

More than Chatbots: Multimodal Large Language Models in geisteswissenschaftlichen Workflows

Oberbichler, Sarah

oberbichler@ieg-mainz.de
Leibniz-Institut für Europäische Geschichte, Mainz,
Deutschland
ORCID: 0000-0002-1031-2759

Pollin, Christopher

christopher.pollin@dhcraft.org
Digital Humanities Craft OG, Österreich
ORCID: 0000-0002-4879-129X

Rastinger, Nina C.

ninaclaudia.rastinger@oeaw.ac.at
Austrian Centre for Digital Humanities and Cultural
Heritage, Österreichische Akademie der Wissenschaften,
Österreich
ORCID: 0000-0002-3235-5063

Schneider, Stefanie

stefanie.schneider@itg.uni-muenchen.de
Ludwig-Maximilians-Universität München, Deutschland
ORCID: 0000-0003-4915-6949

Störiko, Johanna

johanna.stoeriko@uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland
ORCID: 0009-0009-3853-4757

Ewerth, Ralph

ralph.ewerth@tib.eu
TIB - Leibniz Information Centre for Science and
Technology, Hannover; Leibniz University Hannover,
Deutschland
ORCID: 0000-0003-0918-6297

Einleitung

„Being under construction“ bekommt im Zeitalter generativer Künstlicher Intelligenz (KI) eine völlig neue Bedeutung: Ähnlich wie Städte, die ständig erweitert, umgebaut oder umgestaltet werden, befindet sich auch die KI-

Technologie in einem permanenten Zustand des Umbaus. Bestehende Strukturen und Funktionen werden angepasst, erweitert und verbessert. Dieser kontinuierliche Umbauprozess spiegelt die Dynamik der KI-Entwicklung wider, bei der Modelle und Anwendungen ständig optimiert werden. So wie eine Stadt auch während des Umbaus weiter genutzt werden kann, bleibt auch die generative KI trotz – oder gerade wegen – ihres ständigen Umbaus ein leistungsfähiges Werkzeug.

Gleichzeitig sind im Prozess des technologischen Umbaus nicht alle Verbesserungen und Anpassungen gleich wichtig. Entscheidend ist vielmehr, wann Technologien bestimmte Schwellenwerte erreichen, die ihre breite Anwendbarkeit in der Arbeitswelt oder Forschung ermöglichen. Für den Einsatz von generativer KI in geisteswissenschaftlichen Workflows dürfte das Frühjahr 2024 einen solchen Wendepunkt markieren. Diese kritische Schwelle wurde erstmals mit der Einführung neuer Modelle von Unternehmen wie Anthropic, OpenAI, Google und Meta überschritten: Sie haben die generative KI von einem „spielerischen Experiment“, wie es Ethan Mollick, Co-Direktor des Generative AI Lab in Wharton, treffend beschreibt, in ein unverzichtbares Werkzeug verwandelt, das die zukünftige Forschungspraxis maßgeblich verändern wird (Mollick 2024). Eine weitere bedeutende Wende für die Forschung - Entwicklung hin zu mehr Transparenz und rechtlicher Compliance bei der Nutzung von Trainingsdaten - steht bereits vor der Tür.

Das Überschreiten bestimmter Schwellenwerte bringt aber auch erhebliche Unsicherheiten mit sich, da technische Veränderungen auch gesellschaftliche und kulturelle Umwälzungen nach sich ziehen. Fragen zum ethischen Einsatz von KI, also zu Datenschutz, Urheberrecht und Nachhaltigkeit sowie zu potenziellen Auswirkungen auf die Forschungstätigkeit und zur Verlässlichkeit von KI-generierten Informationen sorgen zudem für wissenschaftlichen wie gesellschaftlichen Debatten (Azamfirei, Kudchadkar und Fackler 2023; Bär 2024; Bender et al. 2021; Navigli, Conia und Ross 2023; Vyas 2022). Diese Unsicherheit spiegelt sich in der Natur der generativen KI wider: stets „in Arbeit“, „im Umbau“ – mit einem Potenzial, das wir noch nicht vollständig erfassen können. Die Entwicklung von Leitfäden und Frameworks, die eine LLM-bezogene Daten-, Quellen- und Methodenkritik sowie Explainable AI einschließen, ist ein wichtiger Schritt beim Abbau von Unsicherheiten (Schneider et al. 2022; Oberbichler 2024).

