Debating AI in archaeology: applications, implications, and ethical considerations

Tenzer, M.; Pistilli, G.; Brandsen, A.; Shenfield, A.

Debating AI in Archaeology: Applications, Implications, and Ethical Considerations

MartinaTenzer[1], Giada Pistilli[2], Alex Brandsen[3]and Alex Shenfield[4]

[1]University of York, Department of Archaeology, York, UK
[2]Sorbonne University, Department of Philosophy, Paris, France
[3]Leiden University, Faculty of Archaeology, Leiden, The Netherlands
[4]Sheffield Hallam University, Institute of Engineering and Mathematics, Sheffield, UK
Author for correspondence: martina.tenzer@york.ac.uk

Abstract

Artificial Intelligence (AI) is not a recent development. However, with increasing computational capabilities, AI has developed into Natural Language Processing and Machine Learning, technologies particularly good at detecting correlations and patterns, and categorising, predicting, or extracting information. Within archaeology, AI can process big data accumulated over decades of research and deposited in archives. By combining these capabilities, AI offers new insights and exciting opportunities to create knowledge from archaeological archives for contemporary and future research. However, ethical implications and human costs are not yet fully understood. Therefore, we question whether AI in archaeology is a blessing or a curse?

Keywords

Artificial Intelligence, Archaeology, Heritage Studies, Ethics

Introduction

Although it might seem so, given the current AI hype around Large Language Models (LLMs) and generative AI models for content generation (such as ChatGPT), Artificial Intelligence is not a recent development. Deployment of the technology in the fields of archaeology and heritage studies with both object and remote sensing applications has been widely documented (Bickler 2021). With recent developments and advances of AI tools in the field of text-based analysis, this will be the primary focus of this paper.

The term Artificial Intelligence was coined in 1956 (Russell and Norvig 2016) describing a hypothetical computer technology developed by Alan Turing (Turing 1950). Following the first AI hype of the 1950s and 60s – over-promising the capabilities of AI technology but under-performing due to the lack of computational power – AI research was interrupted by the AI winter of the 1970s and early 1980s. However, after 60 years of exponential growth, AI tools have now entered the mainstream. Examples include chess computers, recommendation systems, and spam filters. Other applications are now leveraging the recent developments in LLMs, for example, the Google search function, instant translations, and closed captioning.

Increasing computational capabilities enabled the development of Machine Learning (ML) and Neural Networks (NN). In particular, Deep Learning with its ability to learn features of interest in parallel, e.g., the attention mechanism in LLMs, pushed AI capabilities. These systems are particularly good at detecting correlations and patterns, and can categorise, predict, or extract data in the context of natural language processing. LLMs, such as Google's BARD, OpenAI's ChatGPT, or Meta's LLaMA now form the basis of a new generation of Open Source LLMs, such as Open Assistant (Köpf et al. 2023). These tools can learn and draw from extensive datasets that are based on the wide knowledge of the Internet, including data from, for example, Wikipedia, GitHub, and Google data search.

Following an early adoption of AI technologies in archaeology for objects and remote sensing applications (Bickler 2021; Argyrou and Agapiou 2022), NLP, ML and DL are now being used for processing vast amounts of data accumulated over decades of research. This knowledge deposited in archives and grey literature can be efficiently analysed, structured, and disseminated using AI technologies – an approach that offers new insights and knowledge extraction from archaeological archives as never before.

However, while the deployment of AI technologies based on LLMs are capable of processing big data in archaeology and other fields, their application also has ethical implications. The lack of transparency of content and quality of the training data has been shown to reinforce social inequalities, misinformation, privacy issues, racial discrimination, risk to natural resources, and human workforce exploitation. Some of these are the same concerns across the discipline of archaeology and CHM, specifically

regarding sensibilities around privacy, bias, and model creation in the context of policy and decision-making.

