

Fachartikel aus:
Zeitschrift für digitale Geisteswissenschaften, Heft 10 (2025)

Titel:
When it was 2024 – Generative AI in the field of digital scholarly editions

Autor*in:
Christopher Pollin

Kontakt: christopher.pollin@uni-graz.at
Institution: Institut für Digitale Geisteswissenschaften, Universität Graz | Institut für Dokumentologie und Editorik | Digital Humanities Craft
GND: [1267937505](#) ORCID: [0000-0002-4879-129X](#)
Contribution (CRediT): [Writing - original draft](#)

Autor*in:
Franz Fischer

Kontakt: franz.fischer@unive.it
Institution: Ca' Foscari Università Venezia | Institut für Dokumentologie und Editorik
GND: [1140931482](#) ORCID: [0000-0002-2162-5531](#)
Contribution (CRediT): [Writing - original draft](#)

Autor*in:
Patrick Sahle

Kontakt: sahle@uni-wuppertal.de
Institution: Bergische Universität Wuppertal | Institut für Dokumentologie und Editorik
GND: [1034443909](#) ORCID: [0000-0002-8648-2033](#)
Contribution (CRediT): [Writing - original draft](#)

Autor*in:
Martina Scholger

Kontakt: martina.scholger@uni-graz.at
Institution: Institut für Digitale Geisteswissenschaften, Universität Graz | Institut für Dokumentologie und Editorik
GND: [121749541X](#) ORCID: [0000-0003-1438-3236](#)
Contribution (CRediT): [Writing - original draft](#)

Autor*in:
Georg Vogeler

Kontakt: georg.vogeler@uni-graz.at
Institution: Institut für Digitale Geisteswissenschaften, Universität Graz | Institut für Dokumentologie und Editorik
GND: [1035297884](#) ORCID: [0000-0002-1726-1712](#)
Contribution (CRediT): [Writing - original draft](#)

DOI des Beitrags:
[10.17175/2025_008](#)

Nachweis im OPAC der Herzog August Bibliothek:
[1923105256](#)

Erstveröffentlichung:
10.07.2025

Lizenz:
Sofern nicht anders angegeben 

Letzte Überprüfung aller Verweise:
15.04.2025

Format:
PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:
[Digitale Edition](#) | [Experiment](#) | [Künstliche Intelligenz](#) | [Optische Zeichenerkennung](#) | [Transkription](#)

Empfohlene Zitierweise:

Christopher Pollin / Franz Fischer / Patrick Sahle / Martina Scholger / Georg Vogeler: When it was 2024 – Generative AI in the field of digital scholarly editions. In: Zeitschrift für digitale Geisteswissenschaften 10 (2025). 10.07.2025. HTML / XML / PDF. DOI: [10.17175/2025_008](https://doi.org/10.17175/2025_008)

Christopher Pollin / Franz Fischer / Patrick Sahle / Martina Scholger / Georg Vogeler

When it was 2024 – Generative AI in the field of digital scholarly editions

Abstract

This contribution examines the current state of research on generative AI applications in digital scholarly editing. Drawing from experiments presented at the DHd 2024 workshop and additional literature, it identifies eight key application areas for Large Language Models (LLMs): (1) documentation of textual transmission, (2) post-processing of retro-digitized editions, (3) text establishment (transcription, layout analysis, OCR/HTR post-processing, markup), (4) normalization, (5) named entity recognition (NER) and other semantic annotation, (6) information enrichment, (7) translation, and (8) summarization. While among these areas, NER garnered the most experimental attention at the workshop, a comprehensive architecture for integrating generative AI across the full editorial stack was proposed. The paper concludes by identifying critical areas for future research: from a practical perspective, the field needs standardized workflow orchestration and evaluation protocols; from a theoretical perspective, researchers must systematically assess the strengths and weaknesses of LLMs in digital scholarly editions while addressing their inherent biases and ethical implications.

Der Beitrag reflektiert den aktuellen Forschungsstand zur Anwendung generativer KI im Bereich der digitalen wissenschaftlichen Edition. Er basiert auf Experimenten, die 2024 bei einem DHd-Workshop präsentiert und diskutiert wurden, und bezieht weitere Publikationen zum Thema ein. Die Untersuchung identifiziert acht Anwendungsfelder generativer KI, vorwiegend in Form von Large Language Models (LLMs), in der wissenschaftlichen Editorik: (1.) Dokumentation der Textüberlieferung, (2.) Nachbearbeitung retrodigitalisierter Editionen, (3.) Textgewinnung und -codierung (Transkription, Dokument-Layout-Analyse, OCR / HTR-Nachbearbeitung, grundlegendes Markup), (4.) Normalisierung, (5.) Erkennung von Named Entities (NER) und tiefere semantische Annotation, (6.) Informationsanreicherung, (7.) Übersetzung und (8.) Zusammenfassung. In den durchgeführten Experimenten wurde der Einsatz generativer KI im gesamten Editionsprozess konzeptualisiert, auch wenn sich NER als Bereich mit den meisten Beiträgen erwies. Im abschließenden Teil des Beitrags werden Erfordernisse an die weitere Forschung benannt: Aus praktischer Perspektive müssen standardisierte Workflow-Orchestrierungen und Evaluationsprotokolle entwickelt werden. Aus theoretischer Sicht bedarf es systematischer Untersuchungen möglicher Stärken und Schwächen von LLMs bei deren Verwendung in digitalen Editionen sowie einer kritischen Reflexion über inhärente Schieflagen und ethische Fragen.

1. Introduction

New technologies are leading to fundamental upheavals both in scientific practices and their epistemological foundations. Where else could this be a greater cliché than in the digital humanities? After machine learning and artificial intelligence have stood on the horizon for many years as ›the next big thing‹, and machine learning methods have been used in various very special fields of application in the humanities such as scholarly editing,¹ the revolution was finally proclaimed for the wider audience. With the publication of ChatGPT by OpenAI in November 2022, generative AI based on large language models (LLM) became so accessible and versatile that ›everyone‹ sooner or later started experimenting with it and developing visions for possible applications. Further developments from GPT-3.5 to GPT-4 and other foundation models (such as Claude, Bard / Gemini, Llama, Mistral) quickly followed, and applications (BingChat, NotebookLM, Perplexity etc.) were added in steadily growing competition. [1]

In response to these developments, scholarship in general and the digital humanities in particular are fundamentally called upon to examine the applicability of new technologies, to adopt and further develop them as tools for their own purposes. Applications of generative AI naturally include scholarly editing, a fundamental practice in humanities research. The Institute for Documentology and Scholarly Editing (IDE) has been stimulating discussions on digital editing through various initiatives since its establishment in 2006. Therefore, the IDE set out to investigate the applicability of generative AI in the field and to critically reflect on its potential for supporting editorial practices and processes at the intersection of literary studies, archival research, and textual criticism. To this end, a workshop was organized at the Passau 2024 annual conference [2]

¹ E.g. Henny-Krahmer et al. 2022.

of the Association for Digital Humanities in the German-speaking countries (DHd), which was intended to further structure the field via a call for experiments and bring together the actors who have already dealt with the possibilities and challenges of AI-based tools in digital editions at this early stage.² Specific use cases that move along the editorial workflow confronting LLMs with editorial materials and tasks were presented and discussed. Experimental tests for a set of given scenarios were examined with regard to their potential, limits, and shortcomings, as well as their ethical and theoretical implications.

This article³ draws on a workshop report to document tentative initial attempts – likely to have a limited lifespan – aimed at clarifying key aspects of applying generative AI to scholarly editing. The main aim is to continue the synthesis work that has already begun before the workshop with, for instance, first experiments in summer 2023 by Elisa Beshero-Bondar on performing text collation⁴ or the XML annotation of unstructured text by Christopher Pollin, Christian Steiner, and Constantin Zach.⁵ The questions developing from these experiences are: What are the areas of application for generative AI in the context of digital scholarly editing? How can it be used as efficiently as possible? What best practices are emerging? Where does it make sense to actively strive for further development? Where do the evolving practices converge in a common methodology? How do we address bias and ethical risks in AI systems – from the initial training data, through reinforcement learning by human feedback (RLHF), to the fine-tuning stages of the training data? How can the results be evaluated? [3]

In addition, an immediate, contemporary historiography of scientific development is particularly important in this area. In moments when new technologies arise, the contrast to previous technical possibilities is very sharp. These periods offer a high potential for fresh thinking, conceptual innovation, and critical reflection. They are the most dynamic times of upheaval, whose subsequent reflections provide a deeper insight into the shape and development of scientific practices. This is particularly evident in the numerous contributions at events focusing on LLMs and generative AI in the digital humanities, where the topic is omnipresent. The interest comes from all directions, including discussions on methodologies, legal aspects, ethics, pedagogy, editing, text and image processing, analysis, and more. An immediate documentation helps to prevent the oblivion of the discussions held and alternative paths of development that were still open at the time – a risk that is always inherent to the implementation of practices that eventually prevail. With this article, we want to anticipate the past of the future and write the history of a transitioning scholarship ›in real time‹ – when it was 2024. [4]

2. Basics: LLMs and their Application

AI-based applications such as OpenAI's Generative Pre-trained Transformer (GPT) have demonstrated transformative capabilities in text processing. Their strength lies in their capability to be used in dialogue systems (chatbot), natural language text generation in general, the inclusion of both local context (specific to the current interaction or document) and ›world knowledge‹ (as patterns learned by the Transformer architecture from vast amounts of training data), multilingual support, and the possibility of fine-tuning to specific tasks.⁶ They can assist humans in (as of 2024) developing coded algorithms for processing information, though their output requires verification and they serve as aids rather than independent developers of algorithmic solutions. Related to scholarly editing, they can be applied in many activities, from the recording and digitization of the historical tradition to the creation of deeply indexed, critically annotated forms of representation. Despite the diversity of subjects and approaches, we do have a reasonably [5]

² Pollin et al. 2023a.

