 ISSN: 2349-6002

Cultural Memory in the Age of AI Archives

Shailja Nimavat
Ph.D. Scholar, Veer Narmad South Gujarat University, Surat, India

Abstract- This paper examines how generative Artificial Intelligence (AI) systems that are trained mostly on Western literary corpora function as the only agents of cultural memory whose operations reproduce longstanding colonial hierarchies. As AI mediates how readers encounter, interpret, and understand literature more and more everyday it acquires power over what is culturally visible and what is obscured. Through the lens of postcolonial theory, digital humanities, and memory studies, the paper argues that non-Western literary traditions appear in AI outputs as spectral presences i.e. flattened, misrepresented, or entirely absent. These "ghosts of the canon" emerge from structural imbalances in digitization, translation, and data availability, which skew AI's narrative frameworks toward Anglo-European norms. The study analyzes how algorithmic summarization, recommendation systems, and stylistic generation reinforce Western conceptions of literary value while marginalizing Indigenous, African, South Asian, Middle Eastern, and diasporic traditions. It further considers the consequences of relying on AI-generated explanations in classrooms, where such omissions risk shaping a monocultural understanding of global literature. Ultimately, the paper contends that AI does not simply reflect the biases of existing canons but rather actively participates in rewriting them. By highlighting the stakes of algorithmic erasure, the study calls for decolonized data practices, community-led corpus building, and transparent archival infrastructures that make room for the plural, entangled histories of world literature.

Keywords: AI archives; cultural memory; postcolonial literature; algorithmic erasure; global literary canon; data bias

I. INTRODUCTION

Questions concerning the distribution of cultural wealth and who controls how people use and interpret books have become increasingly pertinent as generative AI technology becomes a central part of this process. Currently, LLMs (large language models) such as GPT-5 and Gemini are trained on vast digitized text databases that generally consist of English-language texts from Western sources. When LLMs use these databases as training data, they create a whole culture based on Western literature and literary knowledge. One of the consequences of this situation is that LLMs have been used to create reading lists, provide analysis, create literary histories, and even imitate the writing styles and techniques of well-known authors from the West. It is worth noting that the training data on which the AI systems are built is missing large parts of the world's literary history. In this way, AI becomes what Achille Mbembe calls a "technology of command" or a structure that decides "who is remembered, who is forgotten, and how both are constituted as subjects" (Mbembe, 2001).

As a result, as a result of the AI's reliance on Western digitised text collections, many non-Western literary cultures and traditions exist as 'ghosts' within the AI's mind – faint, partial, and sometimes skewed representations that exist only at the outer reaches of the algorithm's memory. Drawing from theories of Postcolonial Studies, Digital Humanities, and Cultural Memory Studies, this paper will explore the manner in which AI technologies have inherited and replicated colonial power relationships through the processes of selection, classification, and generation of data. It will also look at the impact of LLMs on educational and cultural systems by examining the implications of creating a decolonial dataset design, creating a decolonial archival practice, and establishing a decolonial framework for algorithms.

II. AI AND CANON FORMATION: A NEW CULTURAL GATEKEEPER

The AI models that researchers have labelled as neutral tools. They only reflect the text data that they have been trained on. However, Humanities Scholars argue that nothing is innocent from any one dataset.

Wendy Chun argues that Big Data will always reflect the past. As such it cannot exceed that particular time frame (Chun, 2016). In this way, when training data contains a disproportionate number of Western Novels, Western Essay papers, Western Popular culture/cinema/television, the large language models will reflect the Western Canon formation structure.

Throughout history, literary canons have occurred based on institutionally sanctioned forms such as University syllabi, Anthologies, Colonial Education Systems, Developing neo-colonial societies by Western Educational Systems, and Academic Criticism. Canon formation is closely tied to a "distribution of Cultural Capital" (Guillory, 1993). As a result, today's Artificial Intelligence Systems replicate algorithmically the same power structures as previous forms of Canon formation. In the instances where users request an Artificial Intelligence software, an “important world novels”; the AI systems respond based primarily on the data included in the training databases by suggesting authors such as Dickens, Austen, Tolstoy, Hemingway, Orwell, etc., who are represented most often in the training data and are also easily digitized, translated, and distributed.

This results in a feedback loop whereby algorithmic repetition makes the Western canon more apparent and ingrained. Literature from Africa, South Asia, the Caribbean, the Pacific, or Indigenous communities, on the other hand, is rarely found and is frequently represented by a small number of already well-known individuals like Rabindranath Tagore or Chinua Achebe. The outcome is a structural silencing of the subaltern within systems that assert universality, which Gayatri Spivak terms "epistemic violence" (Spivak, 1988).

