

Wie könnte KI die Genealogie verändern?

Von Hermann Hartenthaler

In den vergangenen Jahren hat die Forschung im Bereich der Künstlichen Intelligenz (KI) beeindruckende Fortschritte verzeichnet. Die daraus hervorgegangenen Werkzeuge haben mittlerweile ein Reifestadium erreicht, das ihren Einsatz auch in der Genealogie nicht nur möglich, sondern äußerst fruchtbar macht.

Angesichts der durch die fortschreitende Digitalisierung stetig wachsenden Datenmengen aus genealogischen und historischen Quellen wird die Entwicklung effizienter Analyse- und Interpretationsmethoden zunehmend wichtiger. Der Einsatz von KI bietet hier revolutionär neue Möglichkeiten, um Erkenntnisse in der Genealogie zu gewinnen, die bisher nicht oder nur mit hohem Aufwand erreichbar waren.

Auf Seite 14 dieses Hefts stellt sich die KI „ChatGPT“ selbst vor. Diese Anwendung hat vor über einem Jahr eindrucksvoll die Potenziale der Künstlichen Intelligenz aufgezeigt. Ihre leichte Zugänglichkeit hat ChatGPT zu einem Wendepunkt in der Entwicklung der KI gemacht. Doch neben ChatGPT existieren mittlerweile zahlreiche andere KI-Anwendungen, die in verschiedensten Bereichen zum Einsatz kommen (siehe S. 22). Dieser Artikel stellt spezifische Anwendungen in der Genealogie kurz vor. Exemplarisch wird die Verwendung von ChatGPT zur Analyse von Todesanzeigen detailliert beleuchtet. Die Integration von KI-Technologien in die Ahnen- und Familienforschung verändert die Art und Weise, wie wir forschen, grundlegend. Dabei stellt sich die entscheidende Frage: Werden wir von diesen Entwicklungen profitieren, oder riskieren wir, die Freude an unserem Hobby zu verlieren?

KI-Werkzeuge für die Ahnenforschung

Bei der Ahnen- und Familienforschung gibt es eine ganze Reihe von Tätigkeiten, die durch KI-Werkzeuge effizient unterstützt werden können. Der genealogische Prozess kann grob in die folgenden drei Tätigkeiten gegliedert werden:

- Suchen und Sammeln (Auswerten von Quellen aller Art);
- Zuordnen, Behaupten und Bewerten;
- Präsentieren und Archivieren.

