The Augmented Qualitative Researcher: Using Large Language Models for Interpretive Text Analysis

Elida Izani and Andrea Voyer
Department of Sociology, Stockholm University

INTRODUCTION

Artificial intelligence and large language models (LLMs) like ChatGPT have fast gained traction and emblazoned further debate on the use of artificial intelligence (AI) in academic research. For some, ChatGPT represents the bogeyman in its deceptively clever way of posing as human. In this paper, we take a more optimistic view and argue that ChatGPT and other large language models are technological developments that can play a role in interpretive qualitative research. Just as previous advances in statistical computing revolutionized quantitative research by dramatically extending the speed and power of statistical analyses, so do recent developments in artificial intelligence promise to advance qualitative research by *augmenting* (Do et al., 2022) the ability of qualitative researchers to conduct thorough, systematic, transparent, and reproducible interpretive analysis.

In this article, we model augmented qualitative research through an application of the large language model ChatGPT[1] in interpretive analysis. We present practical guidelines for the use of ChatGPT and other similar LLMs. Given ongoing and rapid technological developments, this paper aims to start a conversation on incorporating LLMs in interpretive work and propose ways to harness the best aspects of LLMs while being mindful of the pitfalls of these innovations. In particular, we model two approaches to using LLMs in qualitative research, with a focus on qualitative coding and higher-level analysis. Interpretive research necessarily involves qualitative or interpretive coding[2]: the systematic labeling or tagging of specific elements within a text, such as themes, sentiments, entities, or events. This coding

[1] At the time of this writing, ChatGPT-3 is the latest and most accessible LLM tool. Our assumption is that other developments are in short order.

[2] The term ‘coding’ can have multiple meanings when one starts engaging in interdisciplinary work. In the data sciences, such activity is often called an ‘annotation’ task, but for the purposes of this paper, the word ‘coding’ refers to the qualitative activity of annotating data. Computer coding is described as ‘programming’ in this paper. For a longer list of potential ambiguous terms, refer to the Glossary in the Appendix.

process serves as a foundation for subsequent analysis, enabling researchers to identify patterns, observe changes over time, and draw meaningful insights from text data. Through interpretive coding, researchers can effectively categorize and organize information, enhancing the accuracy and efficiency of subsequent research tasks. Additionally, coding facilitates data sharing and collaboration among researchers, as it provides a common framework for understanding and interpreting textual data and contributes to the generation of transparent and reliable findings (Kirgil & Voyer, 2022; Nelson, 2021). AI-driven technology can augment the researcher in interpretive analyses by performing tasks that are both fine-grained and scalable.

Interpretive qualitative research design that leverages AI in interpretive work is possible. The time is now right for qualitative and interpretive researchers to take advantage of AI-driven technologies like ChatGPT and other LLMs. First, advances in the availability of digital text and processing power have resulted in large language models trained on a vast corpus of diverse digital text data, including books, articles, websites, and other publicly available sources, making these models particularly useful for dealing with text. Second, the accessibility of ChatGPT and some other models to those without a background in data science makes augmented interpretive research feasible for researchers without specialized training. Although researchers should understand the basics of ChatGPT or whatever LLM they work with, using these tools requires no prior computer programming knowledge and there is no need for researchers to be competent in computer and data science.

However, while ChatGPT and other AI-driven technologies can augment and extend researchers' ability to analyze text (Do et al., 2022), there is very little guidance in the literature about how these technologies can be integrated into qualitative research methods. This article addresses the need for more detailed information on the utility of LLMs in qualitative work and describes how one such model, ChatGPT, can be used in practice.

INTEPRETATIVE ANALYSIS USING LLMs

The interpretive analysis of text is crucial sociology as it allows researchers to extract valuable information and gain a deeper understanding of complex social phenomena. For example, interpretive text analysis has been used to answer questions about gender- and race-based stereotypes (Boutyline et al., 2020; Garg et al., 2018; Bolukbasi et al., 2016), political elites (Bonikowski & Gidron, 2016; DiMaggio et al., 2012; Kirgil & Voyer, 2022), class distinction (Kozlowski et al., 2019; Voyer et al., 2022), and social movements (Almquist &

Bagozzi, 2019; Nelson, 2021), among other topics. Technological developments, in particular the increased power provided by advances in AI, extend the possibilities for the interpretive analysis of textual data.

Researchers have long recognized that AI-driven technologies, such as natural language processing (NLP) and machine learning, can automate and streamline fundamental tasks in the research process, including extracting keywords, sentiments, or themes from large volumes of text (Do et al., 2022; Mohr 1998). These developments have contributed to the rise of an entire subfield, the Digital Humanities (see for example Distant Horizons). However, with some notable exceptions (c.f. Taylor & Stoltz, 2020; Kozlowski et al., 2019; Bodell, Magnusson, & Keuschnigg, 2022), AI driven-technologies have not been widely adopted by qualitative sociologists, and instead tend to be used as an add-on or supplement to social statistics in quantitative research instead of interpretive analysis (Nelson, 2021).

An interpretive analysis, at its core, positions human meaning-making at the center of the research endeavor, implying that meaning is neither given nor fixed but constructed and dynamic (Schwartz-Shea, 2014). Consequently, the discourses, actions, artifacts, and text constitute the social phenomena of interest. Interpretation-based analysis is distinct from dictionary-based methods, which, for example, use the presence or frequency of key words to determine what category a selection of text may belong to (Grimmer & Stewart, 2013). While valuable for the scalability of qualitative work, our recommendations are not targeted towards this type of analysis. Rather, we propose a framework for using LLMs to augment the interpretation of the meaning of texts.