In this paper, we focus on archaeology as part of that wider debate and present examples of successful AI applications in archaeology with text-based analysis as primary focus. We then provide insight into the ethical implications associated with AI before discussing the implications and applications of AI in a safe, sustainable, and socially just way in future. Finally, we want to open the discussion to the question if AI is a blessing or a curse for the discipline.

Applications of AI in archaeology and CHM

Archaeologists have a long tradition of adopting, adapting, and introducing technologies from other disciplines. For example, the pantograph preceded digital photography or survey methods (Novaković 2018) while Lidar has proved useful for detecting sites particularly across difficult terrain (Cohen, Klassen and Evans 2020). AI image recognition techniques were introduced in archaeology for remote sensing (Verschoof-van der Vaart et al. 2020) and object recognition (Anichini et al. 2021).

However, adopting AI technology for text analysis is more challenging. Language is complex with ambiguities and hidden meaning beyond the pure text structure. Yet, NLP has immensely benefited from the integration of LLMs. Machine and Deep Learning have been applied, for example, to archaeological prediction and detection (Resler et al. 2021) and CNN to translate cuneiform tablets of old Sumerian and Akkadian languages (Gutherz et al. 2023). Generative AI is helping to recreate the landscapes of the past for more immersive research of the past (Cobb 2023). Big data has been successfully linked in the project 'Unpathe'd Waters (Eagles 2022).

A current cultural heritage project applied NLP and in particular Topic Modelling (TM) and ML to explore the values attributed by people to familiar cultural landscapes (Tenzer 2022; Tenzer and Schofield 2023). Social media data, online surveys, and interviews provided sufficiently large datasets to infer heritage values from a "bottom-up" or people-centred perspective. TM allows the identification of patterns as themes latent in or emerging from the data, which guarantees an assumption-free approach to empirical data.

AI can also deal with the data deluge being experienced by archaeologists (Bevan 2015). The AGNES project facilitates large-scale synthesising research in The Netherlands, by integrating ML into a search engine which aims to index all the texts about archaeology in the region, some 200,000 documents. Specifically, it uses Named Entity Recognition to automatically detect all time periods, artefacts, and place names, which can then be used in search queries. This allows for more exhaustive and more precise searches, and in a case study on Early Medieval cremations, led to 30% more cremations being found in the literature than were previously known (Brandsen and Lippok 2021).

As well as AI-assisted search and TM, recent advances in the application of LLMs in NLP have shown promise in the identification of personally identifiable information (PII) and potential copyright infringements in digital publishing of archival data from modern historical periods. Legislative requirements (including those imposed by the EU's General Data Protection Regulations and extensions of copyright terms) mean that publishers of historical and heritage archives currently need to spend significant amounts of time and manual effort on ensuring compliance in these fields. Supporting publishing and editorial teams in this process has significant benefits in terms of both the amount of material that can be digitised and published and in catching cases of infringing content that might have otherwise been missed.

However, as useful as the technology seems to be it comes with a human and environmental cost. In the next section, we will present the challenges and risks of AI deployment from an ethical and environmental view as a counterbalance to the advantages and opportunities.

Ethical considerations – exclusion, limitation, bias

The latest AI advancements have given rise to several ethical considerations that warrant thorough examination. In particular, concerns have been raised regarding the transparency of the content and quality of the training data used in AI applications (Bender et al. 2021). These factors have been shown to perpetuate social inequalities (Casilli 2019), propagate misinformation (Wilner 2018), and compromise privacy (Véliz 2021). Furthermore, the use of AI technologies has been linked to instances of racial discrimination (Raji et al. 2020), the endangerment of natural resources, and the

exploitation of human labour (Crawford 2021).

Within the discipline, concerns surrounding privacy, bias, and model creation, are critical for formulating policies and decision-making. For instance, AI algorithms in analysing archaeological data could inadvertently lead to biased interpretations of historical events or the reinforcement of existing power structures if the models used are not designed with these ethical considerations in mind. Specifically, potential harms of fostering a linguistic monoculture, unintentionally strengthening existing power structures, and becoming a monocultural value carrier (Johnson et al. 2022; Pistilli 2022). Archaeology being also about understanding human history through material remains, language becomes a key component of cultural heritage and identity. If archaeological narratives are dominated by a single language or cultural perspective, this can lead to a skewed understanding of the past, privileging certain histories over others.

