI erstens eine Rolle zu. Wir statten sie zweitens mit Kontext- und Hintergrundinformationen zu unserem Prompt aus. Wir legen drittens eine ganz konkrete Aufgabe für die KI fest und teilen der KI viertens mit, in welche Arbeitsschritte diese Aufgabe sequenziert werden soll. Fünftens geben wir ihr präzise Vorgaben, legen die Länge des Outputs fest und vermeiden dabei unnötige Füllworte wie maximal, höchstens oder ungefähr. Sechstens definieren wir ein ganz konkretes Format für die Ausgabe der KI. Wir unterstützen siebtens die Gliederung unseres Textes mit Hilfe von Markdown, um die einzelnen Abschnitte in unserem Prompt auch für die KI möglichst schnell erkennbar zu machen. Schließlich formulieren wir, achtens, sauber, indem wir auf eine konsistente Wortwahl achten und Negationen vermeiden. Das war es für dieses Video. Weitere Empfehlungen findet ihr im Video zu Prompting-Workflows und Strategien, wo wir vor allem auch besprechen, wie wir unsere Arbeitsworkflows mit der KI möglichst effizient gestalten. Und das wars auch schon wieder für dieses Video. Ciao… und bis gleich!

Transkript Video 11 „Prompting Workflows und Strategien“

Hallo und herzlich willkommen zum Prompt-Labor.

In diesem Video schauen wir uns Workflows und Strategien an, die uns dabei helfen, unsere Konversationen mit KI-Tools noch effektiver zu gestalten. Wir werden verschiedene Workflows kennenlernen, um möglichst schnell zum bestmöglichen Prompt zu kommen. Dabei schauen wir auch noch einmal auf die Bereitstellung einer Faktenbasis und lernen den Wert von Musterbeispielen kennen. Wir überprüfen die Idee der Wiederholungen und werfen einen Blick auf exzentrische Prompts.

Das alles führt zu sechs weiteren Empfehlungen, die unsere acht Empfehlungen aus dem ersten Video zum effektiven Prompting ergänzen. Weitere Empfehlungen zum Prompting gibt es in Deep Dive-Videos. Das erste dieser Deep Dive-Videos folgt direkt im Anschluss und beschreibt vier komplexe Prompting Werkzeuge. Lasst uns loslegen!

Im Video zum effektiven Prompting hatten wir acht Empfehlungen kennengelernt. Deshalb setzen wir unsere Zählung hier fort mit Empfehlung neun – und das ist der flexible Wechsel des Prompting-Workflows. Grundsätzlich haben wir beim Prompting drei verschiedene solcher Workflows.

Den ersten bezeichnen wir als Edit-Workflow. Damit meinen wir die Überarbeitung desselben Prompts, bis er das gewünschte Ergebnis liefert. Wir führen hier also noch keine Konversation mit der KI, sondern editieren wieder und wieder denselben Prompt und optimieren so die Antwort der KI. Wie wir das konkret tun können, sehen wir gleich.

Der zweite Workflow ist der Chat-Workflow. Hier spielen wir Ping-Pong mit der KI und führen mit ihr ein Gespräch, um kollaborativ zum gewünschten Ergebnis zu kommen. Der Chat-Workflow ist die klassische Konversation mit einer KI, an deren Ende wir hoffentlich unser Ziel erreicht haben. Solche Konversationen können aber unter Umständen sehr lang und unübersichtlich werden.

Um das zu vermeiden, haben wir als dritten Workflow den Restart. Hier beginnen wir einen neuen Chat, aber mit einem Prompt, der die bisherigen Arbeitsergebnisse aus dem bestehenden Chat aufnimmt. Wir fassen also die Ergebnisse eines längeren Chats zusammen und starten mit diesen eine neue Konversation.

