

Voll automatisiert die Natur in historischen Reiseberichten erkennen? Entzauberung von KI-Werkzeugen und ihr Nutzen für die Geisteswissenschaften

Vignoli, Michela

michela.vignoli@ait.ac.at

AIT Austrian Institute of Technology, Österreich

ORCID: 0000-0002-9495-5697

Gruber, Doris

doris.gruber@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

ORCID: 0000-0002-0512-100X

Einleitung

Der rasante Vormarsch Künstlicher Intelligenz (KI) macht auch vor den Geisteswissenschaften nicht halt. Die vielversprechende Hoffnung lautet, große, digitalisierte Daten-Korpora voll automatisiert zu analysieren und darin Strukturen und Beziehungen mithilfe komplexer statistischer Modellvorhersagen zu erkennen. Modernste Ansätze des maschinellen Lernens (ML), insbesondere große Sprachmodelle (*Large Language Models, LLMs*) wie sie ChatGPT¹ verwendet, erlauben es, in einem nie dagewesenen Umfang multimodale Inhalte mit Rücksicht auf den semantischen und semiotischen Kontext zu durchsuchen, zu ordnen und zusammenzufassen (z.B. Yang et al., 2024). Doch inwieweit treffen diese Versprechen zu und können maschinelle Lernmodelle als unterstützende Werkzeuge für historisch-wissenschaftliche Argumentationsprozesse bei historischen Bildern und nicht standardisierten Texten eingesetzt werden? Sind sie in der Lage, relevante Quellen aufzufinden oder gar interpretative Aufgaben zu lösen?

Deep Learning Ansätze finden bereits in unterschiedlichen kunstgeschichtlichen und historischen Forschungskontexten Anwendung (z.B. Saleh und Elgammal, 2016; Cetinic et al., 2019; Im et al., 2022). Auch das *Ottoman Nature in Travelogues* Projekt² (*ONiT*; FWF: P 35245) entwickelt eine KI-gestützte Methode, die auf die Analyse von

„Naturdarstellungen“ in Reiseberichten über das Osmanische Reich zugeschnitten ist (Vignoli et al., 2023a; Vignoli et al., 2023b). Das Projekt untersucht, welche Rolle die westlichen Darstellungen „osmanischer Natur“ in den zwischen 1501 und 1850 gedruckten Berichten spielen. Dafür kommen KI-Werkzeuge zum Einsatz, die durch einen Distant Reading Zugang (Moretti, 2013) zur Beantwortung dieser Forschungsfrage beitragen. Als „Naturdarstellung“ fasst das Projekt jegliche piktoriale sowie textuelle Darstellungen von Flora, Fauna, Landschaften und -karten. Zur computergestützten Identifikation relevanter Bild- und Textteile in einem heterogenen Korpus von über 1.700 digitalisierten Werken optimieren wir sogenannte Transformer-Modelle (Vaswani et al., 2017) zur Informationsgewinnung (*Information Retrieval*).

Methode

Zunächst untersuchen wir Bild- und Textmaterial in getrennten Verfahren, um die Analyse beider semantisch unterschiedlichen Kodierungen von Informationen anschließend zusammenzuführen. Für die Bildanalyse konnten wir aus den Reiseberichten bisher über 22.000 piktoriale Darstellungen extrahieren. Über 8.000 davon umfassen Naturdarstellungen, wobei die Flora, Fauna, Landschaften und -karten entweder alleinstehend abgebildet sind oder als Detail im Vorder- oder Hintergrund auftreten können. Um das historische Bildmaterial anhand von Text- und Bildabfragen durchsuchbar zu machen, wird ein *Contrastive Language-Image Pre-Training Modell* (*CLIP*) mittels Transfer Learning eingesetzt und feinabgestimmt (Radford et al., 2021).

