

Generative KI, LLMs und GPT bei digitalen Editionen

Czmiel, Alexander

czmiel@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften,
Deutschland

Dumont, Stefan

dumont@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften,
Deutschland
ORCID: 0000-0002-6923-0950

Fischer, Franz

franz.fischer@unive.it
Venice Centre for Digital and Public Humanities, Ca'
Foscari Università Venezia, Italien

Pollin, Christopher

christopher.pollin@uni-graz.at
Zentrum für Informationsmodellierung, Universität Graz,
Österreich
ORCID: 0000-0002-4879-129X

Sahle, Patrick

sahle@uni-wuppertal.de
Lehrstuhl für Digital Humanities, Bergische Universität
Wuppertal, Deutschland
ORCID: 0000-0002-8648-2033

Schaßan, Torsten

schassan@hab.de
Herzog August Bibliothek Wolfenbüttel, Deutschland
ORCID: 0000-0002-8902-4775

Scholger, Martina

martina.scholger@uni-graz.at
Zentrum für Informationsmodellierung, Universität Graz,
Österreich
ORCID: 0000-0003-1438-3236

Vogeler, Georg

georg.vogeler@uni-graz.at
Zentrum für Informationsmodellierung, Universität Graz,
Österreich
ORCID: 0000-0002-1726-1712

Roeder, Torsten

dh@torstenroeder.de
Zentrum für Philologie und Digitalität, Universität
Würzburg, Deutschland
ORCID: 0000-0001-7043-7820

Fritze, Christiane

christiane.fritze@wienbibliothek.at
Wienbibliothek, Österreich
ORCID: 0000-0001-5099-8970

Henny-Krahmer, Ulrike

ulrike.henny-krahmer@uni-rostock.de
Institut für Germanistik, Universität Rostock, Deutschland
ORCID: 0000-0003-2852-065X

Kurzzusammenfassung

Dieser Workshop konzentriert sich auf die Erforschung der Anwendungsmöglichkeiten und Herausforderungen von KI-basierten Anwendungen wie GPT und Large Language Models (LLMs) im Kontext digitaler Editionen. GPT-4, mindestens bis Juli 2023 das führende Modell, bietet erhebliche Potenziale, z.B. für die Umwandlung von unstrukturiertem Text in strukturierte Daten und die Erkennung von benannten Entitäten. Dennoch liefert es bislang noch unbefriedigende Ergebnisse, weshalb sorgfältige Überwachung und Feedbacksysteme unerlässlich sind. Die Integration von LLMs in Arbeitsabläufe und Webentwicklungsprojekte ist vielversprechend, erfordert jedoch noch konzeptionelle und dann auch technische Vorstudien. In Anbetracht der rasanten KI- und LLM-Entwicklungen lädt der Workshop dazu ein, zu experimentieren und Strategien für den effektiven Einsatz dieser Modelle in digitalen Editionsprojekten zu diskutieren.

Generative KI, LLMs und GPT bei digitalen Editionen

KI-basierte Anwendungen wie Generative Pre-trained Transformer (GPT) sind revolutionär im Umgang mit Text. Ihre Stärke liegt in der Dialogfähigkeit, der natürlichsprachigen Textgenerierung, der Einbeziehung von Kontext, der Nutzung als semantischer "Wissensbasis", der Mehrsprachigkeit und der Möglichkeit der Feinabstimmung auf spezifische Aufgaben (*Chen et al. 2023, 1–2*). Außerdem können sie für die Entwicklung von Algorithmen zur Verarbeitung von Information eingesetzt werden. Zum Zeitpunkt der Erstellung dieses Abstracts (Stand Juli 2023) ist GPT-4 (*OpenAI 2023*) das leistungsfähigste Werkzeug. Zum Zeitpunkt des Workshops, im März 2024, wird GPT-4 seit einem Jahr eingeführt und seitdem für die verschiedensten

Einsatzgebiete getestet worden sein. Vielleicht werden bis dahin auch alternative Modelle und Szenarien im KI-Kontext verwendet.

