Information Research - Vol. 30 No. iConf (2025)

How does digital humanities research talk about AI? A bibliometric analysis

Wenyi Shang, Rongqian Ma, and Heather Moulaison-Sandy

DOI: https://doi.org/10.47989/ir30iConf47242

Abstract

Introduction. Despite scholarly work that looked into artificial intelligence (AI)'s impact on digital humanities (DH), there has been little comprehensive investigation into how AI has been reflected and adopted in DH scholarship. This paper addresses this gap by asking: What patterns can be observed in AI integration in the DH scholarly discourse?

Method. A bibliometric analysis was conducted on 2,488 abstracts of articles published in three influential DH journals: *digital scholarship in the humanities*, *digital humanities quarterly*, and *journal of cultural analytics*.

Analysis. Focusing on abstracts containing the term 'AI' or '*artificial intelligence*,' this study conducted four analyses: (1) temporal analysis of AI mentions, (2) collocation analysis, (3) word vector analysis, and (4) topic modeling.

Results. AI has been discussed in DH scholarship well before its recent rise in prominence. The recurring themes in these discussions encompass both technological and human-centred aspects.

Conclusion. AI has been integral to DH, and the discourse around AI in DH reflects the field's dual focus on technology and the humanities. The bibliometric analysis presented in this paper illustrates how information science can inform and guide methodological reflections in DH, offering new insights into the future development of the field.

Introduction

Advances in computing technology have brought changes to multiple aspects of humanities scholarship and research culture, including data practices, methodologies, knowledge production practices, as well as scholarly communication processes (Berry, 2022). In particular, the recent surge in generative artificial intelligence (AI) has garnered attention for its ability to automate creative processes, posing new challenges and opportunities for scholars in digital humanities (DH) while creating a valuable moment to revisit the long history of AI in DH (Gefen et al., 2021). Despite scholarly works that looked into AI's impact on DH both conceptually and empirically, there has been little comprehensive investigation into how AI has been reflected and adopted in DH scholarship. This type of examination of AI adoption in DH, however, can be of great importance, offering insights into, and guidance on, the future integration of AI in DH.

Our paper aims to address this gap by asking: What patterns can be observed in AI integration in DH scholarly discourse? Using a bibliometric approach, we conduct an exploratory analysis of AI terms ('AI' and '*artificial intelligence*') in article abstracts published across three influential DH journals: *digital scholarship in the humanities* (*DSH*, formerly known as *literary and linguistic computing*), digital *humanities quarterly* (*DHQ*), and *journal of cultural analytics* (*JCA*). More specifically, our analysis answers the following questions:

(1) What is the temporal trend of AI mentions represented in the abstracts?

(2) How do AI terms collocate with other concepts in these abstracts?

(3) What are the semantic underpinnings of AI when discussed in these abstracts?

(4) What are the major topics or domains within DH that are at the forefront of AI mentions, as seen in the abstracts?

By investigating AI mentions in DH scholarship, our work aims to empirically illustrate an evolving picture of AI integration in DH, bridging the historical and contemporary practices.

Literature review

AI integration in DH

The integration of AI technology and AI-based tools (e.g., machine learning (ML), natural language processing (NLP)) has accelerated the '*computational turn*' of the field of DH, expanding its capacity to handle more complex datasets, uncover hidden patterns, and automate processes that enhanced both the scale and depth of the scholarly inquiry (Berry, 2022; Gefen et al., 2021). In text analysis, for instance, AI-powered tools, such as NLP, have facilitated the examination of large text corpora, enabling the practice of '*distant reading*' (Moretti, 2013). Furthermore, a growing focus on data and algorithms—particularly on how to conceptualize and work with data, and how to understand as well as address the opaque nature of algorithmic processes and their underlying data transformations—has brought important questions to the forefront of DH research (Bode, 2019; D'Ignazio & Klein, 2020; Drucker, 2011; Lavin, 2021; Schmidt, 2016). In addition to enabling methodological transformation, AI-driven automation has reshaped how labor is conceptualized and executed in DH, introducing new opportunities for interdisciplinary collaborations and research innovation (Smithies, 2017). Building on AI's longstanding presence in DH, the recent rise of generative AI technology presents both new possibilities and challenges for the field. Researchers have begun exploring how this technology can be integrated into DH research and pedagogical practices in ways that are effective, critical, and ethical (Chun & Elkins, 2023; Dedema & Ma, 2024; Guo, 2024; Liu et al., 2023).