³ All authors contributed equally to the text. We would like to thank all contributors for their experiments and feedback on the respective paragraphs of this article. English proof reading was supported by Thea Schaaf. Chiara Citro provided assistance in formatting the bibliography.

⁴ Beshero-Bondar 2023.

⁵ Pollin et al. 2023b.

⁶ Zhu-Tian et al. 2023, p. 1–2.

consensual theoretical, methodological and technical common ground for digital scholarly editing. Due to the lively exchange on editions in recent years, the field has a good foundation for exploring the possible uses and effects of the new technology. The IDE has been actively engaged in shaping this exchange through the structured, reflective, and quality-assuring discourse of criticism and evaluation of editions as editorial research projects, most visibly through the establishment of the review journal **RIDE** (since 2014) and the catalog of criteria for evaluation of digital editions on which it is based.⁷

Christopher Pollin's introduction to the Workshop⁸ established a common ground and clarity about the situation we are facing. After machine learning based methods of text generation had shown promise with OpenAI's GPT-2 and GPT-3.5 models, it was the release of GPT-4 in March 2023 that marked the true breakthrough. While GPT-3.5 had already attracted widespread attention from both scholars and laypeople, GPT-4 demonstrated unprecedented capabilities across a broad range of tasks. Though initial results often fell short of expectations, leading some to quickly abandon their work with GPT, others became fascinated and driven to understand how to achieve better results through experimentation and refinement. This is precisely where the workshop comes in, emphasizing the importance of prompt engineering, which has now almost developed into a discipline in its own right. Prompt Engineering is the process of designing and optimizing prompts to effectively communicate with an LLM. This is an iterative process, in which prompts are adjusted based on the model's output. Crafting precise prompts improves the accuracy and controls bias of the response. Therefore, knowledge of the behavior and limitations as well as the difference in prompting techniques of the model used is crucial. Examples of key prompting strategies include (a) ›few-shot prompting‹, providing examples enabling in-context learning, (b) ›chain-of-thought‹, where the model is instructed to break down problems into step-by-step reasoning, (c) ›persona / expert prompting‹, where the model is assigned a specific role or expertise, and (d) ›reflection prompting‹, which encourages self-evaluation and refinement of responses. [6]

A study on the critical analysis of the correctness and improvement across various LLMs on the ATLAS and other datasets⁹ had shown that GPT-4 outperforms other LLMs in both correctness and quality. It improves the correctness of responses by 75-85 % and the quality of responses by 40-80 %. In fact, the generic GPT-4 outperforms even fine-tuned LLMs. Despite all surprising idiosyncrasies of some LLMs, research is establishing LLMs as best practice for many NLP (Natural Language Processing) tasks. [7]

⁷ Sahle 2014.

⁸ Pollin 2024. See the slide set [IDE-Workshop-Dhd24.pdf](#) in the package.

⁹ Nori et al. 2023; Bsharat et al. 2023.

3. Scholarly Editing as Processing Steps: Expectations ex ante

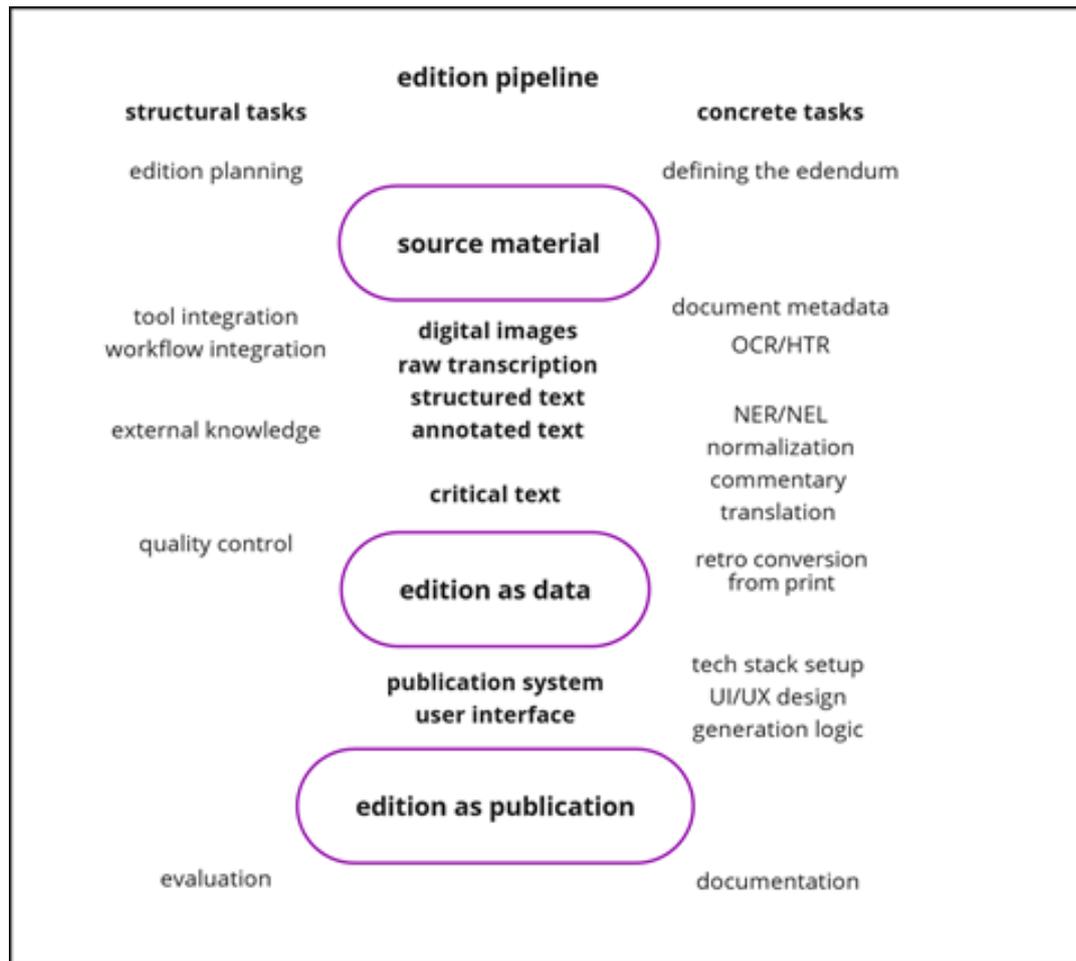


Fig. 1: General model of scholarly editing as knowledge production workflow: from source material to data and publication, through intermediate ›products‹ (middle), concrete steps (right), and more general challenges (left). [Pollin et al. 2025]

The transformative potential of LLMs makes their integration into scholarly editing practices highly desirable. [8] For which editorial use cases should we find the appropriate combination of generative AI based tools? We suggest describing scholarly editing as special scientific practice that includes recurring steps.¹⁰ Figure 1 illustrates one potential graphical representation of this practice. The graphic can be read from top to bottom. The middle column shows three main components framed in purple – source material, edition as data, and edition as publication – linked by intermediate processing steps. From the source material, the pipeline proceeds through digital images, raw transcription, structured and annotated text, leading to the critical text. All these outputs are part of the edition as data, which is fed into a publication system and perceived through the user interface. This creates the edition as a publication. The left and the right column break the pipeline of objects down to tasks. The left column gives generic labels to structural tasks, while the right column gives concrete tasks related to this: Defining the research object and research goals of the edition, i.e. the ›edendum‹,¹¹ is a major practice in planning the edition. Editors document the textual transmission by creating descriptive metadata of the sources. They apply automatic text recognition methods, recognize named entities, link them to identifiers, and normalize the transcription. All these tasks, executed by digital tools, can be integrated into one workflow. When scholars add commentaries, describe

¹⁰ Vogeler 2017; Sahle 2013.

¹¹ Fritze 2022.

named entities semantically, or translate texts, they incorporate knowledge external to the source material into the edition and enrich the information in the source. The conversion of printed editions into digital formats creates data that must meet quality standards as part of the edition's output.

This edition data is processed through a specific technology stack into a user interface, which should be implemented deliberately along with a user experience design. The published digital edition is reviewed through the external scholarly evaluation procedures and should include documentation as scholarly reflection and justification of its creation – and as support for re-use and long term preservation. These steps are certainly not the only way to conceptualize digital scholarly editing, but they help to integrate the concrete examples of application of generative AI to digital scholarly editing. None of the experiments presented later discuss edition planning, but several integrate the technology into the processing of source material into edition as data. [9]

The experiments presented at the workshop and described below can be related to the editorial steps and allow for a critical examination of generative AI's application. We need to identify the particular strengths and potentials of the new tools within the tension between the need for problem-solving in science and the specific affordances of these tools for practical application. This includes the conversion of unstructured text (e.g., transcriptions) into structured text (via structural and semantic markup), the conversion of structured text into explicit data structures, named entity recognition, normalization and enrichment, context-specific annotations, error management, and data control. We are particularly interested in three complementary aspects: first, the integration of AI into the editorial workflow, including planning, design, and evaluation of digital editions as well as code development for web applications and user interfaces; second, prompt engineering for effective model interaction through natural language; and third, AI Engineering – encompassing technical enhancements like retrieval-augmented generation (RAG), vector databases, and tool integration to extend the base capabilities of LLMs. [10]

4. Experiments

While figure 1 is just one possible rough map of the field of editorial practices, the case studies we received in response to our *Call for Experiments*¹² show where first applications of generative AI try to tackle specific editorial tasks. The case studies can be organized along the typical editorial workflow, though not always in strict sequence: [11]

- documentation of textual transmission,
- post-processing of retro-digitized editions,
- text establishment (transcription, Document Layout Analysis, OCR/HTR post-processing, basic markup),
- normalization,
- named entity recognition (NER) and deeper semantic annotation,
- information enrichment,
- translation, and
- summarization.

