Löse dein Problem mit einem Augenzwinkern

Die KI lieferte Quellen. Eine davon existiert sogar.

Wie man Zitate und Quellen überprüft, die von KI-Tools generiert wurden



Max Paradox

Die KI lieferte Quellen. Eine davon existiert sogar.

Wie man Zitate und Quellen überprüft, die von KI-Tools generiert wurden

Max Paradox

Rozwiąż swój problem z przymrużeniem oka

ISBN: 9788384906279

ALL-IN Rafał Płonkowski



INHALTSVERZEICHNIS

INTRO 5

Kapitel 1 - Ein Literaturverzeichnis aus dem Reich der Fantasie: Warum KI Quellen erfindet 11

Kapitel 2 - Das Zitat steht in Anführungszeichen, die Wahrheit macht Urlaub 22

Kapitel 3 - Der Link funktioniert, aber wohin führt er? Anatomie eines falschen Beweises 33

Kapitel 4 - „Die KI hat das doch überprüft!“: Die teuersten Ausreden der Welt 44

Kapitel 5 - Suche nicht nach der Nadel in der Bibliothek: Wie du Quellen findest, ohne den Verstand zu verlieren 57

Kapitel 6 - Wenn die Antwort viel zu gut passt: Die Falle der eigenen Erwartungen 69

Kapitel 7 - Stille Post im Anzug: Wie Quellen zwischen Menschen und Dokumenten verloren gehen 80

Kapitel 8 - PDF ist geduldig: Wenn KI echte Dokumente falsch liest 92

Kapitel 9 - Fünf Minuten für die Wahrheit: So führst du eine erste Prüfung von KI-Antworten durch 105

Kapitel 10 - Vom Zitat zum Original: Eine Ermittlung mit einem echten Abschluss 117

Kapitel 11 - Zahlen lügen nicht, aber ihre Beschreibungen haben eine blühende Fantasie 129

Kapitel 12 - Erstelle ein Literaturverzeichnis, für das du dich nicht schämen musst 141

Kapitel 13 - Bring der KI bei, Beweise zu zeigen, statt nur einen guten Eindruck zu machen 153

Kapitel 14 - Von der KI-Antwort zu einer Entscheidung, die sich begründen lässt 165

Kapitel 15 - Zwei Quellen, drei Wahrheiten und ein Mensch kurz vor der Verzweiflung 178

Kapitel 16 - Wenn die Quelle plötzlich Befehle erteilt: Vertraue nicht jeder Anweisung in einem Dokument 190

Kapitel 17 - Wenn du nicht alles überprüfen kannst: Die Kunst, ehrlich mit Wissenslücken umzugehen 203

Kapitel 18 - Eine Überprüfung, die nicht mit dem Klick auf „Senden“ endet 215

Kapitel 19 - Dein eigener Radar für Unsinn: Wie du nicht in alte Gewohnheiten zurückfällst 228

Kapitel 20 - Die KI kann selbstsicher schreiben. Du kannst wissen, worauf ihre Aussagen beruhen 240

INTRO

Stell dir vor, du arbeitest an einem wichtigen Bericht. Die Abgabefrist rückt näher, und zwar mit der Geschwindigkeit eines Paketboten, der genau dann vor deiner Tür steht, wenn du gerade unter die Dusche gestiegen bist. Du hast Daten, du hast Argumente, dir fehlen nur noch ein paar solide Quellen. Also öffnest du ein KI-Tool und bittest um Veröffentlichungen, die eine bestimmte Behauptung belegen. Wenige Augenblicke später erhältst du fünf elegant formatierte Literaturangaben. Die Autoren haben Namen, die Artikel haben Titel, die Fachzeitschriften haben Ausgabennummern, und einige Veröffentlichungen besitzen sogar DOI-Kennungen. Alles sieht so professionell aus, dass du die Antwort am liebsten ausdrucken, in Leder binden und neben dem Lexikon ins Bücherregal stellen würdest.

Es gibt nur ein winziges Problem. Der erste Artikel existiert nicht. Der zweite existiert, behandelt aber ein ganz anderes Thema. Beim dritten ist der Autor echt, der Titel dagegen frei erfunden. Der vierte Verweis führt zu einer Webseite, die schon vor langer Zeit ihren Betrieb eingestellt hat. Die fünfte Veröffentlichung gibt es tatsächlich, allerdings vertreten ihre Autoren beinahe das Gegenteil dessen, was die KI behauptet.

Herzlichen Glückwunsch. Du hast gerade ein Literaturverzeichnis erhalten, in dem die Wirklichkeit lediglich einen Gastauftritt hat.

Das Beunruhigendste daran ist allerdings nicht, dass künstliche Intelligenz eine Quelle erfinden kann. Viel bemerkenswerter ist, wie überzeugend sie diese präsentiert. Es erscheint keine Warnung: „Achtung, die folgenden

Namen könnten das Ergebnis literarischer Fantasie eines Prozessors sein.“ Du hörst kein nervöses Räuspern. Du siehst auch kein digitales Erröten. Stattdessen bekommst du eine ruhige, ordentlich strukturierte Antwort, die aussieht, als käme sie von jemandem, der gerade aus der Bibliothek zurückgekehrt ist, nachdem er dort drei Tage lang wissenschaftliche Fachliteratur studiert hat.