Despite the recent progress in exploring AI applications within DH and the acknowledgement of the importance of data and related ethical implications in DH research (Gonzalez & Rodrigues, 2022), a comprehensive, empirical investigation of how AI has been reflected in DH scholarship remains limited. This is especially true when considering the historical integration of AI technologies into DH publications and research. With the recent surge in generative AI technologies, a historical perspective on AI's use in DH is increasingly valuable for identifying patterns, gaps, and emerging challenges within the field. Our paper responds to this gap by examining AI use in DH scholarship, tracing its evolution from the 1980s to the present.

Bibliometric analysis of DH

Bibliometric analysis offers valuable methods for illustrating broad trends and usage patterns in DH scholarship through empirical data. Previous studies have highlighted the strengths of bibliometric approaches—such as citation analysis, network analysis, and topic modeling—in examining DH as an interdisciplinary field and research community. For instance, Luhmann and Burghardt (2021) analysed articles from DH journals alongside those from 15 other academic disciplines, addressing the conceptual identity of DH—one of the field's enduring debates—and underscored how DH connects with neighbouring fields like computational linguistics and information science. Similarly, Ma and Li (2021) introduced the '*battleground*' metaphor, emphasizing the tensions and collaborations between humanities and STEM research practices and conventions. Earlier works demonstrated that bibliometric analysis also sheds light on the early history and emergence of DH as a field (Leydesdorff & Akdag Salah, 2010; Sula & Hill, 2019), as well as its gradual evolution into a '*Big Tent*' discipline (Weingart & Eichmann-Kalwara, 2017). Furthermore, citation and network analysis using DH journal article data has provided insights into global DH research collaboration patterns and community structures (Gao et al., 2017, 2018; Li et al., 2024; Nyhan & Duke-Williams, 2014; Tang et al., 2017).

In this paper, we leverage the unique methodological advantages of bibliometric analysis and apply it to examine a publication corpus of three influential DH journals, highlighting the broad patterns of AI adoption in DH scholarship.

Methods

Data collection

We collected metadata from all articles published in three influential, '*exclusively DH*' journals: *DSH*, *DHQ*, and *JCA* (Spinaci et al., 2020). Published since 1986, *DSH* allows us to consider nearly 40 years of DH history. Metadata for *DSH* articles were parsed from HTML files collected by one of the authors using their university library's subscription and access to the journal; the *DHQ* metadata were extracted from XML files available in the journal's GitHub repository (https://github.com/Digital-Humanities-Quarterly/dhq-journal/tree/main/articles). For *JCA*, the metadata were obtained by web scraping the HTML source code of the journal's online open-access articles.

In this study, we analysed abstracts rather than full-text articles because abstracts provide higher information density. We expect that articles focused on AI will include relevant terms in their abstracts. Therefore, we excluded articles without abstracts or publication years, as well as those with metadata information indicating that the articles are not in English. This resulted in a corpus of 1,587 abstracts from *DSH* (69.3% of the total 2,290 articles published between 1986 and 2023), 720 abstracts from *DHQ* (95.2% of the total 756 articles published between 2007 and 2024), and 181 abstracts from *JCA* (99.5% of the total 182 articles published between 2016 and 2024). After data collection, four data analysis experiments were conducted to examine the discourse surrounding AI in these abstracts: temporal analysis of AI mentions, collocation analysis, word vector analysis, and topic modeling.

Temporal analysis of AI mentions

For the temporal analysis of AI mentions, we measured the presence of AI terms in each abstract and tracked their variation over time. We defined '*AI terms*' as 'AI' and '*artificial intelligence*' only, rather than creating an extended thesaurus, to ensure that all terms discussed in this study specifically refer to AI. Abstracts were identified as containing AI mentions if they included either the word '*ai*' or the 2-gram '*artificial intelligence*' (where '*artificial*' is immediately followed by '*intelligence*' within the same sentence). We then calculated the proportion of articles with AI mentions in their abstracts published each year.