RESULTS

Using the example of finding political lifestyle bundles in Twitter bios and exploring norms around smoking across time in etiquette books, we found that using the LLM in both deductive and inductive research tasks, augmented the interpretive research process, making it possible to achieve more and be more confident in our findings. We determined that the research process can incorporate LLMs when it comes to qualitative coding, including both index coding, classification, and thematic analysis, and in responding to the researchers' analytical queries.

More precisely, when it came to qualitative coding and classification tasks, the accessibility and power of the model streamlined and accelerated the research process, and the iterative nature of the human-machine interaction prompted transparency and clarity in the construction of concepts and qualitative codes. In determining how to prompt the LLM to produce the results, we needed to be explicit about background ideas and concept that might problematically remain more implicit in fully human interpretive analysis. In addition to requiring clarity, ChatGPT took on the role of a research assistant and a research interlocuter supplying additional insights into the texts we were analyzing by observing things we missed in our human analyses. In this way, LLMs also provide a robustness check similar to inter-coder reliability by corroborating and challenging the researcher's interpretations and conclusions.

In addition to coding, classification and thematic analysis, LLMs can be prompted to provide a higher-level analysis of the data. In our experience with higher-level analytical queries, we found that LLMs are more limited and cannot be relied on to provide a full analysis. The human researcher is still the key to good interpretive research. Lacking specific subject area expertise and trained to produce answers without an eye to their accuracy, LLMs' responses to analytical queries must be treated with caution and be subject to substantial empirical verification. Only the researcher is in a position to conduct an analysis that takes full account of the broader academic literature, weighs elements of the text data on the basis of their significance, incorporates key theories and context, and keeps in mind the practical and ethical concerns that are fundamental to interpretive research in sociology and the other social and human sciences.

CONCLUSION

The emergence of large-language models has unlocked vast potential in various domains of natural language processing. These models, such as GPT-3, possess the ability to understand and generate human-like text, making them valuable tools for a wide range of applications. In this paper, we argue that LLMs can be one way of augmenting the qualitative researcher and perform tasks much like a research assistant.

We argue LLMs are best suited for interpretive research when they are working on the classification of specific texts. LLMs excel in text-related tasks, such as answering questions, summarizing information, or even engaging in conversation. They can be particularly useful for classifying data with codes specified by the researcher beforehand. They can also illuminate latent patterns in text in an efficient manner. However, the results of prompts asking LLMs to produce a higher level or second order analyses through analytical queries are less reliable must be subject to substantial verification and coupled with the researchers' own careful and independent analysis of the data. It is important to note that LLMs are not designed for producing factual information or performing complex data analysis tasks. For higher level analysis and critical decision-making, we still need to rely on human researchers. LLMs are tools to assist in the process rather than to replace human expertise.

In conclusion, LLMs present significant potential in the realm of interpretive research. Their cost and time effectiveness make them valuable assets for text-related tasks. By utilizing these models responsibly and leveraging their strengths in reading and recognizing text, we can benefit from their capabilities while being cautious not to rely on them for factual information or complex analyses. With careful consideration, LLMs can enhance various stages of interpretive research and complement the researcher's domain-specific, expert knowledge.

References

Almquist, Z. W., & Bagozzi, B. E. (2019). Using radical environmentalist texts to uncover network structure and network features. *Sociological Methods & Research*, *48*(4), 905-960.

Bodell, M. H., Magnusson, M., & Keuschnigg, M. (2022). Seeded Topic Models in Digital Archives: Analyzing the Swedish Understanding of Immigration, 1945--2019.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29* (pp. 4349–4357). Curran Associates, Inc.

Bonikowski, B., & Gidron, N. (2016). The Populist Style in American Politics: Presidential Campaign Discourse, 1952–1996. *Social Forces*, *94*(4), 1593–1621.

Boutyline, A., Arseniev-Koehler, A., & Cornell, D. J. (2020). School, Studying, and Smarts: Gender Stereotypes and Education Across 80 Years of American Print Media.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, *41*(6), 570-606.

Do, S., Ollion, É., & Shen, R. (2022). The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. *Sociological Methods & Research*. https://doi.org/10.1177/00491241221134526\

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635-E3644.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Kirgil, Z. M., & Voyer, A. (2022). "Do your part: Stay apart": Collective intentionality and collective (in) action in US governor's COVID-19 press conferences. *Poetics*, *93*, 101668.

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Mohr, J. W. (1998). Measuring Meaning Structures. *Annual Review of Sociology*, *24*(1), 345–370. https://doi.org/10.1146/annurev.soc.24.1.345

Nelson, L. K. (2021). Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century U.S. South. *Poetics*, *88*, 101539. https://doi.org/10.1016/j.poetic.2021.101539

Taylor, M. A., & Stoltz, D. S. (2020). Concept class analysis: A method for identifying cultural schemas in texts. *Sociological Science*, *7*, 544-569.

Voyer, A., Kline, Z. D., & Danton, M. (2022). Symbols of class: A computational analysis of class distinction-making through etiquette, 1922-2017. Poetics, 94, 101734.