Dabei kann ein Sprachmodell einen Text erzeugen, der wie ein korrektes Literaturverzeichnis aussieht, ohne über ausreichende Grundlagen zu verfügen, um die Richtigkeit jedes einzelnen Bestandteils zu bestätigen. Das ist ein entscheidender Unterschied. Eine glaubwürdig wirkende Quellenangabe formulieren zu können, bedeutet noch lange nicht, die entsprechende Veröffentlichung gefunden, gelesen und überprüft zu haben, ob sie die angeführte Behauptung tatsächlich stützt. Genauso wenig bedeutet die Fähigkeit, eine Schatzkarte zu zeichnen, dass unter der eingezeichneten Palme wirklich eine Truhe voller Gold liegt. Manchmal liegt dort lediglich die Rechnung für den letzten Urlaub.

Und dabei ist die Nutzung künstlicher Intelligenz tatsächlich ausgesprochen bequem. Sie hilft dabei, Ideen schnell zu ordnen, komplizierte Zusammenhänge zu verstehen, einen ersten Textentwurf vorzubereiten oder eine Richtung für die weitere Recherche zu finden. Das Problem beginnt, wenn wir Bequemlichkeit mit einem Beweis verwechseln. Eine Antwort klingt überzeugend, also hören wir auf zu fragen, woher die Information stammt. Wir bekommen einen Link und nehmen deshalb an, dass er zu einem vertrauenswürdigen Inhalt führt. Wir sehen Anführungszeichen und glauben, dass die zitierten Worte tatsächlich genau so gesagt oder geschrieben wurden.

Auf diese Weise kann man ein Zitat veröffentlichen, das niemals jemand geäußert hat, sich auf eine nicht existierende Studie berufen oder Kollegen eine falsche Information mit dem Zusatz „laut Experten“ weitergeben. Die betreffenden Experten erfahren später möglicherweise, dass sie angeblich an einer Untersuchung beteiligt waren, von der sie selbst noch nie gehört haben. Vermutlich sind sie ungefähr so überrascht wie ein Einhorn, das plötzlich im amtlichen Nutztierregister auftaucht.

Dieses Buch ist nicht entstanden, um dich davon zu überzeugen, dass künstliche Intelligenz nutzlos ist. Das wäre ungefähr so vernünftig, wie einen Taschenrechner wegzuwerfen, weil jemand die falschen Zahlen eingegeben hat. Es geht vielmehr darum, den Unterschied zwischen Unterstützung bei der Informationssuche und der Bestätigung von Informationen zu verstehen. Du kannst KI intensiv nutzen, Zeit sparen und trotzdem die Kontrolle darüber behalten, was du letztlich als Tatsache akzeptierst.

Zunächst lohnt es sich, eine unbequeme Wahrheit zu verstehen: Nicht alle fehlerhaften Quellenangaben sind vollständig erfunden. Wenn jeder falsche Verweis zu einem Artikel mit dem Titel „Eine außerordentlich wichtige Studie, die ich mir gerade ausgedacht habe“ führen würde, wäre das Leben erheblich einfacher. Viel häufiger begegnen dir jedoch subtile Fehler. Eine echte Veröffentlichung wird dem falschen Autor zugeschrieben. Ein tatsächlich existierender Bericht erhält ein falsches Erscheinungsjahr. Eine sinngemäße Wiedergabe wird als wörtliches Zitat ausgegeben. Ein Link führt zu einer authentischen Webseite, doch der dort veröffentlichte Inhalt bestätigt die Schlussfolgerung überhaupt nicht.

Deshalb reicht die einfache Frage „Existiert diese Quelle?“ nicht aus. Du musst außerdem feststellen können, ob das richtige Dokument angegeben wurde, ob dessen tatsächlicher Inhalt korrekt wiedergegeben wird, ob sich der Autor der Quelle überhaupt zu der betreffenden Frage äußert und ob die gezogene Schlussfolgerung nicht über die vorhandenen Belege hinausgeht. Das sind mehrere unterschiedliche Arbeitsschritte, auch wenn man sie in der Eile gern in eine einzige Schublade mit der Aufschrift „Überprüft“ steckt. Die Schublade sieht ordentlich aus. Bei ihrem Inhalt ist das keineswegs garantiert.

In der Praxis werden wir eine KI-Antwort deshalb als Ausgangspunkt für die Überprüfung betrachten und nicht als automatisch ausgestelltes Wahrheitszertifikat. Du wirst lernen, Warnsignale in Literaturverzeichnissen zu erkennen, Originaldokumente aufzuspüren, Autoren und Veröffentlichungsdaten zu kontrollieren, Publikationskennungen zu überprüfen und Zitate in ihrem ursprünglichen Zusammenhang zu finden. Außerdem wirst du verstehen, warum scheinbar kleine Unterschiede zwischen Formulierungen wie „Die Studie hat nachgewiesen“, „Die Autoren vermuten“ und „Unter bestimmten Bedingungen wurde beobachtet“ die Bedeutung einer Information vollständig verändern können.

Das bedeutet allerdings nicht, dass du ein professioneller bibliografischer Detektiv werden musst. Du brauchst weder ein mit Büchern vollgestelltes Arbeitszimmer noch eine Lupe oder dramatische Hintergrundmusik, während du den nächsten Browser-Tab öffnest. Was du brauchst, ist eine Methode, mit der du feststellen kannst, was zuerst überprüft werden sollte, wo du nach einer Bestätigung suchen musst und wann die vorhandenen Belege nicht ausreichen.

