P-ISSN 2708-6453	Pakistan Languages and Humanities Review	Nov-Decembe2026, Vol. 10, No.6
O-ISSN 2708-6461	https://doi.org/10.47205/plhr.2026(10-VI)03	[46-65]

Pakistan Languages and Humanities Review
www.plhr.org.pk

RESEARCH PAPER

Algorithmic Orientalism: A Conceptual Framework and Critical Discourse Analysis Protocol for Examining Generative AI Interpretations of Canonical English Texts from a South Asian Perspective

Hossain Ahmed

Lecturer, Department of English Language and Literature, International Islamic University Chittagong, Bangladesh

*Corresponding Author	hossainahmed.iiuc@gmail.com

ABSTRACT

This article develops a framework and protocol for examining how ChatGPT, Claude and Gemini interpret six canonical English texts taught in Bangladesh, limited to English-language outputs. Students obtain literary interpretations from AI systems. Whether these reproduce Orientalist relations of knowledge is unexamined in South Asian literary education. The qualitative design applies Fairclough's Critical Discourse Analysis, informed by Said and Spivak, to 36 responses to paired neutral and Bangladesh-positioned prompts. This protocol paper reports no data. Algorithmic Orientalism is defined through four indicators: Othering, stereotyping, epistemic hierarchy and universalism. Decision rules specify when the hypothesis is rejected. Four propositions predict that AI readings appear universal, apply Western frameworks, merely mention South Asian traditions and stereotype Bangladeshi readers across models. Literature curricula should teach students to read AI interpretations as located discourse. Researchers should apply the protocol and extend it to Bangla prompts and non-US systems.

KEYWORDS Algorithmic Orientalism, Generative AI, Critical Discourse Analysis, Postcolonial Theory, Research Protocol, Literary Education, Bangladesh

Introduction

A student in Dhaka, Chattogram or Rajshahi who meets *The Tempest* or *Heart of Darkness* for the first time can now ask a conversational AI system what the text means. Generative large language models (LLMs) such as ChatGPT, Claude and Gemini return fluent, confident readings of literary works on demand, and their uptake in education has outpaced the research that might evaluate it (Kasneci et al., 2023). Early experimental work has asked whether such systems are reliable guides to ambiguous literary texts (Frederking, 2025). A prior question has received less attention: whose interpretive tradition does the machine speak for?

That question has a long history in South Asia. Viswanathan (1989) showed that English literary study was institutionalised in colonial India as an instrument of moral and political management. The departments that inherited this history, including those in Bangladesh, have long debated how far their curricula and theoretical vocabularies remain tied to it (Al-Quaderi & Al Mahmud, 2010; Hamid, 2021). The canon debate was therefore already an argument about epistemic authority. Generative AI adds a new participant to it.

OPEN ACCESS © 2026 Author/s. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0).

There are grounds to suspect that this participant is not neutral. LLMs are trained on corpora in which English-language, Global North sources are over-represented, and they reproduce the regularities of that material (Bender et al., 2021). Tao et al. (2024) found that the cultural values expressed by leading models sit closest to those of English-speaking and Protestant European societies. Agarwal et al. (2025) found that AI writing suggestions moved Indian participants towards Western styles of expression. If this happens with everyday writing, it may also happen with interpretation, where what counts as a legitimate reading has always been culturally and institutionally regulated.

This article examines that possibility under the name *algorithmic Orientalism*. The term is not new. Dahman and Saou (2025) use it for anti-Muslim bias in AI systems, and related work speaks of digital and computational Orientalism in LLM outputs about Palestine and the Middle East (King et al., 2026; Shahid, 2026). These studies concern how models represent peoples and regions. The present study carries the concept into a different domain, literary interpretation, and a different region, South Asia. It also treats the concept as a hypothesis to be tested, which is why the title poses it as a question. Following Said (1978) and Spivak (1988), the concern is whether AI readings of canonical English texts present Western critical frameworks as the universal language of interpretation, and whether they represent South Asian readers and traditions through stereotype or as a supplement.

Four research questions follow, and the protocol set out in this article is designed to answer them. The first asks what linguistic patterns and interpretive assumptions characterise the readings that ChatGPT, Claude and Gemini produce of canonical English texts commonly taught in Bangladesh. The second question is what Western and Eastern critical lenses are used by the models; and how, if at all, a hierarchy is developed between these two lenses. The third considers how the models reflect South Asian points of view, cultural backgrounds, and interpretive practices and what that looks like when the question asks the reader to think in a South Asian context (specifically in Bangladesh). The fourth question examines how these patterns change in the three models of teaching reading and across British-authored and South Asian-authored texts.

