Reflections on AI in Humanities: Amplifying marginalised voices of women

Raad Khair Allah

Department of English and Comparative Literary Studies, University of Warwick, Coventry, UK
Correspondence: Raad.Khair-Allah@warwick.ac.uk
ORCID: 0009-0002-1440-9563

Editorial review: This article has been subject to an editorial review process.

Abstract

As artificial intelligence (AI) continues advancing rapidly, there is growing potential for its application in the humanities to uncover new insights and perspectives from the historical archives. However, it is also important to consider how AI tools themselves may unintentionally perpetuate existing biases if not developed conscientiously. This critical reflection reflects on the opportunities and challenges of utilising AI to amplify marginalised voices that have been traditionally excluded or underrepresented in mainstream historical narratives, with a focus on women. Through natural language processing and computer vision techniques, AI shows promise in automating the analysis of large volumes of text, image, and multimedia sources to bring to the surface female narratives previously overlooked due to limitations of manual research methods. However, issues such as training data bias, problematic stereotypes learned from legacy sources, and a lack of diversity among AI researchers threaten to replicate the very inequities they are seeking to overcome if not addressed proactively. Collaborative frameworks and design principles centred on representation, accountability and community oversight are needed. By critically examining its social responsibilities and impacts, this reflection argues that AI possesses great potential in the service of feminist and intersectional scholarship when guided appropriately. It calls for continued multidisciplinary dialogue to help ensure technologies amplify marginalised voices rather than risk their further marginalisation.

Keywords: AI; humanities; marginalised voices; women; amplify

Artificial intelligence (AI) is the study and development of computer systems' ability to perform tasks normally requiring human intelligence, such as visual perception, decision-making, translation between languages, and scientific research. If used correctly, digitalisation can help enhance transparency and accountability, amplify marginalised voices, and disrupt entrenched systems of power and patriarchy. While we are seeing an AI revolution now that its research and development are advancing rapidly, AI's conceptual and research origins can be traced all the way back over seventy years to the pioneering work of scientists, such as Alan Turing, John McCarthy, and others. AI has deep-seated roots even if recent advances have been staggering. AI is a new way of doing research, where massive data processing is made possible by techniques of machine and deep learning, offering new perspectives for analysis.

There are many ways that AI is being applied in the humanities. Natural language processing (NLP) allows AI to analyse large text corpora, such as books, articles, and historical documents to identify themes, topics, and ideas, as well as how these evolve over time. In this respect, AI assistants answer questions from digitised archives and suggest related resources, which help disseminate humanities knowledge more broadly and thereby aid humanities research. Computer vision (AI) also facilitates the digitisation and analysis of visual elements, such as old photos, paintings, manuscripts and diagrams for easier access and preservation of humanities resources. A further noteworthy point in relation to the visual features of AI is that its generative models also produce new texts, images, and music in styles of historical periods or individual artists, which in turn extends creative works and enhances cultural exploration. Furthermore, AI is used in sentiment analysis and emotion recognition by analysing literary texts, such as novels, poems and plays to understand how authors portray and evoke different emotions over time. In doing so, AI not only provides insights into cultural and historical trends, but also studies viewers' emotional responses to artworks including paintings, sculptures, performances, and films to better understand human experiences and cultural impacts over eras.

Moreover, it helps determine emotions conveyed through images, ornamentation and symbolism in artefacts from different eras and cultures including by tracking emotional arcs and character development in plays as well as novels in order mechanically to validate or challenge literary theories. This can be done by using facial recognition and physiological sensors, which help interpret artistic intent and its impact. For instance, digital humanities scholars often employ AI tools and methodologies to study and analyse cultural artefacts. They may use AI to examine emotional patterns, symbolism, and character development in literature, plays, and the visual arts, aiming to gain insights into cultural

and historical contexts. Literary theorists and critics may also collaborate with AI researchers to validate or challenge literary theories using computational approaches. They provide the theoretical framework and expertise in interpreting literary works while leveraging AI tools to analyse large volumes of texts and identify patterns related to emotions, symbolism, and character development. Art historians and curators may also utilise AI to analyse visual artworks, including paintings, sculptures, and artefacts from different eras and cultures. Therefore, AI-based techniques can help to identify and interpret emotional cues, symbolism, and stylistic features, providing insights into the artistic intentions and cultural contexts of the works. Yet, collaborations between these professionals and others from related fields allow for interdisciplinary investigations into the emotional and artistic aspects of cultural artefacts. By combining AI technologies with domain-specific expertise, researchers can gain new perspectives and insights into the emotional dimensions of literature, plays, visual arts, and other forms of cultural expression; since AI helps to survey public reaction on social media to cultural events, exhibitions, films, and suchlike gauging sentiments and how they vary across demographics.

However, there are some key constraints on an AI's ability fully to recognise and understand emotional dimensions. One of the significant problems that AI perpetuates and faces is data limitations because marginalised groups have often left fewer records. Historical records and archives, which are valuable sources of data for AI research, may have significant gaps and underrepresentation of marginalised groups, such as black women, immigrants, and refugees. This can be due to factors such as limited access to education, cultural biases, and exclusionary practices. As a result, AI models trained on historical data may fail to capture the full range of experiences and perspectives of marginalised communities, perpetuating biases and reinforcing existing power imbalances.