Andernfalls verbringst du möglicherweise genauso viel Zeit damit, den Titel eines unwichtigen Artikels zu kontrollieren, wie mit der Überprüfung einer Information, von der eine bedeutende Entscheidung abhängt.

Schließlich erfordert nicht jede Situation dieselbe Genauigkeit. Wenn du eine lockere Notiz für den eigenen Gebrauch vorbereitest, kannst du ein höheres Maß an Unsicherheit akzeptieren, solange du diese Unsicherheit nicht verschweigst. Wenn du dagegen eine wissenschaftliche Veröffentlichung, Unterrichtsmaterial, ein berufliches Dokument oder einen Text über Gesundheit, Recht oder Finanzen erstellst, können die Folgen eines Fehlers erheblich schwerwiegender sein. In solchen Fällen ist die Überprüfung von Quellen keine übertriebene Vorsichtsmaßnahme. Sie gehört zur Verantwortung für die Informationen, die du weitergibst.

Ebenso wichtig wird die Fähigkeit sein, zuzugeben, dass sich etwas nicht bestätigen ließ. Das klingt banal, doch moderne Werkzeuge haben uns an sofortige Antworten gewöhnt. Eine leere Stelle im Literaturverzeichnis erscheint beinahe wie eine persönliche Niederlage. Dabei ist die ehrliche Feststellung „Ich konnte keine zuverlässige Bestätigung für dieses Zitat finden“ wesentlich wertvoller als eine perfekt formatierte Quellenangabe, die ins Leere führt. Das Fehlen eines Beweises beweist nicht zwangsläufig, dass eine Behauptung falsch ist. Es berechtigt aber ganz sicher nicht dazu, sie als überprüfte Tatsache darzustellen.

Mit der Zeit wirst du feststellen, dass die wichtigste Fähigkeit nicht darin besteht, die KI bei Fehlern zu ertappen. Viel entscheidender ist es, die richtigen Fragen zu stellen und eine Antwort von einem Beleg unterscheiden zu können. Ein Sprachmodell kann dir helfen, einer Spur nachzugehen,

mögliche Veröffentlichungen vorzuschlagen, geeignete Suchbegriffe zu formulieren und Rechercheergebnisse zu ordnen. Eine endgültige Bestätigung erfordert jedoch die Prüfung der entsprechenden Materialien, möglichst direkt beim ursprünglichen Herausgeber oder an einer anderen vertrauenswürdigen Stelle, die das betreffende Dokument zugänglich macht.

Falls du also jemals eine Quellenangabe kopiert hast, ohne die Quelle zu öffnen, einem Zitat nur deshalb geglaubt hast, weil es klug klang, oder die bloße Existenz eines Links bereits als ausreichenden Beweis betrachtet hast, bist du hier genau richtig. Wir werden deinem Browser keinen öffentlichen Prozess machen. Stattdessen werden wir etwas wesentlich Nützlicheres lernen: die Möglichkeiten künstlicher Intelligenz zu nutzen, ohne ihr die Entscheidung darüber zu überlassen, was wahr ist.

Denn künstliche Intelligenz kann innerhalb weniger Sekunden eine Liste mit zehn Quellen zusammenstellen. Deine Aufgabe besteht nicht darin, die Länge dieser Liste zu bewundern. Deine Aufgabe ist es, herauszufinden, wie viele dieser zehn Einträge tatsächlich einen Platz in deinem eigenen Text verdienen.

Und sollte sich herausstellen, dass nur eine einzige Quelle existiert, weißt du wenigstens, welche es ist.

Kapitel 1 - Ein Literaturverzeichnis aus dem Reich der Fantasie: Warum KI Quellen erfindet

Montag, 9:17 Uhr. In einem Konferenzraum läuft eine Besprechung, bei der du die Ergebnisse einer Analyse präsentierst, die du am Vorabend vorbereitet hast. Alles läuft hervorragend. Die Diagramme sind verständlich, die Schlussfolgerungen vernünftig, und deine Präsentation enthält sogar eine Folie mit Quellenangaben, was in der Welt der Unternehmenspräsentationen beinahe als Zeichen aristokratischer Erziehung gilt. Irgendwann fragt jemand, woher die Information über die Wirksamkeit der besprochenen Lösung stammt. Ohne zu zögern verweist du auf eine wissenschaftliche Veröffentlichung, die dir am Vortag eine künstliche Intelligenz vorgeschlagen hat.

Dein Gesprächspartner gibt den Titel des Artikels in eine Suchmaschine ein. Nichts. Er überprüft den Namen des Autors. Er findet tatsächlich einen Wissenschaftler, allerdings beschäftigt sich dieser mit einem völlig anderen Fachgebiet. Anschließend versucht er, die Fachzeitschrift aufzuspüren. Die existiert immerhin, doch die angegebene Ausgabe enthält keine Veröffentlichung, die auch nur entfernt zu deiner Quellenangabe passt. Im Raum wird es still. Selbst der Projektor, der bisher Geräusche von sich gegeben hat, die an den Start eines kleinen Flugzeugs erinnern, scheint plötzlich etwas diskreter zu arbeiten.