This article does not report data. It makes the conceptual and methodological case for such a study and specifies it in full. The protocol applies Fairclough's (1995) Critical Discourse Analysis (CDA) to 36 AI-generated responses: three models, six texts, and two prompts per text. Its scope is bounded. It analyses the discourse of outputs in English. It supports no claims about model internals or training data, or about how students use such outputs. Bangladesh is the standpoint from which the prompts are written, and South Asia is the regional frame in which results are to be interpreted.

The contribution is therefore very particular. Previous audits of cultural bias in LLMs have largely focused on value surveys and stereotype benchmarks, while political and regional representation in LLM studies have been focused on studies on Orientalism. This article provides an operational definition of algorithmic Orientalism for literary interpretation, clearly articulated criteria for disaffirming it, four propositions for empirical investigation, and a paired-prompt protocol (from Bangladeshi perspective) which can be performed by others and repeated.

The contribution is therefore very particular. Previous audits of cultural bias in LLMs have largely focused on value surveys and stereotype benchmarks, while studies of Orientalism in LLM outputs have concentrated on political and regional representation in the Middle East. This article provides an operational definition of algorithmic Orientalism for literary interpretation, clearly articulated criteria for disaffirming it, four propositions

for empirical investigation, and a paired-prompt protocol from a Bangladeshi perspective that can be replicated and extended by others.

Literature Review

Algorithmic Bias and the Cultural Politics of Large Language Models

It has been shown through critical scholarship that algorithmic systems reflect hierarchies that exist in the data and institutional processes that create it. Although search engines might seem to be ranking content neutrally, Noble (2018) has shown that they actually reinforce racist and sexist narratives under the cover of neutrality, and Birhane (2021) has suggested that this is a relational harm and can't be resolved by a simple technical tweak. As LLM's came along, this critique became extended to the language itself. Bender et al. (2021) argued that models trained on large corpora, which are generally uncurated, must express dominant views, and that if the model is very fluent in the language, its apparent understanding may mask the presence of biases. Gallegos et al. (2024) also showed that bias is present and can be seen across different aspects of the design of LLMs (data, architecture, and evaluation), and that post-hoc correction is hard.

The 'cultural' aspect of this bias is directly relevant to the current research. The results of the leading models are very similar to that of the English-speaking and Protestant European societies and culturally specific prompting does not completely eliminate this congruence, as concluded by Tao et al. (2024). When models were asked about the stereotyped associations they created in Arabic and with Western contexts, Naous et al. (2024) reported that they still preferred Western entities and created stereotyped associations. Two additional research studies about the identity prompting further confirm this concern. Cheng et al. (2023) used the sociolinguistic notion of markedness to show that model generated descriptions of non-default groups are systematically different from those of the unmarked (default) group that is, they are made via difference. Wang et al. (2025) have shown that when models are trained to generate speech from a perspective of identity group, they often overgeneralize or stereotype the group and ignore intra-group variation. As a whole, these studies indicate that use of a prompt outside the context of the West does not consistently lead to a "situated" answer and might lead to the production of a stereotype.

South Asia and Bangladesh in AI Research

A study of this area shows that even the concepts of bias that are studied are often western. According to Sambasivan et al. (2021), the principles of fairness, when framed by race-gender in the United States, do not easily travel to India where the principles of inequality are different and are embodied in the caste, religion and regional system. Khandelwal et al. (2024) discovered that popular LLMs generated caste and religious stereotypes in the Indian context. Qadri et al. (2023) documented their experience of generated images as exoticising and outsider looking in a community-centred study of text-to-image models, with South Asian participants. Agarwal et al. (2025) also found a behavioral impact, as AI recommendations gradually led to a shift in Indian users towards western stylistic patterns in writing.

There are two major drawbacks in the amount of literature currently available as it relates to the present study. It has a distinctively Indian focus. Bangladesh is a place where language nationalism is unique, the public culture is Muslim with its distinctive nature, and the educational and political experiences are separate.In the context of language nationalism, Muslim majority public culture, and unique educational and political

trajectory, Bangladesh is seldom looked upon as an autonomous entity. Besides, the previous studies focus on values, social stereotypes, images and daily writing. This scholarship is rather sparse when it comes to the major activity in the humanities – interpretation – and its centrality to understanding LLM engagement with culturally specific practices of literary reading and meaning making are largely unfulfilled.