Collocation analysis

Collocation analysis was conducted to identify the meaningful terms that co-occur with AI terms in each abstract. We first generated a word count table using the CountVectorizer from the Python module Scikit-learn (Pedregosa et al., 2011). To reduce morphological variants of words to a common stem, we applied the SnowballStemmer for English from the NLTK library (https://www.nltk.org/api/nltk.stem.SnowballStemmer.html). Next, we calculated the number of abstracts in which each stemmed word co-occurs with AI terms. Following the ideas of Algee-Hewitt et al. (2020) and Gilkison and Kurzynski (2024), we employed Fisher's exact test (Fisher, 1922) to identify statistically significant collocates of AI terms, excluding common words like '*the*' and '*a*'. Additionally, referencing Gilkison and Kurzynski (2024), we calculated the ratio of the actual number of abstracts where each stemmed word co-occurs with AI terms to the expected number of such co-occurrences (r-exp), which is defined as:

$$r - exp_i = \frac{\#\ of\ Abstracts\ Containing\ the\ Stemmed\ Word\ i\ and\ AI\ Terms}{\#\ of\ Abstracts\ Containing\ the\ Stemmed\ Word\ i \cdot \frac{\#\ of\ Abstracts\ Containing\ AI\ Terms}{Total\ \#\ of\ Abstracts}}$$

Word vector analysis

In addition to words that co-occur with AI terms, we also identified words with similar semantic meanings with AI terms through word vector analysis, which quantifies the semantic similarity between terms based on their contextual usage. Using the Python library Gensim (Řehůřek & Sojka, 2010), we pre-processed the corpus with simple_process and converted each stemmed word into a vector with the Word2Vec model (Mikolov et al., 2013). For better performance with our relatively small dataset, we adjusted the '*min_count*' parameter to 1 to ensure no word was excluded, and switched the '*sg*' parameter from 0 to 1 to use the skip-gram algorithm instead of the CBOW algorithm. We then calculated the cosine similarity between each stemmed word's vector and the vector of the word '*ai*', ranking their semantic similarity based on these cosine similarity scores.

Topic modeling

Topic modeling is a technique used to '*discover hidden thematic structure in large collections of texts*', and is widely used in DH (Blei, 2012). For this study, we applied topic modeling to examine the themes in abstracts and their relationship with AI terms. We used the MALLET (MAchine Learning for LanguagE Toolkit) for this purpose, which offers efficient sampling-based implementations of latent Dirichlet allocation (LDA), Pachinko allocation, and hierarchical LDA in its topic modeling toolkits (https://github.com/mimno/Mallet) and is commonly used by digital humanists (e.g., Graham et al., 2012). We imported our corpus into MALLET and generated 30 topics. When importing data into MALLET, we specified the '*--token-regex*' as '[\p{L}] +,' since the default regular expression removes words with 1 or 2 characters, including "ai." We also combined the default English stop word list with the French stop word list from https://github.com/stopwords-iso/stopwords-fr (excluding the word '*ai*'). The French stop word list was necessary because some abstracts contain French words without metadata indicating their language, making it impossible to filter them out during data collection. In determining the number of topics, we tested various topic numbers in increments of 10 (from 10 to 50) and observed that the log-likelihood initially rose, then declined, reaching its peak at 30. We then compared the average weight of each topic in abstracts with AI terms to those without. Finally, the three co-authors independently labelled each topic based on the 20 keywords for each topic, and then synthesized their results to determine the final label for each topic.

Results

Temporal analysis of AI mentions

Out of a total of 2,488 abstracts, 28 (1.1%) contain AI terms. The temporal variation in the proportion of articles with AI mentions in their abstracts published each year is illustrated in Figure 1.