Du hast gerade entdeckt, dass dein Literaturverzeichnis mehr Fantasie besitzt als die gesamte Kreativabteilung.

Die Situation ist besonders unangenehm, weil der Fehler überhaupt nicht wie ein Fehler aussah. Es gab keine Tippfehler, keine zufällig zusammengewürfelten Wörter und keine verdächtigen Formulierungen. Die Quelle hatte einen Autor, ein Datum, einen Titel, den Namen einer Fachzeitschrift und eine Bandnummer. Sämtliche Bestandteile standen an der richtigen Stelle. Es fehlte lediglich eine Kleinigkeit: die tatsächliche Veröffentlichung, die man hätte nachlesen können.

Es lohnt sich zu verstehen, warum solche Situationen überhaupt entstehen. Ohne dieses Verständnis ist es leicht, eine von zwei extremen Positionen einzunehmen. Die erste besagt, dass künstliche Intelligenz ein genialer Bibliothekar ist, dem man einfach vertrauen sollte. Die zweite stellt sie als digitalen Betrüger dar, der bewusst versucht, uns bloßzustellen. Beide Vorstellungen sind eindrucksvoll, helfen aber nicht dabei, das Werkzeug sinnvoll einzusetzen.

Ein Sprachmodell muss nicht die Absicht haben, dich zu täuschen, um eine falsche Information zu erzeugen. Es kann einfach eine Antwort formulieren, die zu den während des Trainings erlernten Mustern passt, obwohl sie nicht ausreichend mit einem tatsächlich existierenden Dokument verknüpft ist. Genau darin liegt der grundlegende Unterschied zwischen dem korrekten Erscheinungsbild einer Antwort und der Wahrheit ihres Inhalts.

Wenn eine elegante Quellenangabe einen Beweis vortäuscht

Stell dir eine Aufgabe vor, bei der du folgenden Satz vervollständigen sollst: „Laut einer Studie, die in der Fachzeitschrift ... veröffentlicht wurde“. Ein Mensch, der die Sprache und die grundlegenden Regeln wissenschaftlichen Schreibens kennt, kann problemlos eine natürlich klingende

Fortsetzung vorschlagen. Das bedeutet jedoch nicht, dass er gerade das Archiv des entsprechenden Verlags überprüft hat. Auf ähnliche Weise kann ein Sprachmodell bibliografisch überzeugende Angaben zusammenstellen, indem es erkannte Zusammenhänge zwischen Namen, Titeln, Fachgebieten und den üblichen Formen wissenschaftlicher Quellenangaben verwendet.

Je nach Aufbau und verfügbaren Funktionen kann ein solches Werkzeug auch das Internet durchsuchen, Dokumente lesen oder externe Datenbanken nutzen. Diese Möglichkeiten verbessern die Voraussetzungen für eine Überprüfung, garantieren aber keineswegs, dass die endgültige Antwort korrekt ist. Das Modell kann gefundenes Material falsch interpretieren, Informationen aus zwei verschiedenen Veröffentlichungen miteinander verbinden oder einer echten Quelle eine Schlussfolgerung zuschreiben, die deren Autoren überhaupt nicht formuliert haben.

Das Problem besitzt also mehrere Ebenen. Manchmal existiert die Quelle nicht. Manchmal existiert sie, wurde jedoch falsch beschrieben. In anderen Fällen sind sämtliche bibliografischen Angaben korrekt, während der Fehler in der Interpretation des Textes liegt. Die letzte Variante ist besonders tückisch, weil eine einfache Suche nach der Veröffentlichung den Fehler nicht aufdeckt.

Nehmen wir ein frei erfundenes Beispiel. Du bittest eine KI um wissenschaftliche Untersuchungen zum Einfluss kurzer Pausen auf die Konzentration während der Büroarbeit. Du erhältst den Titel „The Impact of Microbreaks on Cognitive Performance in Office Environments“, dazu Autorennamen, ein Erscheinungsjahr und den Namen einer psychologischen Fachzeitschrift. Alles klingt ausgesprochen plausibel. Das

Thema passt zum Fachgebiet, der Titel besitzt wissenschaftliche Würde, und die Namen wirken, als würden ihre Träger regelmäßig an Fachkonferenzen teilnehmen.

Das bedeutet allerdings nicht, dass der angegebene Artikel existiert. Der Titel in diesem Beispiel dient ausschließlich der Veranschaulichung und ist keine bestätigte bibliografische Quelle. Genauso solltest du jeden ungeprüften Quellenvorschlag behandeln: als Information, die überprüft werden muss, selbst wenn sie ausgesprochen professionell aussieht.

Interessanterweise kann eine Antwort umso glaubwürdiger wirken, je detaillierter sie ist. Seitenzahlen, Veröffentlichungsdaten und DOI-Kennungen erinnern an amtliche Stempel. Man betrachtet sie und geht unbewusst davon aus, dass irgendjemand die entsprechenden Angaben bereits kontrolliert hat. Dabei kann Detailgenauigkeit auch einfach nur Detailgenauigkeit sein. Ein äußerst präziser Fehler bleibt ein Fehler, er ist lediglich besser angezogen.