Während diese Unsicherheiten fortbestehen werden, zeichnen sich auch bemerkenswerte Chancen für die geisteswissenschaftliche Forschung ab (Pollin 2024). So zeigen immer mehr Forschungsvorhaben, die generative KI einsetzen, vielversprechende Anwendungsmöglichkeiten (Armaselu 2024; Oberbichler 2024; Rastinger 2024; Springstein et al. 2024). Insbesondere Multimodal Large Language Models (MLLMs) markieren einen Meilenstein in der Verarbeitung komplexer geisteswissenschaftlicher Daten und eröffnen bislang ungeahnte Perspektiven für die Analyse und Interpretation kultureller Artefakte.

Ziele und Leitfragen des Panels

Dieses Panel hat zum Ziel, anhand konkreter und interdisziplinärer Forschungsbeispiele den Nutzen von (M)LLMs und Large Vision Models (LVMs) für geisteswissenschaftliche Forschungsworkflows aufzuzeigen sowie den Gebrauch dieser Modelle kritisch zu reflektieren. Angesichts der rasanten Fortschritte in der KI-Forschung stehen dabei folgende Fragen im Mittelpunkt:

Wie können (M)LLMs und LVMs effektiv in geisteswissenschaftliche Forschungsworkflows implementiert werden?

Welche konkreten Effizienzsteigerungen sind durch den Einsatz von KI-Modellen in der Datenverarbeitung und -analyse realisierbar?

Wie kann der Fokus von der technischen Datenerzeugung (Normalisierung, Modellierung, Generierung) stärker auf die geisteswissenschaftliche Analyse und Interpretation verlagert werden?

Welche möglichen Bedenken, Konsequenzen und Nachteile ergeben sich aus der Integration von KI in die geisteswissenschaftliche Forschungspraxis?

Was bedeutet der Einsatz von (M)LLMs und LVMs für die Nutzbarkeit in Workflows?

Besonderes Augenmerk wird den jüngsten Entwicklungen in der KI-Forschung gewidmet. Dazu gehören verbesserte Reasoning-Fähigkeiten, die Erweiterung von Kontextfenstern für eine umfangreichere Verarbeitung längerer Texte, Fortschritte im KI-Engineering für robustere und anpassungsfähigere Systeme, die Entwicklung und Evaluation neuer, komplexer Prompt-Strategien, sowie die Leistungsfähigkeit aktueller proprietärer Frontier-Modelle. Das Panel diskutiert dabei grundlegend, welche Möglichkeiten, aber auch Schwierigkeiten der Einsatz generativer KI für geisteswissenschaftliche Forschungsprojekte bietet und welche Richtungen hier zukünftig eingeschlagen werden können. Zur Diskussion stehen zudem, neben grundlegenden Fragen zu quellen-, daten- und methodenkritischen Aspekten, die erweiterten multimodalen Fähigkeiten der neuesten Modelle, die eine ganzheitliche Analyse von Text, Bild und anderen Datentypen ermöglichen.

Statements der Panelist:innen

Das Panel wird von Prof. Ralph Ewerth moderiert und setzt sich aus folgenden Kurzstatements zusammen:

MLLM in den digitalen Geisteswissenschaften: Auswirkungen und Entwicklungen (Pollin)

Im ersten Impulsvortrag wird gezeigt, wie die Entwicklung von Workflows, die auf Multimodalen Large Language Models (MLLMs) basieren, die Digital Humanities beeinflusst. Die Digital Humanities durchlaufen gegenwärtig eine Phase signifikanter Veränderung. Als interdisziplinä-

res Forschungsfeld, das sich der digitalen Verarbeitung und Analyse multimodaler geisteswissenschaftlicher Daten – darunter Texte, Bilder, Audio, Video und deren Forschungskontexte – widmet, sind die Digital Humanities unmittelbar von den Fortschritten im Bereich der MLLMs betroffen. Diese technologischen Innovationen wirken sich weitreichend auf Kernbereiche wie Datenmodellierung, -transformation und Programmierung aus.

