Revisiting Historicism with Language Models: Bridging 19th-Century Perspectives and Machine Learning Approaches for Language Analysis

Tobias Hodel, University of Bern

Historiography underwent a profound transformation in the 19th century thanks to the intellectual current known as historicism. On a scholarly as well as on an architectural and broader level, understood as a «way of contextualizing the world historically»,[1] scholars, architects, and politicians created a movement that is still very well visible and legible in the public sphere, in scholarship, and in politics. These approaches fostered a focus on the world's historicity, critical assessments of the past, and the introduction of new methodologies.

Today, new technologies – particularly Large Language Models (LLMs) – offer potent approaches to textual sources, allowing, at the same time, a new way of encountering the various documents and artifacts of the past. LLMs are trained on extensive textual corpora and can help researchers mine themes, structures, and insights from vast amounts of 19th-century writing. By analyzing the digitized documents of these historicist sources, it is possible to create new connections and become aware of similarities and dissimilarities at scale while also confronting ethical challenges related to data representation and historical interpretation.

In this short article, I highlight the potential to leverage LLMs critically. Since this is more of an appeal, five theses will be presented and introduced, aiming at fostering critical discussion. The broad concerns of historicism are ideal for such a discussion, as many publications suggest that ideally interdisciplinary approaches and rich, informed source material are to be considered.[2] The 19th century is ideal, as most documents are out of copyright and not concerned with data protection regulations. At the same time, most language forms of the century are still very «understandable» for LLMs.

This is not intended as an introduction to historicism or large language models. Both deserve attention on their own and are only dealt with in a rather superficial way here.[3] Large Language Models are based on trans-

---

1 See as an introduction Müller 2014.
2 See e.g. Bon Gloor 2020, 44.
3 In this article, I will not provide any evaluation of the proposed applications and/or models.

former architectures.[4] They are often used in generative settings, e.g., to produce text based on inputs, so-called prompts. Most generative models leverage textual representation of inputs to predict further strings. In that sense, they are able to reproduce the «knowledge» (or rather a very frequent textual representation of it) that they have encountered in the training process.[5] Therefore, we must be extremely careful with the output of such models. Although the generated text sounds plausible, it is highly biased, and it's impossible to have an overview of the used training material.[6] Accordingly, we shouldn't rely on the embedded representation (again, «the knowledge») but instead use the capabilities of such models to deal with unstructured data (= text), in our case, source material.

First Thesis: Researching Historicism requires combining many different sources from a variety of disciplines

When dealing with historicism, especially with recent perspectives on the currently growing field, we identify many different domains and fields of study. While architecture and history build the foundation, it's at the same time about decoration and furniture, political argumentation, and archival studies. In a sense, the field is almost too large to be researched wholesomely.

Technical approaches are a way to circumvent or at least supplement this issue. Mainly due to the (self-)historicizing ideas of the protagonist, who not only commented on their own oeuvre but also left massive amounts of papers and documents. This leads us to the second thesis:

Second Thesis: Digitization (of image and text) is key when LLMs are to be in the loop

If we want to use LLMs, we will need digitized material. This task becomes increasingly hidden as libraries, archives, and museums (short LAMs) understand it to be one of their core tasks. However, digitization is expensive and requires time and resources (material and personnel).

The professionalism of LAMs sometimes stands in the way of pragmatic solutions. Image quality does not need to be on the highest level if it's about some textual witness only of interest to a specific question. There will never be a complete digitized library of the past, or at least not of uniform quality and according to identical standards. Therefore, digitizing «quick and dirty» is sometimes more helpful than digitizing only after years (or not at all).

---

4 Transformer Architectures are a form of neural networks which are built to store certain information while processing input (it could be called «remembering»). See als Vaswani et al. 2017.
5 See as introduction Seemann 2023.
6 See Bender et al. 2021.

Digitization in the form of producing images is also not enough. To deal with textual representation, the recognized text is a must for further processing. Approaches like Optical Character or Automatic Text Recognition need to be leveraged to gain texts in an actionable form.[7]

Third Thesis: LLMs trained in the language of the source material are capable of identifying similarities and differences

As has been mentioned before, it is not only highly critical but plain out reckless to prompt an LLM for specific answers. This would be comparable to using Wikipedia articles for research. Since the latest releases of LLMs are capable of dealing with larger inputs (so-called attention windows), it's possible to not only prompt («pose questions») but also feed large amounts of text into the systems. Such inputs can either be categorized piece by piece, or they can be put in comparison right away by the language model.

For example, we can use notebookLM provided by Google. In the system, it's possible to upload up to 50 documents (e.g. PDFs) and interact with the content, relying on the Google-trained Gemini LLM. Of course, it is advised not to use it for sensitive data, as always Google is collecting data of the input and how we interact with the system.

Working with openly published content, it becomes tangible how interactions with an LLM can support researchers. Based on the input of this journal (the volumes from 2020 to 2024), we get quite a helpful overview of the different tackled perspectives, including mentions of overlaps and different domains (fig. 1).