Das optimierte Modell erreicht bei den von uns abgefragten Inhalten stark unterschiedliche Genauigkeiten und verbessert die Präzision des vortrainierten Modells nur teilweise. Während die Abfragen nach drei der Hauptklassen – „Tier“, „Pflanze“ und „Landschaft“ – unter den ersten 1.000 Ergebnissen höhere Präzisionswerte als das zugrundeliegende Modell erzielt (62–95%), verschlechtert das optimierte Modell die Erkennung der Hauptklasse „Landkarte“ (72%). Zudem kommen wir bei einzelnen Tierspezies, welche wir nebst den Hauptklassen ebenfalls annotiert haben, zu einem ähnlich heterogenen Ergebnis (vgl. Vignoli et al., 2024). Dies lässt sich zum einen auf die im für das *Transfer Learning* verwendeten Datensatz stark variierende Anzahl der Beispiele pro Klasse zurückführen (Liu et al., 2023), zum anderen auf die Bildqualität sowie die Platzierung und Prominenz der zu erkennenden Klassen im Bild (Vignoli et al., 2024).

Beim Einsatz der maschinellen Textanalyse, die im Zentrum dieses Vortrags steht, sind mehrere methodische wie inhaltliche Herausforderungen hervorzuheben. Zum einen betrifft das die Qualität der von uns verwendeten Volltexte, zum anderen den sprachlichen Zuschnitt der Algorithmen. Die automatische Textanalyse des ONiT-Projekts basiert größtenteils auf zum Teil fehlerhaften Volltexten, die mit *Optical Character Recognition* (OCR)-Verfahren im

Rahmen des Austrian-Books-Online-Projekts (ABO), einer Kooperation der Österreichischen Nationalbibliothek mit Google Books, erstellt wurden (Fritze und Krickl, 2020). Unsere bisherigen Experimente zeigen, dass die zugegebenermaßen ausbaufähige Qualität der vorliegenden Daten unsere Analysen zwar beeinträchtigt, jedoch nicht negiert.

Dennoch verfolgen wir einen *Information Retrieval* Ansatz, der sich aktueller LLMs bedient, um die Qualität der Rohdaten zu verbessern und anzureichern (vgl. Singh et al. 2024). Wir experimentieren für die sogenannte Vorverarbeitung (*Preprocessing*) derzeit mit einem open source LLM (LlaMA3.1-70B³), um die maschinenlesbaren Volltexte automatisiert zu bearbeiten. Dadurch sollen insbesondere diejenigen durch den OCR-Prozess entstandenen Artefakte und Fehlerkennungen behoben werden, die sich nicht mit regelbasierter Datenbereinigung (*Regular Expressions*) beheben lassen. (Abb. 1).

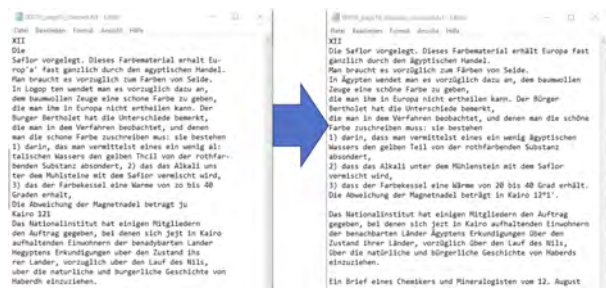


Abb. 1: Beispiel einer von der bereinigten OCR-Textversion aus Bergks Übersetzung von Sonninis *Reisen in Ober- und Niederägypten* (S. 16, links im Bild) generierten Korrektur von LlaMA3.1-70B anhand des Masterprompts „You are a historian expert in historical German texts. Correct the following faulty OCR texts generated from historical travelogues printed from the 17–19th century. [...]“ (rechts im Bild).

Reisen in Ober- und Niederägypten (S. 16, links im Bild) generierten Korrektur von LlaMA3.1-70B anhand des Masterprompts „You are a historian expert in historical German texts. Correct the following faulty OCR texts generated from historical travelogues printed from the 17–19th century. [...]“ (rechts im Bild).

