Vol. 06, No. 2, pp.582-590| JUNE 2026
DOI: https:// DOI 10.34118/jskp.v6i2.4650

Article history (leave this part):
Submission date: 27.03-2026
Acceptance date: 30-05-2026
Available online: 10-06-2026
Keywords:
artificial intelligence, legal regulation, Ethics, Fourth Industrial Revolution, Robots
Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Competing interest:
The author(s) have declared that no **competing interests** exist.
Cite as (leave this part):
Hanan Abufares Elkhimry; . (2024). Title. Journal of Science and Knowledge Horizons: 4(1), 283-293. https://doi.org/10.34118/jskp.v2i02.2727

Journal of Science and Knowledge Horizons
ISSN 2800-1273-EISSN 2830-8379

Digital Humanities and AI Ethics

1 Hassina Benreguia, University of constantine3-Sallah Boubnider (Algeria) *, hassina.benreguia@univ-constantine3.dz
https://orcid.org/0000-0001-5926-3348
2 David Berry, University of Sussex (United Kingdom) *, D.M.Berry@sussex.ac.uk
https://orcid.org/0000-0002-7737-5586

Abstract:

This article offers an in-depth review of Chapter Forty-Two, 'Artificial Intelligence, Ethics, and Digital Humanities', from 'The Bloomsbury Handbook of Digital Humanities' (2023). In this chapter, James O'Sullivan, together with a number of scholars, seeks to reassess the present and future of the digital humanities, laying the groundwork for a conceptual understanding of the field while outlining its perspectives, methodologies, tools, contexts, and future directions.
The article is structured into three sections, mirroring the organization of the chapter itself. The first section analyses how automated research methodologies are reshaping humanities scholarship and redefining the role of the researcher within digital environments. The second section examines existing and emerging applications of artificial intelligence and machine learning in the digital humanities. The third section addresses the ethical and methodological challenges raised by these technologies, with particular emphasis on transparency, accountability, and algorithmic justice. The article ultimately argues for the development of ethical frameworks to guide the responsible use of artificial intelligence in the digital humanities.

***Hassina Benreguia**

Introduction

This article examines the role of artificial intelligence (AI) applications and machine learning (ML) in the development and subsequent transformation of the methodologies, tools, subjects, and research practices of the humanities, leading to what has become known as the digital humanities. This transformation is the result of the expansion of digital networks and the increasing reliance on automation in data analysis, developments that have altered research practices within the humanities and shifted interpretative authority from the human researcher to the machine. These changes have been accompanied by new research challenges that directly affect the ethics of humanities scholarship.

The growing reliance on machine learning and algorithmic systems, with all the biases they entail, has led to distortions in the subject matter, methodology, and outcomes of humanities research. In this context, humanities research risks becoming an endeavor that reflects the intelligence and language of the machine rather than human scholarly judgement. This situation raises critical questions about the ethics of research in the humanities in light of the expanding use of artificial intelligence across all stages of humanities research, the erosion of its human character, and its transformation into a digitally automated enterprise.

Through a critical review of Chapter Forty-Two of 'The Bloomsbury Handbook of Digital Humanities', including the research and studies it presents in the fields of algorithms, intelligent software, machine learning, big data processing, and AI ethics, this article seeks to offer a comprehensive account that enables an understanding of the mechanisms of automation within the humanities, the role of artificial intelligence in this process, and the ethical challenges it raises at multiple levels methodological, academic, and moral. In doing so, the article aims to provide well-grounded responses to the following central questions: How does artificial intelligence affect the ethics of humanities research? How do algorithms, with their inherent biases, shape objectivity in the digital humanities? And who bears ethical responsibility in machine-driven digital humanities research?

METHODOLOGY

This article relies on a thorough, in-depth analysis of the chapter Field of Study, structured around a detailed examination of its three sections:

1. Section one explores how automation affects humanities research and scholars' roles.
2. Section two examines current uses of AI, especially machine learning, in digital humanities.
3. Section three considers the ethical and practical challenges AI poses for the field.

Including research on algorithms and software such as software studies and cultural analytics, new media studies and machine learning, ethics and algorithmic bias, data and surveillance studies.

AUTOMATION OF THE HUMANITIES

The chapter presented the field of study by outlining the historical background to the emergence of the digital humanities and the early adoption of automation in the humanities in 1981. It also linked this development to Steve Jobs's ideas, in which he described computers as "bicycles for the mind", justifying this view by stating that: "created the bicycle, [they] created a tool that amplified an inherent ability … The Apple personal computer is a 21st century bicycle if you will, because it's a tool that can amplify a certain part of our inherent intelligence… [it] can distribute intelligence to where it's needed" (Jobs, 1981, in Berry, 2023) .