Also, there is a need for explainability and transparency in the approach to data collection in qualitative research. As shown in the heritage case study, AI can help analysing vast amounts of social media data or survey responses. However, generating models based on such data can introduce or reinforce biases, for example, excluding already marginalised groups. Shaping policies on models trained on such data would introduce these societal inequalities into systems of governance. The public also needs to have the option to opt-out with regard to data privacy, particularly in the case of vast data sets that are scraped or mined from the internet for training purposes.

While AI has the potential to analyse vast amounts of data and is particularly good at pattern detection (e.g., Casini et al. 2023), the technology has the potential to replace human volunteers in citizen science projects (Ponti and Seredko 2022). This can lead to a decrease of inclusive and engaging projects within archaeology. Excluding the public from the process of data collection and knowledge creation and instead reducing participation to the final product of archaeological investigations can lead to an alienation of archaeology.

Finally, garbage in, garbage out and black box effects carry the risk of creating new content from already flawed data and in an opaque process (Huggett 2021). Kansteiner (2022) and Clavert and Gensburger (2023) warn about the risk of using ChatGPT to reshape historical narratives: ‘If we think that the stories and images we

consume influence our memories, identities, and future behaviour, we should be very wary about letting AI craft our future entertainment on the basis of our morally and politically deeply flawed cultural heritage' (Kansteiner 2022, 124). Similarly, the GenAI technology will take realities of cultural heritage into a new dimension with challenges for authenticity and speculative interpretation in a new era of knowledge production and presentation (Spennemann 2023). A similar effect can be expected in the analysis of large archaeological datasets, shaping a narrative of the past based on weights in hidden layers (Cobb 2023).

Four key messages around ethical considerations result from these observations:

(1) The issue of biases emerging from the data used for training AI models is serious. Therefore, it is crucial to ensure data are as representative as possible. Researchers across the discipline of archaeology and CHM should work closely with data scientists and social scientists to design representative sampling strategies and data gathering methods, and to develop protocols for assessing and correcting for bias in datasets.

(2) The intersection of data science, philosophy, and archaeology suggests the advent of a new kind of archaeological specialism. Within this area of practice, archaeologists will need to understand the nuances of AI and Machine Learning and be well-versed in ethical considerations. Furthermore, users of the new technology have to understand the agency and autonomy of the new technology. Huggett (2021, 428) argues that "in some cases the system can appear to replace human expertise".

(3) The use of AI in shaping historical narratives is controversial. While AI has the potential to analyse large datasets and reveal patterns not always discernible to human eyes, it also carries the risk of propagating flawed interpretations of the past, particularly if the underlying data are biased. Therefore, stringent checks will be needed on the application of AI in this context. This includes the implementation of explainable AI (XAI) techniques to make the decision-making processes of these systems understandable to humans. However, the implementation of XAI techniques - even in simple application domains - is challenging. Two contrasting XAI philosophies exist (Barredo Arrieta et al. 2020)

- 1) designing inherently interpretable AI/ML systems, and 2) applying post-hoc explainability models (such as SHAP (Lundberg and Lee 2017)) to try and explain decisions made by AI models. A key disadvantage of inherently interpretable AI models is that it limits the power and complexity of such approaches - particularly in leveraging the latest generations of generative AI systems; however, criticism has been levelled at post-hoc methods regarding how closely their explanations relate to the decisions made by AI algorithms.

