Introduction to the Fourth Workshop on Humanities-Centred Artificial Intelligence

Sylvia Melzer[1,2,3], Hagen Peukert[4], Stefan Thiemann[4] and Erik Radisch[5]

[1] *University of Hamburg, Institute for Humanities-Centered AI (CHAI), Warburgstraße 28, 20354 Hamburg, Germany*

[2] *University of Hamburg, Centre for the Study of Manuscript Cultures (CSMC), Warburgstraße 26, 20354 Hamburg, Germany*

[3] *University of Hamburg, Cluster of Excellence 'Understanding Written Artefacts' (UWA), Warburgstraße 26, 20354 Hamburg, Germany*

[4] *University of Hamburg, Center for Sustainable Research Data Management, Monetastraße 4, 20146 Hamburg, Germany*

[5] *Sächsische Akademie der Wissenschaften zu Leipzig, Karl-Tauchnitz-Str. 1, 04107 Leipzig, Germany*

Abstract

Artificial Intelligence (AI), as the science of agents acting in the world, offers significant support to research in the Humanities by enhancing efficiency and effectiveness. By adopting a Humanities-centered approach, scholars can tailor AI methods to specific needs. AI methods, developed within the science of human-machine interaction, can assist in interpreting ancient cultural traditions from written artefacts, optimizing processes such as text mining and linguistic analysis. The practical implementation of methods, derived from the science of AI, requires focused development to address specific Humanities challenges and optimize human-machine interaction in this field.

1. Organising Committee

- Dr Sylvia Melzer, Universität Hamburg & Universität zu Lübeck
- Dr Stefan Thiemann, Universität Hamburg
- Dr Hagen Peukert, Universität Hamburg
- Dr Erik Radisch, Sächsische Akademie der Wissenschaften zu Leipzig

2. Program Committee

- Thomas Asselborn, Universität Hamburg
- Prof Dr habil Meike Klettke, Universität Regensburg
- Dr Sylvia Melzer, Universität Hamburg & Universität zu Lübeck
- Dr Hagen Peukert, Universität Hamburg
- Dr Erik Radisch, Sächsische Akademie der Wissenschaften zu Leipzig
- Dr Stefan Thiemann, Universität Hamburg

3. Preface

Our view of artificial intelligence (AI) is the science of agents acting in the world. [1]. Agents receive precepts from the environment and take action. [2] An intelligent agent does an action with the aim to achieve a local optimum. Achieving a local optimum implies the focus on maximizing performance

within a specific environment or task while recognizing that global optimization may be impractical or unnecessary in some cases. In many real-world scenarios, the goal is not absolutely perfect but "good enough" solutions that meet the needs of the current context.

AI is currently on a grand triumphant advance in all parts of society. This advance does not stop at the humanities. Humanities-Centred Artificial Intelligence (CHAI) was suggested as an emerging paradigm in the article [3], and in the fourth CHAI workshop, we will highlight human-machine interactions through a series of current research projects that emphasise the role of data usability, computational methods and the use of large language models (LLMs) [4, 5] in various research areas, especially in the humanities and law.

The article [6] deals with the application of (intelligent) agents in the digital humanities, especially in the field of text analysis. It emphasises that such agents offer new possibilities for analysing and interpreting texts that complement and support the work of humanities scholars rather than replacing it. The focus of the CHAI contributions is also on complementing and supporting the work of humanities scholars. The use of intelligent agents in the humanities requires close collaboration between humanities scholars and AI experts, which can lead to new insights and methods in both fields, as the following articles also demonstrate.

The first paper *Automate Text Processing for Schematically Analyzing Legal Texts* presents an innovative approach to the use of LLMs for processing legal texts and addresses their limitations. Given the complexity and constant evolution of legal documents, the authors propose a method for automatically extracting schematic representations that enables intelligent agents to make informed decisions based on structured information. The method includes a legal case study and outlines a process for modelling and extracting these schemas using LLMs. The paper also evaluates the capabilities of ChatGPT and Gemini in this context. While the authors focus primarily on legal texts, they suggest that their approach could be adapted for different types of natural language texts to improve decision making in different domains.

The second article *From Data Acquisition to Latent Semantic Analysis: Developing VERITRACE's Computational Approach to Tracing the Influence of Ancient Wisdom in Early Modern Natural Philosophy* focuses on the application of latent semantic analysis (LSA) [7] to uncover historical connections and influences. This computational approach not only contributes to a better understanding of ancient philosophies, but also illustrates the broader implications of LSA when analysing large text corpora.

The third article *Retrieving Information Presented on Webpages Using Large Language Models: A Case Study* demonstrates the potential of LLMs in improving information retrieval from digital sources. This is in line with ongoing research on the potential of LLMs to improve the accessibility and usability of data in various domains.