III. ALGORITHMIC AMNESIA: THE STRUCTURAL ERASURE OF NON-WESTERN TRADITIONS

The creation of AI training data heavily relies on digitization. However, digitization itself is uneven: Western texts were digitized in much larger quantities, earlier, and more methodically. "Digitization is not neutral; it reproduces the inequalities of the colonial archive," observes Roopika Risam (Risam, 2018). AI systems are unable to recognize many oral, Indigenous, and precolonial literary forms because they are either undigitized or lack standardized metadata.

This type of structural erasure is referred to in this paper as algorithmic amnesia. The lack of non-Western texts is not coincidental; rather, it is linked to linguistic hierarchy, colonial extraction histories, and unequal infrastructure. For example:

- African oral epics like the *Sundiata* or *Epic of Mwindo* often survive in community performance rather than print archives.
- South Asian regional literatures—Assamese, Meitei, Tamil Sangam poetry, Sindhi folktales—remain poorly represented in global digital repositories.
- Indigenous North and South American literatures may be protected by tribal sovereignty rules that restrict digitization.

AI models interpret a lack of data as a lack of culture. Therefore, an AI system might reduce an entire continent to "Chinua Achebe, Ngũgĩ wa Thiong'o, and Nadine Gordimer" when asked to describe African literature. It frequently creates generic stories full of stereotypes—eagles, wise elders, spiritual journeys—when asked to create a "folktale in a Native American style," exposing the shallowness of its representational knowledge.

Algorithmic colonialism is how these distortions work. They erase marginalized cultures rather than just misrepresenting them.

IV. DISTORTION, STEREOTYPE, AND THE COLONIAL IMAGINARY

AI frequently reverts to Western narrative conventions when it tries to produce or analyze non-Western literature. For example, African novels are framed as "postcolonial struggles," South Asian epics are reduced to "heroic journeys," and Orientalist tropes of "mysticism" or "desert landscapes" are used to describe Middle Eastern poetry. The "Orientalist gaze," which Edward Said describes as a means of "dominating, restructuring, and having authority over the Orient," is reflected in these patterns (Said, 1978).

The issue is that AI interprets non-Western literary traditions through the prism of Western cultural memory rather than just making factual errors.

Examples of common distortions include:

- Equating all East Asian poetry with haiku.
- Treating African literature as monolithic and uniformly political.
- Erasing women writers in regions where translation and preservation disproportionately favored male authors.
- Imposing Western plot structures—rising action, climax, resolution—onto cyclical or fragmentary oral forms.

The underlying presumptions of the algorithm are exposed by these distortions: that Western literary values are normative, that Western narrative forms are universal, and that other literary cultures must be translated into those terms in order to be "recognizable."

V. AI AS A COLONIAL INTERPRETER

AI interprets rather than just storing data. AI thus emerges as a new type of colonial middleman. Colonial administrators translated and categorized colonized cultures in the nineteenth century, frequently misrepresenting them. These days, AI models carry out a similar task by producing summaries, justifications, and interpretations for audiences around the world.

"Can the subaltern speak?" asks Spivak (Spivak, 1988). The question that arises in the context of AI is whether a machine that has been trained not to listen can hear the subaltern.

AI's interpretive power is vast. Students increasingly rely on AI for literary summaries that frame how they understand authors, genres, and historical contexts. Teachers turn to AI to generate reading lists, comparative analyses, and instructional materials that guide classroom priorities. Publishers deploy AI tools to classify manuscripts, determine market categories, and even predict reader interest, while researchers use AI-driven search tools to locate primary texts and secondary scholarship.

When AI systematically omits or misrepresents non-Western traditions, it shapes not only individual readers' interpretations but also global cultural memory. As theorist Andreas Huyssen observes, "What we remember is shaped by the media through which we remember" (Huyssen, 2003). AI has become one such medium—but it remembers selectively.

VI. PEDAGOGICAL CONSEQUENCES: WHAT LITERATURE STUDENTS LEARN AND FORGET

In schools and universities, AI-generated literary explanations are increasingly used as shortcuts. This poses significant risks for the study of world literature. If AI suggests Western texts as the default examples of literary concepts, students learn to associate "literature" with Western forms.

For example:

- Ask an AI for "examples of stream of consciousness," and it returns Joyce, Woolf, and Faulkner, rarely mentioning Bapsi Sidhwa or Qurratulain Hyder.
- Ask for "epic narratives," and the responses prioritize *The Odyssey*, *Beowulf*, and *The Aeneid*, rather than *Mahabharata*, *Shahnameh*, or *Epic of Sundiata*.

This technology thus becomes complicit in what Ngũgĩ wa Thiong'o calls "the domination of the mental universe of the colonized" (Ngũgĩ, 1986). When Western texts dominate both the dataset and the classroom, non-Western literary traditions become peripheral—visible only as token examples.