Für die folgende Indizierung werden die vorhandenen Volltexte in Teilstücke geteilt, mit einem *Sentence Encoder* (*Flax Sentence Embedding*⁴) vektorisiert und in eine Vektor-Datenbank (*Marqo*⁵) geladen. Die einzelnen Buchseiten werden eingelesen und automatisch segmentiert, die dann jeweils vektorisiert und für die Ähnlichkeitssuche indiziert werden. Mit diesem Ansatz sollen die Texte – ähnlich wie die Bilder – durch Textabfragen semantisch durchsuchbar werden. Im Gegensatz zum von uns für die Bildanalyse angewandten *Transfer Learning*, ist es in diesem *Zero-Shot* Ansatz nicht notwendig, zuvor einen Trainingsdatensatz mit viel Ressourcenaufwand zu erstellen. Zudem ermöglicht uns dieser Ansatz, Möglichkeiten zur Erschließung von Themenzusammenhängen und Bedeutungskontexten innerhalb unseres nicht zuletzt aufgrund der unterschiedlichen Sprachen stark heterogenen Textkorpus mittels moderner LLM-Technologie zu evaluieren.

Da die meisten verfügbaren Algorithmen auf modernes Englisch zugeschnitten sind, ist die Verarbeitung anderer Sprachen (insbesondere nicht standardisierter historischer Sprachen) herausfordernder (z.B. Ehrmanntraut 2024). Mögliche Lösungsansätze für die übrigen Sprachen in unserem Korpus (d.h. Französisch und Latein) planen wir exemplarisch zu erkunden. In diesem Zusammenhang erproben wir das Potenzial aktueller Large Language Models wie LlaMA3, die immer besser mit (modernen) mehrsprachigen Quellen, etwa auf Deutsch und Französisch, umgehen können (Dhakal, 2024).

In einem ersten Schritt wenden wir die Methode auf ausgewählte Referenzberichte an. Es handelt sich um solche Reiseberichte, die aufgrund ihrer zeitgenössischen Wirkmacht ausgewählt wurden, da sie in besonders vielen Auflagen und Übersetzungen in den von uns anvisierten Sprachen nachweisbar sind. Hierzu zählt Charles Sigisbert Sonninis (1751–1812) *Voyage dans la Haute et Basse Égypte* (Reisen in Ober- und Niederägypten). Ein mehrbändiges Werk, das zunächst auf Französisch erschien (1798/99), und bald darauf von Henry Hunter (1741–1802) ins Englische (1799) und von Johann Adam Bergk (1769–1834) ins Deutsche (1800) übersetzt wurde. Für die unten zusammengefasste Analyse beschränken wir uns auf den ersten Band.

Die Volltexte der ausgewählten Berichte werden in einem iterativen Prozess anhand eigens durch Expert:innen im ONiT-Team generierter Richtlinien manuell annotiert und überprüft. Hier liegt der Fokus auf jenen Textpassagen, die Flora, Fauna und Landkarten betreffen. Landschaften bleiben aufgrund ihrer semantischen Mehrdeutigkeit bei der Textanalyse außen vor. Die einzelnen Tierspezies erschließen wir hingegen auch hier in die Tiefe. In den Volltexten werden nicht einzelne Schlagwörter, sondern ganze Textpassagen annotiert (d.h. zumindest ein ganzer Satz, teilweise erstrecken sich solche Passagen aber auch über mehrere Seiten). Ausnahmen sind Paratexte wie Inhaltsverzeichnisse, Überschriften und Druckmarginalien, in denen aufgrund ihrer semantischen Struktur auch einzelne oder wenige Wörter annotiert werden. Bei der Annotation operieren wir mit dem standardisierten Vokabular ICONCLASS, wobei bislang über fünfzig relevante Klassen zum Einsatz kamen. Die Anwendung von ICONCLASS in diesem Kontext ist eine innovative Leistung unseres Projekts, da das für Bilder entwickelte Vokabular unseres Wissens erstmals für Texte zum Einsatz gelangt.⁶

In einem zweiten Schritt werden wiederkehrende Konzepte und Terminologien lokalisiert und verwandtes Beispielmaterial extrahiert, um die eingesetzten maschinellen Lernmodelle zur Anreicherung und Identifikation weiterer Textpassagen zu validieren. Hierbei experimentieren wir mit verschiedenen natürlichsprachlichen Abfragen (*Prompt Engineering*), um die Fähigkeiten der vortrainierten Modelle anhand der einzelnen ausgewählten Reiseberichte zu testen und zu optimieren. Das Ziel ist, dieselben Modelle auf das übrige Korpus des ONiT-Projekts anzuwenden, um weitere Naturdarstellungen automatisiert zu erkennen.