Bei der Erschließung von historischen Dokumenten im digitalen Paradigma richten sich Editionen auf Texte, deren Inhalte und Strukturen etwa durch Markup-Sprachen aufbereitet werden. Aber auch Planung, Modellierung und Umsetzung digitaler Editionen manifestieren sich üblicherweise in textueller Form, z.B. in User Stories, Code für Daten und Algorithmen oder Dokumentationen. Sind generative Sprachmodelle dann das ultimative Werkzeug zur Umsetzung digitaler Editionen?

Diese Frage lässt sich aufgrund der rasanten Entwicklung in diesem Bereich noch nicht eindeutig beantworten. Wir befinden uns in einer Übergangsphase, in der die Möglichkeiten und Potenziale, aber auch die Herausforderungen, Grenzen und Risiken dieser neuen Technologie noch ausgelotet werden müssen. Dazu sollten Experimente durchgeführt werden, bei denen typische Aufgaben in Editionsprojekten gepromptet, getestet, angewandt, überprüft, eingeordnet und reflektiert werden. Mit diesem Ziel vor Augen möchten wir einen Aufruf zu Experimenten (Call for Experiments) veröffentlichen, um in einem Workshop zusammen mit der Fachgemeinschaft Versuche im Zusammenhang mit Large Language Models (LLMs) (*Zhao et al. 2023*), GPT oder alternativen Modellen und digitalen Editionen zu erörtern. Folgende Themenbereiche scheinen aktuell von Interesse zu sein. AI-unterstützte Experimente müssen sich aber nicht auf diese beschränken, sondern können sich auf jegliche Aspekte im Zusammenhang mit digitalen Editionen beziehen.

- **Überführung von unstrukturiertem Text (Transkription) in strukturierten Text (Markup)**

Unstrukturierte oder semistrukturierte Texte lassen sich derzeit besonders effizient durch LLMs in strukturierte Daten umwandeln. Beispielsweise kann GPT-4 zur automatischen Annotation, Klassifikation, Erzeugung von Textstrukturen, sowie zur Erstellung von Zusammenfassungen und Abstracts eingesetzt werden. Dies ermöglicht die Implementierung von Workflows, in denen Texte halbautomatisch in Interaktion eines Menschen mit einem LLM oder sogar vollautomatisch in einer Skript-Pipeline nach TEI-XML oder anderen Metadatenstandards und Auszeichnungssprachen konvertiert werden können.

- **Überführung von strukturiertem Text in explizite Datenstrukturen**

Die überwiegende Mehrheit wissenschaftlicher Editionen liegt in gedruckten Ausgaben vor. Das alte Ziel, diese Wissensschätze in standardkonform strukturierte Daten zu überführen, die z.B. nach dem TEI-Standard kodiert sind, besteht weiterhin. Viele Editionen sind inzwischen bilddigitalisiert und stehen oft mit OCR-Daten (z.B. PAGE XML) oder in PDFs zur Verfügung. Der weitere Prozess der Digitalisierung müsste aus einer Mischung regelgeleiteter "Lesung" der typografischen Strukturen und der Berücksichtigung der Un-

schärfe in der Einhaltung der typografischen Regeln (insbesondere zur Trennung von Text und Paratext) sowie der Bereinigung von OCR-Fehlern bestehen.

- **Named Entity Recognition, Normalisierung und Anreicherung**

Eine weitere potentielle Anwendung ist die Erkennung, Extraktion und Annotierung von Entitäten wie Personen, Organisationen, Orten und Werken, deren Identifizierung und Normalisierung ebenso wie die historischer Wert- und Datumsangaben. In diesem Zusammenhang erscheint auch die Verknüpfung mit Normdaten (Named Entity Linking, NEL) interessant, wie die Gegenüberstellung zu anderen Tools wie OpenRefine, oder die Einbindung von Reconciliation.

- **Kontextspezifische Annotationen**

LLMs und insbesondere GPT-4 zeichnen sich durch ihre Fähigkeit aus, Texte zu analysieren und auf Kontextwissen zuzugreifen, wodurch semantische Strukturen in Texten (semi-)automatisch annotiert werden können. Diese annotierten, semantischen Schichten, auch als "assertive Schichten" (*Vogeler 2019*) bezeichnet, können verschiedene Aspekte wie Transaktionen, Kommunikationsprozesse oder Rechtsmittel in historischen Quellen, sowie geografische Phänomene wie Reise- und Transportinformationen umfassen. GPT-4 ist in der Lage, z.B. geografischen Kontext in die Generierung von Antworten einzubeziehen, einschließlich von Informationen über Ländergrenzen, Flussverläufe und Gebirgsketten (*Roberts et al. 2023*). Könnte GPT-4 basierend auf einer historischen Quelle einen Kommentar generieren, der geografischen Kontext nutzt, um die Machbarkeit einer in der Quelle erwähnten Reiseroute zu bewerten?