Man braucht übrigens nicht einmal eine erfundene Fachzeitschrift, um in Schwierigkeiten zu geraten. Es genügt der echte Name eines angesehenen Verlags, kombiniert mit einem Artikel, den es nicht gibt. Das authentische Element verleiht der gesamten Beschreibung den Anschein von Seriosität. Das ist ungefähr so, als würde dir jemand eine erfundene Wohnungsadresse in einem tatsächlich existierenden Hochhaus geben. Das Gebäude steht, die Straße stimmt, aber hinter der angegebenen Wohnungstür befindet sich eine Wand.

Warum wir uns so leicht überzeugen lassen

Unsere Schwierigkeiten entstehen nicht ausschließlich durch die Technologie. Die andere Hälfte des Problems liegt

in der Art, wie wir Informationen beurteilen. Wenn ein Text flüssig, ordentlich strukturiert und in einem selbstsicheren Ton geschrieben ist, erscheint er uns häufig kompetent. Das gilt besonders dann, wenn er ein Thema behandelt, mit dem wir selbst nicht ausreichend vertraut sind.

Hinzu kommt, dass wir eine KI normalerweise gerade dann um Unterstützung bitten, wenn wir Zeit sparen wollen. Wir brauchen schnell verwertbares Material, und das erhaltene Literaturverzeichnis entspricht genau unseren Erwartungen. Eine Überprüfung bedeutet zusätzliche Arbeit. Veröffentlichungen öffnen, Autoren kontrollieren und Zitate vergleichen klingt wesentlich weniger verlockend, als einfach auf „Kopieren“ zu klicken.

Dazu kommt die subtile Versuchung, die eigene Meinung bestätigt zu sehen. Wenn das Werkzeug fünf Quellen präsentiert, die unsere These unterstützen, empfinden wir Zufriedenheit. Würde es dagegen fünf Veröffentlichungen nennen, die dieselbe These infrage stellen, würden wir wahrscheinlich zusätzliche Fragen stellen. Gerade deshalb sollten wir Quellen, die exakt das bestätigen, was wir hören wollten, mit besonderer Aufmerksamkeit betrachten.

Angenommen, du bereitest eine Argumentation für die Einführung einer bestimmten Methode der Arbeitsorganisation vor. Du fragst: „Welche Studien beweisen, dass diese Methode die Produktivität steigert?“ Bereits diese Frage setzt voraus, dass entsprechende Beweise existieren. Das Modell kann auf eine Weise antworten, die der so formulierten Erwartung entspricht, anstatt deren Grundlage zu hinterfragen.

Vernünftiger wäre die Frage: „Welche zuverlässigen Veröffentlichungen untersuchen die Wirksamkeit dieser Methode? Nenne sowohl positive als auch negative

Ergebnisse und kennzeichne deutlich alle Informationen, die du nicht bestätigen kannst.“ Eine solche Formulierung garantiert keine fehlerfreie Antwort, verändert aber die Richtung der Recherche. Du bestellst nicht länger Beweise für eine Schlussfolgerung, die bereits feststeht.

Das ist auch dann wichtig, wenn du ein Werkzeug verwendest, das tatsächlich Zugriff auf eine Suchmaschine hat. Das Internet enthält unzählige Materialien von sehr unterschiedlicher Qualität. Allein die Tatsache, dass eine Webseite gefunden wurde, bedeutet nicht, dass die dort enthaltene Behauptung wahr ist. Informationen zu finden und ihre Glaubwürdigkeit zu beurteilen sind zwei unterschiedliche Tätigkeiten.

Vier Erscheinungsformen falscher Quellen

In der Praxis ist es hilfreich, vier grundlegende Situationen unterscheiden zu können, denn jede verlangt eine andere Reaktion.

Die erste ist die vollständig erfundene bibliografische Angabe. Die Veröffentlichung, ihre Autoren oder die Fachzeitschrift entsprechen keinem tatsächlichen Dokument. Du kannst eine halbe Stunde damit verbringen, nach einem Artikel zu suchen, der ausschließlich in einer generierten Antwort existiert. Wenn du nach einer angemessenen Überprüfung der Kataloge und Datenbanken keine Bestätigung findest, solltest du ihn nicht als Tatsache in dein Literaturverzeichnis aufnehmen. Du musst nicht beweisen, dass dieses Dokument in der gesamten Menschheitsgeschichte niemals existiert hat. Es reicht aus, ehrlich festzustellen, dass es sich nicht verifizieren ließ.

Die zweite Situation besteht in der Vermischung von Angaben. Der Artikel existiert, aber das Erscheinungsjahr,

der Titel oder die Autoren wurden falsch angegeben. Mitunter setzt ein Modell einen einzigen bibliografischen Eintrag aus Bestandteilen mehrerer echter Veröffentlichungen zusammen. Um diesen Fehler zu erkennen, muss die vollständige Quellenangabe mit dem ursprünglichen bibliografischen Datensatz verglichen werden. Einzelne Angaben sollte man nicht nach Gefühl korrigieren, weil man dadurch versehentlich eine weitere Mischform erschaffen könnte.

Die dritte Erscheinungsform betrifft verfälschte Zitate. Ein Autor hat tatsächlich einen bestimmten Satz geschrieben, aber die KI verändert einige Wörter, lässt eine wichtige Einschränkung weg oder verbindet zwei Textstellen zu einer einzigen Aussage. Die Bedeutung kann ähnlich bleiben, doch ein Text in Anführungszeichen sollte dem Original entsprechen. Wenn du die Bedeutung einer Aussage mit eigenen Worten wiedergeben möchtest, verwende eine Paraphrase und gib die entsprechende Quelle an. Verkleide eine Paraphrase nicht als wörtliches Zitat.