The case studies whose discussion now follows are documented in *Generative KI, LLMs und GPT bei digitalen Editionen*¹³ as single slide sets within the larger Zenodo package. [12]

¹² Pollin et al. 2023b.

¹³ Pollin et al. 2024.

Regarding the practice of *documentation of textual transmission*, the conversion of spreadsheets into TEI data appears to be an easily automatable task. However, Gerrit Brüning and Felix Schenke¹⁴ demonstrate for the **Goethe LYRIK** project how much human understanding is needed to handle the documentation of the transmission of all known poems by Goethe that has been previously organized in a tabular format. These spreadsheets contain basic information about manuscript witnesses as well as poems within these witnesses alternating in consecutive rows but in shared columns. With TEI / XML as target format, the Oxygen XML Editor was used with the AI Positron plugin, providing access to the paid OpenAI API. It helps to execute not only standard tasks like summary, correction, and translation, but also conversions described by examples. The application of the prompt on manageable chunks facilitates human supervision and the creation of data, including identifiers designed to interlink the multitude of data fragments, which is extremely close to the planned target format and requires only minor refinement. Even though the TEI requires rather detailed descriptions and larger inputs are less stable, Brüning and Schenke demonstrate how an integration of AI-based activities in established tools such as Oxygen can be effectively used to handle and transform messy table data whose structure is not fully captured by the columnar format. [13]

Generative models offer two promising capabilities for manuscript research: the transformation of human-written queries for use in structured datasets, and access to semantic searches through vector embeddings of manuscript descriptions and metadata. However, these applications are still in early stages of exploration for scholarly editing. The application of generative AI to discovery service – as, for instance, proposed by Xu¹⁵ – is under development in the library community, while the manuscript description community has not yet taken this up. [14]

Experiments in the testbed of *retro digitization of printed scholarly editions* show promising results regarding the evaluation and improvement of OCR output. In 2009, the University of St. Gall started to digitize the collection of Swiss Law Sources. The scanned books have been processed by OCR software and published in PDF on the **SSRQ Online** platform without correction or post-processing. At various points of the digitization workflow, Bastian Politycki¹⁶ tested the application of generative AI models (GPT-3.5/4) to create an integrated OCR post-processing pipeline.¹⁷ Politycki identifies five relevant areas: error correction of OCR results, structural TEI mark-up, annotation of dates and named entities, integration of data sets, and integration in the digital edition. [15]

Based on a training set of three volumes (two editions, one index), two different approaches have been compared. The first approach started from the PDF files, providing the respective edition texts and basic information about the layout and structure in order to train the correction of errors, the classification of segments as edition text, annotation and paratext, and the encoding of structural and semantic units. A zero-shot prompting approach has been refined by enriching the prompts through RAG and the provision of contextual knowledge.¹⁸ Still, the results remained problematic, especially regarding text classification due to the unstructured nature of the OCR text and its prevalent character as print pages obscuring the actual document structure. The second approach starts from an HTML export of the PDF instead of plain text, adding further input such as titles, page numbers, dates, and item numbers from the table of content in XML format as well as layout information regarding typography of the original print. The pipeline combines a series of Python scripts with LLM few-shot prompts. The results of each step are validated by a set of tests. This provides significantly higher accuracy regarding both the recognized text and the classification of its sections and semantic units. [16]

¹⁴ *Umwandlung von tabellarischen Daten in TEI-XML mithilfe des Oxygen AI Positron* (Pollin et al. 2024).

¹⁵ Xu 2024.

¹⁶ *Anwendung generativer KI zur Digitalisierung gedruckter Editionen* (Politycki 2024b).

¹⁷ Politycki 2024a.

¹⁸ Chae / Davidson 2023; Møller et al. 2023; Sun et al. 2023; Zhou et al. 2023.

For a long period, the label ›AI‹ was reserved in scholarly editing for the application of text transcription through *automatic text recognition*, particularly that associated with the activities on recognizing handwritten texts since 2015.¹⁹ Although neural networks were already introduced into the scholarly editing discourse through an explorative paper by Daniele Fusi,²⁰ handwritten text recognition (HTR) was the first successful application of this technology. The successful application of recurrent networks to optical text recognition by Enrique Vidal's team in the tranSkriptorium project (2013–2015) made large scale automatic transcription in ready-made tools available to every scholar. Generative AI applications in this field aim at specific subtasks, such as the creation of synthetic data for training purposes²¹ or the reconstruction of damaged texts.²² HTR is becoming a ›conservative‹ field, in which classical methods of deep learning (training and fine-tuning of existing models) dominate the research. Even if the recently discussed TrOCR methodology²³ is based on the same transformer architecture as the generative machine learning of large language models, it has not yet reached the same level of semantic richness and quality that enables advanced generative AI applications in text processing. However, HTR can be seen as a first example of multimodal enrichment of LLMs, combining the CLIP style prediction of text from images and vice versa.²⁴ The role of the implicit language models in the visual part of HTR are not yet studied in sufficient detail, although many pre-digital edition practices show the close connection between linguistic and paleographical interpretation made by editors when they ›normalize‹, for instance, punctuation or ›u‹ and ›v‹. The multimodal extensions of generic models and ›AI‹ services have not yet challenged the established HTR algorithms, but the developments in multimodal LLMs seem to be a starting point in doing so.²⁵ Transformer-based models play a well established role in post-OCR correction,²⁶ leading to recent tests with LLMs.²⁷ However, as demonstrated by Jacob Möhrke, Sandra Balck and Anna Ananieva in their workshop contribution²⁸, this approach carries a fundamental risk of ›over-normalizing‹,²⁹ discussed in detail below.

[17]

Normalization of texts transcribed from historical documents to improve readability, searchability, and not least, the application of NLP tools, is yet another important step of the editorial workflow. Over the past decade, several tools and approaches for the post-processing and automated error correction of OCR-generated text have been developed based on reference materials, word lists or linguistic tools³⁰. Two experiments tested the potential of LLM-based approaches to facilitate normalization of historical orthography. The first experiment, conducted by Yannic Bracke³¹, presented the newly developed *transnormer* (sic) application (available as [github repository](#)). This application follows the basic concept of using an LLM that was pre-trained on a massive dataset, fine-tuned on a domain-specific dataset – specifically, a manually evaluated subset of the *DTA corpus* containing 5 million tokens of historical texts from 1780–1901. Intermediate results show a good accuracy (98.93 % word accuracy), encouraging enough to train more specific models for smaller historical time periods.

[18]

The other experiment, presented by Kay-Michael Würzner and Robert Sachunsky³², attempted to address the needs of three different user profiles: the corpus linguist with no interest in historical spelling, hyphenation or line breaks; the textual scholar with an interest in historical spelling without hyphenation or line breaks;

[19]

¹⁹ Mühlberger et al. 2019.

²⁰ Fusi 2009.

²¹ Sousa Neto 2024.

²² Assael et al. 2022.

²³ Sommerfeld 2022; Ströbel et al. 2022.

²⁴ Radford et al. 2021.

²⁵ E.g. Liu et al. 2023.

²⁶ Nguyen et al. 2021.

²⁷ Thomas et al. 2024.

²⁸ *Einsatz von GPT-4 für NER* (Möhrke et al. 2024).

²⁹ Möhrke et al. 2024.

³⁰ E.g. Cascaded Analysis Broker - CAB, cf. Jurish 2011.

³¹ *LLM-basierte Normalisierung historischer Schreibweisen* (Bracke 2024).

³² *Korrektur und (De-)Normalisierung historischer Volltexte* (Würzner / Sachunsky 2024).

and the OCR specialist with an interest in faithful and accurate transcriptions as training data.³³ In the first case, ChatGPT-4 can increase the quality of full texts derived from large digitization projects by addressing errors and ensuring consistent normalization. For the other two user profiles, ChatGPT was tasked with denormalizing previously standardized text versions (e.g., from Wikisource or other digital libraries) to better align these texts with original historical documents. The denormalization attempts demonstrated relatively higher historical and material accuracy compared to standard normalization methods, particularly when guided by detailed prompts and supplemented with scans of the originals. Nevertheless, the resulting quality remained insufficient for training data purposes. Significant improvements were achieved through iterative prompting with more specific instructions and the inclusion of original document scans, though the outputs still did not meet the standards required for ground truth data.

A major benefit of LLMs is their ability to process semantic relationships and contextual patterns in texts. Named Entity Recognition (NER) is a prominent application for generative AI, but it has to compete with established methods. Pia Schwarz, Florian Barth, and Lennart Keller³⁴ attempt to map academia in newspapers and compare SpaCy transformer-based fine-tuned models with OpenSource LLMs (Llama 2 13B, OpenOrca Platypus-2). Their results indicate the need for computational power, the length of context, and the risk of unsystematic and invented results.³⁵ It is well known that *Named Entity Linking* (NEL) is problematic for LLMs, which tend to invent formal identifiers. Inspired by the results of Tanti Kristianti and Laurent Romary³⁶ and Delpeuch,³⁷ the **HIPE 2022 experiment** could demonstrate that models trained on Wikidata or DBpedia sets are already successful in this task.³⁸ Consequently, they suggest fine tuning LLMs on structured data from Wikidata that includes explicit references to the identifiers. This comparative analysis reveals two critical requirements for effective LLM implementation in entity recognition tasks: integration with structured data sources to minimize hallucination of identifiers, and sufficient context window size to maintain accurate entity recognition and linking. [20]

Traija Nisha, Franziska Pannach, and Jörg Wettlaufer³⁹ tackle a more complex task involving NER and NEL: extracting information on the itineraries from travelogs.⁴⁰ This task needs deeper semantic understanding because the system has to distinguish between narrative on the travel itself and mentioning of other places. The **corpus of Middle East travelers** lists ca. 800 travelogs with ca. 180 full texts. While a previous study on grounding characters and places in the narrative⁴¹ supports results by Schwarz, Barth, and Keller⁴² that plain LLM based methods were then not yet performing better than a fine-tuned BERT model, the Göttingen team built two custom GPTs – specialized versions of GPT configured for specific tasks through custom instructions and knowledge bases – on OpenAI's platform (**Travelogue Annotator** and **Itinerizer**) for few-shot prompting and including graphical user interfaces with a spatial display on a map. [21]

Jacob Möhrke, Sandra Balck, and Anna Ananieva⁴³ hypothesize that NER in historical texts can be reformulated as a text generation task using the inherent strengths of LLM.⁴⁴ Encouraged by experiments with modern language,⁴⁵ they created a custom GPT called **DH Assistant** to insert special characters marking named entities in a travelog by F. X. Bronner from 1810 / 1817. The GPT-based model performs text [22]

³³ Würzner / Sachunsky 2024.