- **Fehlermanagement und Datenkontrolle**

Wie jedes generative KI-Modell erzeugt auch GPT-4 häufig unbefriedigende Ergebnisse. Diese können durch fehlende oder tendenziöse Trainingsdaten, schlechtes Prompting, "Halluzinationen", fehlerhaftes Reasoning, übermäßig komplexe Aufgaben, eine begrenzte Anzahl von Ein- und Ausgab tokens oder eingeschränkte "Aufmerksamkeitsspannen" entstehen. Es ist von entscheidender Bedeutung, die erzeugten Daten gründlich zu prüfen und Feedback sowie redaktionelle Entscheidungen miteinzubeziehen, wie beispielsweise Studien aus dem medizinischen Bereich nahelegen (*Nori et al. 2023*). Ggf. ließe sich auch GPT-4 selbst zur Validierung der eigenen Resultate einsetzen.

- **Integration in Arbeitsabläufe**

Workflows (*Møller et al. 2023*) können iterativ gestaltet sein, wobei mehrere Prompts nacheinander eingesetzt werden, um verschiedene Aspekte der Annotation, Normalisierung oder des Fehlermanagements zu übernehmen. Ein erster Prompt erstellt z.B. die Textstruktur, ein zweiter verwendet die Ergebnisse des ersten Prompts für die Named Entity Annotation, ein dritter annotiert semantische Phänomene basierend auf den Ergebnissen des zweiten Prompts und der vierte führt eine automatische Datenqualitätsprüfung durch, basie-

rend auf “Ground Truth” TEI-XML-Dateien. Zwischen den Iterationen könnte ein:e Editor:in involviert sein, der bzw. die Fehler korrigiert und Feedback gibt. Allerdings sind hierfür geeignete Infrastrukturen und Softwarelösungen erforderlich, die derzeit noch nicht existieren.

- **Planung, Konzeption und Evaluierung digitaler Editionen**

Vor der Entwicklung eines digitalen Editionsprojekts ist es wichtig, die Anforderungen sowie die Forschungsfragen genau zu verstehen. Dies beinhaltet die Sammlung und Analyse von Anforderungen, User Stories von Stakeholdern sowie Kontextwissen von Expert:innen.

- **Webentwicklung und Benutzeroberflächen**

LLMs können auch in Webanwendungen integriert werden, um dynamische und interaktive Inhalte zu erstellen. Dies kann z.B. durch APIs geschehen, die es Entwicklern ermöglichen, auf die Fähigkeiten von z.B. GPT-4 zuzugreifen und diese in ihre Webanwendungen zu integrieren. Darüber hinaus unterstützen generative Sprachmodelle die Implementierung von digitalen Editionen, da sie sehr effizient im Schreiben von Quellcode sind.

- **Fine Tuning , Prompt Tuning , Prompt Engineering und Vektordatenbanken**

Fine Tuning ist ein Prozess, bei dem ein vortrainiertes Modell weiter trainiert wird, um es für eine spezifische Aufgabe zu optimieren. Prompt Tuning (*Wu et al. 2023*) hingegen steuert ein Sprachmodell durch gezielte Auswahl und Gestaltung von Eingabeaufforderungen (Prompts), ohne das Modell selbst zu ändern. Prompt Engineering (*White et al. 2023*) bezieht sich auf die Entwicklung effektiver Eingabeaufforderungen für Sprachmodelle. Vektordatenbanken (*Windsor and Choi 2023*) können zur Speicherung von Embeddings genutzt werden, die bei der Feinabstimmung, dem Prompt Tuning und dem Prompt Engineering hilfreich sind. Sie erlauben es, eine Sammlung von effektiven Prompts zu speichern und zu verwalten, wobei die Speicherung als Vektoren einen schnellen Zugriff auf ähnliche oder verwandte Prompts ermöglicht. Experimente könnten zeigen, wie und für welche Aufgaben diese Methoden im Bereich digitaler Editionen angewendet werden können.