Orientalism, Digital Colonialism and Decolonial Approaches to AI

A second strand of scholarship supplies the political vocabulary for analysing these dynamics. Kwet (2019) described digital colonialism as the control of the Global South's digital infrastructure by corporations based in the United States, and Ricaurte (2019) argued that data-centred epistemologies impose particular ways of knowing as universal. Mohamed et al. (2020) proposed decolonial theory as a means of identifying sites of coloniality in algorithmic systems, and Adams (2021) cautioned that calls to decolonise AI risk becoming merely rhetorical unless they confront the Eurocentric conception of reason built into the field. Zembylas (2023) extended this argument to AI in higher education. An older body of work on techno-Orientalism had already examined how Asia is imagined through technology in Western culture (Roh et al., 2015).

Recent empirical studies have begun to test these ideas directly on model outputs. King et al. (2026) analysed ChatGPT outputs in English and Arabic and reported Orientalist framing of Palestinians. Shahid (2026), in a preprint, scored conversations with two models about the Middle East on dimensions of Orientalist discourse. McKenna and Tshuma (2026), also in a preprint, submitted identical prompts to four LLMs and read the outputs through postcolonial theory. Dahman and Saou (2025) used the term algorithmic Orientalism for anti-Muslim bias across AI systems. These studies collectively establish that output-level analysis of Orientalism in AI is feasible and productive. None, however, addresses literary interpretation or South Asia, which defines the specific contribution of the present study.

Canon, Colony and English Studies in Bangladesh

The canonical texts that the models are asked to interpret are themselves products of the colonial encounter (Viswanathan, 1989). The texts selected for this study have been central to that debate. Achebe's (1977) reading of *Heart of Darkness* fundamentally altered how the novella could be taught; Said (1993) read the Western novel against the imperial histories it presupposes; and postcolonial criticism made *The Tempest* and *A Passage to India* standard sites for examining colonial discourse (Loomba, 2015). That postcolonial readings of these texts are now mainstream in the Western academy is itself significant for interpreting what models produce, since it raises the possibility that such readings may be reproduced without necessarily reflecting a situated South Asian perspective.

In Bangladesh specifically, Al-Quaderi and Al Mahmud (2010) argued that English-medium schooling leaves students more familiar with English literature than with their own literary traditions, and Hamid (2021, 2023) contended that decolonising the curriculum requires interrogating English itself and reconceiving it as a Southern language. Bain (2026) examined introductory poetry syllabi in Bangladeshi English departments and found that poets' national identity, race and gender influence curricular inclusion. Chambers and Kamal (2026) demonstrated that the anthologised canon of literary theory continues to marginalise Global South voices. These studies indicate that Euro-American frameworks already dominate the human curriculum in Bangladesh; any assessment of AI

must therefore be made against that institutional reality rather than against an idealised alternative.

South Asian Interpretive Traditions

Any claim that AI privileges Western interpretive frameworks presupposes the existence of substantive alternatives, and South Asia has several. The Sanskrit tradition developed the theory of *rasa* over more than a millennium, constituting one of the world's most sophisticated accounts of aesthetic experience (Pollock, 2016). Bengal has its own literary-critical history, shaped by Vaishnava devotional poetics and by a modern vernacular criticism (Kaviraj, 2003). The Persian and Urdu tradition, central to Muslim literary culture in the region, developed a distinctive poetics in the *sabk-e hindi*, or Indian style (Faruqi, 2004). Devy (1992) argued that colonial education induced a cultural amnesia in which Indian criticism came to treat Western theory as the only legitimate theory available. These traditions must be named separately. To allow Sanskrit poetics to stand for South Asia as a whole would reproduce, within this study, the India-centred reduction that it examines in the models.

Generative AI and Literary Interpretation

Research on generative AI in literature education is recent and concentrated primarily on questions of effectiveness and reliability (Frederking, 2025; Kasneci et al., 2023). The closest precedent for the present study is Soraya et al. (2026), who compared AI interpretations of folktales with those of undergraduate readers and found the AI readings coherent but culturally generalised—a finding that lends preliminary empirical support to the concerns motivating this protocol. Fish (1980) provides an important theoretical reason why reliability cannot be the only criterion: meaning is produced by interpretive communities whose assumptions determine what counts as a valid reading. An LLM can be understood as a statistical condensation of the interpretive communities best represented in its training data. The critical question is which communities those are, and by what discursive operations their assumptions are naturalised as universal.