Besides, most AI systems are trained on texts, images, and other data that do not necessarily capture the full complexities and subtleties of human emotion and social context. Therefore, more diverse data is needed. Another problem is the subjectivity of emotions; since what emotions mean can vary greatly between individuals, cultures, and situations based on personal experiences. Yet, AI generalises from patterns, which limits its ability to recognise nuance. In addition, emotions are often expressed and experienced via multiple channels such as speech, facial expressions, and physiology which current AI is not sufficiently integrated to apprehend holistically. Moreover, the same emotion expressed may carry very different meanings depending on cultural norms, relationships between people, and the events that give rise to it, which AI struggles with. It is worth considering here that emotions in real life are often ambiguous,

mixed, conflicting, or evolving – harder characteristics for AI reliably to discern compared to clear labels. Thus, for an AI truly to understand how emotions function and impact humans it would likely need a kind of internal experiential model and self-awareness that has yet to be achieved. Hence, while AI can identify surface expressions, it does not have access to the deeper interconnected web of emotional knowledge that humans intuitively possess from living as emotional, social beings. Thus, more human-aligned data and progress on general AI is still needed for machines to recognise emotional realities properly.

Despite its inability to understand complex human emotions and while many histories and stories from marginalised groups have been underrepresented or overlooked in traditional scholarly works, AI has the potential to amplify marginalised voices including women's voices. This is done by analysing and processing historical documents shared by women in different languages and so exposing narratives that have remained untapped, such as letters, diaries and newspapers at a scale that exceeds human capabilities. In the last few decades, a significant body of work has been done to expose these narratives through gender studies and queer studies. However, AI algorithms can analyse large volumes of text data much faster than human researchers. While a human researcher may take a significant amount of time to read and understand a single document, AI models can process thousands or even millions of documents in a relatively short period. This enables researchers to examine a broader range of materials and identify patterns, themes, and narratives more efficiently. In addition, AI-powered natural language processing (NLP) techniques allow algorithms to understand and interpret text in multiple languages. With the ability to analyse diverse linguistic sources, AI can overcome language barriers and enable researchers to access historical documents written in different languages, including those shared by women. This expands the scope of available data and provides a broader understanding of untapped narratives across cultures and regions. AI models can be also trained to categorise and extract relevant information from documents, for example by identifying dates, locations, names, and themes. This automated process can help researchers organise and index large collections of historical documents, making it easier to search for specific narratives or topics of interest. Furthermore, AI algorithms can identify connections and relationships between documents, uncovering hidden narratives that may have been overlooked by human researchers. In so doing, AI can facilitate the cross-referencing and linkage of information across various sources. By connecting related documents and data points, AI algorithms can provide a more comprehensive understanding of historical events, movements, and individuals. This capability allows researchers to uncover connections and narratives that

go beyond the limitations of human memory and manual analysis. While AI offers significant advantages in processing historical documents and exposing untapped narratives, it is important to note that human interpretation and contextual understanding remain essential. Hence, the use of AI in historical research should be complemented by human expertise to ensure accurate interpretation, critical analysis, and nuanced understanding of the historical context.

AI serves as a tool that enhances human capabilities, enabling researchers to explore a broader range of historical narratives and shed light on previously overlooked perspectives. The technologies of optical character recognition, machine translation and named entity recognition are often used to do so. Amplifying marginalised women's voices allows for more inclusive discovery and thus diverse telling of history that acknowledges the experiences and contributions of all people, for AI generates both structured metadata and summaries synthesising AI insights from vast distributed collections in order to guide new feminist and intersectional scholarship. Additionally, it identifies sentiments, speech patterns, and emotional arcs within first-person women's narratives using stylistics, sentiment analysis, and natural language generation techniques. AI also helps to classify genres of music based on decoded emotions to explore their historical evolution and how styles reflect lived experiences, thereby detecting emotion in marginalised voices of women in oral histories to better understand oppressions of the past from an empathetic standpoint. Consequently, it democratises access to women's histories globally through multilingual and interdisciplinary digital frameworks grounded in feminist practices.

However, there are also potential downsides to consider when it comes to the use of AI in this context. The digitisation and analysis of historical documents, particularly personal materials such as letters and diaries, raise ethical concerns regarding privacy, consent, and the ownership of personal narratives. Thus, care must be taken to ensure that women's histories are handled with sensitivity and respect, obtaining appropriate permissions and considering the potential impact on individuals and communities connected to those narratives. Additionally, AI models may struggle with capturing the full range of cultural and linguistic nuances present in historical documents, especially across different languages and regions. These complexities of language, context, and cultural references may be challenging for AI algorithms to interpret accurately, and so can result in misrepresentations or oversimplifications of women's histories, potentially erasing important subtleties and nuances from the narratives. AI algorithms may also struggle with the nuanced interpretation and analysis of historical documents, particularly those with complex symbolism, metaphor, or cultural context. Human researchers often bring

contextual knowledge, critical thinking, and interpretive skills that AI models may not fully replicate. Thus, it is important to consider AI as a tool that augments human expertise rather than replacing it entirely. To mitigate these downsides, it is crucial to adopt transparent and inclusive practices in AI development. This includes ensuring diverse representation in data collection, involving domain experts and historians in the curation and interpretation of data, and regularly evaluating as well as addressing biases and limitations in AI models. This can be done by engaging in ongoing dialogue and collaboration with communities and stakeholders, who can help ensure that the digitisation and analysis of women's histories are conducted ethically and responsibly.

