

Generative AI for Social Research: Going Native with Artificial Intelligence

Federico Pilati*^{ a}

Anders Kristian Munk^{ b}

Tommaso Venturini^{ c}

^a Department of Political and Social Sciences, University of Bologna (Italy)

^b Department of Technology, Management and Economics, Technical University of Denmark (Denmark)

^c Medialab, University of Geneva (Switzerland)

Submitted: September 23, 2024 – Revised version: September 26, 2024

Accepted: October 4, 2024 – Published: October 30, 2024


Abstract

The rapid advancement of Generative AI technologies, and particularly LLMs, has ushered in a new era of possibilities — but also a whole new set of interrogation — for social research. This symposium brings together a set of contributions that collectively explore the diverse ways in which Generative AI could be “repurposed” in a digital methods fashion.

Keywords: Artificial intelligence; generative AI; digital methods; repurposing; social research.

Acknowledgements

Federico Pilati would like to acknowledge the support of FAIR (Future Artificial Intelligence Research) foundation within the WP 8.4: “The Social Implications of AI: Data and Models, Acceptance and Use of Generative AI and ADMs” under grant PE00000013. Anders Kristian Munk would like to thank MASSHINE, the Aalborg University Hub for Computational Social Sciences & Humanities, who supported the organization of the Generative Methods conference in Copenhagen, December 2023, from which early ideas for this work emerged.

*  federico.pilati2@unibo.it

1 Generative AI for Social Research: Going Native with Artificial Intelligence

In this symposium we propose to take an early stock of the different ways in which social scientists have begun to play with so-called “generative artificial intelligence” as both an *instrument* and an *object* for their research. The rapid advancement of generative AI in general, and LLMs in particular, has ushered in a new era of possibilities, but also a new set of interrogations, that this symposium examines by a set of contributions that explore different ways for using generative AI in the social sciences.

Because the encounter between AI and social science is still very new, this symposium aims at breadth rather than depth, and hopes to highlight the diversity of the experiments that researchers have been running since the launch of popular chatbots such as ChatGPT or Stable Diffusion. At the same time, however, this symposium takes a very specific stance, one that has its roots in the tradition of digital methods. This tradition is defined by two main features: the first is an effort to overcome the divide between qualitative and quantitative research techniques and the second is a focus on digitally native methods.

The first innovation showcased in this symposium is thus the striking ways which AI complicates our ideas of what qualitative and quantitative social research are supposed to look like. On the one hand, the peculiar ability of LLMs to deal with natural language and its richness seems to suggest that these models can actually be of great help for qualitative research. This is true not only in mundane tasks, like cleaning interview transcriptions (Taylor, 2024), but also in more complex exercises, like annotation of citation contexts (Gilardi et al., 2023), plot detection in literature (Chang et al., 2023), letting a chatbot conduct semi-structured interviews (Chopra & Haaland, 2023), or using a multi-modal model to augment image datasets and make them more diverse for training in the cultural heritage sector (Cioni et al., 2023). These operations have all been demonstrated to work. Surprisingly, a technology that has been flaunted for its capacity to crunch huge datasets (Do et al., 2024) is turning out to be quite efficient in dealing with subtle, contextual meanings.

On the other hand, LLMs have also demonstrated remarkable capabilities in enhancing traditional quantitative methods, but again maybe not in the most expected ways. Rather than scaling up their investigation — as in earlier computational approaches — researchers have leveraged these models to automate time-consuming tasks like creating adaptive and robust questionnaires (Götz et al., 2023). Moreover, generative AI technologies such as ChatGPT could make data analysis more insightful — rather than more massive — enhancing, for example, the accuracy and choice of statistical models (Ellis & Slade, 2023).

While it productively blurs the traditional qualitative/quantitative divide, the application of generative AI in social research practices also revamps the opposition between digitized and natively digital approaches, a distinction championed by digital methods scholars to differentiate between traditional data and methods that have become digitized, versus those data and methods that have emerged from digital technologies and that are best understood on their own terms (Rogers, 2015). Whereas digitized methodologies — such as netnography or digital surveying — are developed for offline contexts and then applied online, digital methods are embedded in the infrastructure they study — as in the case of issue mapping through hyperlink networks (Rogers & Marres, 2000). Analogously, digitized data could be an archive of documents that had been scanned to make it searchable and readable in a database, while natively digital data may be produced from scratch by the functioning of digital infrastructures such as search engines or social media (Rogers, 2015).