Zwischenergebnisse

Erste Zwischenergebnisse unserer Experimente zeigen, dass das für *Sentence Encoding* verwendete Modell ohne *Transfer Learning* zwar in der Lage ist, anhand der nachbearbeiteten Texte relevante Ergebnisse zu liefern. Letztere stimmen jedoch nur zum Teil mit den von uns manuell identifizierten Textstellen überein. Für eine erste Validierung der gewählten Methode führen wir zwei Varianten derselben Abfrage durch: erstens mit den mit den regelbasiert-bereinigten OCR Texten sowie zweitens mit der mittels LLaMA3 generierten Korrekturfassung der OCR Texte. Für die Hauptklassen „Tier“ und „Pflanze“, welche wir in unserem annotierten Sonnini-Text jeweils 347 mal auf 215 Seiten (= 1,61 Annotationen pro relevanter Seite), respektive 186 mal auf 148 Seiten (= 1,26 Annotationen pro relevanter Seite) identifizierten, fördert der vortrainierte *Retriever* anhand der bereinigten Texte unter den ersten zweihundert gereihten Ergebnissen nur etwa 25% der Seiten mit den von uns mit „Tier“ annotierten Textpassagen (insgesamt 66 Vektoren auf 53 Seiten, welche 1,25 Vektoren pro relevanter Seite entsprechen) und 32% der mit „Pflanze“ annotierten Seiten zutage (49 Vektoren auf 47 Seiten, = 1,04 Vektoren pro relevanter Seite).

Ernüchternderweise lassen die Ergebnisse mit der LLM-korrigierten Fassung quantitativ keine Verbesserungen für die Erfolgsrate erkennen, da die Ergebnisse für beide Klassen mit jeweils 22% und 29% unter der Erfolgsrate mit den regelbasiert-bereinigten Texten liegen.⁷ Die äußerst niedrige Erfolgsrate bei der Suche nach Tieren und Pflanzen im Allgemeinen lässt sich darauf zurückführen, dass die Ergebnisse der hybriden Suche eine starke Tendenz haben, diejenigen Sätze weiter vorne zu reihen, die ebenfalls die gesuchten Stichwörter „Tier, Tiere“ bzw. „Pflanze, Pflanzen“ enthalten. Da in unserem annotierten Datensatz jedoch alle Nennungen von einzelnen Tier- und Pflanzenspezies mit den jeweiligen Hauptklassen „Tiere“ und „Pflanzen“ annotiert sind, fallen die Resultate dieser generellen Suchanfragen einseitig aus.

Bessere Ergebnisse konnten die von uns näher analysierten Tierklassen „Vogel“ (91 Annotationen auf 72 Seiten) und „Pferd“ (48 Annotationen auf 40 Seiten) erzielen. Bei diesen spezifischeren Suchanfragen vermag der *Retriever* in den regelbasiert-korrigierten Texten 56% der Seiten mit den von uns annotierten „Vögeln“ und 67% der Seiten mit „Pferden“ zu identifizieren. Jedoch verbessert auch hier die LLM-korrigierte Textfassung die Leistung des *Retrievers*, der lediglich 44% (Vögel) resp. 60% (Pferde) der von uns annotierten Seiten identifiziert, nicht. Obschon die mit spezifischen Tiernennungen erzielten Ergebnisse deutlich besser sind als allgemeine Suchanfragen mit demselben Modell, scheint der verwendete *Sentence Encoder* für die Textsprache nicht optimal geeignet zu sein, sofern er nicht nachtrainiert wird. Daher planen wir weitere Experimente mit anderen Sprachmodellen zur Vektorisierung der Texte.