NotebookLM is a retrieval augmented generation (RAG) system that relies on vectorized input (the chunks of text). Consequently, it is more reliable than the prompt-generated output from LLMs.[8] Depending on the form of storage (vectorization), it is possible to work with multi-lingual input. In the example, German, French, and English texts were fed into the system. Most LLMs (and their embedding models) were trained in various languages, and they are, therefore, capable of dealing with multilingual input.

Fourth Thesis: Leveraging LLMs is but a method to start the research process; it is not an end in itself

Although the output of LLMs seems to be well-versed and impeccable with regard to grammar and style, it's still what we call «slop»: automatically generated output following, on the one hand, the input, but on the other hand, the most likely combination of tokens (sequence of words). This leads to a form of text that is quite legible. At the same time, the output feels artificial and lacks a personal note.[9]

---

7 With regard to Automatic Text Recognition, especially for handwriting, see Hodel 2023.
8 See as one of the earliest papers about RAG: Lewis et al. 2021.
9 For AI Slop, see the Wikipedia Article online: ‹https://en.wikipedia.org/wiki/AI_slop›.

Historismus·ch

> Wie wird die Thematik des Historismus in den unterschiedlichen Ansätzen aufgegriffen? und was trennt die Ansätze? Lässt sich daraus eine Definition von "Historismus" ableiten

Die Auseinandersetzung mit dem Historismus erfolgt in den Quellen auf vielfältige Weise, wobei sowohl Gemeinsamkeiten als auch trennende Aspekte in den unterschiedlichen Ansätzen erkennbar sind. Eine explizite, umfassende Definition des "Historismus" wird in den Texten nicht formuliert, aber durch die verschiedenen Betrachtungen lassen sich wesentliche Merkmale ableiten.

Gemeinsame Aspekte in den Ansätzen:

- **Zeitlicher Rahmen:** Die meisten Quellen konzentrieren sich auf das **19. und frühe 20. Jahrhundert** als die Hauptperiode des Historismus [1] … . El-Wakil erweitert diesen Rahmen in einem Artikel und bezieht auch das **18. Jahrhundert** mit ein [4] .
- **Geografischer Fokus:** Ein **starker Fokus auf die Schweiz** ist in allen Quellen vorhanden, entweder als zentraler Untersuchungsgegenstand oder als wichtiger Bezugspunkt im europäischen Kontext [1] … . Das **Schweizerische Netzwerk für Historismus** bildet den institutionellen Rahmen für viele der Beiträge [2] … .
- **Kunsthistorische und denkmalpflegerische Perspektive:** Viele Beiträge untersuch[en] und betonen die Bedeutung dieser Epoche für das kulturelle Erbe [1] … . Gleichzeitig [wird?] der Erforschung und Valorisierung des Historismus hervorgehoben, insbesondere im [Erhalt?] historischer Bausubstanz [1] … .
- **Vielfalt der untersuchten Objekte:** Das Interesse gilt einer **breiten Palette von Obj[ekten]** … , Inneneinrichtungen [1] … , Mobiliar [3] … , Glasmalerei [3] … und Eisenkunst [1[5]]
- **Betonung der Transferprozesse:** Mehrere Beiträge thematisieren **kulturelle, künstl[erische]** **Transferprozesse** im Kontext des Historismus, sowohl innerhalb Europas als auch da[rüber hinaus?]
- **Rezeption historischer Stile:** Die Auseinandersetzung mit der **Aneignung und Inter[pretation]** wiederkehrendes Thema [1] … .

> Die Idee zur Gründung eines Schweizer Netzwerks für Historismus entstand im Rahmen des von 2015 bis 2018 an der Universität Zürich angesiedelten Publikationsprojektes Der Orient in der Schweiz, in dem ein Team aus Kunst-historiker*innen, Denkmalpfleger*innen und Architektinnen die orientali-sierende Architektur der Schweiz als ein wichtiges Zeugnis des Historismus untersucht hat.1 Wie die Auseinandersetzung mit dem Thema gezeigt hat, sind es nicht nur die exotisch anmutenden neo-islamische Bauten und Interieurs, die aufgrund ihrer im 20.

For the purpose presented here, this is a minor issue if we consider the output the first step in research. Based on the provided comparison or the gained summaries, we make the second step by either close-reading the singled-out material or further refining the forms of comparison and maybe even switching to other distant reading approaches.[10]

As has been shown in different research scenarios (mainly focusing on coding), careful use of LLMs leads to very usable results but can also be highly time-consuming. Especially if the output leads to new insights, the certainty of the claims that the LLM makes needs to be established.

Fifth Thesis: Working with and on LLM should also happen as Open Science

Although it's primarily commercial LLMs like the above-mentioned Google product or the notorious ChatGPT, which are getting featured in media and explanatory pieces, it's possible and even necessary to consider using open-source tools for many of the presented tasks. It is equally important to contribute to open science. This is because the more openness a system provides, the easier it becomes to assess the output critically. Only a few of the so-called foundational models (the largest and most expensive category) of LLMs are freely available, and even fewer of the models are available in open-source. LLMs like the Llama family allow at least some traceability since the weights (the skeleton of the LLM) are openly available. However, the training material is not monitored, which makes it hard to assess the inherent biases.