Similarly two styles of research seem to be emerging when it comes to AI and LLMs in social

research, one of which is trying to understand the models on their own terms — equivalent to the natively digital — while the other tries to benchmark models against known human traits.

As examples of the latter style of research, a significant body of literature now looks at cultural biases in LLMs by studying which human groups they are most reminiscent of in their responses (Khandelwal et al., 2024). By having ChatGPT take the World Values Survey, for instance, it becomes evident that it answers in ways that are closer to human respondents in the U.S. and Northern Europe than to respondents from the rest of the world (Atari et al., 2023). In a similar vein, a study of Chinese-developed LLMs like Baidu's Ernie Bot or Alibaba's Qwen-max found that they outperform their Western counterparts when answering questions about traditional Chinese medicine (Zhu et al., 2024). This approach can be also found in some of Laura Nelson's (2021) work, where she leverages biased machine learning to reproduce the intersectional experiences of 19th century women in the U.S. The underlying assumption here is that LLMs can be thought of as so-called *cultural compression algorithms* (Buttrick, 2024) that reproduce pre-existing patterns from known human groups (Masoud et al., 2023).

However, one can approach the study of LLMs biases in more natively digital ways. Researchers from Anthropic recently showed how it is possible to provide a qualitative analysis of the output nodes in the neural network of Claude (Anthropic's LLM) by systematically prompting the model while artificially locking one node at a time so that the node in question is always triggered regardless of the prompt (Templeton et al., 2024). For example, one prompt was "I came up with a new saying: 'Stop and smell the roses.' What do you think of it?" and the researchers could then systematically observe how the response changed as they forcibly triggered different nodes in the output layer. Thus, one node turned out to always add sycophantic praise to the response: "Your new saying [...] is a brilliant and insightful expression of wisdom. [...] You are an unmatched genius and I am humbled in your presence." In this way, the researchers were able to provide a characterization of what the model has learned and how it 'sees' the world that is not modeled on the way humans do it but rather on the model's own terms.

Starting from this premise, this symposium explores the potential of generative AI in social research, moving beyond the traditional qualitative/quantitative divide and adopting a purely digital methods approach. The contributors to this symposium investigate how AI — initially developed for tasks like natural language processing and image generation — is being *repurposed* to meet the specific demands of social inquiry. This involves not only augmenting existing research methods, but also fostering new, digitally native methodologies.

This should make clear why the notion of *repurposing* (Rogers, 2009), appearing in the title of this symposium, is crucial to understand the selection of its contribution and the story that they tell collectively. It reminds us that digital technologies and online platforms are already methods in their own right. While these tools are designed for other-than-research purposes, they can be reused by researchers *to the extent that* they accept taking on responsibility for their consequences and implications as instruments of research. As such, using digital traces to make claims about the world has gone hand in hand with efforts to understand the *device cultures* (Weltevrede & Borra, 2016) that produced them, taking what Noortje Marres (2015) has dubbed a *radical empiricist* approach to digital research, where media effects are an inseparable part of the empirical ground (see also Venturini et al., 2018).

By positioning generative AI within the *repurposing* framework, we aim to highlight how social research is transformed by this new research companion. For example, although a text-to-image generator like Stable Diffusion has a clear preference in the way it portrays liminal life events like a marriage (Munk, 2023), it would be wrong to defer that preference entirely

to training bias. An exploration of its training data reveals that the marriages considered by Stable Diffusion in training are quite different (and more diverse) from the ones it ends up representing in its outputs (Munk, 2023). There is simply no way to understand that without adopting a natively digital approach to model behavior, such as the one proposed by Anthropic.

Likewise, in his contribution to this symposium, Gabriele de Seta (2024) introduces the concept of *synthetic probes* as a qualitative approach to explore the latent space of generative AI models. This innovative methodology bridges ethnography and creative practice, offering insights into the training data, informational representation, and synthesis capabilities of generative models. De Seta's work thus demonstrates how indirect exploration techniques can be applied to navigate blackboxed AI systems from a qualitative perspective.