Diese drei Workflows klingen unmittelbar einleuchtend, allerdings kommt es ausgerechnet beim ersten davon – je nach KI-Tool – zu unterschiedlichen Verhaltensweisen. Zunächst mal können alle kommerziellen KI-Tools Prompts editieren. Das ist die gute Nachricht. Claude etwa zeigt zu jedem unserer Prompt direkt ein Edit-Symbol an. Bei ChatGPT und Gemini ist dieses Symbol erst sichtbar, wenn wir mit dem Mauszeiger über dem Prompt hovern. Nach einem Klick auf das Icon kommen wir in den Edit-Modus und können dann einen bestehenden Prompt editieren. Was aber passiert beim Bearbeiten? Hier verhalten sich die KI-Systeme durchaus unterschiedlich. Lasst uns das mal mit ChatGPT demonstrieren. Ich schreibe als Prompt: “Kann ich Prompts bearbeiten?”, schicke diesen Prompt ab und erhalte eine Antwort. Ich glaube aber, dass ChatGPT meine Anfrage nicht richtig verstanden hat, weil die Antwort so allgemein klingt. Also editiere ich den Prompt, indem ich nun schreibe: “Kann ich Prompt nachträglich bearbeiten?” Wir sehen, dass direkt unterhalb meines Prompts nun ein Zähler aufgetaucht ist und zusätzlich die Möglichkeit existiert, über die Pfeile zwischen den Prompt-Versionen hin und her zu schalten. Die Antwort von ChatGPT ist eine andere, aber ich habe weiterhin das Gefühl, dass die KI meine Anfrage nicht richtig verstanden hat. Also noch einmal bearbeiten, indem ich frage, ob ich abgeschickte Prompts bearbeiten kann. Nun antwortet ChatGPT:

“Nein, abgeschickte Prompts können nicht bearbeitet werden.” Wenn ich etwas ändern möchte, muss ich einen neuen Prompt abschicken. Ganz offensichtlich stimmt das nicht, denn ich kann Prompts nicht nur editieren, sondern auch zwischen den Bearbeitungs-Versionen hin und her springen. ChatGPT kennt also seine eigenen Fähigkeiten nicht. Diese Bearbeitungs- Versionen sind übrigens vollständig unabhängig voneinander. Dinge, die ich in einer Version geschrieben habe, sind in den anderen Versionen nicht existent und für die KI nicht zugreifbar. Ich kann jede der Bearbeitungs-Versionen wie gewohnt fortsetzen, indem ich zum Beispiel schreibe: “Das stimmt nicht, ich kann abgeschickte Prompts bearbeiten.” ChatGPT antwortet nun, dass ich natürlich einen neuen Prompt schreiben kann, aber basierend auf einem vorherigen und ihn dann erneut abschicken kann. Erneut beharrt die KI darauf, dass ein Editieren nicht möglich ist. Auch spätere solche späteren Prompts in einem bereits editierten Chat können wir bearbeiten. Zum Beispiel könnten wir ChatGPT ganz explizit darauf hinweisen: “Nein das stimmt nicht. Du hast eine Edit-Funktion.” Mit dieser Bearbeitung entsteht sozusagen eine Baumstruktur aller Bearbeitungs-Versionen und die Äste dieses Baumes sind, wie gerade geschrieben, vollständig unabhängig voneinander. Und auch zwischen den Bearbeitungs-Versionen späterer Prompts können wir wieder hin und her schalten. Richtig gut!

Bei Google Gemini funktioniert das anders. Wenn ich dort einen Prompt bearbeite, sehen wir erstmal, dass auch Google Gemini über die eigenen Fähigkeiten nicht wirklich gut informiert ist. Aber noch viel wichtiger ist, dass der Prompt, den ihr dort vor Euch seht, bereits die zweite Version des ursprünglichen Prompts ist. Zu dem kann ich aber nicht zurückkehren. Google Gemini speichert also frühere Versionen des Prompts nicht, sondern überschreibt diese bei jeder Bearbeitung. Es entsteht also keine Baumstruktur, sondern frühere Ergebnisse werden schlicht und einfach gelöscht. Claude hingegen verhält sich genauso wie ChatGPT. Auch hier entsteht die besagte Baumstruktur mit dem einzigen Unterschied, dass Claude uns nicht darüber informiert, welche Bearbeitungs-Versionen wir gerade anschauen. Etwas weniger nutzerfreundlich also, aber trotzdem fast genauso gut wie ChatGPT.