³⁴ *Klassifikation und Linking von Entitäten* (Schwarz et al. 2024).

³⁵ Schwarz et al. 2024.

³⁶ Kristianti / Romary 2020.

³⁷ Delpeuch 2019.

³⁸ Ehrmann et al. 2020.

³⁹ *Itinerare erkennen in Reiseberichten* (Nisha et al. 2024).

⁴⁰ Nisha et al. 2024.

⁴¹ Soni et al. 2023.

⁴² Schwarz et al. 2024.

⁴³ *Einsatz von GPT-4 für NER* (Möhrke et al. 2024).

⁴⁴ Möhrke et al. 2024.

⁴⁵ De Toni et al. 2022.

processing in three steps: decomposition, annotation, and recomposition. Untrained standard NER like flair⁴⁶ resulted in F1 scores of 0.18 for this multiclass annotation task. In a zero-shot-experiment, the custom GPT performed significantly better with a F1-score of 0.43, rising in both precision and recall. Their experiment also included normalization tasks for dates and currencies, in which the LLM identified problems to be solved that were not explicitly stated in the prompt. However, this ability also resulted in a hypercorrect output of the annotation task, where the reference text for the NER labels were normalized, although this task was not requested – an effect which might be controlled in a prompt.

Nina C. Rastinger⁴⁷ has experimented with GPT-3.5 in her project to extract information about people arriving in Vienna in the 18th century as documented in the *Wienerisches Diarium*⁴⁸. In a standard prompting scenario, she tested a one-shot-prompt to process these semi-structured lists. The results showed poor accuracy in identifying where named entities began and ended within the text, with the system generating 139 false entities, resulting in a ›hallucination quote‹ of 0.35 %. Most of these errors followed a ›hyper-correct‹ pattern we have also seen in other experiments: change in case, graphematics, and punctuation, but also semantic inferences, such as resolving a co-reference to the name of the entity. For historians, such cases would not necessarily count as false positives. Overall, the experiment shows potential, particularly in its good semantic referencing and the ability of the generic GPT model to adapt to the source material, despite using only few examples in the prompt. [23]

While the NER experiments show promising results in entity identification, they do not address the subsequent challenge of information enrichment through entity linking. Entity linking can be addressed with RAG, a method being applied in many industry tasks such as customer support or health care. In the workshop discussion, Andreas Kuczera reported that the *Regesta Imperii* team is leveraging their internal knowledge base stored in a Neo4j graph database for this task. [24]

Machine learning can also be part of a pipeline not using any of the LLMs used for generative AI applications. Carina Geldhauser, Ipek Tuncel, and Saahil Sundaresan⁴⁹ build an eScriptorium-based HTR pipeline to annotate nomina sacra in greek majuscule manuscripts, incorporating specialized XML markup. The application uses the implicit language model created by fine-tuning the kraken model, which uses an established deep learning architecture combining convolutional layers with long-short-term-memory (LSTM). Like Rastinger's approach, they enhance the system with manually curated lists while exploring how modern language processing can support this specialized task. [25]

Christopher Pollin⁵⁰ demonstrates the capabilities of LLMs in TEI annotation through a semantic markup of plain text letters. He used the [Hugo Schuchardt Archive](#) as a case study⁵¹ to trace the evolution of a process for supporting the creation of digital editions of correspondences, progressing from simple ›bad prompts‹ to sophisticated workflows. Pollin's experiment compares GPT-3.5 and GPT-4 models. A simple task description in a ›one shot‹ setting yields unsatisfactory results. It produces well-formed XML with some reasonable tagging but almost never valid TEI. However, a more sophisticated prompt including examples – a so-called *few-shot* approach – changes this significantly. Building on research in prompting techniques,⁵² Pollin refines this further by using persona modeling (›You will act as a skilled expert automaton ...‹), leveraging context (›...that is proficient in transforming unstructured text, specifically multilingual letters from or to Hugo Schuchardt (1842–1927), into well-formed TEI XML‹), assigning a clear task (›Analyze the provided text based on the mapping rules I have shared and then execute the transformation to produce TEI XML ensuring you [26]

⁴⁶ Akbik et al. 2019.

⁴⁷ *Informationsextraktion aus frühneuzeitlichen Ankunftslisten* (Rastinger 2024).

⁴⁸ *Resch: DIGITARIUM* and the *Wiener Zeitungin ANNO*.

⁴⁹ *Halbautomatische Annotierung antiker Handschriften* (Geldhauser et al. 2024).

⁵⁰ *Von ›bad prompts‹ mit ChatGPT-3.5 zu Workflows mit GPT-4-Agenten* (Pollin 2024).

⁵¹ Hurch 2024.

⁵² Bsharat et al. 2023.

adhere to the guidelines and only annotate if certain), and implementing few-shot prompting by adding examples for in-context learning. This approach is supported by TEI mapping rules that list elements for correspondence modeling along with their application, such as ›<salute> Salutations within the letter‹, and modeling instructions, such as ›Preserve the original text or produce well-formed TEI XML according to the TEI standards‹. Finally, emotional prompting⁵³ improves quality by adding urgency, such as ›This is very important for my career!‹.

Comparing the two models, GPT-4 showed stronger ›reasoning‹ capabilities in ›understanding‹ task logic and followed TEI Guidelines more closely to produce valid XML. The model also successfully identified and normalized names and dates. [27]

Combining these findings with the best performing LLMs on reasoning tasks, including a human-in-the-loop and integrating external knowledge bases via RAG, Pollin proposes a workflow that combines several subsystems into an conceptual framework for transforming unstructured text into TEI-compliant XML documents – the underlying data structure required for digital scholarly editions. This follows the concept of interacting LLM-based agents invoked by other AI components.⁵⁴ Pollin’s proposal considers interacting agents for analyzing visual features in digital facsimiles (GPT-4 Vision), adding TEI annotations to unstructured text (*teiCrafter*⁵⁵), conceptualizing the target structure – the text model (*teiModeler*⁵⁶) – and verifying the resulting TEI encoding (*teiVerifier*). Particularly in the modeling and verification steps, the system incorporates knowledge bases such as entity stores, schema descriptions, and glossaries. It delegates analytical processes and decisions to predefined subsystems called ›actions‹. These interacting systems need the human-in-the-loop to check the results, maintain communication, and feed further information into the system. [28]

The final step of an ideal and complete editorial workflow is the *translation and summarization* of the edited text. The experiment presented by Dominic Fischer, Martin Volk, Patricia Scheurer, and Phillip Ströbel⁵⁷ used ChatGPT to produce English translations of the original Latin letters of the reformer Heinrich Bullinger (1504–1575) as part of the *Bullinger Digital* edition project hosted at the University of Zurich. Compared to ›traditionalk machine translation tools incl. Google Translate, results have been satisfactory. In particular, ChatGPT-3.5 and even more so GPT-4 are capable of dealing with text parts mixing German and Latin passages. Additional prompts introducing glossaries (such as ›Translate cesar as emperor‹) helped to secure accuracy for historical terms. Equally useful for the end user of a scholarly edition proved the results of prompting to summarize individual letters of the collection. [29]

As these experiments demonstrate, generative AI can significantly enhance various tasks within the editorial process. This observation has been further supported by experiments conducted since spring 2024. DeRose, in his paper at the Balisage conference 2024, explores how LLMs often disregard important markup and structural information during both training and output generation, emphasizing the implications of this limitation and proposing methods for improved integration and utilization of these features.⁵⁸ Lang, on the other hand, advocates for an editor-in-the-loop approach, where LLM-assisted corrections enhance the quality of Transkribus OCR outputs, enabling editors to efficiently review suggested changes using git diff.⁵⁹ Scholger, Strutz, and Pollin presented experiments on integrating LLMs into scholarly editing workflows, demonstrating experiments with GPT-4 and RAG for TEI encoding, handling editorial interventions, normalization, named entity recognition, and translation, using the example of the Austrian orientalist Joseph [30]

⁵³ Li et al. 2023.

⁵⁴ Wu et al. 2023.

⁵⁵ Pollin et al. 2023c.

⁵⁶ Pollin et al. 2023d.

⁵⁷ *LLMs for Bullinger Digital* (Fischer et al. 2024).

⁵⁸ DeRose 2024.