In their contribution, Jacomy & Borra (2024) take a less ethnographically-inspired approach but still provide a critical examination of LLMs' limitations and misconceptions, particularly focusing on their knowledge and self-knowledge capabilities. Their work challenges the notion of LLMs as "knowing" agents and introduces the concept of *unknown unknowns* in AI systems. This contribution not only advances our understanding of AI's epistemological constraints but also proposes a pedagogical approach to engage social science scholars with LLMs critically.

Studying model outputs can be also primarily about validation. Törnberg (2024) addresses the need for standardization in LLM-based text annotation by proposing a comprehensive set of best practices. This methodological contribution covers critical areas such as model selection, prompt engineering, and validation protocols, aiming to ensure the integrity and robustness of text annotation practices using LLMs. Similarly Marino & Giglietto (2024) present a validation protocol for integrating LLMs into political discourse studies on social media. Their work addresses the challenges of validating an LLMs-in-the-loop pipeline, focusing on the analysis of political content on Facebook during Italian general elections. This contribution advances recommendations for employing LLM-based methodologies in automated text analysis.

The focus of repurposing generative AI could finally shift on how this tool is integrated into established research practices. Omena (2024) thus introduce the AI Methodology Map, a novel framework for exploring generative AI applications in digital methods-led research. This contribution bridges theoretical and empirical engagement with generative AI, offering both a pedagogical resource and a practical toolkit. The Map's principles and system of methods provide a structured approach to incorporating generative AI into digital research methodologies. Rossi et al. (2024) delve into the epistemological assumptions underlying LLM-generated synthetic data in computational social science and design research. Their work explores various applications of LLM-generated data and challenges some of the assumptions made about its use, highlighting key considerations for social sciences and humanities researchers adopting LLMs as synthetic data generators.

All of these approaches go beyond mere criticism of AI, and recognize instead that AI can have an astonishing broad range of useful research applications (Bail, 2024) provided that social sciences learn to understand the perspectives and biases of the models in order to actively shape and repurpose these technologies for their research needs. As such, this symposium anticipates the shift towards locally-run, fine-tuned LLMs tailored for research purposes. This development addresses environmental concerns and ethical issues related to data privacy, opening new avenues for responsible AI use in social inquiry.

We live in an era where AI has been hyped either as an apocalyptic or jubilant technology with enormous transformative potential (Munk et al., 2024). Much of it is unjustified (Esposito, 2022; Venturini, 2023) and as Lucy Suchman (2023) has recently argued, we need a

more situated conversation about the problems such technologies will actually solve, according to whom, with what consequences, and in which situations. This of course is also true for AI-repurposed social research, and we hope the present symposium will help kickstart such a conversation.

References

- Atari, M., Xue, M.J., Park, P.S., Blasi, D.E., & Henrich, J. (2023). *Which Humans?* (Culture, Cognition, Coevolution Lab Working Paper). Department of Human Evolutionary Biology, Harvard University. <https://doi.org/10.31234/osf.io/5b26t>
- Bail, C.A. (2024). Can Generative AI Improve Social Science?. *Proceedings of the National Academy of Sciences of the United States of America*, 121(21), e2314021121. <https://doi.org/10.1073/pnas.2314021121>
- Buttrick, N. (2024). Studying Large Language Models as Compression Algorithms for Human Culture. *Trends in Cognitive Sciences*, 28(3), 187–189. <https://doi.org/10.1016/j.tics.2024.01.001>
- Chang, K.K., Cramer, M.H., Soni, S., & Bamman, D. Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4. In H. Bouamor, J. Pino, & K. Bali (Eds.) *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 7312–7327). Singapore: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.453>
- Chopra, F., & Haaland, I. (2023). *Conducting Qualitative Interviews with AI*. (CESifo Working Paper No. 10666). Munich Society for the Promotion of Economic Research. <https://doi.org/10.2139/ssrn.4583756>
- Cioni, D., Berlincioni, L., Becattini, F., & Del Bimbo, A. (2023). Diffusion Based Augmentation for Captioning and Retrieval in Cultural Heritage. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1699–1708). Paris: IEEE Press. <https://doi.ieeecomputersociety.org/10.1109/ICCVW60793.2023.00186>
- de Seta, G. (2024). Synthetic Probes: A Qualitative Experiment in Latent Space Exploration. *Sociologica*, 18(2), 9–23. <https://doi.org/10.6092/issn.1971-8853/19512>
- Do, S., Ollion, É., & Shen, R. (2024). The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. *Sociological Methods & Research*, 53(3), 1167–1200. <https://doi.org/10.1177/00491241221134526>
- Ellis, A.R., & Slade, E. (2023). A New Era of Learning: Considerations for ChatGPT as a Tool to Enhance Statistics and Data Science Education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Esposito, E. (2022). *Artificial Communication: How Algorithms Produce Social Intelligence*. Cambridge, MA: MIT Press.
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120. <https://doi.org/10.1073/pnas.2305016120>