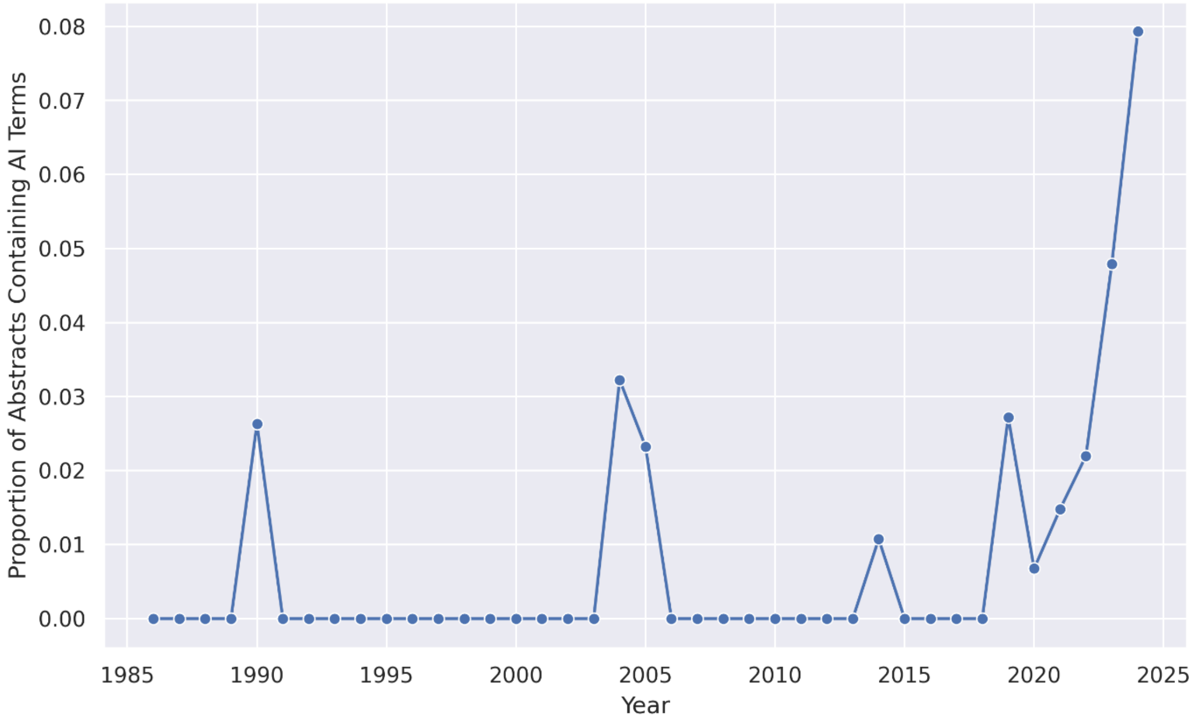

Figure 1. Temporal variation of AI mentions in the abstracts of *DSH*, *DHQ*, and *JCA*

Overall, the leading DH journals show a sharp increase in AI mentions in their abstracts after 2019, with a particularly notable surge beginning in 2023. This trend aligns with the recent development of AI technology, especially generative AI, which has drawn significantly increased attention from DH scholars, potentially leading to more extensive discussions in these journals. Additionally, it is noteworthy that four articles mentioning AI terms were published before this recent rise. With the earliest dating back to 1990, they applied AI to tasks including text interpretation (De Vuyst, 1990; Terras, 2005), text encoding (Terras & Robertson, 2004), and optical character recognition (OCR) (Heckmann et al., 2014). This finding supports Berry's (2022) observation of digital humanists' extended history of using automation (pp. 446–447), which predates the recent surge of AI in public discourse, and further highlights that the term 'AI' was occasionally explicitly used in these early practices.

Collocation analysis

There are 16 stemmed words identified as statistically significant collocates of AI terms ($p < 0.001$ in Fisher's exact test, excluding '*ai*', '*artifici*', and '*intellig*'). These collocates are presented in Table 1, ranked in ascending order of p-values (i.e., more statistically significant collocates appear higher in the list).

Rank	Word	r-exp	p-value
1	ethic	15.55	1.49E-07
2	system	3.29	1.45E-05
3	learn	4.21	5.24E-05
4	big	11.11	6.23E-05
5	machin	4.62	6.74E-05
6	dungeon	88.86	1.22E-04
7	big-fiv	88.86	1.22E-04
8	loop	88.86	1.22E-04
9	openai	88.86	1.22E-04
10	gpt-2	88.86	1.22E-04
11	digitis	22.21	2.63E-04
12	capabl	8.08	2.93E-04
13	recognit	5.92	3.76E-04
14	offlin	44.43	7.23E-04
15	historian	6.53	7.93E-04
16	ecolog	14.81	9.31E-04

Table 1. Statistically significant collocates of AI terms

The 16 statistically significant collocates of AI terms span both technical and humanities-oriented concepts. Technical terms like '*system*', '*machin*', '*openai*', and '*gpt-2*' highlight the technical focus in discussions of AI within DH scholarship. Simultaneously, humanities-oriented terms such as '*historian*' indicate the application of AI in specific humanities disciplines and underscore the prominence of humanistic perspectives in DH work. Additionally, the presence of terms such as '*ethic*' further highlights the field's emphasis on ethical issues. This observation is also echoed in digital humanities scholarship (e.g., Gonzalez & Rodrigues, 2022), especially in the context of teaching with emerging AI technologies (e.g., Chun & Elkins, 2023; Berry, 2022).