Alternative stories and perspectives challenge dominant patriarchal narratives, which have often excluded or misrepresented women's voices. Moreover, AI helps to track the shifting of societal moods and viewpoints regarding critical social and political issues, such as war, sexuality, politics, and social change. As a result, it will enrich understanding of lived experiences from an empathetic perspective and so help to address exclusion and to revise prevailing versions of events. It is worth mentioning here that gaining new perspectives and engaging with voices from diverse backgrounds not only enhance our ability to think broadly and critically about history, society, and culture from multiple vantage points, but also help us to avoid the repetition of biases. If such marginalised voices are not heard, the viewpoints and biases that have traditionally marginalised these groups risk being perpetuated. Hence in many ways, AI expands the discovery, access, analysis, and enjoyment of humanities knowledge through an unbiased lens if a combination of strategies and practices across various stages of AI development has been adopted. Some of these key approaches include ensuring that the training data is diverse and representative; data pre-processing and cleaning; interdisciplinary collaboration; and ongoing monitoring and evaluation. By adopting these approaches, we can strive towards developing AI systems that are more equitable, fair, and unbiased. AI itself is not inherently biased or unbiased. The biases in AI systems arise from the data used to train them, the design choices made during their development, and the potential biases of the human creators involved in those processes. Addressing the issue of bias in AI requires careful attention throughout the entire development pipeline. It involves collecting diverse and representative datasets, ensuring inclusive and fair annotation processes, applying rigorous testing and evaluation procedures, as well as implementing fairness-aware techniques in algorithm design. Ethical considerations, transparency, and accountability are essential in mitigating biases and ensuring that AI systems are developed and deployed in an unbiased manner. It is important to note that achieving

completely unbiased AI systems is challenging and may prove not entirely possible.

Indeed, in some ways AI systems may unintentionally perpetuate gender or ethnic bias. For example, if training datasets are disproportionately made up of information about, or from one gender, algorithms can learn inherent biases that reflect current or historical imbalances. In addition, machine learning systems trained on large corpora can inadvertently pick up on problematic or offensive stereotypical gender associations from language over time. If systems are optimised for metrics like user engagement, they may autonomously leverage known arbitrary preferences that happen to advantage one gender. Also, making decisions based strictly on correlations in data without awareness of societal biases could disadvantage groups subjected to historical discrimination. When systems exhibit unexplained behaviours, it becomes difficult to audit, correct, or mitigate any gender harms that may unintentionally emerge. Compounding this if women are underrepresented in the AI field and in datasets during system development, their unique perspectives on design risks and harms are less likely to be considered Therefore, careful consideration of these factors during development helps to reduce unintended bias, but ongoing monitoring and evaluation of AI systems for biases are still important. This involves auditing the data, assessing the model's performance across different subgroups, and conducting fairness assessments to identify and mitigate any biases that may emerge during deployment.

To conclude, the goal of AI should be to empower women's voices both for the benefit of women and to increase diversity of perspective in general rather than to have a technology that speaks for them. Broader issues of inclusion and power structures still demand human interventions as technology alone cannot undo systemic marginalisation. Ongoing critical examination of how AI systems themselves reflect or perpetuate gender biases is also still needed. AI produces opportunities but also carries responsibilities towards marginalised groups if it is to be developed responsibly by multiple stakeholders who play important roles at different stages of AI development including AI developers and researchers, data providers and curators, as well as regulators and policymakers.

Raad Khair Allah is a PhD candidate at the Faculty of Arts/ Department of English and Comparative Literary, University of Warwick, UK. Her thesis title is 'Contemporary Arab Women Writers, Filmmakers, and Artists in an International Frame'. Her research interests include Women's Studies, Gender, Feminism, Sexuality, War and digital humanities. She is also a Senior Teaching Assistant at the Faculty of Arts/ Department of English Literary Studies and a former member of the seminar series organising committee at CSGW/Center for the Study of Women and Gender at the same institution. She was shortlisted for the Paula Svonkin Creative Art Award at the Pacific Ancient and Modern Language Association (PAMLA) conference in Los Angeles, USA, 2022 for her project on the use of Miro software in Humanities titled "Marginalization of Arab Women and Revolutionizing Patriarchy". Prior to joining the University of Warwick, she worked as an English lecturer at Damascus University (part-time, 2009-2012) and the Syrian Private University (full-time, 2014-2018).

To cite this article:

Khair Allah, R., 2023. Reflections on AI in Humanities: Amplifying the marginalised voices of women. *Exchanges: The Interdisciplinary Research Journal*, 11(1), 77-84. Available at: https://doi.org/10.31273/eirj.v11i1.1453.