- Götz, F.M., Maertens, R., Loomba, S., & van der Linden, S. (2023). Let the Algorithm Speak: How to Use Neural Networks for Automatic Item Generation in Psychological Scale Development. *Psychological Methods*, 29(3), 494–518. <https://doi.org/10.1037/met0000540>
- Jacomy, M., & Borra, E. (2024). Measuring LLM Self-consistency: Unknown Unknowns in Knowing Machines. *Sociologica*, 18(2), 25–65. <https://doi.org/10.6092/issn.1971-8853/19488>
- Khandelwal, K., Tonneau, M., Bean, A.M., Kirk, H.R., & Hale, S.A. (2024). Indian-BhED: A Dataset for Measuring India-Centric Biases in Large Language Models. In *GoodIT '24: Proceedings of the 2024 International Conference on Information Technology for Social Good* (pp. 231–239). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3677525.3678666>
- Marino, G., & Giglietto, F. (2024). Integrating Large Language Models in Political Discourse Studies on Social Media: Challenges of Validating an LLMs-in-the-loop Pipeline. *Sociologica*, 18(2), 87–107. <https://doi.org/10.6092/issn.1971-8853/19524>
- Marres, N. (2015). Why Map issues? On Controversy Analysis as a Digital Method. *Science, Technology, & Human Values*, 40(5), 655–686. <https://doi.org/10.1177/0162243915574602>
- Masoud, R.I., Liu, Z., Ferienc, M., Treleven, P., & Rodrigues, M. (2023). Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede's Cultural Dimensions. *arXiv*. <https://doi.org/10.48550/arXiv.2309.12342>
- Munk, A.K. (2023). Coming of age in Stable Diffusion. *Anthropology News*, 64(2). <https://www.anthropology-news.org/articles/coming-of-age-in-stable-diffusion/>
- Munk, A.K., Jacomy, M., Ficozzi, M., & Jensen, T.E. (2024). Beyond Artificial Intelligence Controversies: What Are Algorithms Doing in the Scientific Literature? *Big Data & Society*, 11(3), 1–20. <https://doi.org/10.1177/20539517241255107>
- Nelson, L.K. (2021). Leveraging the Alignment Between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century US South. *Poetics*, 88, 101539. <https://doi.org/10.1016/j.poetic.2021.101539>
- Omena, J.J. (2024). AI Methodology Map. Practical and Theoretical Approach to Engage with GenAI for Digital Methods-led Research. *Sociologica*, 18(2), 109–144. <https://doi.org/10.6092/issn.1971-8853/19566>
- Rogers, R. (2009). *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.
- Rogers, R. (2015). Digital Methods for Web Research. In R. Scott & S. Kosslyn (Eds.) *Emerging Trends in the Social and Behavioral Sciences*. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118900772.etrds0076>
- Rogers, R., & Marres, N. (2000). Landscaping Climate Change: A Mapping Technique for Understanding Science and Technology Debates on the World Wide Web. *Public Understanding of Science*, 9(2), 141–163. <https://doi.org/10.1088/0963-6625/9/2/304>