⁵⁹ Lang 2024.

von Hammer-Purgstall.⁶⁰ At a conference on ChatGPT and generative AI in medieval research in September 2024, Schonhardt explored the LLM-assisted enrichment of medieval correspondence.⁶¹ Abel and Pultar presented studies on converting indexes into knowledge graphs,⁶² while Armbruster and Kuczera developed a RAG workflow for content processing.⁶³

The feasibility of integrating these applications into real editions has been demonstrated. These include prompt refinement, external knowledge bases for RAG, delegated subtasks feeding back into the main task (*Actions*), and, in particular, the human-in-the-loop guiding the AI system in a conversation. Comparing the experiments with an ideal set of tasks (see figure 1) demonstrates that more detailed work is still necessary to assess the possibilities to apply automatic text generators and AI in several areas relevant to digital scholarly editing. The definition of the subject matter (*edendum*), for instance, can be part of the brainstorming use of generic Chatbots,⁶⁴ but can also include publically available manuscript and archival databases for RAG. Multimodal approaches can add context to an automated text recognition (*ATR*) process in context-specific definition of transcription methods, e.g., by describing more complex transcription practices for orthography and abbreviations. Textual criticism can use code creation to include existing deterministic tools like CollateX. A **brief experiment**⁶⁵ with the stemma codicum of Pier Aimone-Braida's edition of Simon of Bisignano's *Summa Decretorum*⁶⁶ demonstrates that text generators are capable of making informed individual suggestions on textual criticism because we can conceptualize these decisions as a combination of probabilities of a textual sequence and reasoning on textual tradition. RAG technologies were not as much in use in February 2024, but their application for entity linking tasks is obvious. The extraction of documentation from annotation practices constitutes a typical text generation task, constrained only by the text generator's context window size. The use of code generation to set up technical systems for publishing has been demonstrated by Christopher Pollin in September 2024.⁶⁷

[31]

5. Conclusions

These experiments and the related research answer several of the questions asked in the introduction. The areas of application for generative AI in the context of digital scholarly editing are mostly allocated in the processing of source material into editions as data: automatic transcription, structural and semantic annotations, retro-conversion, normalization and translation. Knowledge in prompt engineering methods can increase their efficiency. While Christopher Pollin already suggested a deeper integration in early 2024, the scholarly community has not yet reached best practices, and in the following, we want to suggest some research lines to follow up from the experiments, trying to contribute to the development of a common methodology through a thorough evaluation of the methods tested. These proposals have to highlight biases and ethical risks.

[32]

The global debates on the developments and trends observed in early 2024 can be considered under two perspectives: a practical (productive) and a theoretical (reflective) one, each partitioned into two sub-perspectives.

[33]

⁶⁰ Scholger et al. 2024.

⁶¹ Schonhardt 2024.

⁶² Abel / Pultar 2024.

⁶³ Armbruster / Kuczera 2024.

⁶⁴ Park / Kulkarni 2024.

⁶⁵ Vogeler 2024.

⁶⁶ Aimone-Braida 2007.

⁶⁷ Pollin / Vogeler 2024.

5.1 The Practical Perspective 1: Workflow Orchestration

The practical challenge lies in developing effective workflows for complex editorial tasks. A major challenge is identifying specific steps to navigate the extensive range of available options, including the selection of effective tools, the design of optimized workflows, and the identification of reliable integration methods for generative AI applications. Given the rapid development of foundational models (LLMs), data accessibility, computational resources, and integration capabilities, it is evident that a consensus is needed on the utilization and combination of tools (such as LLMs, generic tools, and specialized applications) best suited for specific editorial and analytical tasks. Prompt engineering continues to play a pivotal role in optimizing these tools, requiring a profound understanding of how particular prompting techniques and tricks lead to optimal results. As the examples above demonstrate, promising results can already be achieved with reasonable effort by combining advanced prompt engineering and retrieval-augmented generation (RAG), enhanced by in-context learning. The latter can be facilitated through structured communication with a human-in-the-loop guidance. Fine-tuning base models allows for further tailoring of AI systems to specific humanities tasks and datasets. An example for this is the chatbot ParzivAI which uses Command R+, an open-source model, fine-tuned for communication about and translation of Middle High German.⁶⁸ This approach, however, requires more advanced technical expertise in machine learning and substantial computational resources. Training new models presents an even higher workload and is currently secondary due to limited humanities training data and high resource requirements. Beyond this, a much more expansive scenario should also be considered, in which individual ›agents‹ optimized for specific tasks could be interconnected to form a cohesive, integrated editorial system, as OpenAI is currently planned with its ›Operator‹.

[34]

5.2 The Practical Perspective 2: Evaluation

A central question, often unanswered in the experimental phase, concerns the evaluation of results and quality control. Following this initial phase of experimentation and knowledge-building, it is now time to consider pipelines for large-scale applications and devising scenarios for their evaluation, thus enabling reliable assessments for LLM quality and efficiency. Only a few NLP tasks are already better off in this field. Xie et. al tested the performance of ChatGPT for named entity recognition (NER) in a zero-shot scenario on both domain-specific and general datasets by employing task decomposition, syntactic and tool-based augmentation, and self-consistency techniques.⁶⁹ Similarly, Nina Rastinger has taken up the challenge⁷⁰ and compared the performance of four LLMs in NER tasks as presented in the workshop. Nevertheless, benchmarks to rigorously evaluate the quality and accuracy of AI-generated results for our specific tasks are still lacking. While [Google's BigBench application](#) provides a set of benchmarks, some of which include humanities-related examples – primarily from the linguistics domain⁷¹ – none are directly associated with the specialized domain of scholarly editing. Consequently, we currently lack detailed insights into what marks a successful experiment in contrast to a less successful one. Finally, it has become evident across all experiments that there is not yet enough experience to realistically assess efficiency: experimentation, by nature, initially ignores the question for the cost-benefit ratio in a productive usage, i.e. they are first investing in an unknown future.

[35]

5.3 Theoretical Perspective 1: General Strengths and Weaknesses

In the second, more theoretical perspective, fundamental questions arise regarding which ›strengths‹ and ›weaknesses‹ of generative AI actually make it viable for productive use in certain areas – or not. Such assessments can only be made in relation to the respective expectations, which then enable a realistic

[36]

⁶⁸ Nieser / Renkert 2024.

⁶⁹ Xie et al. 2023.

⁷⁰ Rastinger 2024.

⁷¹ Srivastava et al. 2023.

assessment of their usability in specific workflows. Fundamental concerns exist regarding the scientific basis of the systems: as they are stochastic processes with some built-in randomness, outputs of text-generating LLMs are only reproducible to a limited extent. The fundamental absence of explainability in generative AI procedures and results aggravates this problem, as standard explainability methods have thus far been applied to digital humanities problems only in image classification.⁷² Additionally, commercial providers frequently alter their engines and configurations without sufficient – or sometimes any – documentation. In applying generative AI to annotation tasks, we observed various issues, such as hallucinated XML elements not provided in the given schema, alterations to the source document's structure, and occasional abbreviations of input text.

Despite these issues, it is evident that LLMs are highly effective in generating, translating and improving texts. They can transform complex texts into simpler language and create summaries. These models are good at recognizing semantic relationships and exhibit a form of common sense that enables them to generate texts that appear logically consistent and sound plausible. With their ability to handle complex tasks that formalization-oriented algorithms could not handle well before, LLMs may now be better suited to understand and process heterogeneous, fuzzy and complex data typical of research in the humanities. Moreover, we can even leverage the stochastic nature of LLMs: multiple requests yield diverse useful interpretations of a given task, from which a human can select the most suitable.

[37]

5.4 Theoretical Perspective 2: Bias and Ethical Issues

The use of generative AI in digital scholarly editing touches on numerous critical issues regarding bias and ethical concerns.⁷³ The notorious ›coded bias‹ in training data represents a foundational challenge, as models trained on broad, often uncurated datasets risk replicating and amplifying existing societal biases, especially regarding gender and ethical perceptions. In scholarly editing, the bias against historic forms of language is exacerbated as digitized material lacks sufficient coverage, often due to neglect or limited resources. OCR correction and normalization tasks may inadvertently favor dominant cultural perspectives, marginalizing less represented voices and thus narrowing the intellectual diversity of generated content. Additionally, the high computational costs associated with training and running these models entail significant energy consumption and its often-overlooked environmental impact. Language bias further complicates these issues, as prompts in English typically yield higher-quality responses than those in other languages.⁷⁴ This discrepancy persists even when examples are provided in a few-shot setting in another language, or when output is explicitly requested in a target language other than English. The reliance on commercial tools, whose training data and algorithms remain largely opaque, adds to the ethical complexity, as researchers have limited control or insight into the models' inner workings. This lack of transparency compromises reproducibility – a key scholarly standard – and places substantial power in the hands of commercial providers. As Nils Reiter provocatively put it ›when you're evaluating #LLMs, and you're including non-local, API-based models, you're not doing model evaluation, but product testing.«⁷⁵ While the availability of the computational resources and skills necessary for employing free and local AI systems currently poses a problem, their use is essential for research that is comprehensible, reproducible, and open.

[38]

When it was (the spring of) 2024, we realized the many opportunities generative AI holds for digital scholarly editing. However, leveraging these opportunities requires a critical and nuanced engagement with the technological, methodological, and ethical challenges they present. We were and are in a phase of ongoing synthesis of practical experimentation and theoretical reflection to shape rigorous, transparent, sustainable, and reusable practices in this field. Ongoing research in the application of generative AI to

[39]

⁷² El-Hajj et al. 2023.

⁷³ Kamocki et al. 2024; Cooper / Grimmelmann 2024; Dornis / Stober 2024; Gervais et al. 2024; W. Scholger et al. 2024.

⁷⁴ Zhang et al. 2023.

⁷⁵ Mastodon. Nils Reiter 2024.

digital scholarly editing refines workflows, establishes benchmarks, addresses biases, and works towards a common methodology. What might the results of this research be? On November 14, 2024, GPT-4o offered the following answer: »This research into the application of generative AI to digital scholarly editing could lead to enhanced efficiency, standardization, bias mitigation, innovative technological advancements, improved accessibility, cost reduction, new training standards, and improved methods for digital preservation and archiving of scholarly works.« This sounds reasonable enough. Yet, as digital humanities scholars with a knowledge cutoff in November 2024, we refrain from carving in stone a final answer to this question ourselves. The future, after all, is not the domain of immediate documentation and contemporary historiography of science.