Word vector analysis

Among the 12,418 distinct stemmed words (excluding '*ai*' itself), it is unsurprising that many of the terms with the highest semantic similarity to 'ai' are technical. For example, '*intellig*' (5th), '*portal*' (6th), and '*microcomput*' (11th) were among the terms ranked the highest in semantic similarity to '*ai*.' This potentially demonstrates an alignment between DH and information science in perceiving these technical elements as integral to the study of AI.

Topic modeling

As shown in Figure 2, the 28 abstracts containing AI terms have substantially higher average weights (at least twice as high) in topics 10 (Ancient Texts), 17 (Digital Cultural Heritage), 23 (Narratology), 24 (Machine Learning), and 25 (Digital Media) compared to the 2,460 abstracts without AI terms.

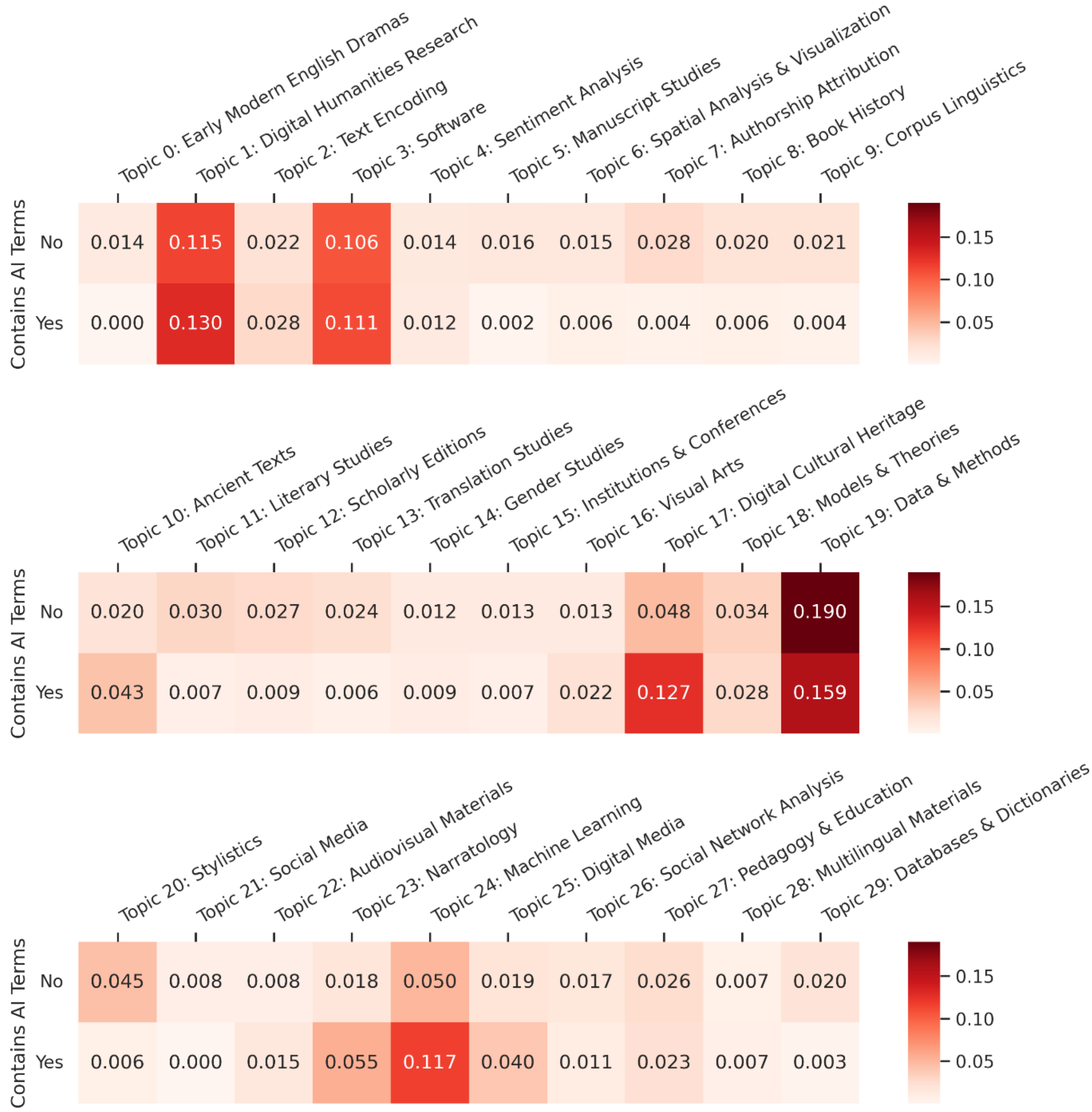

Figure 2. Average topic weights for abstracts with and without AI terms

Topics associated with AI in the topic modeling process aligned with themes previously revealed in the analysis (i.e., technical and humanities-oriented). Technical aspects are strongly demonstrated in Topic 24, which is comprised of keywords related to ML methodology (e.g., ‘*machine*’, ‘*learning*’, ‘*accuracy*’, ‘*classification*’, ‘*features*’). Articles with abstracts where Topic 24 is dominant (weight > 0.4) reflect a particular interest in technical aspects, including the application of AI and ML to tasks like OCR (Heckmann et al., 2014), text classification (Zheng & Jin, 2019), and information extraction (Sánchez-Salido et al., 2023).

AI is also aligned with humanities-oriented themes. Topics 10, 17, 23, and 25 each represent an intersection between DH and a specific humanities field. Topic 10 highlights keywords related to ancient texts and language systems (e.g., ‘*Greek*’, ‘*ancient*’, ‘*language*’, ‘*system*’, ‘*linguistic*’); Keywords of Topic 17 centres on cultural heritage (e.g., ‘*cultural*’, ‘*heritage*’, ‘*historical*’, ‘*archive*’, ‘*collection*’); those of Topic 23 focuses on narratology (e.g., ‘*narrative*’,’ *character*’, ‘*story*’, ‘*novel*’, ‘*plot*’); and those of Topic 25 concerns media studies (e.g., ‘*media*’, ‘*video*’, ‘*game*’, ‘*art*’, ‘*virtual*’). This suggests that certain humanities disciplines—linguistics, heritage science, media studies, and a subfield of literature, narratology—may be at the forefront of incorporating AI advancements, reflecting their active engagement in AI-driven research.