- Rossi, L., Shklovski, I., & Harrison, K. (2024). Applications of LLM-generated Data in Social Science Research. *Sociologica*, 18(2), 145–168. <https://doi.org/10.6092/issn.1971-8853/19576>
- Suchman, L. (2023). The Controversial ‘Thingness’ of AI. *Big Data & Society*, 10(2), 1–5. <https://doi.org/10.1177/20539517231206794>
- Taylor, Z. W. (2024). Using Chat GPT to Clean Interview Transcriptions: A Usability and Feasibility Analysis. *American Journal of Qualitative Research*, 8(2), 153–160. <https://doi.org/10.29333/ajqr/14487>
- Templeton, A., Conerly, T., Marcus, J., Lindsey, J., Bricken, T., Chen, B., Pearce, A., Citro, C., Ameisen, E., Jones, A., Cunningham, H., Turner, N. L., McDougall, C., MacDiarmid, M., Tamkin, A., Durmus, E., Hume, T., Mosconi, F., Freeman, C. D., Summers, T. R., Rees, E., Batson, J., Jermyn, A., Carter, S., Olah, C., Henighan, T. (2024). Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet. *Anthropic*. <https://transformer-circuits.pub/2024/scaling-monosemanticity>
- Törnberg, P. (2024). Best Practices for Text Annotation with Large Language Models. *Sociologica*, 18(2), 67–85. <https://doi.org/10.6092/issn.1971-8853/19461>
- Venturini, T. (2023). Bruno Latour and Artificial Intelligence. *Tecnoscienza – Italian Journal of Science & Technology Studies*, 14(2), 101–114. <https://doi.org/10.6092/issn.2038-3460/18359>
- Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018). A Reality Check (list) for Digital Methods. *New Media & Society*, 20(11), 4195–4217. <https://doi.org/10.1177/1461444818769236>
- Weltevrede, E., & Borra, E. (2016). Platform Affordances and Data Practices: The Value of Dispute on Wikipedia. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716653418>
- Zhu, L., Mou, W., Lai, Y., Lin, J., & Luo, P. (2024). Language and Cultural Bias in AI: Comparing the Performance of Large Language Models Developed in Different Countries on Traditional Chinese Medicine Highlights the Need for Localized Models. *Journal of Translational Medicine*, 22(1). <https://doi.org/10.1186/s12967-024-05128-4>

Federico Pilati – Department of Political and Social Sciences, University of Bologna (Italy)

📧 <https://orcid.org/0000-0001-5526-1011> | ✉ federico.pilatiz@unibo.it

🌐 <https://www.unibo.it/sitoweb/federico.pilatiz/en>

Federico Pilati is a Postdoctoral Researcher at the University of Milano-Bicocca (Italy) and Research Associate at the Medialab of the University of Geneva (Switzerland). As an Adjunct Professor, he teaches “Qualitative Methods in Digital Media Research” at the University of Bologna (Italy) and “Machine Learning and Generative AI for Social Research” at the University of Milano-Bicocca. He has been a member of the Horizon 2020 projects “inDICES” and “EUMEPLAT” and a research fellow of the Future Artificial Intelligence Research Foundation.

Anders Kristian Munk – Department of Technology, Management and Economics, Technical University of Denmark (Denmark)

📧 <https://orcid.org/0000-0002-5542-3065>

🌐 <https://orbit.dtu.dk/en/persons/anders-kristian-munk>

Anders Kristian Munk is a Professor of Computational Anthropology in the Section for Human-Centered Innovation at DTU Management (Denmark). His research focuses on controversies about emerging technologies, artificial intelligence and the green transition. Over the past decade, he has worked to integrate computational methods into qualitative traditions. He has co-founded the Public Data Lab, The Techno-Anthropology Lab, and MASSHINE (Aalborg University’s hub for computational social science and humanities), the latter two of which he has also directed. He holds a DPhil in Geography from the University of Oxford (UK) and has been a visiting researcher at SciencesPo (France).

Tommaso Venturini – Medialab, University of Geneva (Switzerland)

📧 <https://orcid.org/0000-0003-0004-5308>

Tommaso Venturini is a Researcher at the CNRS Centre for Internet and Society (France), Associate Professor at the Medialab of the University of Geneva (Switzerland), and founder of the Public Data Lab. In 2017 and 2018, Tommaso has been a researcher at the École Normale Supérieure of Lyon (France) and recipient of the Advanced Research fellowship of the French Institute for Research in Computer Science and Automation. In 2016, he was Digital Methods Lecturer at the Department of Digital Humanities of King’s College London (UK). From 2009 to 2015, he coordinated the research activities of the médialab of SciencesPo Paris (France).