Die rapide technologische Entwicklung erfordert eine eingehende Auseinandersetzung mit den Möglichkeiten und Grenzen der Automatisierung in diesen Bereichen. Dabei gilt es, nicht nur technische Aspekte zu berücksichtigen, sondern auch methodologische Implikationen für die geisteswissenschaftliche Forschung und den damit verbundenen Ressourceneinsatz zu reflektieren.

Die Entwicklung und Integration von KI-basierten Workflows in den Digital Humanities befindet sich noch in einem frühen Stadium. Fortschritte in Bereichen wie AI Agents, Advanced Prompt Engineering und Prompt Optimization sowie die Gestaltung geeigneter Benutzerschnittstellen (UI) bieten Potenzial für Effizienzsteigerungen. Die Entwicklung von Workflows, die auf MLLMs basieren, könnte die Möglichkeiten der derzeitigen GPT-4-Tier-Modelle erweitern. Und mit der direkten Anbindung an die GPT-5-Tier-Modelle werden diese Arbeitsabläufe voraussichtlich deutlich verbessert – oder auch nicht.

Text- und Tokenklassifikation mit LLMs (Rastinger)

Unstrukturiertes Textmaterial bedeutet für Forschende oftmals die Herausforderung, spezifische Textbestandteile zu identifizieren und/oder linguistisch-semantisch zu klassifizieren. Insbesondere in ressourcenarmen Kontexten, in denen durch Non-Standard-Daten (z.B. historische Sprache, spezifische Domäne, wenig erforschte Textsorte) noch keine annotierten Trainings- oder Testdaten vorliegen, zeigt sich hier das hohe Potenzial von LLMs (Wang et al. 2023): Anstatt mit erheblichem Aufwand umfangreiche Trainingssets zu erstellen, können über niedrigschwellige Zugänge wie Prompt Engineering oder Retrieval-Augmented-Generation (RAG) je nach Aufgabe bereits sehr gute Ergebnisse erzielt werden, die den Prozess von Datenaufbereitung und -analyse wesentlich beschleunigen.

Der Impulsvortrag behandelt Fallbeispiele aus der Arbeit mit LLMs und frühneuzeitlichen Zeitungstexten, genauer mit periodisch publizierten Listen (Rastinger 2024), wobei unterschiedliche Modelle (z.B. GPT-4o-mini, LLaMA3) und unterschiedliche Klassifikationsaufgaben (z.B. Erkennung Abkürzungen und benannten Entitäten) in den Blick rücken. Hieran soll diskutiert werden, wo Potenziale, aber auch Grenzen der Tokenklassifikation mit generativer KI liegen und inwiefern der Einsatz bestimmter Techniken, wie In-Context-Learning (Loukas et al. 2023) oder Self-Consistency (Xie et al. 2023), die Output-Qualität von LLMs verbessern bzw. stabilisieren und so ihre Integration in geisteswissenschaftliche Workflows erleichtern kann.

Erstellen von Bildkorpora mit MLLMs (Störiko)

Für die digitale Untersuchung geisteswissenschaftlicher Fragestellungen ist ein gutes Korpus unerlässlich. Modelle wie CLIP (Contrastive Language-Image Pre-Training; Radford et al. 2021), BLIP2 (Bootstrapping Language-Image Pre-training; Li et al. 2023), oder GPT-4o (OpenAI 2024) sind dabei Meilensteine für die Arbeit mit großen, unstrukturierten Bildsammlungen. Sie ermöglichen, diese semantisch zu durchsuchen oder mit inhaltlichen Tags zu versehen (Smits und Wevers 2023). Mittlerweile stehen für Aufgaben wie Image Captioning oder Image Question Answering verschiedene Modelle frei und niederschwellig zur Verfügung.