Discussion and conclusion

In summary, the results of this empirical study reveal that AI has been discussed in DH scholarship well before its recent rise in prominence. These discussions encompass a wide range of perspectives, from technical approaches to those aligned with specific humanities disciplines, as well as efforts by information professionals to facilitate access to collections through ethical engagement and professional practice. The recurring themes of both technological and human-centred aspects present no surprise, given the dual focus of the DH field on technology and the humanities.

However, it is surprising that only 1.1% of abstracts in the corpus explicitly mention 'AI' or '*artificial intelligence*'. Even when considering only publications from the 2020s—a period when AI has become an extremely prominent topic in public discourse—this proportion increases only slightly to 2.7%. This may lie precisely in the close association between AI and the computing technologies that enable DH work—DH is so inherently linked to AI that explicit mentions of it might seem unnecessary. For authors, the need to explicitly identify AI technologies might not be as strongly felt in this fundamentally technology-based field.

The low proportion of abstracts mentioning AI terms can also be explained by the possibility that, in some cases, AI is considered as synonymous with the computational aspects integral to DH. In future work, we plan to consider additional AI-related terms (e.g., ML, NLP, large language models) to provide a comprehensive view of the use of AI technologies. Future research could also use the full texts of articles and a wider range of DH journals. This approach would enable a more comprehensive and nuanced interpretation of AI integration in DH, complementing the insights gained from abstract analysis.

In conclusion, this paper represents preliminary work to address the gap in empirical research on AI adoption in DH scholarship. Our findings demonstrate that algorithmic processes have significantly impacted the landscape of DH research, with practitioners actively leveraging recent advancements in AI technology. This underscores the importance of reflecting on AI integration in DH. Our bibliometric analysis illustrates how information science can inform and guide such methodological reflections in DH, offering new insights into the future development of the field.