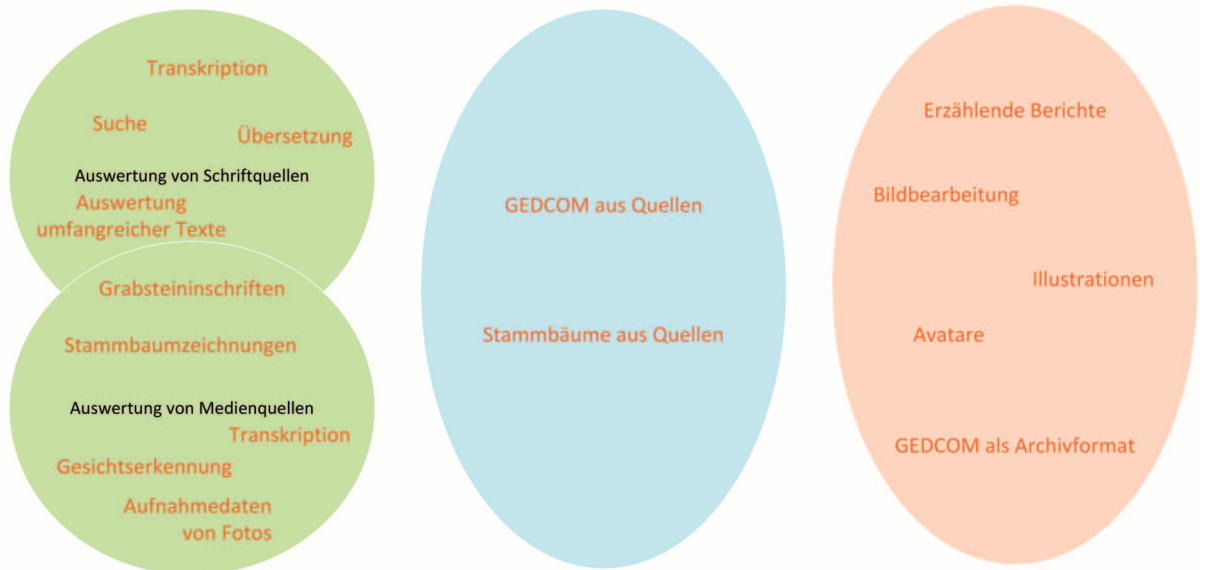
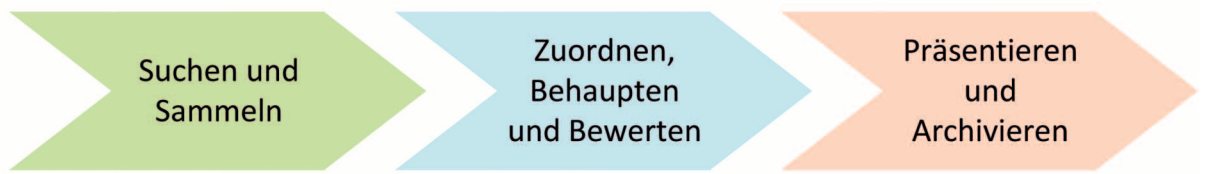
Im Folgenden werden bereits verfügbare KI-Anwendungsszenarien kurz vorgestellt, die diese Prozessschritte unterstützen. Die Linkliste (siehe S. 3) zu diesem Artikel enthält Verweise zu weiterführenden Informationen.

Auswerten von schriftlichen Quellen

- Suche nach genealogischen Quellen: Durch die Verknüpfung von KI-Werkzeugen wie ChatGPT und Datenbanksuchmaschinen wie Open Archives lassen sich Quellen finden und flexibel auswerten. Auch große Plattformen wie FamilySearch setzen KI zur Erschließung genealogischer Quellen ein.

- Transkription inkl. Handschriftenerkennung: Dazu gibt es hochspezialisierte KI-Systeme wie Transkribus (siehe S. 18), aber auch ChatGPT kann Text in Fotos erkennen und auswerten. Da bei einer Transkription immer Erkennungsfehler vorkommen können, ist es sehr sinnvoll, diese durch ChatGPT korrigieren zu lassen, wobei alle Änderungen natürlich sehr sorgfältig geprüft werden müssen.
- Übersetzung fremdsprachiger Dokumente: Da ChatGPT mit Texten aus sehr vielen Sprachen trainiert worden ist, kann diese KI auch lateinische Kirchenbucheinträge übersetzen.
- Auswertung von längeren, unstrukturierten Texten: Die Ermittlung genealogisch relevanter Informationen etwa in einer Ortschronik oder in einer Biografie ist die Paradedisziplin von ChatGPT. So lassen sich selbst umfangreiche Dokumente schnell auswerten. Entscheidend ist dabei die Informationsdichte: Die Daten zu fünf Personen in 100 Seiten zu finden ist kein Problem; wenn aber Daten zu 50 Personen auf einer Seite stehen, dann muss man sich von ChatGPT zur Auswertung etwas programmieren lassen.

KI im genealogischen Prozess



Auswerten von multimodalen Quellen

- Extraktion genealogischer Daten aus Abbildungen: Die Transkription von Stammbaumzeichnungen oder auch Grabsteininschriften einschließlich der korrekten Interpretation der Angaben ist mit ChatGPT möglich.
- Gesichtserkennung: Zur Identifizierung von Personen in alten Familienfotos gibt es spezialisierte Anwendungen, die allerdings auf Grund der Qualität alter Fotos und des oft sehr unterschiedlichen Lebensalters der dargestellten Personen meist schnell an ihre Grenzen stoßen.
- Erkennung des Aufnahmedatums einer Fotografie: Dazu hat MyHeritage eine KI trainiert, die etwa an Hand der Kleidermode eine Fotografie datieren kann (siehe CG 4/2023).