Diskussion

Unsere bisherigen Ergebnisse zeigen, dass moderne LLMs zumindest teilweise in der Lage sind, die durch das OCR-Verfahren eingeführten „Übersetzungsfehler“ vom digitalen Schriftbild in maschinenlesbare Schriftzeichen zu korrigieren. Jedoch deutet unsere vorläufige Ergebnisevaluation darauf hin, dass dies die Auffindbarkeit der Inhalte nicht verbessert. Es macht sich bemerkbar, dass die Modelle auf „moderne“ Sprachen optimiert sind, denn es besteht die Tendenz, das frühneuzeitliche Vokabular durch ein aktuell gebräuchliches zu ersetzen (Abb. 2). Dies kann zwar durch geeignete Anweisungen (*Prompts*) an das LLM beeinflusst werden, verändert jedoch mitunter die Orthografie und lexikale Elemente, was sich teils auf die semantische Bedeutungsebene auswirkt. Dazu zählen auch von den Modellen generierte Fehler und Ungenauigkeiten, die ohne menschlichen Eingriff unbemerkt bleiben (vgl. Singh et al. 2024). Die schlechteren Ergebnisse bei der Suche anhand der LLM-korrigierten Texte lassen sich unter anderem auf solche neu generierten Fehler zurückführen. Eine genauere qualitative Analyse und Validierung der Ergebnisse stehen diesbezüglich jedoch noch aus.

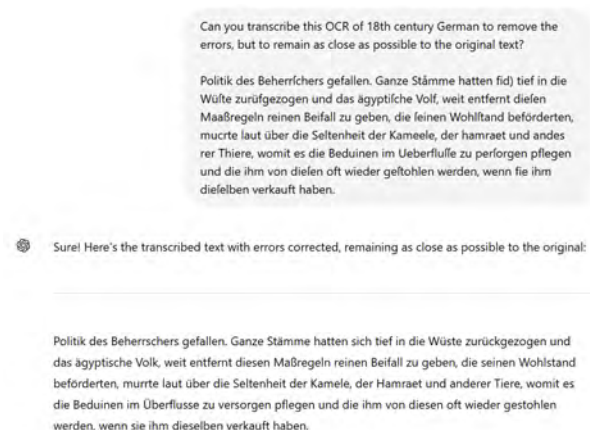


Abb. 2: Screenshot des Prompts mit der fehlerhaften OCR eines Textbeispiels aus Bergks Übersetzung von Sonninis. *Reisen in Ober- und Niederägypten* (Teil 2), S. 14, und die resultierende Transkription von ChatGPT3 (Gratisversion).

Eine weitere Einschränkung hängt mit der Art und Weise zusammen, wie die Resultate von komplexen statistischen Modellen erschlossen werden und wie die dadurch neu entstehende Heuristik und Ordnung der maschinell ausgewerteten Informationen die weitere Forschungsarbeit prägen. Transformer-Modelle wandeln die Merkmale der zu analysierenden Daten in Vektoren um, basierend auf deren Distanz ihre Ähnlichkeit mathematisch berechnet wird. Die dadurch entstehende Ordnung hängt davon ab, wie stark die erschlossene Ähnlichkeit der vektorisierten Merkmale mit unseren vordefinierten Klassen bzw. Sucheingaben übereinstimmen. Unsere Ergebnisse zu den Hauptklassen „Tier“ und „Pflanze“ verdeutlichen, dass der von uns gewählte Ansatz die Resultate dahingehend verzerrt, dass die Textvektoren, in denen der gesuchte Begriff wörtlich enthalten ist, zu einer höheren Übereinstimmung mit den Suchkriterien führt. Die Konsequenz ist, dass Nennungen spezifi-

scher Tierklassen wie „Pferd“ und „Vogel“ bei einer generellen Suche nach „Tier, Tieren“ tendenziell verloren gehen.

Dennoch bieten ML-Modelle eine erhebliche Hilfe zur Erstellung einer ersten Ordnung nach Relevanz von Bild- und Textquellen, was insbesondere zur Exploration großer, bislang nicht erschlossener Daten-Korpora dienlich ist. Besonders bei großen Datensätzen besteht jedoch weiterhin enormer manueller Aufwand, sowohl bei der Annotation von Trainings- bzw. Validierungsdaten als auch bei der Evaluierung und Qualitätskontrolle der erschlossenen Inhalte. Dies wird deutlich, wenn die von unserem Text- *Retriever* erschlossenen Inhalte genauer beleuchtet werden. Obwohl die erstgereihten Ergebnisse der 200 gefundenen Textstellen vielfach gut mit der Suchanfrage übereinstimmen, müssen besonders die weiter hinten gereihten Resultate manuell überprüft und aussortiert werden. Für die hier präsentierte Suche in einem Werk mit Annotationen in der Größenordnung von jeweils 50–350 Textbeispielen pro Klasse bedeutet dies, dass in allen Fällen mehr als die Hälfte der erschlossenen Textvektoren für die Suche irrelevante Ergebnisse liefern und hierbei nicht alle annotierten Beispiele finden.

