

Introduction to the Third Workshop on Humanities-Centred Artificial Intelligence

Sylvia Melzer^{1,2}, Hagen Peukert¹ and Stefan Thiemann¹

¹Universität Hamburg

²Universität zu Lübeck

Since Humanities-Centred Artificial Intelligence (CHAI) was suggested as an emerging paradigm [1], research in this area has revealed methodological advancements and clear directions towards machine training. This implies a methodological focus on the subject of Humanities, which is indeed maintained throughout the current and past workshops. It also entails the underlying question how the machine can help to more efficiently approach research questions in the Humanities or devise new ones that cannot be followed otherwise. Yet, this year, a broad societal discussion on conversational agents and its consequences for the society heralds a second strain of research in this area that is implicitly given in the CHAI-roadmap, but never actively pursued: ideology. This classic of Humanities pushes the human being, its properties, behavior, and needs to the forefront whereas the methodology follows suit.

These two perspectives, ideology and method, do not contradict, but complement each other. They may set forth a positive view of technology and, assumably, positive net effects on the human being. And, chances are high that scientific endeavors in Humanities-Centred AI shift in part more decidedly to the tenets of strong AI. An ideological perspective would embrace ethical issues such as a privacy and accountability paradox that is driven by the inherent inscrutability property of deep nets [2], but also the evaluation of fact and fake as well as other information content brought about by bottom up processes of new media. In a Habermasian understanding [3, 4] the structural change of the public no longer converges on an agreement by discourse [5]. It is completely open as to which truth and commitment can be achieved in the public. Equally relevant to the focus of ideological thinking in Human-Centred AI could be the consequences of speech and language technologies, e.g. conversational agents, to the labor market, education, or cultural industries like (script) writers and media production.

While practical AI applications that we observe in our daily life are developed by keeping a methodological focus in the first place, the consequences of having them will be reflected in the

Humanities-Centred AI (CHAI), Workshop at the 46th German Conference on Artificial Intelligence, September 26, 2023, Berlin, Germany


✉ sylvia.melzer@uni-hamburg.de (S. Melzer); hagen.peukert@uni-hamburg.de (H. Peukert); stefan.thiemann@uni-hamburg.de (S. Thiemann)

🌐 <https://www.csmc.uni-hamburg.de/about/people/melzer.html> (S. Melzer)

🆔 0000-0002-0144-5429 (S. Melzer); 0000-0002-3228-316X (H. Peukert); 0000-0001-8300-2519 (S. Thiemann)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

ideological perspective of Humanities-Centred AI, which, in turn, will require other methods. To illustrate, advances in modeling natural language has evoked a range of applications that change the way, in which text production is taught and examined at schools and universities. These changes now require new technological means of observation and checking plagiarism. Thus, some circular feedback loops emanate from the initial advancement. This, however, is not supposed to mean downsizing the focus on methods. In fact, once a self reinforcing cycle has started, continuous improvement on the methods seem to be the best solution to flatten the spiral over time as it is the case for other technological innovations. Consequently, CHAI also presents seven papers centering around methodological solutions of diverse challenges across the Humanities. In the rapidly evolving research and technology landscape, the contributions in this volume highlight the latest advances and challenges approaches from computer science to be established in the Humanities. The contributions cover a wide range of topics.

The first regular paper addresses the challenges of reusing data in research archives, even when following established guidelines. Innovative solutions are proposed to make research data more accessible and user-friendly.

The second regular paper introduces FrESH, an approach for enhancing Subjective Content Descriptions (SCDs) [6] in a model using human feedback. It focuses on improving models' accuracy by incorporating feedback without the need for complete retraining.

The third regular paper explores the application of latent Dirichlet allocation (LDA) [7] in uncovering hidden thematic structures within a specific domain, academic journals focused on modern and ancient manuscripts. While LDA is commonly used for various types of text data, its behavior in highly domain-specific corpora is less understood. The paper discusses the insights gained from applying LDA to this specialized corpus, shedding light on steps specific to dealing with domain-specific data.