Bibliography

- Christina Abel / Yannik Pultar: Register auf dem Weg zum Knowledge Graph. In: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Universität des Saarlandes (eds.): ChatGPT und generative KI in der mediävistischen Grundlagenforschung. (Saarbrücken, 19.–20.09.2024) Saarbrücken 2024. [\[online\]](#)
- Pier Virginio Aimone-Braida (ed.): *Summa in decretum Simonis Bisinianensis*. Vol. 2. Fribourg 2007. [\[Nachweis im GVK\]](#)
- Alan Akbik / Tanja Bergmann / Duncan Blythe / Kashif Rasul / Stefan Schweter / Roland Vollgraf: FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). 2019. DOI: [10.18653/v1/N19-4010](#)
- Stephan Armbruster / Andreas Kuczera: Mehr als ChatGPT: Inhalterschließung mit Retrieval Augmented Generation (RAG). In: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) / Universität des Saarlandes (eds.): ChatGPT und generative KI in der mediävistischen Grundlagenforschung. (Saarbrücken, 19.–20.09.2024) Saarbrücken 2024. [\[online\]](#)
- Yannis Assael / Thea Sommerschild / Brendan Shillingford / Mahyar Bordbar / John Pavlopoulos / Marita Chatzipanagiotou / Ion Androutsopoulos / Jonathan Prag / Nando de Freitas: Restoring and Attributing Ancient Texts Using Deep Neural Networks. In: *Nature* 603 (2022), no. 7900, p. 280–283. PDF. DOI: [10.1038/s41586-022-04448-z](#)
- Rachel Bawden / Jonathan Poinhos / Eleni Kogkitsidou / Philippe Gambette / Benoît Sagot / Simon Gabay: Automatic Normalisation of Early Modern French. In: Nicoletta Calzolari / Frédéric Béchet / Philippe Blache / Khalid Choukri / Christopher Cieri / Thierry Declerck / Sara Goggi / Hitoshi Isahara / Bente Maegaard / Joseph Mariani / Hélène Mazo / Jan Odijk / Stelios Piperidis (eds.): Proceedings of the Thirteenth Language Resources and Evaluation Conference. (LREC, Marseille, 20.–25.06.2022) Marseille 2022, p. 3354–3366. [\[online\]](#)
- Berlin-Brandenburgische Akademie der Wissenschaften (ed.): Deutsches Textarchiv (DTA). Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. 2007. [\[online\]](#)
- Elisa Beshero-Bondar: Declarative markup in the time of »AI«: Controlling the semantics of tokenized strings. In: Proceedings of Balisage: The Markup Conference 2023 (Washington, DC, July 31.07.–04.08.2023) (= Balisage Series on Markup Technologies 28). 2023. HTML. DOI: [10.4242/BalisageVol28.Beshero-Bondar01](#)
- Yannic Bracke: LLM-basierte Normalisierung historischer Schreibweisen. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. PDF. [\[online\]](#)
- Yannic Bracke: transormer. In: Yannic Bracke (ed.): ybracke. GitHub. 2024. Dataset. [\[online\]](#)
- Sondos Mahmoud Bsharat / Aidar Myrzakhan / Zhiqiang Shen: Principled Instructions Are All You Need for Questioning LLaMa-1/2, GPT-3.5/4. 26.12.2023 (v1). PDF. DOI: [10.48550/arXiv.2312.16171](#)
- Youngjin Chae / Thomas Davidson: Large Language Models for Text Classification: From Zero-Shot Learning to Fine-Tuning. 23.08.2023. PDF. DOI: [10.31235/osf.io/sthwk](#)
- CLEF (ed.): HIPE – Identifying Historical People, Places and other Entities: Shared Task on Named Entity Recognition and Linking in Multilingual Historical Documents. 2022. [\[online\]](#)
- Feder Cooper / James Grimmelman: The Files are in the Computer: On Copyright, Memorization, and Generative AI (April 22, 2024). In: Michael Heise (ed.): Cornell Legal Studies Research Paper Forthcoming / Chicago-Kent Law Review Forthcoming. 22.04.2024. PDF. [\[online\]](#)
- Antonin Delpuech. OpenTapioca: Lightweight Entity Linking for Wikidata. 19.04.2019 (v1). PDF. DOI: [arXiv:1904.09131](#)
- Steven DeRose: Can LLMs help with XML? In: Proceedings of Balisage: The Markup Conference 2024 (Washington, DC, 29.07.–02.08.2024) (= Balisage Series on Markup Technologies 29) 2024. HTML. DOI: [10.4242/BalisageVol29.DeRose01](#)
- Francesco De Toni / Christopher Akiki / Javier De La Rosa / Clementine Fourrier / Enrique Manjavacas / Stefan Schweter / Daniel Van Strien: Entities, Dates, and Languages: Zero-Shot on Historical Texts with T0. In: Association for Computational Linguistics (ed): Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models. Dublin 2022, p. 75–83. DOI: [10.18653/v1/2022.bigscience-1.7](#)
- Tillmann Döncke / Florian Barth / Hanna Varachkina / Caroline Sporleder: MONAPipe: Modes of Narration and Attribution Pipeline for German Computational Literary Studies and Language Analysis in spaCy. In: Robin Schaefer / Xiaoyu Bai / Manfred Stede / Torsten Zesch (eds.): Proceedings of the 18th Conference on Natural Language Processing. (KONVENS, Potsdam, 12.–15.09.2022) Potsdam 2022, p. 8–15. PDF. [\[online\]](#)
- Tim Dornis / Sebastian Stober: Urheberrecht und Training generativer KI-Modelle. (= Recht und Digitalisierung | Digitization and the Law, 9). Baden-Baden 2024. PDF. DOI: [10.5771/9783748949558](#)
- Maud Ehrmann / Matteo Romanello / Simon Clematide / Alex Flückiger: CLEF-HIPE-2020 Shared Task Named Entity Datasets. In: Zenodo. 11.03.2020. Dataset. DOI: [10.5281/zenodo.6046853](#).
- Hassan El-Hajj / Oliver Eberle / Anika Merklein / Anna Siebold / Noga Shlomi / Jochen Büttner / Julius Martinetz / Klaus-Robert Müller / Grégoire Montavon / Matteo Valleriani: Explainability and transparency in the realm of digital humanities: toward a historian XAI. In: *International Journal of Digital Humanities* 5, (2023), p. 299–331. PDF. DOI: [10.1007/s42803-023-00070-1](#)
- Dominic Fischer / Martin Volk / Patricia Scheurer / Phillip Ströbel: LLMs for Bullinger Digital. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. (Passau, 27.02.2024) Passau 2024. [\[online\]](#)
- Christiane Fritze (ed.): Manifest für digitale Editionen. In: *Digital Humanities im deutschsprachigen Raum* (ed.): DHdBlog. 11.03.2022. HTML. [\[online\]](#)
- Daniele Fusi: Aspects of Application of Neural Recognition to Digital Editions. In: Bernhard Assmann / Alexander Czmil / Franz Fischer / Malte Rehbein / Torsten Schaßan / Georg Vogeler / Sabine Büttner / Oliver Duntze / Christiane Fritze / Patrick Sahle / Philipp Steinkrüger / Katharina Weber (eds.): *Kodikologie und Paläographie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age*. (= Schriften des Instituts für Dokumentologie und Editorik, 2). Norderstedt 2009, p. 175–195. [\[Nachweis im GVK\]](#)
- Ulrike Henny-Krahmer / Bernhard Geiger / Fabian Kaßner / Marc Lemke / Gerlinde Schneider / Martina Scholger (eds.): *Machine Learning and Data Mining for Digital Scholarly Editions*, Konferenz. (Rostock, 09.–10.06.2022) Rostock 2022. PDF [\[online\]](#)
- Carina Geldhauser / Ipek Tuncel / Saahil Sundaresan: Halbautomatische Annotierung antiker Handschriften. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. PDF. [\[online\]](#)
- Daniel Gervais / Noam Shemtov / Marmanis Haralambos / Catherine Zaller Rowland: The Heart of the Matter: Copyright, AI Training, and LLMs. 2024. PDF. [\[online\]](#)
- Bernhard Hurch (ed.): *Hugo Schuchardt Archive*. Graz 2022. HTML. [\[online\]](#)
- IDE (ed.): *RIDE. A review journal for digital editions and resources*. 2014. HTML. [\[online\]](#)

Institut für Computerlinguistik und Institut für Schweizerische Reformationsgeschichte der Universität Zürich (eds.): Bullinger Digital. Digitale Erschließung von Heinrich Bullingers Briefwechsel. 2020. HTML. [\[online\]](#)

Itinerizer. 26.03.2025. HTML. [\[online\]](#)

Bryan Jurish: Finite-State Canonicalization Techniques for Historical German. Dissertation, Universität Potsdam. 2011. [\[online\]](#)

Pawel Kamocki / Toby Bond / Krister Lindén / Thomas Margoni / Aleksei Kelli / Andrius Puksas: Mind the Ownership Gap? Copyright in AI-generated Language Data. In: Krister Lindén / Thalassia Kontino / Jyrki Niemi (eds.): Selected papers from the CLARIN Annual Conference 2023. Linköping electronic conference proceedings. (Leuven, 16.–18.10.2023) Linköping 2024, p. 102–113. PDF. DOI: [10.3384/ecp210008](#)

Klassik Stiftung Weimar (eds.): Goethe Lyrik. Neue Weimarer Ausgabe. [\[online\]](#)

Tanti Kristanti / Laurent Romary: DeLFT and Entity-Fishing: Tools for CLEF HIPE 2020 Shared Task. In: Linda Cappellato / Carsten Eickhoff / Nicola Ferro / Aurélie Névéol (eds.): CEUR Workshop Proceedings 2696, CLEF 2020 CLEF 2020 Working Notes, Working Notes of CLEF 2020 – Conference and Labs of the Evaluation Forum. (22.–25.09.2020, Thessaloniki) Wien / Budapest. 2020. PDF. [\[online\]](#)

Sarah Lang: LLM-powered-OCR-correction. In: Sarah Lang (ed.): sarahalang. GitHub. 2024. Dataset. [online](#)

Cheng Li / Jindong Wang / Yixuan Zhang / Kaijie Zhu / Wenxin Hou / Jianxun Lian / Fang Luo / Qiang Yang / Xing Xie: Large Language Models Understand and Can Be Enhanced by Emotional Stimuli. 12.11.2023 (v7). PDF. DOI: [10.48550/arXiv.2307.11760](#)

Yuliang Liu / Zhang Li / Biao Yang / Chunyuan Li / Xucheng Yin / Cheng-lin Liu / Lianwen Jin / Xiang Bai: OCRBench: On the Hidden Mystery of OCR in Large Multimodal Models. 26.08.2024 (v7). PDF. DOI: [arXiv:2305.07895](#).