Genug mit den Workflows und weiter zu Empfehlung zehn. Die lautet, eine Faktenbasis bereitzustellen, und zwar auch dann, wenn wir kein Retrieval Augmented Generation nutzen können. Was RAG ist, haben wir schon im Video zur Faktentreue besprochen. Um auch ohne RAG eine Faktenbasis bereitzustellen, haben wir grundsätzlich drei verschiedene Wege.

Erstens können wir – sofern das gewählte KI-Tool dies erlaubt – eine Internetrecherche anfragen. Das tun wir entweder als Teil einer Custom Instruction oder eines System Prompts oder auch direkt aus unserem Prompt heraus.

Die zweite Möglichkeit ist, Dokumente hochzuladen, wenn dies technisch möglich ist. Das können PDF-Dateien sein, bei einigen KI-Systemen aber auch andere Dateitypen.

Die dritte Möglichkeit ist, Textpassagen, aus denen die KI Fakten extrahieren soll, direkt in den Prompt einzukopieren, sodass wir in unserem Prompt einen eigenen Abschnitt mit der Faktenbasis haben. Das ist aber die deutlich schlechteste Möglichkeit, weil viele KI-Tools die Länge eines Prompts stark beschneiden und unsere Prompts damit nicht beliebig lang werden können. Auch die Internetrecherche finde ich nicht wirklich zielführend, weil wir nie wissen können, ob das KI-Tool eine zuverlässige Quelle gefunden hat.

Deswegen ist meine klare Empfehlung, für die Bereitstellung einer Faktenbasis auf jeden Fall eigene Dokumente hochzuladen. Neben der Faktenbasis haben sich auch Musterbeispiele als unheimlich hilfreich für die Steigerung der Ergebnisqualität erwiesen – und genau das ist unsere Empfehlung elf. Musterbeispiele sind Beispiele für Formulierungen, die wir von der KI erwarten. In aller Regel würden wir im Rahmen auf die Existenz solcher Musterbeispiele im Rahmen des Kontexts verweisen – zum Beispiel indem wir schreiben: “Liebe KI, du verwendest für den erstellten Text einen typischen Sprachstil. Musterbeispiele für Texte in diesem Sprachstil findest du am Ende dieses Prompts unter der

Überschrift ‘Musterbeispiele für die Formulierung’.” Warum liefern wir diese Musterbeispiele nicht auch als PDF-Dokument? Nun, die Antwort darauf ist, dass noch nicht alle KI-Tools den Upload von mehreren PDFs erlauben, sodass wir die Faktenbasis als PDF-Dokument hochladen und von den Musterbeispielen als Teile des Prompts trennen müssen.

Empfehlung zwölf lautet, dasselbe einfach mal mehrfach zu machen und zum Glück sind wir bei KI-Tools hier nicht im Bereich der Definition von Wahnsinn, die gemeinhin Albert Einstein zugeschrieben wird… nämlich dasselbe wieder und wieder zu tun, aber unterschiedliche Ergebnisse zu erwarten. Wobei Albert Einstein hier vermutlich genau eines erwidern würde. “I never said that.” Das hat er nämlich nie gesagt. Jedenfalls lohnt sich das leicht wahnsinnige Mehrfachmachen bei KI-Tools absolut und immer.

Dabei können wir das Ganze in zwei Varianten tun. Variante eins ist, im gleichen KI-Tool denselben Prompt mehrfach zu wiederholen, dies aber jeweils in neuen Chats, damit die alten Ergebnisse die neue Textgenerierung nicht beeinflussen. Was wir damit überprüfen, ist die sogenannte Intra-Tool-Variabilität, also die Variation der Antwortgüte desselben KI-Tools auf denselben Prompt. Damit kriegen wir heraus, ob eine konkrete Antwort nur zufällig sehr gut oder sehr schlecht war oder ob unser Prompt reproduzierbar die beobachteten Ergebnisse liefert.