Cultural Memory and the Ghost Archive

Cultural memory is shaped by the institutions that transmit and preserve it; it is never neutral. Though they carry over the colonial logic of the archive, AI models operate as memory systems. Archives "do not simply preserve memory—they produce it," according to Michelle Caswell (Caswell, 2014). In a similar vein, AI uses probabilistic selection to curate memory, making what it repeats culturally visible and forgetting what it leaves out.

Thus, non-Western literature often appears in AI outputs as:

- Ghosts (faint, distorted echoes)
- Specters (invoked but not understood)
- Silences (total erasure)

This spectral presence is systemic rather than coincidental. It reveals the underlying cultural politics of AI, i.e. long-standing global inequality, capital, digitization, and translation all influence memory.

VII. TOWARD DECOLONIZING AI LITERARY SYSTEMS

Deep structural changes in educational, cultural, and technological institutions are necessary to address algorithmic erasure. First, it is crucial to create reliable non-Western literary datasets. To produce multilingual, locally controlled corpora that accurately reflect a variety of literary traditions, communities, academic institutions, and cultural institutions must work together. Oral texts, Indigenous stories, precolonial manuscripts, regional literatures, and the extensive corpus of women's writing that has historically been marginalized or left out of previous archives are all being digitized as part of this effort.

Second, systems for classification and metadata need to be rethought. Metadata should be based on Indigenous knowledge frameworks, postcolonial theory, and community-specific taxonomies that respect local epistemologies rather than Western genre labels or colonial classifications.

Third, it's critical that AI training data be transparent. Businesses creating AI models should reveal the linguistic, cultural, and geographic makeup of their datasets so that researchers and users can recognize and challenge preexisting biases.

Fourth, a key component of any endeavor to decolonize AI systems should be community governance. The representation, accessibility, and interpretation of non-Western literary traditions in digital infrastructures must be shaped by these communities.

Lastly, pedagogical intervention is essential. In order to help students understand algorithmic bias as a type of cultural power and to challenge how AI-produced outputs affect their comprehension of world literature, educators must provide them with critical AI literacy.

VIII. CONCLUSION

Although generative AI has grown to be a significant player in international literary culture, its perception of the world is not uniform. AI systems replicate the colonial hierarchies ingrained in historical archives because they are primarily trained on Western texts. As a result, non-Western traditions are reduced to ghostly presences that are rarely seen, frequently distorted, and frequently erased. These algorithmic omissions are not accidental; rather, they are a reflection of long-standing differences in what has been valued, preserved, translated, and digitalized within global knowledge systems. Because of this, reading lists, literary evaluations, and historical accounts produced by AI often recur. Because of this, Anglo-European viewpoints are often repeated in AI-generated reading lists, literary analyses, and historical narratives, perpetuating a limited understanding of what constitutes "world literature." This essay has argued that AI contributes to literary interpretation, cultural memory construction, and canon formation in ways that subtly but effectively uphold Western dominance and the epistemic disparities of previous colonial and imperial infrastructures.

However, this future is not certain. AI is shaped by historical biases that can be rectified. We can envision literary systems that truly reflect the diversity of global culture by implementing decolonial practices in data design, digitization, metadata creation, and AI governance. This entails creating multilingual corpora, digitizing Indigenous and oral literary traditions, and promoting community-led archival practices that subvert Western institutions' authority. Additionally, it calls for ethical cooperation with underrepresented literary communities, critical transparency regarding training data, and educational interventions that encourage students to challenge algorithmic authority. AI has the capacity to both shed light on long-forgotten histories and elevate voices that have been marginalized in cultural memory. But only if we consciously change the technological and cultural

frameworks that dictate what AI can learn, remember, and eventually represent will this potential be fulfilled.

REFERENCES

[1] Benjamin, Walter. *The Work of Art in the Age of Its Technological Reproducibility.* Harvard University Press, 2008.

[2] Caswell, Michelle. "Toward a Survivor-Centered Approach to Records Documenting Human Rights Abuse." *Archival Science* 14, no. 3 (2014): 273–293.

[3] Chun, Wendy Hui Kyong. *Updating to Remain the Same: Habitual New Media.* MIT Press, 2016.

[4] Guillory, John. *Cultural Capital: The Problem of Literary Canon Formation.* University of Chicago Press, 1993.

[5] Huyssen, Andreas. *Present Pasts: Urban Palimpsests and the Politics of Memory.* Stanford University Press, 2003.

[6] Mbembe, Achille. *On the Postcolony.* University of California Press, 2001.

[7] Ngũgĩ wa Thiong'o. *Decolonising the Mind: The Politics of Language in African Literature.* Heinemann, 1986.

[8] Risam, Roopika. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy.* Northwestern University Press, 2018.

[9] Said, Edward W. *Orientalism.* Vintage Books, 1978.

[10] Spivak, Gayatri Chakravorty. "Can the Subaltern Speak?" In *Marxism and the Interpretation of Culture*, edited by Cary Nelson and Lawrence Grossberg, University of Illinois Press, 1988.