Im Workshop sollen konkrete Anwendungsfälle vorgestellt werden, in denen GPT-4 oder andere LL-Modelle für die oben genannten Szenarien auf der Basis unterschiedlichen Quellenmaterials experimentell erprobt und im Hinblick auf Potenziale, Grenzen und Probleme sowie ethische (*Baktash and Dawodi 2023*) und theoretische Implikationen diskutiert werden. In einem weiteren Schritt soll auf Basis der Erkenntnisse aus den Experimenten reflektiert werden, welchen Einfluss diese Technologien auf die digitalen Editionen der Zukunft nehmen können. Dabei sollte GPT-4 nicht allein betrachtet werden, sondern insbesondere andere zukünftige Open Source-Alternativen wie bei-

spielsweise Orca (*Mukherjee et al. 2023*) berücksichtigt werden. Diskutiert werden sollen aber auch die immer noch bestehenden Grenzen und zu überwindenden Hürden, wie z.B. die Differenz zwischen einem rein auf Zeichen und Tokens aufbauenden Textverständnis und der Visualität handschriftlicher oder typografischer Dokumente, die eigentlich ein “bildliches” Lesen erfordern würden.

Das Format des Workshops, der acht Stunden (zwei Tageshälften) dauert, besteht aus der Vorstellung von Experimenten und einer kritischen Diskussion. Die Experimente werden in vier Blöcken zusammengefasst, in denen nach der Vorstellung der Experimente mindestens 30 Minuten für die Diskussion und Ergebnissicherung reserviert sind. Die Ergebnisse der Blöcke des ersten halben Tages werden in einer Zusammenfassung nach der Mittagspause sowie in einer abschließenden Runde gesammelt. Das Ergebnis des Workshops soll ein Experimentbericht sein, der zur Veröffentlichung zeitnah als Blogbeitrag und anschließend in einer einschlägigen Fachzeitschrift (ZfdG, DHQ, magazin) vorgesehen ist. Die Teilnehmer:innen setzen sich aus Beitragenden (Einzelpersonen oder Gruppen, die Experimente präsentieren), aktiv mitdiskutierenden Teilnehmer:innen und den Workshopleiter:innen aus dem Kreis der Einreichenden zusammen.

Zeitplan

- 09:00 - 09:30 Uhr: Einführung
- 09:30 - 10:45 Uhr: Block 1 - Experimentvorstellung und Diskussion
- 10:45 - 11:00 Uhr: Pause
- 11:00 - 12:15 Uhr: Block 2 - Experimentvorstellung und Diskussion
- 12:15 - 12:45 Uhr: Mittagspause (30 Minuten)
- 12:45 - 13:15 Uhr: Wrap-Up
- 13:15 - 14:30 Uhr: Block 3 - Experimentvorstellung und Diskussion
- 14:30 - 14:45 Uhr: Pause (15 Minuten)
- 14:45 - 16:00 Uhr: Block 4 - Experimentvorstellung und Diskussion
- 16:00 - 17:00 Uhr: Abschlussdiskussion und Festhalten von Ergebnissen

Workshop

- Format: Diskussionen anhand von präsentierten Experimenten
- Dauer: zwei halbe Tage
- Zielgruppe: Forscher:innen, Entwickler:innen und Fachleute im Bereich digitaler Editionen
- Maximale Teilnehmer:innenzahl: 20
- Erforderliche technische Ausstattung: Beamer, Computer, Internetzugang

Call for Experiments: Generative KI, LLMs und GPT bei digitalen Editionen

Im Zuge unseres bevorstehenden Workshops “Generative KI, LLMs und GPT bei digitalen Editionen” auf der Dhd24, lädt das Institut für Dokumentologie und Editorik (IDE) zur Einreichung experimenteller Beiträge ein, die sich mit den Anwendungsmöglichkeiten und Herausforderungen KI-basierter Tools wie GPT und Large Language Models (LLMs) in digitalen Editionen auseinandersetzen. In diesem Workshop sollen konkrete Anwendungsfälle präsentiert und diskutiert werden, in denen GPT-4 oder alternative LL-Modelle auf Basis unterschiedlichen Quellenmaterials für die vorgegebenen Szenarien experimentell erprobt und hinsichtlich ihrer Potenziale, Grenzen, Probleme sowie ethischen und theoretischen Implikationen diskutiert werden, um darauf aufbauend anhand der gewonnenen Erkenntnisse zu erörtern, welchen Einfluss diese Technologien auf die digitalen Editionen der Zukunft haben könnten.