Mastodon. Nils Reiter. @nielsreiter. 12.9.2024. [\[online\]](#)

Jacob Moehrke: DH Assistant. Expert in Digital Humanities, assisting with code formatting, data curation, TEI-XML, and NLP tasks. 26.03.2025. HTML. [\[online\]](#)

Jacob Möhrke / Sandra Balck / Anna Ananieva: Einsatz von GPT-4 für NER. In: DHD (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. HTML. [\[online\]](#)

Anders Giovanni Møller / Jacob Aarup Dalsgaard / Arianna Pera / Luca Maria Aiello: Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks. 26.04.2023 (v1). PDF. DOI: [arXiv:2304.13861](#)

Guenter Muehlberger, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, Hervé Déjean, Markus Diem, Stefan Fiel, Basilis Gatos, Albert Greinöcker, Tobias Grüning, Guenter Hackl, Vili Haukkovaara, Gerhard Heyer, Lauri Hirvonen, Tobias Hodel, Matti Jokinen, Philip Kahle, Mario Kallio, Frederic Kaplan, Florian Kleber, Roger Labahn, Eva Maria Lang, Sören Laube, Gundram Leifert, Georgios Louloudis, Rory McNicholl, Jean-Luc Meunier, Johannes Michael, Elena Mühlbauer, Nathanael Philipp, Ioannis Pratikakis, Joan Puigcerver Pérez, Hannelore Putz, George Retsinas, Verónica Romero, Robert Sablatnig, Joan Andreu Sánchez, Philip Schofield, Giorgos Sfikas, Christian Sieber, Nikolaos Stamatopoulos, Tobias Strauß, Tamara Terbul, Alejandro Héctor Toselli, Berthold Ulreich, Mauricio Villegas, Enrique Vidal, Johanna Walcher, Max Weidemann, Herbert Wurster, Konstantinos Zagoris: Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study. In: Emerald Publishing (ed.): Journal of Documentation 75 (2019), no. 5, p. 954–976. 1. 12.09.2019. HTML. DOI: [10.1108/JD-07-2018-0114](#).

Thi Tuyet Hai Nguyen / Adam Jatowt / Mickael Coustaty / Antoine Doucet: Survey of Post-OCR Processing Approaches. In: Albert Zomaya (ed.): ACM Computing Surveys 54, Issue 6 (31.07.2022), no. 124, p. 1–37. 2021. PDF. DOI: [10.1145/3453476](#)

Florian Nieser / Thomas Renkert: KI Showcase: Der Chatbot »ParzivAl«. In: Heidelberg School of Education (ed.): Fokus Lehrerbildung. 11.09.2024. DOI: [10.58079/12aa9](#)

Tarjia Alam Nisha / Franziska Pannach / Jörg Wettlaufer: Itinerare erkennen in Reiseberichten. In: DHD (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. PDF. [\[online\]](#)

Tarjia Alam Nisha: Travelouge Annotator. Expert in annotating travelogue excerpts, focusing on character-place relationships. Letzter Zugriff: 26.03.2025. HTML. [\[online\]](#)

Harsha Nori / Yin Tat Lee / Sheng Zhang / Dean Carignan / Richard Edgar / Nicolo Fusi / Nicholas King / Jonathan Larson / Yuanzhi Li / Weishung Liu / Renquian Luo / Scott Mayer McKinney / Robert Osazuwa Ness / Hoifung Poon / Tao Qin / Naoto Usuyama / Chris White / Eric Horvitz: Can Generalist Foundation Models Outcompete Special-Purpose Tuning? Case Study in Medicine. 28.11.2023 (v1). PDF. DOI: [10.48550/arXiv.2311.16452](#)

Österreichische Nationalbibliothek (ed.): ANNO. Historische österreichische Zeitungen und Zeitschriften. 2022. HTML. [\[online\]](#)

Soya Park / Chinmay Kulkarni: Thinking Assistants: LLM-Based Conversational Assistants that Help Users Think By Asking rather than Answering. 29.02.2024 (v2). PDF. DOI: [arXiv.2312.06024](#)

Bastian Politycki (2024a): SSRQ-Retro-Lab: Exploring the Application of LLMs for the Digitization of Printed Scholarly Editions (v2.0.0). Zenodo. 2024. DOI: [10.5281/zenodo.10683209](#)

Bastian Politycki (2024b): Anwendung generativer KI zur Digitalisierung gedruckter Editionen. In: DHD (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. PDF. [\[online\]](#)

Christopher Pollin: Von »bad prompts« mit ChatGPT-3.5 zu Workflows mit GPT-4-Agenten. In: DHD (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024.

Christopher Pollin / Alexander Czmiel / Stefan Dumont / Franz Fischer / Patrick Sahle / Torsten Schaßan / Martina Scholger / Georg Vogeler / Torsten Roeder / Christiane Fritze / Ulrike Henny-Krahmer: Generative KI, LLMs und GPT bei digitalen Editionen. In: DHD (ed.): DHd2024 – »Quo Vadis DH?«, (Passau, 26.02.–01.03.2024) Zenodo. Passau 2024. DOI: [10.5281/zenodo.10893761](#)

Christopher Pollin / Alexander Czmiel / Stefan Dumont / Franz Fischer / Christiane Fritze / Patrick Sahle / Thorsten Roeder / Torsten Schaßan / Markus Schnöpf / Martina Scholger / Georg Vogeler (2023a): Call for Experiments: Generative KI, LLMs und GPT bei digitalen Editionen (DHd2024) In: DHdBlog. 18.12.2023. HTML. [\[online\]](#)

Christopher Pollin / Christian Steiner / Constantin Zach (2023b): New Ways of Creating Research Data. Conversion of Unstructured Text to TEI XML using GPT on the Correspondence of Hugo Schuchard with a Web Prototype for Prompt Engineering. In: FORGE 2023. Tübingen 2023. DOI: [10.5281/zenodo.8425163](#)

Christopher Pollin / Sabrina Strutz / Georg Maximilian Reiter / Christian Steiner / Helmut Klug (2023c): teiCrafter. In: Digital Edition Creation Pipelines: Tools and Transitions (DigEdTnT). 23.02.2024. HTML. [\[online\]](#)

Christopher Pollin / Sabrina Strutz / Georg Maximilian Reiter / Christian Steiner / Helmut Klug (2023d): teiModeler. In: Digital Edition Creation Pipelines: Tools and Transitions (DigEdTnT). 22.02.2024. HTML. [\[online\]](#)

Christopher Pollin / Georg Vogeler: Digital History mit LLMs: Modellieren, Extrahieren, Transformieren, Visualisieren. In: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Universität des Saarlandes (ed.): ChatGPT und generative KI in der mediävistischen Grundlagenforschung (Saarbrücken, 19.–20.09.2024) Saarbrücken 2024. [\[online\]](#)

Alec Radford / Jong Wook Kim / Chris Hallacy / Aditya Ramesh / Gabriel Goh / Sandhini Agarwal / Girish Sastry / Amanda Askell / Pamela Mishkin / Jack Clark / Gretchen Krueger / Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. 26.02.2021 (v1). PDF. DOI: [10.48550/arXiv.2103.00020](https://doi.org/10.48550/arXiv.2103.00020)

Nina Rastinger: Informationsextraktion aus frühneuezeitlichen Ankunftslisten. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. (Passau, 27.02.2024) Passau. 2024. PDF. [\[online\]](#)

Nina Rastinger: Named Entity Recognition mit LLMs. In: DHd-AG (ed.): Angewandte Generative KI (AKGI-AG). 25.09.2024. HTML. [\[online\]](#)

Rechtsquellenstiftung des Schweizerischen Juristenvereins (ed.): Sammlung Schweizerischer Rechtsquellen. 2009. HTML. [\[online\]](#)

Claudia Resch (ed.): Das Wien[n]erische Diarium. 2018. HTML. [\[online\]](#)

Patrick Sahle: Digitale Editionsformen: Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Vol. 2: Befunde, Theorie und Methodik. (= Schriften des Instituts für Dokumentologie und Editorik, 8). Norderstedt 2013. [\[online\]](#)

Patrick Sahle: Kriterienkatalog für die Besprechung digitaler Editionen. In: Institut für Dokumentologie und Editorik (ed.). 2014. HTML. [\[online\]](#)

Martina Scholger / Sabrina Strutz / Christopher Pollin: Empowering Text Encoding with Large Language Models: Benefits and Challenges. In: Universidad del Salvador (ed.): TEI Buenos Aires 2024 (Buenos Aires, 07.-11.10.2024) Buenos Aires 2024. PDF. DOI: [10.5281/zenodo.13969082](https://doi.org/10.5281/zenodo.13969082)