Zum Abschluss möchten wir noch ein paar Überlegungen für die maschinell gestützte Suche nach intermedialen Inhalten anführen. Multimodale Modelle wie das von uns für die Bildanalyse verwendete CLIP-Modell ermöglichen die Suche nach Ähnlichkeiten über textuelle und piktoriale Grenzen hinweg (vgl. Vignoli et al., 2024). Durch die Integration eines solchen Modells in die von uns für die Textanalyse eingesetzten Vektordatenbank, ließen sich auch intermediale Zusammenhänge und wiederkehrenden Verbindungen zwischen den Bildern und Texten unseres Korpus maschinell erkunden. Inwieweit die dergestalt maschinell erschlossenen Zusammenhänge mit den für historische Fragestellungen relevanten Aspekten übereinstimmen, etwa ob bestimmte Tiere in den Texten oder Bildern prominenter behandelt werden, gilt es noch zu erkunden. Ein wichtiger Aspekt hierbei sind Verzerrungen (*Bias*), welche die Modelle aufgrund ihrer Trainingsdaten einbringen. Aktuelle Transformer-Modelle, die auch in unserem Projekt zur Anwendung kommen, wurden ursprünglich anhand Milliarden moderner, aus dem Internet stammender Daten trainiert (z.B. Smits und Wevers, 2023; Shtedritski et al., 2023: 11994; Birhane et al., 2021). Entsprechend sind die anhand dieser Modelle erzielten Ergebnisse – sowohl bei *Information Retrieval* als auch bei von LLMs generierten Inhalten – geprägt von den in den Trainingsdaten widerspiegelten Weltbildern, Semantik und Symbolik, und eben nicht den historischen Gegebenheiten zur Zeit der Entstehung der von uns untersuchten Reiseberichten (vgl. Mozilla Foundation 2022).

Conclusio und Ausblick

Das Team des ONiT-Projekts ist überzeugt, dass der Einsatz von KI in den Geisteswissenschaften nicht aufzuhalten ist und in vielen Bereichen neue, bislang ungeahnte Möglichkeiten eröffnet. Unsere Experimente zeigen aber

auch, dass das große Versprechen der Revolution der Geisteswissenschaften kritisch zu betrachten ist und der Einsatz von KI immer im Bewusstsein der Möglichkeiten und Grenzen der Methodik erfolgen muss. Die Algorithmen gelangen derzeit nur mit signifikantem Ressourceneinsatz – der mitunter das übersteigt, was eine klassische geisteswissenschaftliche Untersuchung desselben Quellenmaterials bedürfte – in die Nähe eines Vollständigkeitsanspruchs, den es in allen Wissenschaften zu bestreben gilt. Die Algorithmen operieren mit Wahrscheinlichkeiten, die durch Feinabstimmung (*Finetuning*) mittels manuell kreierter, *Ground Truth* teils zwar eine hohe Genauigkeit erreichen können. In den Fällen, wo dies etwa aufgrund von Daten- oder Ressourcenmangel nicht gelingt, wird nur ein Teil des anvisierten Quellenmaterials automatisiert erschlossen. Folglich verzerrt dies die Ergebnisse der Untersuchungen und die historiographische Repräsentation der Vergangenheit. Auch präzise Modelle sind weit davon entfernt, alle Vorkommnisse bestimmter Inhalte zu erschließen. Die Notwendigkeit manueller Arbeit zur Überprüfung der Ergebnisse bleibt vorhanden, insbesondere um falsche Positive auszusortieren und falsche Negative aufzuspüren. Welche Informationen nicht aufgefunden werden, bleibt vielfach noch ein weißer Fleck auf der Landkarte.