Still, it's probable (and this author hopes at least so) that more and more large and smaller models become available in a more or less open fashion.

Fig. 1. Screenshot of NotebookLM, provided with the articles of historismus.ch (2020–2023), prompted for similarities in perspectives. The webpage also provides insights into the chunk (a piece of text) that led to the statement. This is an intriguing feature, allowing for means to check all the insights in detail (see the pop-up at the bottom to the right). Screenshot by the author.

---

10 See as an example Hitz et al. 2024.

Such models would allow for adaptability and (partial) replicability. This seems unnecessary if we understand the LLM as a mere tool; if we want to understand it as part of our methodology, it becomes key.

Challenges in Applying LLMs to Historicist Texts

While the integration of AI promises to enrich our comprehension of historicism, it also raises distinct challenges, like **Language and Context Preservation:** Even well-tuned language models may struggle with outdated stylistic conventions and context-specific references. Historical terms that carried particular weight in the 1800s may be misunderstood by LLMs if not carefully annotated. For example, political terms like «liberal» or «conservative» had different connotations in the 19th century than they do today.

Interpretative Fidelity: One of Ranke's legacies is the push for objective, source-based accounts of the past. The output of LLMs can inadvertently introduce errors if the training set includes modern political or cultural biases. Scholars must apply a critical lens to the LLM's interpretations, comparing them against other documentary evidence and historical scholarship.

Ethical Considerations: Digitizing material often involves the potential of (mis)representation. It's archives, libraries, and museums that choose what to digitize at what moment in time. As a result, we might only have access to what is deemed important today.

Conclusion

19th-century historicism stood at the forefront of a new, methodical approach to writing history. Scholars like Ranke laid the bedrock for the discipline as we know it by valuing archival evidence, contextual nuance, and critical analysis. Today, Large Language Models give us an opportunity to rethink and renovate our methodology, offering novel means to explore and interpret historicist texts—while also posing questions about interpretative fidelity, ethical uses of data, and the risk of anachronism. This convergence of the old and the new speaks to a shared quest: pursuing a deeper understanding of the past. LLMs can help illuminate aspects of the 19th century that might have been overlooked otherwise, while scholars can ensure that digital analyses are grounded in rigorous source criticism.

Tobias Hodel is a tenure-track Assistant Professor in Digital Humanities at the University of Bern. He studied history and German literature in Zurich. Currently, he researches machine learning applications for the humanities, focusing on pre-modern documents and information extraction tasks.

Historismus·ch

Bibliography

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and others, «On the Dangers of Stochastic Parrots», *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA 2021 (FAccT '21), 2021, 610–623, <https://doi.org/10.1145/3442188.3445922>.

Henriette Bon Gloor, «Vergessenes Mobiliar des Historismus: eine Spurensuche in der Schweiz», *Zeitschrift des Schweizer Netzwerks für Historismus* 1, 2020, 38–50,<https://doi.org/10.5169/seals-919721>.

Benjamin Hitz, Ismail Prada Ziegler, Aline Vonwiller, «From Record Cards to the Dynamics of Real Estate Transactions: Working with Automatically Extracted Information from Basel's Historical Land Register, 1400–1700», Jérôme Baudry, Lucas Burkart, Béatrice Joyeux-Prunel, Eliane Kurmann, Moritz Mähr, Enrico Natale, Christiane Sibille, and Moritz Twente (eds.), *Digital History Switzerland 2024: Book of Abstracts*, 2024, <https://digi-histch24.github.io/submissions/462>.

Tobias Hodel, «Konsequenzen der Handschriftenerkennung und des maschinellen Lernens für die Geschichtswissenschaft. Anwendung, Einordnung und Methodenkritik», *Historische Zeitschrift* 316, nᵒ 1, 2023, 151–180, <https://doi.org/10.1515/hzhz-2023-0006>.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, and others, «Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks», *arXiv,* 12.04.2021, <https://doi.org/10.48550/arXiv.2005.11401>.

Bertrand Müller, «Historismus (Geistesgeschichte)», *Historisches Lexikon der Schweiz (HLS)*, Version vom 22.05.2014, <https://hls-dhs-dss.ch/de/articles/027834/2014-05-22>.

Michael Seemann, «Künstliche Intelligenz, Large Language Models, ChatGPT und die Arbeitswelt der Zukunft», Hans Böckler Stiftung, Düsseldorf, 2023 (Forschungsförderung Working Paper 304), <https://www.boeckler.de/fpdf/HBS-008697/p_fofoe_WP_304_2023.pdf>.

Ashish Vaswani, Noam Shazeer, Niki Parmar, and others, «Attention Is All You Need», *arXiv,* 05.12.2017, <https://doi.org/10.48550/arXiv.1706.03762>.