Three bodies of scholarship thus converge without fully connecting. Audits of cultural bias in LLMs seldom address interpretive practice. Studies of Orientalism in AI outputs address political and regional representation, primarily of the Middle East. Research on AI and literary reading has begun to compare human and machine responses but has not subjected canonical English texts to CDA informed by postcolonial theory. No study located in searches conducted for this article in November 2026 analyses AI interpretations of canonical English literature from a Bangladeshi standpoint. That narrower gap is the one addressed here.

Conceptual and Theoretical Framework

The framework has three parts. Said supplies an account of representation, Spivak an account of voice, and Fairclough the analytic procedure. Chakrabarty and Mignolo are used for one specific purpose, the distinction between universal and located knowledge.

Orientalism as a discourse of representation

Said (1978) claimed that Orientalism was a disciplined collection of statements which resulted in the construction of the knowledge of the Orient by the European culture. Here he uses three features of his account. The first is exteriority: The Orient is talked about and talked about. The other is essentialism, which appears as a deconstruction of the

heterogeneous societies into their fixed attributes, like spirituality or timelessness. The third is the Western superiority as knower and norm, which is superior.The third is higher rank of the West: its superiority as knower and norm. This stereotype is a fixing of difference, to be repeated, for which there is no proof, Bhabha (1994) went on. A stereotype doesn't have to be negative. While a disparaging description of 'Eastern spirituality' can stifle the imagination, so can an appreciative description.

Two cautions govern the application of Said's framework here. His analysis focused primarily on the Arab and Islamic world and cannot be transposed to South Asia without critical mediation; it must be discussed and not simply assumed. Ahmad (1992) further cautioned against treating the West as a unified, continuous entity, and the same caution applies to analyses that posit a stable binary between Western and Eastern critical frameworks. The research questions retain this binary because it reflects the terms in which the debate is commonly conducted, but the analytic classification adopted in the methodology is more granular, and the binary itself is treated as an object of examination rather than as a stable analytic given.

Voice, representation and epistemic hierarchy

Spivak (1988) shifted attention from how the Other is represented to whether the Other can be heard. She distinguished two senses of representation: speaking for (*Vertretung*) and portraying (*Darstellung*). Both are in play when a model answers a reader who has identified as a student in Bangladesh, since the response may portray what such a reader thinks and may also speak in that reader's place. Her term *epistemic violence* names the disqualification of a group's ways of knowing by a dominant system of knowledge.

Chakrabarty (2000) argued that European thought is both indispensable and inadequate for understanding non-European experience, and that its claim to universality has a local history. Mignolo (2011) made a related argument from decolonial theory: Western modernity presents its own epistemology as universal and assigns other knowledges the status of culture or tradition. Postcolonial and decolonial theory differ in genealogy and are sometimes at odds, and no synthesis is attempted here. The study takes from both one shared and observable distinction, between knowledge treated as *theory* and knowledge treated as *culture*. The sociolinguistic concept of markedness, as applied to LLMs by Cheng et al. (2023), gives that distinction a textual form: the unmarked position is the one that needs no label.

Critical Discourse Analysis and machine-generated text

CDA treats language use as social practice that reflects and reproduces relations of power (Wodak & Meyer, 2016). Fairclough (1995) proposed that a discursive event be analysed in three dimensions. As *text*, it is described through vocabulary, grammar, modality and cohesion. As *discursive practice*, it is interpreted through its production and consumption, including the other texts and discourses on which it draws. As *social practice*, it is explained by reference to ideology and hegemony. Fairclough (2003) focused on a number of features of importance in this context: legitimation by appeal to authority, assumptions that are taken for granted in a text, and modality. CDA was created for texts written by humans, and should be used with machine output with justification. An LLM does not have an intention nor an ideology. Bender et al. (2021) describe its output as text that is stitched together from sequences of linguistic form in the training data, without reference to meaning. However, there are three factors that support the method. First, CDA is about the text and its social conditions, not about the author's mind; Fairclough's analysis isn't based on the recovery of intention. Second, the features of the models are textually identifiable and have a clearly traceable intertextual resources, and students read them as

authoritative speech. Third, training can be done on a skewed archive, tuning by a corporation and prompting by a user can be specified sociotechnically. The analysis consequently does not assume any motivation for the models. It is a description of the action of the discourse. Because the production process is observed only through its outputs, claims at the level of social practice are stated as interpretations.