The chapter noted that Steve Jobs was influenced by the ideas of Doug Engelbart and Vannevar Bush. about "man- computer symbiosis" and "augmenting human intellect" using technology (Ekbia and Nardi, 2017, in berry, 2023) .

These concepts, which emphasized the role of computers and machines in enhancing human cognitive and creative capacities—both in general and specifically for humanities researchers—provided the foundational framework for the early development of digital humanities. Initially, digital humanities emerged as the use of computers in humanities research as an assisting tool, a mechanism to augment the researcher's ability to carry out tasks with greater speed and precision. This included the use of technology for information and data collection, text processing, precise statistical analysis of quantitative data, classification and analysis of large datasets, and the development of methods for presenting results through various software applications.

However, the current orientation in digital humanities goes beyond merely employing computers as auxiliary tools for researchers, moving instead towards the replacement of the researcher by machines, which perform all research tasks. This constitutes a form of full automation in digital humanities, enabled by artificial intelligence, machine learning, and algorithms, and has given rise to ethical problems that have distorted the essence of humanities research .

Artificial Intelligence (AI) has also entered the world of Digital Humanities, particularly with machine learning and natural language processing. AI has forced researchers to change their approach to data analysis and data organization. Many projects around the world deal with DH nowadays, and many of them have started applying AI techniques (Guido, Yahya, Sohail, 2025).

In response, a number of studies have advocated for the regulated and deliberate use of automation, which should remain a supportive mechanism for researchers in organizing and analyzing large datasets that cannot otherwise be processed due to human limitations of capacity and time. This involves the deployment of algorithms to sort, organize, and classify large datasets, or to apply complex statistical methods.

In this context, (McCarty, 2003, in Berry, 2023) proposed a division of digital humanities labor in terms of three branches working respectively on algorithmic (e.g., "mechanical" data analysis), metalinguistic (e.g., encoding texts in TEI and related markup languages), and representational (e.g., visual and presentational techniques) aspects of digital humanities work.

PHASES OF AUTOMATION IN DIGITAL HUMANITIES

The chapter presented the field of study with three phases of automation in digital humanities, namely (Berry, 2023):

The first involved bespoke software, often programmed by a humanist-coder, to solve a particular problem. The second was the generalization of these early software tools into packages which provide a "one-stop shop" and which involve little or no programming requirement, such as Voyant Tools. In the latest phase, we see the beginning of much more sophisticated machine- learning tools, again by humanist-coders who are required to customize the machine-learning software they use for a specific digital humanities project, for example the MALLET (MAchine Learning for LanguagE Toolkit) software.

ALGORITHMS AND AMBIGUITY IN DIGITAL HUMANITIES

It has become apparent from the studies discussed in this chapter that researchers in the humanities engage with websites, software packages, smart applications, and algorithms to process texts, images, and big data, perform complex statistical operations, and carry out many other tasks, often without understanding the nature of these software and algorithms, how they function, or the complexity underlying them. This leads to a significant degree of ambiguity in digital humanities research, which is difficult to overcome.

Among the complex software and systems used in digital humanities discussed in this chapter is 'Voyant Tools', an open-source system that provides various text-processing functions within a single package built on complex code accessible via GitHub. It has become clear that researchers who rely on it rarely concern themselves with the code or examine the source code, due to the inability of many humanities scholars to understand how these programmes work or to comprehend the meaning of the code itself. Consequently, they are unable to fully grasp how this code processes their research data.

This raises a critical question: how can we trust the results of automated humanities research that depends on complex software which the researcher cannot understand, cannot follow in its data processing, and cannot identify who makes the decisions or issues the judgments regarding its code, treating it effectively as a black box?

That is to raise the issue of software mediation which becomes a potentially new and interesting research question for the digital humanist (Berry, 2011, in Berry, 2023). This process of automation can also increase the tendency towards formalization made possible by “algorithmization” (Allison et al, 2011, in Berry, 2023) .

This also raises the question of the creation and maintenance of digital humanities’ intersubjective, historical, and philosophical concepts as manifested in and supported by these technologies (Liu, 2012, in Berry, 2023). So, for example, assumptions can be made about gender or race, or about which books are more important than others, such as a literature canon (Benjamin, 2019, in Berry, 2023) .