Walter Scholger / Kim Nayyer / Korijalka Kuzman-Slogar / Pawel Kamocki: Legal and Ethical Aspects of Generative AI and their Impact on Digital Humanities. In: ADHO DH2024 Conference (Washington, DC, 06.-08.08.2024). 2024. PDF. DOI: [10.5281/zenodo.13270747](https://doi.org/10.5281/zenodo.13270747)

Michael Schonhardt: Briefe vom Meister Sepp: Generative KI in der Erschließung gelehrter Briefwechsel der frühen Mediävistik. In: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Universität des Saarlandes (eds.): ChatGPT und generative KI in der mediävistischen Grundlagenforschung (Saarbrücken, 19.-20.09.2024). Saarbrücken 2024. [\[online\]](#)

Pia Schwarz / Florian Barth / Lennart Keller: Klassifikation und Linking von Entitäten: Spezifischer Klassifikator vs. Large Language Model. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. Passau 27.02.2024. PDF. [\[online\]](#)

Jasha Sohl-Dickstein: BIG -bench. In: google (ed.): Sohl-Dickstein. GitHub. 2024. Dataset. [\[online\]](#)

Romeo Sommerfeld: Handwritten Character Recognition - an unofficial implementation of the paper TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models. In: Romeo Sommerfeld (ed.): rsummerfeld. GitHub. 30. November 2022. Dataset. [\[online\]](#)

Sandeep Soni / Amanpreet Sihra / Elizabeth Evans / Matthew Wilkens / David Bamman: Grounding Characters and Places in Narrative Text. In: Association for Computational Linguistics (ed.): Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Vol. 1: Long Papers. Toronto 27.05.2023 (v1), p. 11723-11736. DOI: [10.48550/arXiv.2305.17561](https://doi.org/10.48550/arXiv.2305.17561)

Arthur Flor de Sousa Neto / Arthur Flor de / Byron Leite Dantas Bezerra / Gabriel Calazans Duarte de Moura / Alejandro Héctor Toselli: Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review. In: SN Computer Science 5 (2024), no. 258. DOI: [10.1007/s42979-023-02583-6](https://doi.org/10.1007/s42979-023-02583-6).

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hossain, Amanda Askell, Amanda Souza, Ambrose Slone, Ameet Rahane, Anantharaman Iyer, Anders Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew Dai, Andrew La, Andrew Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholami-davoodi, Arfa Tabassum, Arul Menezes, Arun Kirubakaran, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakaş, Benjamin Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinon, Cameron Diao, Cameron Dour, Catherine Stinson, Cedrick Argueta, César Ferri Ramírez, Chandan Singh, Charles Rathkopf, Chenlin Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Chris Waites, Christian Voigt, Christopher Manning, Christopher Potts, Cindy Ramirez, Clara Rivera, Clemencia Siro, Colin Raffel, Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman, Dan Roth, Daniel Freeman, Daniel Khoshabi, Daniel Levy, Daniel Mosegui González, Danielle Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Derek Yuret, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodola, Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan Chi, Ethan Dyer, Ethan Jerzak, Ethan Kim, Eunice Engeru Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar, Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra, Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio Mariani, Gloria Wang, Gonzalo Jaimovitch-López, Gregor Betz, Guy Gur-Ari, Hana Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Shevlin, Hinrich Schütze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap Monelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac, James Simon, James Koppel, James Zheng, James Zou, Jan Kocoń, Jana Thompson, Janelle Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan Waweru, Joan Burden, John Miller, John Balis, Jonathan Batchelder, Jonathan Berant, Jörg Froberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones, Joshua Tenenbaum, Joshua Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh Dhole, Kevin Gimpel, Kevin Omondi, Kory Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt, Luheng He, Luis Oliveros Colón, Luke Metz, Lütfi Kerem Şenel, Maarten Bosma, Maarten Sap, Maartje ter Hoeve, Maheen Farooqi, Manaal Faruqi, Mantas Mazeika, Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramirez Quintana, Marie Tolkiehn, Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał Śwędrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimeo Xu, Mirac Suzgun, Mitch Walker, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdah Gheini, Mukund Varma T, Nanyun Peng, Nathan Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nicole Martinez, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter Chang, Peter Eckersley, Phu Mon Htut, Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei, Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ramon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sotnikov, Ronan LeBraz, Rosanne Liu, Rowan Jacobs, Rui Zhang, Ruslan Salakhutdinov, Ryan Chi, Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif Mohammad, Sajant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel Bowman, Samuel Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah Rous, Sarik Ghazarian, Sayan Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshiwal, Shyam Upadhyay, Shyamolima (Shammie)Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislava Dehaene, Stefan Divic, Stefano Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven Piantadosi, Stuart Sieber, Summer Misherghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tatsu Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Sunduny, Tobias Gerstenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Ramasesh, Vinay Uday Prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zirui Wang, Ziyi Wu: Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. 12.06.2023 (v3). PDF. DOI: [10.48550/arXiv.2206.04615](https://doi.org/10.48550/arXiv.2206.04615)

- Phillip Benjamin Ströbel / Simon Clematide / Martin Volk / Tobias Hodel: Transformer-based HTR for Historical Documents. 21.03.2022 (v1). PDF. DOI: [arXiv:2203.11008](https://arxiv.org/abs/2203.11008)
- Xiaofei Sun / Xiaoya Li / Jiwei Li / Fei Wu / Shangwei Guo / Tianwei Zhang / Guoyin Wang: Text Classification via Large Language Models. 15.05.2023 (v1). PDF. DOI: [arXiv:2305.08377](https://arxiv.org/abs/2305.08377)
- Alan Thomas / Robert Gaizauskas / Haiping Lu: Leveraging LLMs for Post-OCR Correction of Historical Newspapers. In: Rachele Sprugnoli / Marco Passarotti (eds.): Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024. Torino 25.05.2024, p. 116–21, 2024. PDF. [\[online\]](#)
- Enrique Vidal (ed.): tranSkriptorium. 2013–2015. 26.11.2024. HTML. [\[online\]](#)
- Georg Vogeler: What are we waiting for? State of the art in digital editing. Blogpost zum Day DH. In: Helmut Klug / Selina Galka / Elisabeth Steiner (eds.): HRSM Projekt Kompetenznetzwerk Digitale Edition. 2017. [\[online\]](#)
- Georg Vogeler: SummaSimonisBisanensis. In: Georg Vogeler (ed.): GVogeler. GitHub. 2024. Dataset. [\[online\]](#)
- Jörg Wettlaufer / Deniz Kılınçoğlu: Travels in the 19th-Century Ottoman Empire: A Digital History Research Project. 26.11.2024. HTML. [\[online\]](#)
- Qingyun Wu / Gagan Bansal / Jieyu Zhang / Yiran Wu / Beibin Li / Erkang Zhu / Li Jiang / Xiaoyun Zhang / Shaokun Zhang / Jiale Liu / Ahmed Hassan Awadallah / Ryan W White / Doug Burger / Chi Wang: AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. 16.08.2023 (v1). PDF. DOI: [arXiv:2308.08155](https://arxiv.org/abs/2308.08155)
- Kay-Michael Würzner / Robert Sachunsky: Korrektur und (De-)Normalisierung historischer Volltexte. In: DHd (ed.): Workshop: Generative KI, LLMs und GPT bei digitalen Editionen. (Passau, 27.02.2024) Passau. PPTX. 2024. [\[online\]](#)
- Tingyu Xie / Qi Li / Jian Zhang / Yan Zhang / Zuozhu Liu / Hongwei Wang: Empirical Study of Zero-Shot NER with ChatGPT. 16.10.2023 (v1). PDF. DOI: [10.48550/arXiv.2310.10035](https://arxiv.org/abs/10.48550/arXiv.2310.10035)
- Chenchen Xu: Construction of library subject information intelligent perception system integrating large language model. In: IEE (ed.): 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference. (IMCEC, 24.–26.05.2024, Chongqing) Chongqing 2024, p. 1671–1675. DOI: [10.1109/IMCEC59810.2024.10575069](https://doi.org/10.1109/IMCEC59810.2024.10575069)
- Linting Xue / Aditya Barua / Noah Constant / Rami Al-Rfou / Sharan Narang / Mihir Kale / Adam Roberts / Colin Raffel: ByT5: Towards a token-free future with pre-trained byte-to-byte models. 2021. PDF. DOI: [arXiv:2105.13626](https://arxiv.org/abs/2105.13626)
- Xiang Zhang / Senyu Li / Bradley Hauer / Ning Shi / Grzegorz Kondrak: Don't Trust ChatGPT when your Question is not in English: A Study of Multilingual Abilities and Types of LLMs. In: Houda Bouamor / Juan Pino / Kalika Bali (eds.): Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. (EMNLP, Singapore, 06.12.–10.12.2023.) Singapore 2023, p. 7915–7927. PDF. DOI: [10.18653/v1/2023.emnlp-main.491](https://arxiv.org/abs/10.18653/v1/2023.emnlp-main.491)
- Wenxuan Zhou / Sheng Zhang / Yu Gu / Muhao Chen / Hoifung Poon: UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition. 07.08.2023 (v1). PDF. DOI: [10.48550/ARXIV.2308.03279](https://arxiv.org/abs/10.48550/ARXIV.2308.03279)
- Chen Zhu-Tian / Chenyang Zhang / Qianwen Wang / Jakob Troidl / Simon Warchol / Johanna Beyer / Nils Gehlenborg / Hanspeter Pfister: Beyond Generating Code: Evaluating GPT on a Data Visualization Course. 06.10.2023 (v2). PDF. DOI: [10.48550/arXiv.2306.02914](https://arxiv.org/abs/10.48550/arXiv.2306.02914)

List of Figures

Fig. 1: General model of scholarly editing as knowledge production workflow: from source material to data and publication, through intermediate products (middle), concrete steps (right), and more general challenges (left). [Pollin et al. 2025]