(4) Ethical guidelines for AI applications in archaeology and heritage practice need to be drafted and widely adopted to prevent misses and to promote the responsible use of these powerful technologies. However, crafting ethical guidelines for AI use in archaeology requires a balance between preventing misuse and adapting to the varied legal and practical contexts of global research environments. Discussions at the World Archaeological Congress (WAC 2023) and studies on remote sensing practices (Fisher et al. 2021) stress the challenge of developing standards that accommodate the distinct local regulations and the particularities of conducting research across different cultures and regions. Nevertheless, Davis (2020, 1) argues, that a high level of automation based on algorithms has the potential to create 'consistent definitions which permit reproducible research designs', which shows the advantages of automation for compatibility and reproducibility of data.

Discussion

Recent developments and the rapid adoption of AI technology into archaeology and heritage practice, as presented in this paper, show the importance of a debate around ethical implications and sustainable applications of AI. To enable the discourse, we have presented the advantages and capabilities of the applications, which allow more time and resource efficient workflows (Tenzer 2022; Tenzer and Schofield 2023), and enable the analysis and reuse of 'big data' accumulated over decades of archaeological investigations lying dormant in archives and grey literature (Brandsen and Lippok 2021). Furthermore, we provide different views on the implications of AI applications from archaeology, heritage studies, data science and philosophy, showing inherent challenges regarding limitation, bias, and social impact (Bender et al. 2021; Casilli

2019; Crawford 2021; Véliz 2021).

Interdisciplinary/cross-disciplinary research and collaboration will be necessary in the near future to apply this technology to a wide variety of disciplines. Collaboration between data science, sociology, philosophy, and archaeology is becoming increasingly important. Understanding how AI technology can influence epistemology and hermeneutics has to focus the discussion on the agency and cognitive artefacts of the technology in view of the output (Huggett 2021, 421). University courses bridging the complex knowledge of the various disciplines will be increasingly necessary. The projects presented here and the collaboration of the authors of this paper exemplify how cooperation can work to foster mutually beneficial collaboration.

Furthermore, the discipline needs to understand how AI deployment will impact on future employment for archaeologists and the changing work environment. What are the prospects for future archaeologists as a professional and academic career? Do we need to become computer scientists ourselves, and teach this to our students? Ultimately, will AI replace archaeologists? Harari (2017) argues that there is 'only a 0.7% chance'. However, it can replace the monotonous tasks of daily work, and carry out the large-scale analyses that precede archaeological work. However, the technology is evolving with increasing speed and predictions of future impact on the profession, especially after the pandemic, are difficult going forward.

AI deployment in the discipline needs to run alongside the development of strategies and best practice guidelines safeguarding the responsible, fair, and sustainable use of this new technology. Exploitation of human and natural resources with a cost for the environment needs to be highlighted and potential risks to reinforce social inequality must be considered.

Archaeology and CHM scholars are well equipped to study and deal with these societal effects of AI, looking at large scale influences on society for decades, and having the theories, methods, and background for these analyses. But to do so, they first need to understand the AI methods and their implications.

Conclusion

In post-phenomenological ontology, humans are experiencing the world with and

through technology (Gattiglia 2022; Ihde 2009). While we are at a point where machines not only assist humans (first machine revolution) but replace humans in the production or creative workflow (second machine revolution), we need to reorientate and redefine objectives. AI is here to stay, and the question will be how to use it responsibly and sustainably.

This means alignment: where does the technology work towards humanities values and goals and where are the dangers and risks of losing control, and therefore the benefits for society and humanity as a whole; not for the benefit of a few, but for the improvement of the environment, health, and society of the many?

Where does the development go from here? How can AI shape the future of the past – increasing our understanding of the past, using the vast amount of data from archaeology and history to create material that promotes and conveys this knowledge? Where does the future of the discipline lie regarding cooperation and education? We are at a point where archaeology and heritage practice cannot only benefit from these technological developments and advances but must also contribute to the ethical and practical discussion of AI in human culture and societies. Coming back to the initial question if AI in archaeology and CHM is a blessing or a curse, we provided examples of advantages and beneficial applications of the technology, but also highlighted challenges that need to be resolved before AI can be used safely and democratically. The debate is wide open.

Funding statement

This paper is part of an AHRC/UKRI WRoCAH-funded PhD project at the University of York. Grant reference number: AH/R012733/1.