Wir starten daher einen Call for AI-Experiments zu jeglichen Themenbereichen die im Zusammenhang mit digitalen Editionen stehen, einschließlich (aber nicht beschränkt auf):

- Überführung von unstrukturiertem Text (Transkription) in strukturierten Text (Markup)
- Überführung von strukturiertem Text in explizite Datenstrukturen
- Named Entity Recognition, Normalisierung und Anreicherung
- Kontextspezifische Annotationen
- Fehlermanagement und Datenkontrolle
- Integration in Arbeitsabläufe
- Planung, Konzeption und Evaluierung digitaler Editionen
- Webentwicklung und Benutzeroberflächen
- Fine Tuning, Prompt Tuning, Prompt Engineering und Vektordatenbanken

Um einen Beitrag einzureichen, senden Sie bitte eine halbseitige Skizze und Link zum Chat-Protokoll (wenn möglich) an ki@i-d-e.de bis zum 14. Januar 2024. Die Vorschläge werden von den Organisator:innen (Mitgliedern des IDE) des Workshops begutachtet, die Benachrichtigung über die Annahme erfolgt bis zum 26. Januar 2024.

Bibliographie

Baktash, Jawid Ahmad, and Mursal Dawodi. 2023. ‘Gpt-4: A Review on Advancements and Opportunities in Natural Language Processing’. arXiv. <https://doi.org/10.48550/arXiv.2305.03195>.

Chen, Zhutian, Chenyang Zhang, Qianwen Wang, Jakob Troidl, Simon Warchol, Johanna Beyer, Nils Gehlenborg, and Hanspeter Pfister. 2023. ‘Beyond Generating Code: Evaluating GPT on a Data Visualization Course’. arXiv. <https://doi.org/10.48550/arXiv.2306.02914>.

Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, and Luca Maria Aiello. 2023. ‘Is a

Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks’. arXiv. <https://doi.org/10.48550/arXiv.2304.13861>.

Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 2023. ‘Orca: Progressive Learning from Complex Explanation Traces of GPT-4’. arXiv. <https://doi.org/10.48550/arXiv.2306.02707>.

Nori, Harsha, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. ‘Capabilities of GPT-4 on Medical Challenge Problems’. arXiv. <https://doi.org/10.48550/arXiv.2303.13375>.

OpenAI. 2023. ‘GPT-4 Technical Report’. arXiv <https://doi.org/10.48550/ARXIV.2303.08774>.

Roberts, Jonathan, Timo Lüddecke, Sowmen Das, Kai Han, and Samuel Albanie. 2023. ‘GPT4GEO: How a Language Model Sees the World’s Geography’. arXiv. <http://arxiv.org/abs/2306.00020>.

Vogeler, Georg. 2019. ‘The “Assertive Edition”’. In *International Journal of Digital Humanities* 1 (2), 309–22. <https://doi.org/10.1007/s42803-019-00025-5>.

White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. ‘A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT’. arXiv. <https://doi.org/10.48550/arXiv.2302.11382>.

Windsor, Brad, and Kevin Choi. 2023. ‘Thistle: A Vector Database in Rust’. arXiv. <https://doi.org/10.48550/arXiv.2303.16780>.

Wu, Junda, Tong Yu, Rui Wang, Zhao Song, Ruiyi Zhang, Handong Zhao, Chaochao Lu, Shuai Li, and Ricardo Henao. 2023. ‘InfoPrompt: Information-Theoretic Soft Prompt Tuning for Natural Language Understanding’. arXiv. <https://doi.org/10.48550/arXiv.2306.04933>.

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, et al. 2023. ‘A Survey of Large Language Models’. arXiv. <https://doi.org/10.48550/arXiv.2303.18223>.