- Transkription von Audio- und Videoaufzeichnungen: Die Umwandlung von Interviewaufzeichnungen, etwa mit Verwandten, in durchsuchbare Texte erleichtert die Erschließung solcher Mediendokumente; je nach Aufnahmequalität, Wortwahl und Sprechweise (Dialekt) muss man die Transkripte aber mehr oder weniger intensiv nachbearbeiten.

Zuordnen, Behaupten und Bewerten

- Zuordnung von Quellenangaben zu Personen in Stammbäumen oder Weltbäumen: Eine KI hat für FamilySearch über 35 Millionen Quellendatensätze analysiert und daraus automatisch einen Stammbaum mit rund 9,3 Millionen Personen erstellt.
- Erstellung von GEDCOM-Dateien:

ChatGPT kann aus strukturierten und unstrukturierten Quellen Angaben zu Personen ermitteln und daraus direkt eine GEDCOM-Datei erzeugen, die dann in einem beliebigen Genealogieprogramm weiterbearbeitet werden kann.

Präsentieren und Archivieren

- Erstellung von Büchern und Artikeln
 - Erstellung eines erzählenden Berichts: ChatGPT kann (kleinere) GEDCOM-Dateien analysieren und daraus einen erzählenden Bericht erstellen; dabei kann ggf. auch der zeitgeschichtliche Hintergrund einbezogen werden. Die Erzählperson und der Stil können dabei frei gewählt werden.
 - Bearbeitung von Fotos: MyHeritage hat eine ganze Reihe von KI-Werkzeu-

gen zur Kolorierung und Animation alter Fotos entwickelt.

- Erstellung von Illustrationen: Es gibt inzwischen sehr viele KI-Anwendungen, die künstlerische oder fotorealistische Bilder oder auch Grafiken erzeugen können, die sich zur Illustration in Ahnenberichten einsetzen lassen. Auch mit ChatGPT lässt sich so etwas realisieren, wobei das Zeichnen von Stammbäumen dabei nur in sehr einfachen Fällen gelingen wird.

- Erzeugung von Avataren: Wenn man zu einer verstorbenen Person Dokumente, Video- und Tonaufzeichnungen hat, kann man mit einer speziellen KI einen Avatar erzeugen, sodass man mit der verstorbenen Person einen Dialog zu ihrem Leben wie in einer Videokonferenz führen kann. Sicherlich eine eher gruselige KI-Anwendung.

Nutzung von ChatGPT in praktischen Beispielen

Die Integration von KI-Werkzeugen wie ChatGPT in die genealogische Arbeit bietet dank einer benutzerfreundlichen Schnittstelle, die Dialoge in natürlicher Sprache unterstützt, eine relativ einfache Handhabung. Eine gewisse Einarbeitungszeit für das Erzielen optimaler Ergebnisse ist aber unumgänglich. Das Entwickeln effektiver Anfragen, bekannt als „Prompts“, setzt Kenntnisse über die Fähigkeiten und Einschränkungen der KI sowie ein klares Verständnis der jeweiligen Aufgabenstellung voraus (siehe Kasten).

Unter den verfügbaren KI-Tools sticht ChatGPT von Open AI besonders hervor, insbesondere wegen seiner Fähigkeit, mit GEDCOM-Dateien – dem Standardformat für Genealogiedaten – zu arbeiten. Dies macht es zu einem besonders nützlich-

chen Werkzeug für Genealogieanwendungen. ChatGPT ist in einer kostenlosen Version nach einfacher Registrierung zugänglich und bietet zusätzlich eine leistungsstärkere, abonnementbasierte Version an. Diese Premium-Version ermöglicht den Einsatz spezialisierter Module und den Zugriff auf einen Marktplatz für GPTs (generative pre-trained transformers, siehe S. 22), wo Benutzer solche maßgeschneiderten KI-Modelle untereinander teilen können. Mit bereits etwa drei Millionen verfügbaren Anpassungen, einschließlich Dutzender speziell für genealogische Zwecke, eröffnet dieser Marktplatz umfangreiche Möglichkeiten für individuelle Forschungsbedürfnisse.