Variante zwei ist das Testen desselben Prompts in verschiedenen KI-Tools. Hiermit prüfen wir die Intratool-Variabilität, indem wir feststellen, wie verschiedene KI-Systeme mit derselben Anweisung umgehen. Wiederholungen also: super Sache!

Empfehlung dreizen wird jetzt etwas seltsam, denn es geht um exzentrisches Prompting. Ich habe hier mal ein wissenschaftliches Beispiel für genau solches dargestellt. Die Autoren dieser Studie haben verschiedene ungewöhnliche Promptformulierungen ausprobiert und geprüft, ob diese zu unerwartetem Verhalten der KI führen. Die Autoren beschreiben unter anderem, dass ein Prompting im Star-Trek-Stil, in dem die KI eine Star-Trek-basierte Custom Instruction erhält und zusätzlich angewiesen wird, jede ihrer eigenen Antworten mit einem Zitat aus Star Trek zu beginnen, in verschiedenen Fachdomänen bessere Resultate produziert. Unglaublich, aber wahr. So scheint Star Trek vor allem im Bereich der Naturwissenschaften und Mathematik bessere Leistung aus der KI herauszukitzeln und zwar reproduzierbar und über mehrere KI-Tools hinweg. Dass so etwas funktioniert, zeigt zwei Dinge: Erstens, dass KI-Tools für uns eine komplette Blackbox sind und zweitens, dass wir als User durchaus mal Dinge ausprobieren sollten, die uns eher abwegig vorkommen. Wir haben Beispiele dafür schon in mehreren Videos gesehen, zum Beispiel indem wir Formulierungen stark übertreiben und der KI sagen: “du bist äußerst sorgfältig” oder “du bist überdurchschnittlich intelligent” oder “du bist die kreativste Person im Raum”. Wir haben auch geklärt, was diese Formulierungen bewirken, nämlich dass wir die KI aus ihrer Grundkonfiguration, eine normale Konversation mit durchschnittlichen Personen zu führen, herausholen und sie auf höherwertige Aufgaben vorbereiten. Weiterhin hat sich gezeigt, dass KI-Antworten unter Umständen besser werden, wenn wir die KI anweisen, höflich zu sein. Schließlich können auch abwegigere Formulierungen wie zum Beispiel: “du hast noch eine Stunde bis zur Deadline” helfen, die KI stärker auf die Aufgabe zu fokussieren. Bei den exzentrischen Prompts gibt es keine Standardempfehlungen, aber ihr wisst jetzt, dass es durchaus lohnenswert sein kann, auch mal Dinge auszuprobieren, die im ersten Augenblick seltsam erscheinen.

Übrigens haben wir bei unserem Übersetzungsprompt aus dem Video zum effektiven Prompting genau diese Empfehlungen berücksichtigt, indem wir die Rolle der KI in gewisser Weise übertrieben definiert haben. Und das waren sechs weitere Empfehlungen für effektives Prompting, nämlich neuntens, nicht bei einem Prompting-Workflow zu bleiben, sondern zwischen Edit, Chat und Restart flexibel hin und her zu wechseln. Zehntens, eine Faktenbasis bereitzustellen und diese gegebenenfalls, elftens, durch Musterbeispiele zu ergänzen. Wir haben gesehen, dass es sich zwölftens lohnen kann, dasselbe mehrfach zu machen, und zwar entweder innerhalb derselben KI in einem neuen Chat oder in verschiedenen KI-

Systemen. Wir haben zudem, Nummer dreizehn, das exzentrische Prompten kennengelernt, mit dem wir gerne experimentieren können. Damit sind wir aber noch lagne nicht am Ende. Eine weitere Empfehlung haben wir nämlich noch, nämlich komplexe Prompting-Werkzeuge zu verwenden, wie zum Beispiel den Megaprompt. Aber darum kümmern wir uns in einem eigenen Video. Denn das war's auch schon wieder für dieses Video. Ciao und tschüss!