Therefore, algorithmic bias has become a major crisis confronting social, cultural, and political concepts, and it threatens the clarity and objectivity of the humanities. Nevertheless, digital humanities can address this bias by developing "approaches to these issues, digital humanities practitioners can helpfully draw on a increasing range of recent work which outlines best practice and principles for responsible use of AI in society" (Padilla, 2019).

where digital humanities can control the mechanisms that produce algorithmic bias and promote ethical approaches to the use of artificial intelligence in their research through a new approach based on an interdisciplinary framework. "the interdisciplinary traditions of digital humanities can make a vital contribution here. Cultural and media specialists can contribute to combatting bias in design; historians and linguists can assist in assessing the linguistic and other contexts that might generate bias " (Andrew, 2023).

MACHINE LEARNING IN THE HUMANITIES

The chapter focused on the way in which humanistic researchers use artificial intelligence and machine learning systems in digital humanities research. These are algorithms that use machine-learning techniques to sort and classify textual data into “buckets” that can represent themes or concepts drawn from the underlying data (Blei, 2012, in Berry, 2023). Machine learning relies on a process whereby the programmer does not necessarily instruct the system by writing an algorithm, rather the system is shown the input and the output and taught to “extract automatically the algorithm for this task” (Alpaydin, 2016, in Berry, 2023) .

The integration of Machine Learning (ML) techniques and methodologies in social and humanities sciences research has emerged as a transformative approach, offering innovative tools for analyzing complex datasets, uncovering hidden patterns and enhancing traditional research methodologies. The convergence of computational advancements and the increasing availability of large datasets have significantly heightened interest in applying ML to address multidimensional inquiries across various disciplines (Bitzenis, Koutsoupias, Nosios, 2025).

The effectiveness of employing machine learning systems in humanities research is linked to their ability to analyze large and complex datasets, to detect correlations and nonlinear relationships within the data, and to generalize their findings to new, previously unseen data. This makes model evaluation and validation crucial(Thabtah, Hammoud, Kamalov, Gonsalves, 2020).

It is essential to select the appropriate evaluation techniques and performance metrics, and strategies for handling imbalanced datasets are necessary to ensure these models' reliability, fairness, and real-world applicability (Liu, Zhang, Yang, Dong, Shen, Song, 2020).

Despite the extensive processing and analytical capabilities of machine learning algorithms, they are also susceptible to overfitting, poor generalization, and biases especially when dealing with imbalanced or unrepresentative data (Takase, Oyama, Kurihara, 2019).

ETHICAL APPROACHES TO THE USE OF ARTIFICIAL INTELLIGENCE IN DIGITAL HUMANITIES

The chapter examined the most recent literature addressing the ethical aspects of the automation of the humanities and the use of artificial intelligence in digital humanities, in order to arrive at an adequate answer to the question: What ought to be the principles or normative frameworks for guiding decisions in digital humanities projects and research units?

It addressed the question by presenting a new ethical vision for the ethics of digital humanities in the context of artificial intelligence, which it divided into three ethical approaches, namely:

Deontological approaches :

This approach is based on the idea of adhering to agreed-upon duties and fundamental principles, whereby decisions must be made in accordance with established moral norms and obligations, rather than on the anticipated outcomes, in "compliance with the highest ethical principle: the Categorical Imperative", that is not to use people as means to ends (van den Hoven, 2010, in Berry, 2023).

This creates the sense of a shared set of rules, perhaps professional guidelines, that also allow others to complain or hold to account those who break

these principles, for example the AoIR ethics guidelines for research (AoIR, 2021, in Berry, 2023) .

Consequentialist approaches:

Unlike the deontological approach, the consequentialist approach assumes that moral compliance is assessed based on the outcomes produced by an action, rather than on binding principles, one "ought to choose that course of action which maximizes utility compared to the alternatives" (van den Hoven, 2010). The more likely the actions that maximize that a given end is realized, that is that the moral definition of a good is achieved, the more right the decision (Anscombe, 1958, in Berry, 2023).

Virtue ethics:

This approach is based on values, ideals, and virtues as determinants for decision-making or the performance of any action. So the "virtuous person also possesses a general intellectual capacity, practical wisdom" and this enables them to "identify the morally relevant features in every situation and determine the right course of action" (MacIntyre, 2007, in Berry, 2023).

Through a review of the studies and literature covered in this chapter, it becomes clear that they focus on deontological and consequentialist approaches to considering the ethics of using artificial intelligence in the digital humanities.

These studies address a range of issues, such as data bias and algorithmic bias, the replacement of the human researcher by machines, problems of interpretation and the dominance of big data, the interpretive authority of algorithms, and other related concerns. The literature links these issues to the problem of inequality in the distribution and classification of data, as well as to the abundance of biased data, and assumes that solutions, along with democracy and objectivity, can be achieved through the production of more objective data in order to realize balance and equality in data processing. From an ethical perspective, this body of work is grounded in regulating researchers' use of artificial intelligence.