Algorithmic Orientalism as an analytic concept

Building on earlier uses of the term (Dahman & Saou, 2025) and on the related notion of computational Orientalism (Shahid, 2026), this article defines algorithmic Orientalism for the domain of interpretation. It is the reproduction, in machine-generated readings of literature, of Orientalist relations of knowledge, in which Western interpretive frameworks function as the unmarked universal and non-Western readers and traditions appear as marked, essentialised or supplementary. Table 1 sets out four indicators.

Table 1
Indicators of algorithmic Orientalism in AI literary interpretation

Indicator	What it looks like in a response	Source
Othering	The South Asian reader or culture is contrasted with an implied normal reader, beyond acknowledging the location the user stated	Said (1978)
Cultural stereotyping	Beliefs, feelings or experiences are attributed to South Asian readers as a group, without grounding in the text under discussion	Said (1978); Bhabha (1994)
Epistemic hierarchy	Euro-American approaches are applied as theory; South Asian traditions are absent, or named without being applied	Spivak (1988); Mignolo (2011)
Universalism	Themes are presented as timeless and human, and the reading is not located	Chakrabarty (2000); Said (1993)

Expectations and decision rules

Stating in advance what would count for and against the hypothesis guards against a framework that can absorb any result. The literature supports two competing expectations. Studies of default cultural alignment (Tao et al., 2024) and of identity prompting (Cheng et al., 2023; Naous et al., 2024; Wang et al., 2025) suggest that responses to the neutral prompt will be unlocated and Euro-American in their frameworks, and that positioned responses will carry some stereotype. Against this, postcolonial readings of several selected texts are mainstream in Western criticism (Loomba, 2015), so the models may well reproduce them unprompted.

Three rules govern the interpretation of results. First, an indicator is treated as *consistent* only if it occurs in at least two-thirds of the relevant responses and in all three models, as *partial* if it occurs in between one-third and two-thirds or in only one or two models, and as *weak or absent* otherwise. Second, the hypothesis is treated as supported only if epistemic hierarchy and universalism are consistent in the neutral condition, or Othering and stereotyping are consistent in the positioned condition. If none of the four indicators is consistent, the hypothesis is rejected for this corpus. Third, some patterns are not counted as evidence for the hypothesis. The use of metropolitan postcolonial theory (Said, Bhabha, Spivak) is recorded as a separate category and reported descriptively. A positioned response that differs little from the neutral one is not counted as hierarchy when the neutral response is itself located and plural. An acknowledgement of the reader's stated location is not counted as Othering.

Counter-evidence is defined just as explicitly: the application of a South Asian framework to a specific feature of a text, differentiation among South Asian positions, the labelling of a Western reading as Western, and acknowledgement of the limits of the response's own standpoint.

Material and Methods

Research design

This section sets out the protocol in full. It is written in the present tense and describes procedures to be followed; no data have yet been generated. The study is qualitative and interpretive. It applies Fairclough's (1995) three-dimensional CDA to a purposively constructed corpus of AI-generated literary interpretations, read through the theoretical framework developed above. Frequency counts are reported only to indicate the distribution of patterns across the corpus; they carry no inferential weight. The study involves no human participants.

Selection of literary texts

Six texts are selected purposively. Four are by British authors: *The Tempest* (Shakespeare), 'Ode to a Nightingale' (Keats), *Heart of Darkness* (Conrad) and *A Passage to India* (Forster). Two are by South Asian authors: *Gitanjali* (Tagore, in his own English version of 1912) and *The God of Small Things* (Roy). 'Canonical' is used in the curricular sense, for texts with an established place in English syllabi, which now includes anglophone South Asian writing. All six are staples of the anglophone curriculum and of postcolonial criticism. Their place in Bangladeshi undergraduate teaching is to be confirmed against current university syllabi at the point of data collection, and any text not found there is replaced by a comparable one.

The choice allows two comparisons. The other is between texts that have colonial content, and a lyric that do not, to see if South Asian perspective is only brought into the text when it calls for it. The other is between written texts by the British and written texts by the South Asians. There are two limits recognised. This is the first comparison that is based on only one poem. It is absent any material from a Bangladeshi writer, since the study's subject is the canon, which is taught as it is.