Die vierte Situation ist die hinterhältigste: Die Quelle existiert, das Zitat ist korrekt, aber es bestätigt die aufgestellte Behauptung nicht. Eine Veröffentlichung beschreibt einen Zusammenhang zwischen zwei Variablen, während die KI daraus einen Beweis für einen ursächlichen Zusammenhang macht. Eine Untersuchung beschäftigt sich mit einer kleinen, besonderen Personengruppe, doch die Antwort formuliert eine Schlussfolgerung über sämtliche Menschen. Die Autoren schreiben von einer Möglichkeit, während die generierte Zusammenfassung Gewissheit verkündet.

Du kannst also einen echten Artikel und einen korrekten Link besitzen und trotzdem ein falsches Argument

vorbringen. Das Literaturverzeichnis wird freigesprochen, doch die Schlussfolgerung hat sich inzwischen vom Tatort entfernt.

Der erste Filter, bevor du mit der Suche beginnst

Du musst nicht sofort jede Veröffentlichung einer vollständigen Analyse unterziehen. Im Arbeitsalltag hilft ein erster Filter, mit dem du entscheiden kannst, welche Einträge genauer geprüft werden müssen. Beginne damit, die Antwort der KI unter der Annahme zu lesen, dass es sich um einen Vorschlag handelt und nicht um eine bereits abgeschlossene wissenschaftliche Recherche.

Achte auf Quellen, die ohne vollständige Angaben präsentiert werden, verdächtig perfekt auf deine Frage zugeschnitten sind oder ausgesprochen entschiedene Behauptungen enthalten, ohne konkrete Textstellen aus dem Dokument zu nennen. Solche Merkmale beweisen nicht, dass die Angaben falsch sind, geben aber Anlass zur Kontrolle. Auch ein fehlender Autorenname oder ein fehlendes Datum bedeutet nicht zwangsläufig mangelnde Glaubwürdigkeit, denn nicht alle wertvollen Materialien besitzen die Struktur einer wissenschaftlichen Veröffentlichung.

Versuche bei jeder wichtigen Quelle drei Fragen zu beantworten. Kann ich das Dokument unabhängig von der KI-Antwort finden? Stimmen die bibliografischen Angaben mit dem tatsächlichen Dokument überein? Weiß ich, welche konkrete Textstelle meine Behauptung bestätigt? Wenn du bereits bei der ersten Frage nicht weiterkommst, solltest du nicht automatisch zur dritten übergehen und so tun, als wäre alles in Ordnung.

Nehmen wir an, du hast fünf Quellen für eine Präsentation erhalten. Zwei davon findest du schnell in vertrauenswürdigen Katalogen. Bei einer weiteren ist das Erscheinungsjahr falsch, aber das Originaldokument lässt sich eindeutig identifizieren. Die beiden verbleibenden Veröffentlichungen kannst du nicht finden. Das bedeutet nicht, dass du deine gesamte Arbeit wegwerfen musst. Du kannst die bestätigten Quellen behalten, die Angaben der dritten korrigieren und die beiden nicht verifizierten Einträge entfernen. Entscheidend ist, dass du nicht alle fünf für gleichermaßen glaubwürdig hältst.

Wenn eine Quelle für eine wichtige Entscheidung ausschlaggebend ist, bildet die Bestätigung ihrer Existenz lediglich den Anfang. Anschließend müssen auch der Inhalt und die Grenzen der Veröffentlichung beurteilt werden. Handelt es sich dagegen um ergänzendes Material, kannst du auf einen unsicheren Verweis einfach verzichten, anstatt viele Stunden mit der Suche zu verbringen.

Es lohnt sich außerdem, die eigene Überprüfung kurz zu dokumentieren. Dafür brauchst du kein kompliziertes System. Eine knappe Notiz wie „beim Verlag bestätigt“, „Volltext nicht gefunden“ oder „Quelle stützt die Schlussfolgerung nicht“ reicht häufig aus, damit du auch eine Woche später noch weißt, warum du eine bestimmte Veröffentlichung verwendet oder verworfen hast.

KI als Assistent, nicht als Kronzeuge

Die nützlichste Veränderung besteht darin, zwei Rollen voneinander zu trennen. Künstliche Intelligenz kann hervorragend dabei helfen, Fragen vorzubereiten, Begriffe zu ordnen und Suchstrategien zu entwickeln. Sie sollte jedoch nicht automatisch zum endgültigen Zeugen für die

Richtigkeit jener Informationen werden, die sie selbst erzeugt hat.

Wenn du fragst: „Sind die Quellen, die du genannt hast, echt?“, erhältst du möglicherweise eine weitere überzeugende Antwort, die ebenfalls überprüft werden muss. Das ist keine unabhängige Kontrolle. Das Werkzeug kann einen Fehler korrigieren, aber auch eine frühere Ungenauigkeit wiederholen. Die bloße Versicherung „Ich habe es überprüft“ ersetzt nicht die Möglichkeit, nachzuvollziehen, was tatsächlich kontrolliert wurde und wo die Überprüfung stattgefunden hat.