An einem einfachen Beispiel soll nun gezeigt werden wie etwa ein „obituary“, eine in den USA übliche Form einer Todesanzeige, mit ChatGPT ausgewertet

WIE FORMULIERT MAN ERFOLGREICHE PROMPTS?

Ein „Prompt“ ist ganz allgemein die Aufforderung eines Computers an einen Benutzer, eine Eingabe zu machen. In einem Dialog (Chat) mit einem KI-System ist ein Prompt die Formulierung einer Aufgabe, die von der KI bearbeitet werden soll.

Ein solcher Prompt sollte knapp, vollständig und präzise formuliert werden. Dabei hat es sich bewährt, die in der Abbildung genannten Punkte anzusprechen. Es empfiehlt sich also, das Themenfeld (etwa „Genealogie“) zu benennen, dazu kann man der KI eine Rolle zuweisen (z. B. Genealoge); man sollte das Ziel und die Schritte zur Erreichung des Ziels vorgeben; und man sollte genau beschreiben, wie das Ergebnis dargestellt werden soll, etwa in der Form einer Tabelle. Und schließlich muss man das Quellmaterial in den Prompt kopieren, etwa in der Form eines Textes oder eines Bildes.

Kleinere Rechtschreibfehler werden verziehen, aber unvollständige Aufgabenstellungen oder solche, die logische Fehler enthalten, führen regelmäßig zu unbefriedigenden Ergebnissen.

Bei komplexen Aufgaben gibt es zwei grundsätzlich verschiedene Vorgehensarten. Bei der einen formuliert man einen komplexen Prompt, beschreibt etwa die Schritte zur Aufgabenlösung sehr detailliert. Wenn man mit dem Ergebnis noch nicht zufrieden ist, verbessert man den Prompt und fängt so jedes Mal wieder von vorne an. Einen solch ausgefeilten Prompt kann man dann auch wiederverwenden, wenn man später ein ähnliches Problem erneut bearbeiten möchte.

Man kann aber auch schrittweise vorgehen: In einem ersten Prompt fordert man die KI auf, ein erstes Zwischenergebnis zu erzielen, und dann bearbeitet man dieses Zwischenergebnis in weiteren Schritten, bis man am Ziel ist. Dies kann dank der Fähigkeit der KI, sich an den Verlauf des Dialogs zu „erinnern“, über viele Schritte gehen. Dieses Vorgehen kann sehr hilfreich sein, wenn man selbst noch nicht genau weiß, was man erreichen kann oder wie man dort hinkommt.

Prompt für ChatGPT



Vorlage

Rolle
Thema

Ziel
Ergebnis

Aufgabe
Schritte

Form
Gestaltung
Stil

Text
Bild
Quelle

Beispiel

Als erfahrener
Genealoge

Erfasse die Daten aus
einer Todesanzeige

Suche die Namen,
die Verwandtschafts-
beziehungen und die
Datumsangaben

Erstelle eine Tabelle
und eine GEDCOM-
Datei

Hier ist die
Todesanzeige für
Julius Meier aus der
Morgenpost vom
1.1.2024

werden kann. Eine solche Todesanzeige enthält in einer sehr verdichteten Form die Namen der Angehörigen und ihre Beziehung zur verstorbenen Person. Die Abbildung zeigt den Auftrag an ChatGPT zur Auswertung, der hier auch gleich den auszuwertenden Text enthält. Alternativ könnte man den Text auch als Datei hochladen oder einen Verweis auf eine Adresse im Internet verwenden. Im Beispiel wird sowohl eine Liste der vorkommenden Personen als auch eine GEDCOM-

Datei angefordert. Diese Datei kann mit einem Klick heruntergeladen und in einem Genealogieprogramm eingelesen werden. Sie enthält alle in der Todesanzeige vorkommenden Angaben. Die Abbildung zeigt eine danach mit dem Programm webtrees erstellte Grafik mit den familiären Beziehungen – und einem Fehler: die Nichte Anne Marie hat den Mädchennamen Shingleton, nicht Keplinger. Der Vorgang unter Nutzung von ChatGPT ist etwas schneller im Ver-

gleich zur Erfassung der Personen direkt im Genealogieprogramm. Wenn man aber nicht nur eine solche Todesanzeige auswerten möchte, sondern sehr viele, dann wird man mit ChatGPT viel effizienter zu guten Ergebnissen kommen, auch wenn man das Ergebnis auf Fehler prüfen muss.