AI models and access

They use Three Generative AI Systems: ChatGPT (OpenAI), Claude (Anthropic) and Gemini (Google). They are selected as they are general purpose systems readily available. The decision is based on availability and there is no data on the extent of the use of each system by Bangladesh students. The three are made by U.S.-based companies. A system developed elsewhere is not included in the study, thus it is not possible to say anything about the pattern, based on the origin of the models. The systems are approached via their public chat interfaces, where students meet them. There's a price to this: such interfaces are given hidden commands to the system and can change to fit past uses or location. To minimize the impact of these influences, each prompt was typecast in a new conversation, from a new user or logged out user, with memory disabled, custom instructions disabled, and web search disabled in the interface (when enabled). Version of the model, access tier, access interface, country of access and dates for model generation are documented for each system and reported.

Prompt design and data generation

Two prompts for each text are given to each model. Both signify student and the only difference between them is the reader's position. The neutral prompt is: 'I am an undergraduate student of English literature' (with the title and author of each text changed accordingly): What does the play, The Tempest, written by William Shakespeare, mean? Talk about the central idea(s) of the text. The wordings of the following poems are the same: 'Ode to a Nightingale' by John Keats, Heart of Darkness by Joseph Conrad, A Passage to India by E. M. Forster, Gitanjali by Rabindranath Tagore and The God of Small Things by Arundhati Roy. The South Asian prompt is merely adding to the reader's location, which is 'I am an undergraduate student of English literature at a university in Bangladesh'. What does the William Shakespeare play The Tempest mean? Talk about the central idea(s) of the text. This is repeated with the other 5 texts. There are three design decisions that address identified risks. Using the phrase 'how can … be interpreted' does not assume one interpretation. The positioned prompt does not designate a perspective, and therefore a framing, is not a stipulation to be performed; rather it is an inference from the model and not a demand of the prompt. The prompt uses the term Bangladesh and not 'South Asia' so that it can be seen if a response refers to Bangladesh or includes it as a part of a larger area.

The first complete response to each prompt is retained without regeneration. This produces the primary corpus of 36 responses (3 models × 6 texts × 2 prompts), which is the object of the full CDA. Because LLM output varies between runs, each of the 36 conditions is generated twice more to form a stability corpus of 72 responses. The stability corpus is not analysed in full. It is screened only for the presence or absence of each code family, to check whether a pattern in the primary corpus recurs. Every response is saved verbatim with a code, date, time and model version.

Analytic procedure

Analysis follows Fairclough's three dimensions. At the level of text, each response is examined for evaluative vocabulary, pronominal choices and reader address, modality and hedging, the naming and application of critics and theoretical frameworks, and the order and proportional allocation of content. At the level of discursive practice, the analysis identifies the critical frameworks each response draws upon and classifies them. At the level of social practice, patterns across the corpus are interpreted through the theoretical framework, and their implications for literary education in South Asia are considered.

Critical frameworks are classified across four categories rather than a Western–Eastern binary: Euro-American (formalist, psychoanalytic, Marxist, Romantic-aesthetic, and related traditions); metropolitan postcolonial (the postcolonial theory canonised in Anglo-American universities, including the work of Said, Bhabha and Spivak); South Asian (Sanskritic, Bengali, and Perso-Arabic or Islamic critical traditions, and criticism produced in the region); and other. A framework is coded as applied when at least one of its concepts is used to make a substantive claim about a specific feature of the text under discussion; it is coded as mentioned when it is named or glossed without being deployed analytically.

The codebook is otherwise deductive, comprising the four indicators in Table 1 and the counter-discursive features set out in the expectations and decision rules above. Each code has a written definition, inclusion and exclusion criteria, and an anchoring example drawn from the literature. Codes are added inductively where patterns emerge that the codebook does not anticipate; each such addition is logged with its rationale. Coding is applied to meaning units of one or more sentences. Comparison proceeds along three axes: neutral versus positioned prompt within each model and text; model versus model; and British-authored versus South Asian-authored texts.

There is no idealised human reader as a reference point for the lack of South Asian frameworks. Previous research has proved that English teaching in Bangladesh is Euro-American based (Bain, 2026; Hamid, 2021). The study thus does not make the assumption that a human teacher or guidebook would provide anything that the models might lack; the significance of what is absent is that it is so large and consistent as to be something new, rather than a novelty.