Besser ist es, die KI zur Erstellung einer Liste der Punkte zu verwenden, die bestätigt werden müssen, und diese anschließend mit den Originaldokumenten zu vergleichen. Du kannst das Modell auch bitten, die Grenzen seiner eigenen Antwort aufzuzeigen. So erhältst du Unterstützung bei der Organisation deiner Arbeit, ohne die gesamte Verantwortung auf das System zu übertragen.

Du musst nicht jeden generierten Satz behandeln wie einen Verdächtigen in einem Kriminalfall. Es genügt, die Sorgfalt an die möglichen Folgen eines Fehlers anzupassen. Eine Information, die lediglich als unverbindliche Inspiration dient, hat ein anderes Gewicht als ein Zitat in einer Abschlussarbeit oder die Grundlage einer geschäftlichen Entscheidung. Gerade in den wichtigeren Fällen brauchst du mehr als eine elegante Quellenangabe.

Wenn du also das nächste Mal ein Literaturverzeichnis erhältst, das wie das Ergebnis einer monatelangen Arbeit eines ganzen Forschungsteams aussieht, solltest du es nicht automatisch verwerfen. Du solltest es aber ebenso wenig ungeprüft übernehmen. Betrachte es als eine Liste möglicher Adressen, bei denen es sich lohnt, einmal nachzusehen.

Denn selbst das professionellste Literaturverzeichnis der Welt hilft dir nicht weiter, wenn die Hälfte seiner Autoren ausschließlich in der Fantasie des Modells wohnt.

Kapitel 2 - Das Zitat steht in Anführungszeichen, die Wahrheit macht Urlaub

Anführungszeichen gehören zu den kleinsten typografischen Zeichen, denen Menschen erstaunlich viel Macht zugestehen. Es genügt, ein paar Wörter zwischen zwei solcher Zeichenpaare zu setzen und den Namen einer bekannten Persönlichkeit hinzuzufügen, und plötzlich sieht ein gewöhnlicher Satz aus wie eine historische Aussage. Wenn daneben noch ein Datum, ein Buchtitel oder der Name einer Konferenz auftaucht, steigt die vermeintliche Glaubwürdigkeit weiter. Man hört auf zu fragen, ob jemand diese Worte tatsächlich ausgesprochen hat, und beginnt stattdessen darüber nachzudenken, wie man sie am besten in der eigenen Präsentation einsetzen könnte.

Eines Tages bereitest du einen Artikel über Entscheidungsfindung vor. Du bittest eine KI um ein kurzes Zitat eines bekannten Denkers, das die Bedeutung kritischen Denkens hervorhebt. Du erhältst einen Satz, der so treffend klingt, dass du ihn am liebsten auf ein Poster drucken und über deinen Schreibtisch hängen würdest. Die Aussage hat den passenden würdevollen Rhythmus, der Autor ist bekannt, und die Botschaft fügt sich perfekt in deinen Text ein.

Als du jedoch versuchst, das Zitat in der ursprünglichen Veröffentlichung zu finden, taucht ein Problem auf. Die Suchmaschine zeigt Hunderte von Webseiten, auf denen dieselbe Aussage wiederholt wird, aber keine davon nennt eine konkrete Stelle, an der sie erstmals erschienen ist. Ein

Portal verweist auf ein Buch, ein anderes auf eine Rede, ein drittes nennt ein Datum, das sich mit keinem bestätigten Ereignis in Verbindung bringen lässt. Die vierte Seite zitiert die erste, die fünfte die vierte, und die sechste versichert, der Satz sei „allgemein bekannt“.

Soeben ist ein digitales Orchester entstanden, in dem sämtliche Musiker vom selben fehlerhaften Notenblatt spielen.

Das Problem falscher Zitate ist älter als künstliche Intelligenz. Schon seit Langem schreiben Menschen bekannten Persönlichkeiten Aussagen zu, die ausreichend klug, witzig oder inspirierend klingen. Das Internet hat ihre Verbreitung erleichtert, und KI-Werkzeuge können die Entstehung neuer Varianten zusätzlich beschleunigen. Allerdings ist nicht jedes falsch zugeschriebene Zitat durch ein Sprachmodell entstanden. Manche Fehler waren bereits in den verfügbaren Materialien enthalten, bevor solche Modelle darauf trainiert wurden.

Gerade deshalb erfordert die Überprüfung von Zitaten mehr Genauigkeit als die bloße Frage nach dem Autor. Man muss feststellen, welche Worte tatsächlich gesagt wurden, in welchem Zusammenhang sie standen, wo sie festgehalten wurden und ob die angegebene Person wirklich ihr Urheber ist. Erst dann erfüllen Anführungszeichen ihre eigentliche Aufgabe.

Wenn eine Paraphrase einen fremden Anzug anzieht

Das häufigste Problem besteht nicht unbedingt darin, dass eine Aussage vollständig erfunden wurde. Manchmal erkennt eine KI den allgemeinen Sinn eines Gedankens, formuliert ihn mit eigenen Worten und präsentiert das Ergebnis anschließend als wörtliches Zitat. So entsteht ein

Satz, der einer bestimmten Ansicht durchaus entsprechen kann, aber nicht dem tatsächlichen Wortlaut des Originals.