References

Anichini, F. et al. (2021). The automatic recognition of ceramics from only one photo: The ArchAIDE app. *Journal of Archaeological Science: Reports*, 36, p.102788. [Online]. Available at: doi:10.1016/j.jasrep.2020.102788.

Argyrou, A. and Agapiou, A. (2022). A Review of Artificial Intelligence and Remote Sensing for Archaeological Research. *Remote Sensing*, 14 (23), p.6000. [Online]. Available at: doi:10.3390/rs14236000.

Barredo Arrieta, A. et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, pp.82–115. [Online]. Available at: doi:10.1016/j.inffus.2019.12.012.

Bender, E. M. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. 1 March 2021. New York, NY, USA: Association for Computing Machinery. pp.610–623. [Online]. Available at: doi:10.1145/3442188.3445922 [Accessed 16 May 2023].

Bevan, A. (2015). The data deluge. *Antiquity*, 89 (348), pp.1473–1484. [Online]. Available at: doi:10.15184/aqy.2015.102.

Bickler, S. H. (2021). Machine Learning Arrives in Archaeology. *Advances in Archaeological Practice*, 9 (2), pp.186–191. [Online]. Available at: doi:10.1017/aap.2021.6.

Brandsen, A. and Lippok, F. (2021). A burning question – Using an intelligent grey literature search engine to change our views on early medieval burial practices in the Netherlands. *Journal of Archaeological Science*, 133, p.105456. [Online]. Available at: doi:10.1016/j.jas.2021.105456.

Casilli, A. (2019). *En attendant les robots: enquête sur le travail du clic*. Paris: Seuil.

Casini, L. et al. (2023). A human–AI collaboration workflow for archaeological sites detection. *Scientific Reports*, 13 (1), p.8699. [Online]. Available at: doi:10.1038/s41598-023-36015-5.

Clavert, F. and Gensburger, S. (2023). *Is artificial intelligence the future of collective memory? Bridging AI scholarship and Memory Studies*. Call for papers for *Memory Study Review* (Brill Publishing). [Online} Available at: https://www.c2dh.uni.lu/sites/default/files/cfp_is_artificial_intelligence_the_future_of_collective_memory_-_2023.pdf

Cobb, P. (2023). Large Language Models and Generative AI, Oh My!: Archaeology in the Time of ChatGPT, Midjourney, and Beyond. *Advances in Archaeological Practice*, 11, pp.363–369. [Online]. Available at: doi:10.1017/aap.2023.20.

Cohen, A., Klassen, S. and Evans, D. (2020). *Ethics in Archaeological Lidar*. 3 (1), pp.76–91. [Online]. Available at: doi:10.5334/jcaa.48.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Davis, D. (2020). Defining what we study: The contribution of machine automation in archaeological research. *Digital Applications in Archaeology and Cultural Heritage*, 18, p.e00152. [Online]. Available at: doi:10.1016/j.daach.2020.e00152.

Eagles, P. (2022). Artificial Intelligence for data enhancement, linking and exploration. *https://unpathdwaters.org.uk*. [Online]. Available at: https://unpathdwaters.org.uk/artificial-intelligence-for-data-enhancement-linking-and-exploration/ [Accessed 2 November 2023].

Fisher, M. et al. (2021). Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project. *Archaeological Prospection*, 28 (3), pp.279–292. [Online]. Available at: doi:10.1002/arp.1816.

Gattiglia, G. (2022). A postphenomenological perspective on digital and algorithmic archaeology. In: *Archeologia e Calcolatori*. 33(2). All'Insegna del Giglio. pp.319–334.

Gutherz, G. et al. (2023). Translating Akkadian to English with neural machine translation. *PNAS Nexus*, 2 (5), p.pgad096. [Online]. Available at: doi:10.1093/pnasnexus/pgad096.

Harari, Y. N. (2017). The rise of the useless class. *ideas.ted.com*. [Online]. Available at: https://ideas.ted.com/the-rise-of-the-useless-class/ [Accessed 17 May 2023].