Der dritte Impulsvortrag zeigt am Beispiel eines Experiments zur Vorschlagwortung von Fotografien auf historischen Bildpostkarten, wie MLLMs zur Erstellung eines Bildkorpus eingesetzt werden können. Das Experiment ist das Ergebnis einer Lehrveranstaltung und eines Studierendenprojektes zu einigen Postkarten der Reihe „Lernt Deutschland kennen!“. Dabei werden die Bilder von mehreren Modellen in mehreren Durchläufen beschrieben und die besten Schlagworte statistisch ermittelt – ähnlich wie bei Projekten mit mehreren menschlichen Annotierenden. Die so entstehenden inhaltlichen Tags können für die Zusammenstellung von Korpora für spezifische Fragestellungen, aber auch zur statistischen Analyse des Datensatzes verwendet werden. Der Ansatz zeigt, dass auch unperfekte, sich noch in der konstanten Entwicklung befindende Modelle bereits produktiv eingesetzt werden können.

LLMs für die qualitative Datenanalyse (Oberbichler)

Der Einsatz von LLMs in der qualitativen Inhalts- bzw. Diskursforschung ist bislang noch kaum erforscht. Neueste Modelle, open-source wie auch closed-source, eröffnen jedoch vielversprechende Möglichkeiten für ihre Anwendung in der qualitativen geisteswissenschaftlichen Forschung. Während herkömmliche digitale Methoden der korpusbasierten Diskurs- oder Inhaltsanalyse Texte als „bag of words“ behandeln und sprachliche Muster anhand von Häufigkeitsverteilungen identifizieren (Marjanen et al. 2020; Rouhana 2023), bieten LLMs eine deutlich tiefgreifendere Fähigkeit zur Verarbeitung von Sprache, Semantik und Kontext (Hayes 2023). LLMs können nicht nur statistische Muster erkennen, sondern auch die Komplexität der natürlichen Sprache sowie die Dynamik der menschlichen Kommunikation und des menschlichen Ausdrucks erfassen (Rouhana 2023).

Dieser Impulsvortrag illustriert anhand konkreter Beispiele aus einem laufenden Forschungsprojekt den Einsatz von LLMs in der qualitativen Datenanalyse. Das Projekt untersucht transnationale Nachrichtenströme in Deutschland, Frankreich, Großbritannien, Italien und der Schweiz im Zeitraum von 1850 bis 1950 und konzentriert sich da-

bei auf Nachrichten und Narrative zu den Themen (Rückkehr-)Migration und Umwelt-/Naturkatastrophen. Unter Berücksichtigung datenethischer Aspekte historischer Zeitungsquellen soll anhand dieses Projektes diskutiert werden, welche Potenziale, aber auch Risiken LLMs für die qualitative Forschung bergen und wie sie bisher wenig automatisierte Forschungsfelder beeinflussen können.

Kunsthistorisches Retrieval mit LVMs (Schneider)

In den letzten Jahren wurden verstärkt e-Research-Werkzeuge entwickelt, die eine multimodale Erschließung kunsthistorischer Daten ermöglichen (Ohm et al. 2023; Schneider et al. 2022). Insbesondere durch die Implementierung von Trainingsmethoden wie CLIP (Radford et al. 2021) haben sich dabei auf neuronalen Netzen basierende Vektorrepräsentationen für verschiedene kunsthistorische Retrievalaufgaben als nützlich erwiesen (Castellano und Vessio 2022; Zhao et al. 2023) und – damit einhergehend – einen Wandel von der historisch etablierten Praxis des Close Viewing – der qualitativen Analyse einzelner Objekte in ihren raum-zeitlichen Kontexten – hin zum quantitativen Distant Viewing großer Bildkorpora angestoßen (Arnold und Tilton 2019).