Ähnlich wie viele Reisende der Frühen Neuzeit sehen wir solche weißen Flecken aber nicht als Hindernis weiterer Explorationen, sondern – im Gegenteil – als ihr Ansporn. Was und warum die Algorithmen bislang nicht auffinden können, gilt es genauso zu reflektieren wie Maßnahmen, die wir ergreifen können, um einem blinden Vertrauen in KI-Lösungen entgegenzuwirken. Je mehr hochqualitative, standardisierte Daten der Forschung frei zur Verfügung stehen, desto schneller und besser können wir die Algorithmen trainieren. Es bedarf an mehr Expertise in den Geisteswissenschaften, um solche Standardisierungen über fachliche, sprachliche und geographische Grenzen hinweg durchzusetzen, sowie einen reflektiert-kritisch und zielgerichteten Einsatz der Methodik, um einen langfristigen und nachhaltigen Umgang mit KI in der geisteswissenschaftlichen Forschung und darüber hinaus zu garantieren.

Fußnoten

1. <https://openai.com/chatgpt/>.
2. <https://onit.oeaw.ac.at/>.
3. <https://llama.meta.com/llama3/>.
4. https://huggingface.co/flax-sentence-embeddings/all_datasets_v4_mpnet-base.
5. <https://www.marqo.ai/>.
6. Eine Übersicht von Projekten, die das ICONCLASS Vokabular einsetzen, findet sich hier: <https://iconclass.org/help/aboutc>.
7. Die Validierung dieser vorläufigen Ergebnisse steht noch aus und wird für die Präsentation an der Konferenz nachgereicht.