About the authors

Wenyi Shang is Assistant Professor in the iSchool, University of Missouri, USA. He received his Ph.D. from the University of Illinois Urbana-Champaign, and his research interests are in the digital humanities. He can be contacted at wenyishang@missouri.edu.

Rongqian Ma is Assistant Professor of Information and Library Science in the Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA. She received her Ph.D. from the University of Pittsburgh, and her research interests focus on computational literary studies, digital archives, and scholarly communication. She can be contacted at rm56@iu.edu.

Heather Moulaison-Sandy is Associate Professor in the iSchool, University of Missouri, USA. She received her Ph.D. from Rutgers University, and her research interests focus on the organization of information, scholarly communication, and technology. She can be contacted at moulaisonhe@missouri.edu.

References

Algee-Hewitt, M., Porter, J. D., & Walser, H. (2020). Representing race and ethnicity in American fiction, 1789-1920. Journal of Cultural Analytics, 5(2). https://doi.org/10.22148/001c.18509

Berry, D. M. (2023). AI, ethics, and digital humanities. In J. O'Sullivan (Ed.), The Bloomsbury handbook to the digital humanities (pp. 445–457). Bloomsbury Academic.

Blei, D. M. (2012). Topic modeling and digital humanities. Journal of Digital Humanities, 2(1). https://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei

Bode, K. (2019). A world of fiction: Digital collections and the future of literary history. University of Michigan Press.

Chun, J., & Elkins, K. (2023). eXplainable AI with GPT4 for story analysis and generation: a novel framework for diachronic sentiment analysis. International Journal of Digital Humanities, 5(2), 507–532. https://doi.org/10.1007/s42803-023-00069-8

De Vuyst, J. (1990). Knowledge representation for text interpretation. Literary and Linguistic Computing, 5(4), 296–302. https://doi.org/10.1093/llc/5.4.296

Dedema, M., & Ma, R. (2024). The Collective Use and Evaluation of Generative AI Tools in Digital Humanities Research: Survey-Based Results (arXiv:2404.12458). arXiv. https://doi.org/10.48550/arXiv.2404.12458

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. The MIT Press.

Drucker, J. (2011). Humanities approaches to graphical display. Digital Humanities Quarterly, 5(1). https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Fisher, R. A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society, 85(1), 87–94. https://doi.org/10.2307/2340521

Gao, J., Duke-Williams, O., & Mahony, S. (2017). The intellectual structure of digital humanities: an author co-citation analysis. The Book of Abstracts of Digital Humanities Conference 2018, 450–452.

Gao, J., Nyhan, J., Duke-Williams, O., & Mahony, S. (2018). Visualising the digital humanities community: a comparison study between citation network and social network. The Book of Abstracts of Digital Humanities Conference 2018, 201–204.

Gefen, A., Saint-Raymond, L., & Venturini, T. (2021). AI for digital humanities and computational social sciences. In B. Braunschweig & M. Ghallab (Eds.), Reflections on artificial intelligence for humanity (pp. 191–202). Springer. https://doi.org/10.1007/978-3-030-69128-8_12

Gilkison, A., & Kurzynski, M. (2024). Vectors of violence: legitimation and distribution of state power in the People's Liberation Army Daily (Jiefangjun Bao), 1956-1989. Journal of Cultural Analytics, 9(1). https://doi.org/10.22148/001c.115481

Gonzalez, M. E., & Rodrigues, M. V. (2022). Digital humanities: ethical implications and interdisciplinary challenges. Humanities Bulletin, 5(1), Article 1.

Graham, S., Weingart, S., & Milligan, I. (2012, September 2). Getting started with topic modeling and MALLET. Programming Historian. https://programminghistorian.org/en/lessons/topic-modeling-and-mallet

Guo, Q. (2024). Prompting Change: ChatGPT's impact on digital humanities pedagogy: a case study in art history. International Journal of Humanities and Arts Computing, 18(1), 58–78. https://doi.org/10.3366/ijhac.2024.0321

Heckmann, D., Frank, A., Arnold, M., Gietz, P., & Roth, C. (2014). Citation segmentation from sparse & noisy data: a joint inference approach with Markov Logic Networks. Digital Scholarship in the Humanities, 31(2), 333–356. https://doi.org/10.1093/llc/fqu061