Hier setzt dieser Impulsvortrag an: Anhand zweier Fallbeispiele wird gezeigt, wie diese Werkzeuge in der kunsthistorischen Praxis effektiv angewendet werden können. In diesem Kontext werden auch – unter dem Gesichtspunkt der Explainable Artificial Intelligence (XAI) – aus einem im Rahmen des DFG-Schwerpunktprogramms Das digitale Bild laufenden Projekt Methoden zur „Ent-Blackboxierung“ von LVMs vorgestellt, um die Entscheidungsprozesse künstlicher neuronaler Netze transparent und somit für bildorientierte Forschungsprozesse besser nutzbar zu machen.

Methodik und Ablauf des Panels

Nach einer Einführung durch den Moderator (Prof. Ralph Ewerth), in der auch die Leitfragen des Panels vorgestellt werden, zeigen die fünf Panelist:innen in jeweils fünfminütigen Statements anhand konkreter Beispiele aus ihrer Forschung bzw. Arbeit auf, wie sie generative KI in ihre geisteswissenschaftlichen Workflows integrieren. Während der Statements hat das Publikum die Möglichkeit, über OnlineTED®, einer Plattform für interaktive Online-Beteiligung,¹ weitere Leitfragen zu formulieren, die gemeinsam mit den bestehenden Leitfragen in einer 30-minütigen Diskussionsrunde erörtert werden. Im dritten Teil des Panels wird die Diskussion dann vollends für das Publikum geöffnet. Statements, Anmerkungen und weitere Fragen können dann sowohl über Wortmeldungen als auch über OnlineTED® in die Diskussion eingebracht werden.

Fußnoten

1. <https://onlinetd.de/de> (zuletzt abgerufen am 20.07.2024).