The fourth regular paper discusses the challenges faced by humanities scholars when fine-tuning Large Language Models (LLMs) [8, 9, 10], such as BERT [11], for domain-specific tasks with limited training data. It emphasizes the increasing availability of research data in information systems as a valuable resource for fine-tuning these models. The paper presents a novel method for fine-tuning BERT models on-demand, using training data from pre-modern Arabic as an example. In addition, the paper presents the development of a Humanities Aligned Chatbot that utilizes the fine-tuned model to make LLMs more accessible in humanities research.

The fifth regular paper proposes the use of a federated cross-domain information system to supplement missing research data in humanities projects. It demonstrates how an indexing approach can be integrated into an federated information system for efficient federated information retrieval, addressing challenges in presenting search results from diverse information sources. In addition, it will be discussed how users can interact with the system by using natural language queries integrated with GPT-4¹ to generate SQL queries. The result is a cross-domain information system that facilitates comprehensive research in the humanities by bringing together multiple sources of information and facilitating efficient, federated information retrieval.

The first short paper presents a new approach to assessing responsible AI that combines the results of a literature review with an evaluation framework. It provides an overview of the

¹<https://gpt4all.io/index.html>

responsible use of AI, presents evaluation metrics tailored to humanities data, and introduces VERIFAI, an example implementation of the evaluation framework.

The second short paper explores how recent advancements in research are enabling the analysis of different modalities in historical artefacts. This work discusses the potential applications of vision-language models in the context of historical research.

In essence, these seven contributions collectively highlight ongoing efforts to utilise advanced technologies, improve data-driven research and bridge the gap between AI capabilities and domain-specific needs. They offer promising solutions for more efficient, accurate and accessible research in the humanities domain.

References

- [1] R. Möller, *Humanities-Centred Artificial Intelligence (CHAI) as an Emerging Paradigm*, De Gruyter, Berlin, Boston, 2021, pp. 245–266. URL: <https://doi.org/10.1515/9783110753301-013>. doi:doi:10.1515/9783110753301-013.
- [2] H. Peukert, *Inscrutability versus Privacy and Automation versus Labor in Human-Centered AI: Approaching Ethical Paradoxes and Directions for Research*, in: M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (Eds.), *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*, volume 35 of *Annals of Computer Science and Information Systems*, IEEE, 2023, p. 1101–1105. URL: <http://dx.doi.org/10.15439/2023B7504>. doi:10.15439/2023B7504.
- [3] J. Habermas, *Strukturwandel der Öffentlichkeit*, Suhrkamp, Frankfurt am Main, 1962.
- [4] J. Habermas, *Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit*, Sonderband Leviathan, 1 ed., Nomos, Baden-Baden, 2021, pp. 470–500.
- [5] J. Habermas, *Der philosophische Diskurs der Moderne. Zwölf Vorlesungen*, Suhrkamp, Frankfurt am Main, 1981.
- [6] F. Kuhr, T. Braun, M. Bender, R. Möller, *To Extend or not to Extend? Context-specific Corpus Enrichment*, in: *Proceedings of AI 2019: Advances in Artificial Intelligence*, volume 11919 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 357–368. URL: https://doi.org/10.1007/978-3-030-35288-2_29.
- [7] D. M. Blei, A. Y. Ng, M. I. Jordan, *Latent dirichlet allocation*, *J. Mach. Learn. Res.* 3 (2003) 993–1022.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, *Attention is all you need*, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30, Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [9] C. Lei, *Unsupervised Learning: Word Vector*, Springer Singapore, Singapore, 2021, pp. 95–149. URL: https://doi.org/10.1007/978-981-16-2233-5_7. doi:10.1007/978-981-16-2233-5_7.
- [10] D. Rothman, *Transformers for Natural Language Processing: Build innovative deep neural*

- network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more, Packt Publishing, 2021. URL: <https://books.google.de/books?id=Cr0YEAAAQBAJ>.
- [11] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: <http://arxiv.org/abs/1810.04805>. arXiv:1810.04805.