Bibliographie

- Birhane, Abeba, Vinay Uday Prabhu und Emmanuel Kahembwe.** 2021. "Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes." *arXiv*, 5. Oktober. doi:10.48550/arXiv.2110.01963.
- Cetinic, Eva, Tomislav Lipic und Sonja Grgic.** 2019. "A Deep Learning Perspective on Beauty, Sentiment, and Remembrance of Art." *IEEE Access* 7: 73694–73710. doi:10.1109/ACCESS.2019.2921101.
- Dhakal, Shreeya.** 2024. *Exploring Multilingual Aspects and Vocabulary of LLaMA 3 Compared to LLaMA 2*. In: Icodeformybbhasa, 22. April. <https://www.icodeformybbhasa.com/p/exploring-multilingual-aspects-and> (zugegriffen: 02. Dezember 2024).
- Ehrmanntraut, Anton.** 2024. "Historical German Text Normalization Using Type- and Token-Based Language Modeling." *arXiv*, 4. September. doi:10.48550/arXiv.2409.02841.
- Fritze, Christiane und Martin Krickl.** 2020. "Austrian Books Online – Acht Jahre Digitalisierung des historischen Buchbestandes der Österreichischen Nationalbibliothek mit Google." *Bibliothek Forschung und Praxis* 44,1: 89–99. doi:10.1515/bfp-2020-0008.
- Im, Chanjong, Yongho Kim und Thomas Mandl.** 2022. "Deep Learning for Historical Books: Classification of Printing Technology for Digitized Images." *Multimedia Tools and Applications* 81,4: 5867–5888. doi:10.1007/s11042-021-11754-7.
- Liu, Yang, Guoping Yang, Shaojie Qiao, Meiqi Liu, Lulu Qu, Nan Han et al.** 2023. "Imbalanced Data Classification: Using Transfer Learning and Active Sampling." *Engineering Applications of Artificial Intelligence* 117, Part B (2023): Article 105621 doi: 10.1016/j.engappai.2022.105621
- Moretti, Franco.** 2013. *Distant Reading*. London/New York: Verso.
- Mozilla Foundation.** 2000. *Internet Health Report 2022. Who Has Power Over AI?* In: Internet Health Report 2022, <https://2022.internethealthreport.org/facts/> (zugegriffen: 02. Dezember 2024).
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al.** 2021. "Learning Transferable Visual Models From Natural Language Supervision." In *Proceedings of Machine Learning Research (PMLR)* 139:8748–8763.
- Saleh, Babak und Ahmed Elgammal.** 2016. "Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature." *International Journal for Digital Art History* 2. doi:10.11588/DAH.2016.2.23376.
- Shtedritski, Aleksandar, Christian Rupprecht und Andrea Vedaldi.** 2023. "What Does CLIP Know about a Red Circle? Visual Prompt Engineering for VLMs." In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 11987–11997 doi:10.48550/arXiv.2304.06712.
- Singh, Ayush, Navpreet Singh und Shubham Vatsal.** 2024. "Robustness of LLMs to Perturbations in Text." *arXiv*, 12. Juli. doi:10.48550/arXiv.2407.08989
- 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. doi:10.1093/llc/fqad008.
- Sonnini, Charles Sigisbert.** 1798/99. *Voyage Dans La Haute Et Basse Égypte, Fait Par Ordre De L'Ancien Gouvernement, Et Contenant Des Observations De Tous Genres: Avec une Collection de 40 Planches, gravées en taille-douce par J. B. P. Tardieu, contenant des Portraits, Vues, Plans, Carte Géographique, Antiquités, Plantes, Animaux, etc. dessinées sur les lieux, sous les yeux de l'Auteur*, 3 Bde., Paris: F. Buisson.
- Sonnini, Charles Sigisbert.** 1799. *Travels in Upper and Lower Egypt: Undertaken by Order of the Old Government of France; by C. S. Sonnini, Engineer in the French Navy, and Member of Several Scientific and Literary Societies. Illustrated with Forty Engravings; Consisting of Portraits, Views, Plans, a Geographical Chart, Antiquities, Plants, Animals, &c. Drawn on the Spot, under the Author's Inspection*. [...] 3 Bde., aus dem Französischen von Henry Hunter, London: John Stockdale und T. Gillet.
- Sonnini, Charles Sigisbert.** 1800. *C. S. Sonnini's, ehemaligen Offiziers und Ingenieurs des französischen Seewesens und Mitgliedes mehrerer gelehrten und litterarischen Gesellschaften, Reisen in Ober- und Niederägypten, auf Befehl der ehemaligen Regierung in Frankreich unternommen. Nebst einem Auszuge aus des Bürger's Carl Norry, Mitgliedes der philotechnischen Gesellschaft, Bemerkungen über Aegypten [...]*, 2 Bde., aus dem Französischen von Johann Adam Bergk, Leipzig und Gera: Wilhelm Heinsius.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser und Illia Polosukhin.** 2017. "Attention is all you need." In *Proceedings of the 31st International Conference on Neural Information Processing Systems (= NIPS'17)*, 6000–6010.
- Vignoli, Michela, Doris Gruber und Michael Seidl.** 2024. "Revolution or Evolution? AI-Driven Retrieval of Nature Representations in Historical Prints." In *Proceedings of the DH2023 Conference (= DSH Special Issue)*, in Begutachtung.
- Vignoli, Michela, Doris Gruber und Rainer Simon.** 2023a. "Revolution or Evolution? AI-Driven Image Classification of Historical Prints." In *Digital Humanities 2023: Book of Abstracts*, hg. von Anne Baillot, Toma Tasovac, Warter Scholger, und Georg Vogeler, 184–185. Graz: Universität Graz. doi:10.5281/zenodo.7961822.
- Vignoli, Michela, Doris Gruber, Rainer Simon und Axel Weissenfeld.** 2023b. "Impact of AI: Gamechanger for Image Classification in Historical Research?" In *Digitale Methoden in der geschichtswissenschaftlichen Praxis:*

Fachliche Transformationen und ihre epistemologischen Konsequenzen: Konferenzbeiträge der Digital History 2023, Berlin, 23.–26.5.2023, hg. von Melanie Althage, Martin Dröge, Torsten Hiltmann, und Claudia Prinz. doi: 10.5281/zenodo.8322398.

Yang, Yixin, Zheng Li, Qingxiu Dong, Heming Xia und Zhifang Sui. 2024. "Can Large Multimodal Models Uncover Deep Semantics Behind Images?" *arXiv*, 20. Juni. doi:10.48550/arXiv.2402.11281.