Nehmen wir an, der Autor einer Veröffentlichung schreibt: „In der untersuchten Teilnehmergruppe waren regelmäßige Pausen mit einer geringeren selbstberichteten Müdigkeit verbunden." Die KI könnte diesen Abschnitt in die wesentlich eindrucksvollere Aussage verwandeln: „Regelmäßige Pausen verringern die Müdigkeit von Arbeitnehmern." Auf den ersten Blick behandeln beide Sätze ein ähnliches Thema. Der Unterschied ist jedoch enorm.

Der erste Satz beschreibt einen beobachteten Zusammenhang innerhalb einer bestimmten Gruppe. Der zweite suggeriert einen allgemeinen ursächlichen Zusammenhang. Die Einschränkung hinsichtlich der untersuchten Personen wurde entfernt, die Information darüber, dass die Müdigkeit auf Selbstauskünften beruhte, ist verschwunden, und eine vorsichtige Formulierung wurde durch eine entschiedene Behauptung ersetzt.

Wenn du den zweiten Satz in Anführungszeichen setzt und mit dem Namen des Autors versiehst, schreibst du ihm Worte zu, die er in diesem Beispiel nie verwendet hat. Darüber hinaus veränderst du die Bedeutung seiner Aussage. Das ist keine harmlose sprachliche Verschönerung mehr.

Denk daran, dass wörtliche Zitate und Paraphrasen unterschiedliche Aufgaben erfüllen. Ein Zitat dient dazu, bestimmte Worte originalgetreu wiederzugeben. Eine Paraphrase ermöglicht es, einen fremden Gedanken in eigenen Worten darzustellen, ohne seine Bedeutung zu verändern, und dabei die entsprechende Quelle anzugeben. In beiden Fällen muss man den Originaltext verstehen. Nur beim wörtlichen Zitat wird jedoch erwartet, dass die

wiedergegebenen Worte tatsächlich mit dem Original übereinstimmen.

Künstliche Intelligenz kann bei der Formulierung einer Paraphrase helfen, sofern du Zugang zum Ausgangstext hast und ihre Richtigkeit beurteilen kannst. Wenn du dagegen um ein genaues Zitat bittest, brauchst du eine tatsächlich vorhandene Textstelle. Das Modell sollte fehlende Worte nicht durch eigene schöpferische Einfälle ersetzen, selbst wenn es dabei bemerkenswertes literarisches Talent beweist.

Bei übersetzten Zitaten entsteht eine zusätzliche Schwierigkeit. Eine deutsche Übersetzung einer englischen Aussage ist keine wörtliche Wiedergabe der ursprünglichen Worte in derselben Sprache. Man kann sie als Übersetzung zitieren, sollte diesen Umstand aber entsprechend kennzeichnen. Wenn du eine veröffentlichte Übersetzung verwendest, ist es sinnvoll, ihre Quelle anzugeben. Übersetzt du selbst, solltest du deine eigene Version nicht als einzig autorisierte Formulierung darstellen.

Dieser Unterschied ist besonders in wissenschaftlichen und beruflichen Texten wichtig, in denen ein einziges Wort den Umfang einer Behauptung verändern kann. Das englische „may“ bedeutet nicht dasselbe wie „must“, und „associated with“ bedeutet keineswegs automatisch „caused by“. Eine sprachliche Glättung kann unbemerkt aus einer vorsichtigen Schlussfolgerung eine kategorische Versicherung machen.

Siebenhundert Webseiten behaupten dasselbe. Wer hat es also überprüft?

Eine der irreführendsten Eigenschaften des Internets besteht darin, dass sich Informationen beliebig oft

vervielfältigen lassen, ohne dass dadurch die Zahl unabhängiger Beweise zunimmt. Ein Zitat kann auf Hunderten von Webseiten, in Tausenden von Beiträgen und in Dutzenden von Präsentationen auftauchen, obwohl letztlich sämtliche Materialien auf einen einzigen ungeprüften Eintrag zurückgehen.

Stell dir vor, jemand veröffentlicht einen eindrucksvollen Satz, den er einem bekannten Wissenschaftler zuschreibt. Ein anderer Autor kopiert ihn in einen Artikel, der nächste setzt ihn auf eine Grafik, und wieder jemand anderes fügt ihn einer Online-Sammlung inspirierender Zitate hinzu. Einige Jahre später liefert eine Suchmaschine unzählige Treffer. Allein durch die ständige Wiederholung erscheint die Information zunehmend vertraut, und Vertrautheit wird leicht mit Wahrheit verwechselt.

Es reicht deshalb nicht aus, die Webseiten zu zählen, auf denen bestimmte Worte vorkommen. Entscheidend ist, ob wenigstens eine davon auf eine glaubwürdige ursprüngliche Quelle verweist.

Bei einer Aussage aus einem Buch kann das eine bestimmte Ausgabe mit Seitenzahl oder einer anderen eindeutigen Fundstelle sein. Bei einer öffentlichen Rede ist eine offizielle Aufzeichnung, eine Aufnahme oder eine zuverlässige Transkription hilfreich. Geht es um einen wissenschaftlichen Artikel, solltest du möglichst den eigentlichen Text der Veröffentlichung heranziehen, anstatt dich mit einer Zusammenfassung zufriedenzugeben, die jemand anderes erstellt hat.

Originalmaterialien sind allerdings nicht immer zugänglich. Frühere Reden wurden möglicherweise nie aufgezeichnet, Dokumente können sich in Archiven befinden, und eine Aussage kann ausschließlich in Berichten von Zeitzeugen