Ausblick

Die gezeigte Vielfalt der KI-Anwendungsszenarien in der Genealogie, zusammen mit der rasanten Weiterentwicklung der KI-Technologie, kann zu grundlegenden Änderungen in der Ahnen- und Familienforschung führen. Während der Fokus heute hauptsächlich auf Kirchenbüchern liegt, kann KI im Lesen historischer Schriften weniger versierten Forschenden auch historische Grundbücher, Gerichtsprotokolle, Häuserchroniken und viele andere Arten historischer Dokumente zugänglich machen, um sie nach Fakten zu ihrer Familiengeschichte zu durchsuchen. In solchen meist nicht indizierten Quellen findet man Angaben zu den Vorfahren oft nur nach dem Lesen von sehr viel Text; eine KI wie ChatGPT ist hier effizienter. Diese Vielfalt an neuen Quellen und die daraus ableitbaren Informationen ermöglichen ein tieferes Verständnis darüber, wie die Vorfahren gelebt haben und was ihnen widerfahren ist, und macht Geschichte lebendiger.

Die Effizienzsteigerung durch den Einsatz von KI in der genealogischen Forschung könnte im Bereich eines Faktors von Zehn bis Hundert liegen. Was heute Tage in Anspruch nimmt, könnte zukünftig in Stunden bewältigt werden. Wo heute Hunderte von Helfern mehrere Monate benötigen, um ein umfangreiches Werk zu indizieren, könnte dies zukünftig einigen wenigen in ein paar Wochen gelingen. Die Art der Tätigkeiten dürfte sich verändern: weniger Routinetätigkeiten, stattdessen als neuer Schwerpunkt die Sicherung der Qualität genealogischer Daten und der Schlussfolgerungen. Es wird dann wohl wichtiger sein, die richtigen Fragen zu stellen, als bloß fleißig Daten zu sammeln. Als Belohnung winkt ein farbenreicheres und umfassenderes Bild der Vorfahren, wodurch die Freude am Hobby „Genealogie“ auch zukünftig bestehen bleibt. Hoffen wir also, dass die eingangs genannte Befürchtung, KI könnte uns um diese Freude bringen, nicht eintreten wird.

You

Hier ist eine Todesanzeige. Erstelle als erfahrener Genealoge eine Tabelle aller vorkommenden Personen und ermittle ihr Verwandtschaftsverhältnis zur Verstorbenen (in deutscher Sprache). Führe auch die erwähnten Ehepartner einzeln auf. Die Tabelle soll Vornamen, Nachnamen und das Verwandtschaftsverhältnis enthalten. Erstelle dann aus der Tabelle eine GEDCOM-Datei.

KEYSER, WV – Jo Anne Keplinger, 85, of Keyser, passed away peacefully with family by her side, on Saturday, May 6, 2023 at Moran Manor in Westernport, MD. Born on February 23, 1938 in Keyser, WV, she was the daughter of the late Harry Henry and Mary Margaret (Pieper) Keplinger. She was also preceded in death by her brother, Robert W. Keplinger, sisters, Helen M. (Richard M.) Smith and Jeanne M. (Blakely V.) Shingleton; her nephew, Jon B. Shingleton and her niece, Anne Marie (Shingleton) Alsoleibi.
###

ChatGPT

Um eine neue GEDCOM-Datei basierend auf der erneuten Anfrage zu erstellen, werde ich zuerst die Personen und ihre Verwandtschaftsverhältnisse zur Verstorbenen in einer aktualisierten Tabelle darstellen. Anschließend erstelle ich die GEDCOM-Datei mit diesen Informationen.