RESULTS and Discussion

The review of the field of study and research presented in this chapter, as reflected in the various studies and bodies of literature examined, shows that the emergence and consolidation of the digital humanities were not accompanied by an ethical framework to regulate the field and to guide researchers towards the optimal use of digital technologies, artificial intelligence, and machine learning. Despite the growing reliance of digital humanities scholars on algorithms and big data, with all the biases they entail, the research community has not yet converged on the establishment of a shared and clearly defined ethical guideline to govern the boundaries of the human researcher's use of machines.

Many studies that have sought to address the ethical problems arising from the use of artificial intelligence in the humanities and digital humanities conclude by urging researchers to cultivate virtue and to rely on self-regulation in accordance with the ethical norms commonly accepted within the scholarly community. This occurs in the absence of precise ethical standards and clearly defined boundaries governing the use of these new technologies, whose negative effects on humanities research outcomes are difficult to avoid. Consequently, reliance on virtue ethics alone is insufficient in contexts where humanities researchers employ software packages whose operational mechanisms are unknown and make use of biased data and algorithms whose structure and intended purposes are neither understood nor transparent.

This lack of formalized standards of good practice is notable considering the close relationships digital humanities have with digital archives, for example (eHumanities, 2010, in Berry, 2023), which tend to raise questions about decolonization, gender, race, class, and sexuality, not just in terms of representation but also in terms of the assumed categories inherited from physical archive practices or in using data scraped from the Internet (Liu, 2012, in Berry, 2023).

Conclusion

After reviewing Chapter Forty-Two of 'The Bloomsbury Handbook of Digital Humanities' (2023), together with the recent studies and literature it presents on the automation of the humanities, digital humanities, algorithms, and machine learning—and the methodological, academic, and ethical issues these developments raise at the core of humanities scholarship—it becomes evident that artificial intelligence has the potential to open new horizons for the humanities and for digital humanities as an emerging field of research .

However, this potential can only be realised through the rationalisation and ethical governance of the use of automation and AI within humanities research. Humanities inquiry is characterised by complexity and interpretative depth; it cannot be completed without a human presence, nor can the role of the researcher be ignored or replaced by algorithmic systems.

This introductory review of the studies included in the chapter further demonstrates that the use of artificial intelligence and algorithmic systems, with all the biases they carry, has generated new ethical challenges that threaten the authenticity of humanities research. Consequently, there is an urgent need to seek ethical frameworks for the use of digital tools and archives, facial recognition technologies, automated cataloguing, and various mechanisms of text generation within the context of digital humanities.

Accordingly, this article proposes a set of principles intended to lay new foundations for the ethical regulation of digital humanities research:

- Research ethics, particularly in the humanities, are well-established values that must be upheld within digital humanities research, including objectivity, neutrality, precision, and academic integrity.
- Respect for privacy must be ensured when using individuals' personal data.
- Integrity in the handling of historical data is essential, alongside an awareness of the risks associated with its algorithmic analysis, as algorithmic bias can distort historical truth.
- Academic misconduct must be avoided, and the principles of proper scholarly attribution must be respected.
- Intellectual property rights must be upheld, with data and information duly attributed to their rightful owners, while fabrication, falsification, and plagiarism are strictly avoided.

References

Berry, D. M. (2023). AI, ethics, and digital humanities. In J. O'Sullivan (Ed.), The Bloomsbury guide to digital humanities (pp. 445–457). University of Sussex.

Bitzenis, A., Koutsoupias, N., & Nosios, M. (2025). Machine learning in social sciences and humanities research: A structured literature review. Proceedings of the International Conference on International Business (ICIB) 2023–2024, 1(1), 1–17.

Liu, B., Zhang, H., Yang, L., Dong, L., Shen, H., & Song, K. (2020). An experimental evaluation of imbalanced learning and time-series validation in the context of CI/CD prediction. In Proceedings of the 24th International Conference on Evaluation and Assessment in Software Engineering (pp. 21–30).

Padilla, T. (2019). Responsible operations: Data science, machine learning, and AI in libraries. OCLC Research.

Prescott, A. (2023). Bias in big data, machine learning and AI: What lessons for the digital humanities? Digital Humanities Quarterly, 17(2).

Russo, G., Momtaz, Y., & Maqsood, S. (2025). Artificial intelligence in the digital humanities. Journal of Artificial Intelligence Research Innovation, 1(1), 42–45.

Takase, T., Oyama, S., & Kurihara, M. (2019). Evaluation of stratified validation in neural network training with imbalanced data. In Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 1–4).

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. Information Sciences, 513, 429–441.