The forth article *Testing the Syntactic Competence of Large Language Models with a Translation Task* includes a discussion of the use of translation tasks as a method for testing the syntactic competence of LLMs, particularly in the treatment of dative ambiguity in Russian. This research emphasises the importance of language processing in the evaluation of agents' LLMs and their ability to process complex linguistic structures.

In the fifth article *Tracing the Palola Shahi Royal Genealogy by Fusing LLMs and Databases?: A Case Study*, research into tracing royal genealogies, such as the Paḷola Ṣāhi lineage (for more details see [8, 9]), through the fusion of LLMs and databases illustrates the innovative applications of agents in historical research. This case study highlights the potential of interdisciplinary collaboration that combines computational techniques with historical research.

The first invited article *Humanities in the Center of Data Usability: Data Visualization in Institutional Research Repositories* sets the stage by emphasizing the critical need for effective data visualization techniques that enhance the reusability and interoperability of research datasets. In addition, an innovative citation approach is presented that makes it possible to refer not only to the entire repository, but also to a specific data set. The second invited article on *Human-Centred Open-Source Automatic Text Recognition for the Humanities with OCR4all* emphasises the need for user-friendly tools that empower researchers in the humanities to a mass data analysis using software tools effectively. Both contributions are in line with the general trend towards the development of open source and generic solutions that improve the findability, accessibility, interoperability and reusability of data.

In summary, these articles give an overview about the significant advances in data utilisation according to FAIR principles, computational methods and the use of LLMs, and demonstrate their impact on different areas of research. The integration of LLMs into legal and humanities research not only streamlines processes but also opens up new ways of study and understanding.

4. Presentations

Abstracts and presentations are available at: https://doi.org/10.25592/uhhfdm.15984

Keynote: Humanities in the Center of Data Usability: Data Visualization in Institutional Research Repositories
Hagen Peukert, Lucas Voges, Sylvia Melzer

From Data Acquisition to Latent Semantic Analysis: Developing VERITRACE's Computational Approach to Tracing the Influence of Ancient Wisdom in Early Modern Natural Philosophy
Jeffrey Wolf Vrije

Automate Text Processing for Schematically Analyzing Legal Texts
Magnus Bender

Retrieving Information Presented on Webpages Using Large Language Models: A Case Study
Thomas Asselborn, Karsten Helmholz, Ralf Möller

Testing the Syntactic Competence of Large Language Models with a Translation Task
Edyta Jurkiewicz-Rohrbacher

Tracing the Palola Shahi Royal Genealogy by Fusing LLMs and Databases?: A Case Study
Hui Xu, Thomas Asselborn, Haiyan Hu-von Hinüber, Oskar von Hinüber, Sylvia Melzer

Invited presentation: Human-Centred Open-Source Automatic Text Recognition for the Humanities with OCR4all
Christian Reul, Maximilian Nöth, Herbert Baier, Florian Langhanki, Kevin Chadbourne

Funding Information

References

[1] S. Melzer, R. Möller, GenAI in Education, Science, and Society (13 Vorlesungen, University of Hamburg), Präsentationen/Dias, 2024. URL: https://www.edit.fis.uni-hamburg.de/ws/files/56466394/GenAI_2024.zip.

[2] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, 3 ed., Prentice Hall, 2010.

[3] 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.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

[5] OpenAI, Better language models and their implications, 2019. URL: https://openai.com/blog/better-language-models/, archived from the original on 2020-12-19.

[6] J. Chun, K. Elkins, The crisis of artificial intelligence: A new digital humanities curriculum for human-centred ai, International Journal of Humanities and Arts Computing 17 (2023) 147–167. doi:10.3366/ijhac.2023.0310.

[7] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman, Indexing by latent semantic analysis, Journal of the American society for information science 41 (1990) 391–407.

[8] W. Luo, O. von Hinüber, News from palola: The jokhang and the yong-he inscriptions of surendrāditya, in: N. Kudo (Ed.), Śāntamatiḥ - Manuscripts for Life, volume 15 of *Bibliotheca Philologica et Philosophica Buddhica*, International Research Institute for Advanced Buddhology, Soka University, Tokyo, 2023, pp. 207–223.

[9] O. von Hinüber, Die Palola ṣāhis: Ihre Steininschriften, Inschriften auf Bronzen, Handschriftenkolophone und Schutzzauber: Materialien zur Geschichte von Gilgit und Chilas, Antiquities of Northern Pakistan, Heidelberg Academy of Sciences and Humanities, Mainz, 2016. URL: https://d-nb.info/1123441529/34.