Die beteiligten Personen und ihre Verwandtschaftsverhältnisse zur Verstorbenen sind:

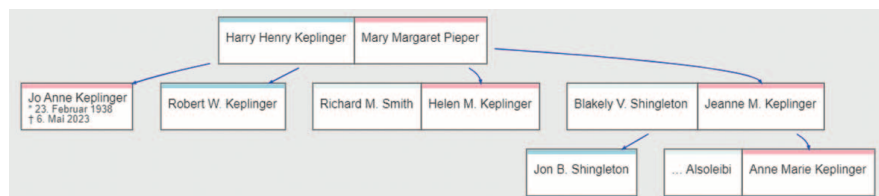
- Jo Anne Keplinger (Verstorbene)
- Harry Henry Keplinger (Vater)
- Mary Margaret Pieper (Mutter)
- Robert W. Keplinger (Bruder)
- Helen M. Smith (Schwester, verheiratet mit Richard M. Smith)
- Richard M. Smith (Schwager)
- Jeanne M. Shingleton (Schwester, verheiratet mit Blakely V. Shingleton)
- Blakely V. Shingleton (Schwager)
- Jon B. Shingleton (Neffe, Sohn von Jeanne M. Shingleton)
- Anne Marie Alsoleibi (Nichte, Tochter von Jeanne M. Shingleton, verheiratet mit einem Alsoleibi)

Ich werde nun eine GEDCOM-Datei mit diesen Daten erstellen.

Die aktualisierte GEDCOM-Datei für die Familie Keplinger wurde erstellt. Sie können die Datei über den folgenden Link herunterladen:

[Download aktualisierte GEDCOM-Datei \[-\]](#)

Dialog mit ChatGPT: Auswertung einer Todesanzeige



Von ChatGPT erstellte Grafik nach Auswertung einer Todesanzeige

Der Weg in die Zukunft der Genealogie könnte jedoch noch Herausforderungen bergen. KI ist ein mächtiges, aber auch komplexes Werkzeug. Sollte KI vorrangig von wenigen, bereits heute marktbestimmenden kommerziellen Genealogieplattformen eingesetzt werden, könnte dies zu einer Abhängigkeit führen, die den Zielen des Vereins für Computergenealogie entgegensteht. Die freie Verfügbarkeit von in gemeinsamer Arbeit zusammengetragenen Informationen wäre dann eventuell nicht mehr gewährleistet. Eine KI-Strategie ist daher essentiell, um die Potenziale von KI-Werkzeugen breit zugänglich zu machen. Eine Bildungsinitiative zur Verbesserung der Datenkompetenz für Interessierte, die Förderung von Open Data und der Aufbau von Partnerschaften zur Festlegung von Richtlinien und Standards für die Nutzung von KI in der Genealogie sollten kurzfristig in Angriff genommen werden.

Dank des intuitiven Zugangs zu KI könnten viele, die sich mit Genealogie befassen, in die Lage versetzt werden, effizient mit diesen neuen Werkzeugen zu arbeiten und neues Wissen zu generieren. Künstliche Intelligenz sollte als Verstärker für die Schwarmintelligenz der Ahnenforscher dienen. ■

WIE SOLLTE MAN CHATGPT BESSER NICHT NUTZEN?

Eine KI wie ChatGPT ist so gebaut, dass sie auf nahezu jede Frage eine Antwort liefert. Ob diese Antwort in einem objektiven Sinn richtig oder falsch ist, kann die KI nicht beurteilen, sie nutzt das von ihr im Training erworbene Wissen über Texte und erzeugt darauf basierend Wörter und Sätze. Als ein sogenanntes „großes Sprachmodell“ (Large Language Model) berechnet ChatGPT die Wahrscheinlichkeit, dass auf ein Zeichen oder ein Wort ein bestimmtes anderes folgt.