Huggett, J. (2021). Algorithmic Agency and Autonomy in Archaeological Practice. *Open Archaeology*, 7, pp.417–434. [Online]. Available at: doi:10.1515/opar-2020-0136.

Ihde, D. (2009). *Postphenomenology and Technoscience: The Peking University Lectures*, SUNY series in the Philosophy of the Social Sciences. New York: Suny Press. [Online]. Available at: https://sunypress.edu/Books/P/Postphenomenology-and-Technoscience [Accessed 31 October 2023].

Johnson, R. L. et al. (2022). *The Ghost in the Machine has an American accent: value conflict in GPT-3*. arXiv. [Online]. Available at: http://arxiv.org/abs/2203.07785 [Accessed 26 April 2023].

Kansteiner, W. (2022). Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially InTELligent? *History and Theory*, 61 (4), pp.119–133. [Online]. Available at: doi:10.1111/hith.12282.

Köpf, A. et al. (2023). *OpenAssistant Conversations -- Democratizing Large Language Model Alignment*. arXiv. [Online]. Available at: doi:10.48550/arXiv.2304.07327 [Accessed 10 May 2023].

Lundberg, S. and Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*.

Novaković, P. (2018). *Impact of the large-scale excavations in the Slovene preventive archaeology*.

Pistilli, G. (2022). What lies behind AGI: ethical concerns related to LLMs. *Revue Ethique et Numérique*. [Online]. Available at: https://hal.science/hal-03607808 [Accessed 17 May 2023].

Ponti, M. and Seredko, A. (2022). Human-machine-learning integration and task allocation in citizen science. *Humanities and Social Sciences Communications*, 9 (1), pp.1–15. [Online]. Available at: doi:10.1057/s41599-022-01049-z.

Raji, I. D. et al. (2020). Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. AIES '20. 7 February 2020. New York, NY, USA: Association for Computing Machinery. pp.145–151. [Online]. Available at: doi:10.1145/3375627.3375820 [Accessed 16 May 2023].

Resler, A. et al. (2021). A deep-learning model for predictive archaeology and archaeological community detection. *Humanities and Social Sciences Communications*, 8 (1), pp.1–10. [Online]. Available at: doi:10.1057/s41599-021-00970-z.

Russell, S. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach, Global Edition*. 3rd edition. Boston: Pearson.

Spennemann, D. H. (2023). *Generative Artificial Intelligence, Human Agency and the Future of Cultural Heritage*. SSRN Scholarly Paper, Rochester, NY . [Online]. Available at: doi:10.2139/ssrn.4583327 [Accessed 31 October 2023].

Tenzer, M. (2022). Tweets in the Peak: Twitter Analysis - the impact of Covid-19 on cultural landscapes. *Internet Archaeology*, (59), Internet Archaeology. [Online]. Available at: doi:10.11141/ia.59.6 [Accessed 19 August 2022].

Tenzer, M. and Schofield, J. (2023). Using Topic Modelling to reassess heritage values from a people-centred perspective – Applications from the north of England. *Cambridge Archaeological Journal*, (First view). [Online]. Available at: doi:https://doi.org/10.1017/S0959774323000203.

Turing, A. M. (1950). I.—computing machinery and intelligence. *Mind*, LIX (236), pp.433–460. [Online]. Available at: doi:10.1093/mind/LIX.236.433.

Véliz, C. (2021). *Privacy is Power: Why and How You Should Take Back Control of Your Data:* Corgi.

Verschoof-van der Vaart, W. B. et al. (2020). Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from The Netherlands. *ISPRS International Journal of Geo-Information*, 9 (5), p.293. [Online]. Available at: doi:10.3390/ijgi9050293.

Wilner, A. S. (2018). Cybersecurity and its discontents: Artificial intelligence, the Internet of Things, and digital misinformation. *International Journal*, 73 (2), pp.308–316. [Online]. Available at: doi:10.1177/0020702018782496.