Lavin, M. (2021). Why digital humanists should emphasize situated data over capta. Digital Humanities Quarterly, 15(2). https://digitalhumanities.org/dhq/vol/15/2/000556/000556.html

Leydesdorff, L., & Salah, A. A. A. (2010). Maps on the basis of the Arts & Humanities Citation Index: The Journals Leonardo and Art Journal versus 'Digital Humanities' as a topic. Journal of the American Society for Information Science & Technology, 61(4), 787–801. https://doi.org/10.1002/asi.21303

Li, K., Ma, R., & Fang, Z. (2024). Explicit or implicit digital humanities? An examination of search strategies to retrieve digital humanities publications from large-scale scholarly databases. Quantitative Science Studies. Advance online publication. https://doi.org/10.1162/qss_a_00321

Liu, C., Wang, D., Zhao, Z., Hu, D., Wu, M., Zhang, H., Lin, L., Liu, J., Shen, S., Li, B., & Zhao, L. (2024). SikuGPT: a generative pre-trained model for intelligent information processing of ancient texts from the perspective of digital humanities. ACM Journal on Computing and Cultural Heritage. Advance online publication. https://doi.org/10.1145/3676969

Luhmann, J., & Burghardt, M. (2021). Digital humanities—a discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. Journal of the Association for Information Science and Technology, 72(12), 1475-1498. https://doi.org/10.1002/asi.24533

Ma, R., & Li, K. (2021). Digital humanities as a cross-disciplinary battleground: an examination of inscriptions in journal publications. Journal of the Association for Information Science and Technology, 73(2). https://doi.org/10.1002/asi.24534

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio & Y. LeCun (Eds.), 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings. http://arxiv.org/abs/1301.3781

Moretti, F. (2013). Distant reading. Verso.

Nyhan, J., & Duke-Williams, O. (2014). Joint and multi-authored publication patterns in the digital humanities. Literary and Linguistic Computing, 29(3), 387–399. https://doi.org/10.1093/llc/fqu018

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-Learn: machine learning in Python. Journal of Machine Learning Research, 12(85), 2825–2830.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 45–50.

Sánchez-Salido, E., Menta, A., & García-Serrano, A. (2023). Seeking information in Spanish historical newspapers: the case of Diario de Madrid (18th and 19th Centuries). Digital Humanities Quarterly, 17(4). https://www.digitalhumanities.org/dhq/vol/17/4/000735/000735.html

Schmidt, B. (2016). Do digital humanists need to understand algorithms? In M. K. Gold & L. F. Klein (Eds.), Debates in the digital humanities 2016. University of Minnesota Press.

Smithies, J. (2017). The digital humanities and the digital modern. Palgrave Macmillan.

Spinaci, G., Colavizza, G., & Peroni, S. (2020). Preliminary Results on Mapping Digital Humanities Research. In C. Marras, M. Passarotti, G. Franzini, & E. Litta (Eds.), La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica: Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD): 15-17 gennaio 2020, Milano, Università Cattolica del Sacro Cuore, 246–252.

Sula, C. A., & Hill, H. V. (2019). The early history of digital humanities: an analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). Digital Scholarship in the Humanities, 34(Supplement_1), i190–i206. https://doi.org/10.1093/llc/fqz072

Tang, M.-C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. Scientometrics, 113(2), 985–1008. https://doi.org/10.1007/s11192-017-2496-6

Terras, M. (2005). Reading the readers: modelling complex humanities processes to build cognitive systems. Literary and Linguistic Computing, 20(1), 41–59. https://doi.org/10.1093/llc/fqh042

Terras, M., & Robertson, P. (2004). Downs and acrosses: textual markup on a stroke level. Literary and Linguistic Computing, 19(3), 397–414. https://doi.org/10.1093/llc/19.3.397

Weingart, S. B., & Eichmann-Kalwara, N. (2017). What's under the big tent?: A study of ADHO conference abstracts. Digital Studies, 7(1), 6. https://doi.org/10.16995/dscn.284

Zheng, W., & Jin, M. (2019). Comparing multiple categories of feature selection methods for text classification. Digital Scholarship in the Humanities, 35(1), 208–224. https://doi.org/10.1093/llc/fqz003