Wenn das Trainingsmaterial nicht ausgewogen war, spiegeln die Aussagen der KI diese Unausgewogenheit wider. Wenn man ChatGPT nach einer Quelle für eine Aussage fragt, wird die KI diese ggf. in überzeugender Form erfinden - man sagt dann, dass die KI „halluziniert“. Man sollte also alle Aussagen einer KI gründlich prüfen, bevor man sie als gesichert übernimmt. Wir vertrauen den Aussagen eines Computers oft mehr als denen eines Menschen, was im Falle einer KI nicht angebracht ist.

ChatGPT nach aktuellen Ereignissen zu fragen oder wie eine Suchmaschine zu nutzen ist weniger sinnvoll. Allerdings kann die aktuelle Version von ChatGPT auf das Internet zugreifen und so Lücken in den Trainingsdaten schließen. Der Zugriff von ChatGPT auf aktuelle Nachrichtendienste ist in Arbeit oder auch schon erfolgt, sodass ChatGPT hier auch auf Informationen zugreifen kann, die sich hinter Bezahlschranken verbergen. Ansonsten ist der Zugriff von ChatGPT auf per Passwort geschützte Inhalte, etwa bei Ancestry, generell nicht möglich.

Es ist nicht sinnvoll, ChatGPT nach konkreten Informationen zu längst verstorbenen Vorfahren zu fragen. Die Trainingsdaten der KI werden diese Informationen höchstwahrscheinlich nicht enthalten haben. Allerdings gibt es Plugins für ChatGPT, die man in der Bezahlversion aktivieren kann, und die direkt auf digitalisierte Quellenarchive zugreifen.

Die Auswertung von umfangreichen Dokumenten, etwa einer Ortschronik mit mehreren Hundert Seiten, ist aufgrund der begrenzten Verarbeitungskapazität nicht möglich. Man kann der KI durchaus zu einem solchen Werk gezielte Fragen stellen, aber das Extrahieren etwa aller genannten Personen muss in kleineren Portionen von jeweils wenigen Seiten erfolgen. Oder man bittet die KI, eigens für so eine Aufgabe ein Programm etwa in der Programmiersprache Python zu erstellen, das dann die Aufgabe löst. Auch das Erzeugen oder Analysieren einer GEDCOM-Datei ist auf etwa drei bis vier Generationen beschränkt; wenn der Stammbaum umfangreicher wird, kommen keine verlässlichen Ergebnisse heraus.

Der Betreiber von ChatGPT behält sich vor, die vom Nutzer eingegebenen Informationen ggf. zur Verbesserung der KI zu verwenden. Daraus folgt, dass man aus Datenschutzgründen auf keinen Fall vertrauliche Daten, etwa zu lebenden Personen, in einem Chat eingeben darf.

Ahnenblatt

SOFTWARE FÜR FAMILIENFORSCHER



**KOSTENLOSE DEMO-VERSION:
WWW.AHNENBLATT.DE**

Linkliste zum Artikel „Wie könnte KI die Genealogie verändern?“

- ChatGPT von Open AI

<https://chat.openai.com/>

- CompGen YouTube Video: ChatGPT in der Genealogie mit praktischen Beispielen

https://www.youtube.com/watch?v=FZqSz-JWh_Q&t=2935s

- CompGen Discourse-Kategorie: Künstliche Intelligenz

<https://discourse.genealogy.net/c/kuenstliche-intelligenz/157>

- Blog-Beiträge von CompGen zum Thema KI, Kategorie "Künstliche Intelligenz (KI)"

<https://www.compgen.de/category/kuenstliche-intelligenz-ki/>

- GenWiki: Künstliche Intelligenz

https://wiki.genealogy.net/K%C3%BCnstliche_Intelligenz

- Facebook-Gruppe: Genealogy and Artificial Intelligence (AI)

<https://www.facebook.com/groups/genealogyandai>

- Blog von Steve Little: AI Genealogy Insights

<https://aigenealogyinsights.com/>

- Einsatz von ChatGPT beim Transkribieren lateinischer Kirchenbuchtexzte:

<https://discourse.genealogy.net/t/lesehilfe-latein/820023>

- Wie nutze ich Chatbots effizient?

https://share.deutschlandradio.de/dlf-audiothek-audio-teilen.html?audio_id=dira_DLF_55833e09