Bibliographie

- Armaselu, Florentina.** 2024. "Small-Scale Testing on Generative AI and Post-OCR Correction in Historical Datasets." In DH Benelux 2024, Leuven, Belgium. Zenodo. 10.5281/zenodo.11403647.
- Arnold, Taylor und Lauren Tilton.** 2019. "Distant Viewing. Analyzing Large Visual Corpora." *Digital Scholarship in the Humanities* 34: i3–i16. 10.1093/llc/fqz013.
- Azamfirei, Razvan, Sapna R. Kudchadkar und James Fackler.** 2023. "Large Language Models and the Perils of Their Hallucinations." *Critical Care* 27 (1): 120. 10.1186/s13054-023-04393-x.
- Bär, Dominik.** 2024. "Über die Verwendung von Large Language Models in der Wissenschaft: Eine überschätzte Gefahr?" 10.13154/294-12469.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major und Shmargaret Shmitchell.** 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. 10.1145/3442188.3445922.
- Castellano, Giovanna und Gennaro Vessio.** 2022. "A Deep Learning Approach to Clustering Visual Arts." In *International Journal of Computer Vision* 130 (11): 2590–2605. 10.1007/S11263-022-01664-Y.
- Hayes, Adam.** 2023. " "Conversing" with Qualitative Data: Enhancing Qualitative Research through Large Language Models (LLMs)." OSF. 10.31235/osf.io/yms8p.
- Li, Junnan, Savarese, Silvio, Hoi, Steven.** 2023. "BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models." *arXiv*. 10.48550/arXiv.2301.12597.
- Loukas, Lefteris, Ilias Stogiannidis, Prodromos Malakasiotis und Stavros Vassos.** 2023. "Breaking the Bank with ChatGPT: Few-Shot Text Classification for Finance." In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, 74–80. *arXiv*. 10.48550/arXiv.2308.14634.
- Marjanen, Jani, Elaine Zosa, Simon Hengchen, Lidia Pivovarova und Mikko Tolonen.** 2020. "Topic Modelling Discourse Dynamics in Historical Newspapers." *arXiv*. 10.48550/arXiv.2011.10428.
- Navigli, Roberto, Simone Conia und Björn Ross.** 2023. "Biases in Large Language Models: Origins, Inventory, and Discussion." *J. Data and Information Quality* 15 (2): 10:1-10:21. 10.1145/3597307.
- Oberbichler, Sarah.** 2024. "Large-Scale Research with Historical Newspapers: A Turning Point through Generative AI." *Billet. Digital Humanities Lab (blog)*. 21 June 2024. 10.58079/11v9i.
- Ohm, Tillmann, Mar Canet Sola, Andres Karjus und Maximilian Schich.** 2023. "Collection Space Navigator. An Interactive Visualization Interface for Multidimensional Dataset." In *Proceedings of the 16th International Symposium on Visual Information Communication and Interaction. VINCI 2023*, 24:1–24:5. 10.1145/3615522.3615546.
- OpenAI.** 2024. "Hello GPT-4o." <https://openai.com/index/hello-gpt-4o/> (zugegriffen: 20. Juli 2024).
- Pollin, Christopher.** 2024. "Workshopreihe 'Angewandte Generative KI in den (Digitalen) Geisteswissenschaften'." Zenodo. 10.5281/ZENODO.10647754.
- Radford, Alex, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger und Ilya Sutskever.** 2021. "Learning Transferable Visual Models From Natural Language Supervision." In *Proceedings of the 38th International Conference on Machine Learning. ICML 2021* 139: 8748–8763.
- Rastinger, Nina C.** 2024. "Re-Reading Lists in Historical Newspapers: Digital Insights into an Overlooked Text Type." In *Selected Papers from the CLARIN Annual Conference 2023*. 10.3384/ecp210016.
- Rouhana, Toni.** 2023. "Critical Discourse Analysis Guided Topic Modeling: The Case of Al-Jazeera Arabic." *Information, Communication & Society* 26 (5): 904–22. 10.1080/1369118X.2023.2166364.
- Schneider Stefanie, Matthias Springstein, Javad Rahnama, Hubertus Kohle, Ralph Ewerth und Eyke Hüllermeier.** 2022. "ART. Eine Suchmaschine zur Unterstützung von bildorientierten Forschungsprozessen." In *8. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e. V. DHd 2022*, 142–147. 10.5281/zenodo.6310.528175.
- Springstein, Matthias, Stefanie Schneider, Javad Rahnama, Julian Stalter, Maximilian Kristen, Eric Müller-Budack und Ralph Ewerth.** 2024. "Visual Narratives. Large-scale Hierarchical Classification of Art-historical Images." In *Proceedings of the IEEE Workshop on Applications of Computer Vision. WACV 2024*, 7220–7230.
- Smits, Thomas und Melvin Wevers.** 2023. "A Multimodal Turn in Digital Humanities. Using Contrastive Machine Learning Models to Explore, Enrich, and Analyze Digital Visual Historical Collections." *Digital Scholarship in the Humanities* 38 (3): 1267–1280. 10.1093/llc/fqad008.
- Vyas, Bhuman.** 2022. "Ethical Implications of Generative AI in Art and the Media." *International Journal For Multidisciplinary Research* 4 (4): 1–11. 10.36948/ijfmr.2022.v04i04.9392.
- Wang, Shuhe, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li und Guoyin Wang.** 2023. "GPT-NER: Named Entity

Recognition via Large Language Models.“ arXiv. 10.48550/arXiv.2304.10428.

Xie, Tingyu, Qui Li, Jian Zhang, Yan Zhang, Zuozhu Liu und Hongwei Wang. 2023. "Empirical Study of Zero-Shot NER with ChatGPT.“ In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 7935–7956. arXiv. 10.48550/arXiv.2310.10035.

Zhao, Yuguang, Jeroen Stumpel, Huib de Ridder und Maarten W. A. Wijntjes. 2023. "Zooming in on Style. Exploring Style Perception Using Details of Paintings.“ International Journal of Computer Vision 23 (2). 10.1167/jov.23.6.2.