Trustworthiness

The criteria by Lincoln and Guba (1985) are used to deal with trustworthiness. Dependability and confirmability is supported through an audit trail: all prompts and responses are recorded as they are and an audit trail of all codebook revisions is maintained in a dated log. The credibility is enhanced by a second coder, who has no involvement in the creation of the codebook and codes at least half of the primary corpus (18 responses, distributed evenly across models, texts and prompts). In keeping with the recommendations of O'Connor and Joffe (2020), the presence of each code family is reported as percent agreement and Cohen's kappa prior to discussion. Disagreements are then addressed through discussion and the codebook is developed to be more refined. It portrays agreement figures and the kind of disagreements. The second coder's role is to test whether the code definitions can be applied consistently; the interpretation at the level of social practice remains the author's. Negative cases are sought and reported. Transferability is supported by full description of the protocol.

Researcher positionality

The interpretation of data is inevitably influenced by the location, training and institutional context of the analyst in relation to English studies in the South Asian context, which will be explicitly stated when reporting the study found in the data. Anyone who reads the corpus with the idea of looking for Orientalism is in danger of encountering it. Those safeguards against that risk are the decision rules established prior to data collection, the second coder, or the stability corpus and the systematic search for the negative cases. They don't remove it, they take it out of the shadows and put it under control.

Ethical considerations and the role of AI in the analysis

No human participants are involved, so the study is expected to fall outside formal ethical review. This is to be confirmed with the analyst's institution before data collection. The AI systems are used within their providers' terms of service, and their responses are treated as data and not cited as scholarship. Coding and interpretation are carried out by human coders. They are not delegated to any generative AI system, since the systems under study cannot be the instrument that assesses them.

Limitations

The primary corpus retains only one response per condition, and the stability corpus provides a coarse rather than systematic check on output variability. Findings should be reported with reference to specific dates and model versions, and should not be generalised to the models as such. Comparative analysis by model origin is unavailable because all three systems are developed in the United States. All prompts are in English, which is the medium of instruction in English departments in Bangladesh; prompts in Bangla or other South Asian languages may engage different discursive registers and are reserved for future work. The public interfaces through which the models are accessed are not fully transparent with respect to system-level instructions. The study addresses only

outputs and makes no claims about how students read, trust or resist the interpretations they receive from AI systems.

Results and Discussion

There are no empirical data produced, the study outlined in the previous section is a protocol to be carried out. The results of this section are therefore reported as 'no findings'. It aims to promote four theorically and empirically supported statements (one for each research question) to be tested by the protocol. The section outlines the textual evidence which would support the proposition, the counter-evidence that would disprove it and the overall analytical implications of supporting or disproving it. The propositions are not findings, but are drawn from literature and the theoretical framework, and may turn out to be false. Combined, they form the context for the data that is to come and how it will be analyzed.

Proposition 1: The Unlocated Default

The first proposition holds that responses to the neutral prompt will present their readings as unlocated: themes will be described as timeless or universal, claims will be advanced with categorical modality, and the reader addressed will be assigned no particular cultural, historical or geographical position. This proposition rests on well-established evidence that the default cultural orientation of leading LLMs clusters with Western, specifically English-speaking Protestant European, societies (Tao et al., 2024), and on the concept of the unmarked default as theorised by Cheng et al. (2023). In Faircloughian terms, categorical modality and unargued presuppositions are textual markers of the common ground a discourse assumes with its reader; universalism, as defined in Table 1, requires both that themes are presented as timeless and human and that the reading is not attributed to any particular perspective or tradition.

The proposition would be disconfirmed if neutral responses routinely label their readings as perspectives, employ substantial hedging, or present a range of located readings in acknowledged relation to one another. This outcome is genuinely plausible, particularly for texts such as *Heart of Darkness*, *A Passage to India* and *The Tempest*, whose postcolonial dimensions are now mainstream in Western criticism (Loomba, 2015); a neutral response that offers a postcolonial reading and names it as such would not meet the definition of universalism. "Ode to a Nightingale" provides the more critical test, since nothing in its subject matter invites a culturally located reading. If neutral responses to the Keats poem are as explicitly perspectival as responses to the Conrad, the proposition fails for the corpus as a whole.

Either would have significant analytical ramifications. The universalist interpretation of literary meaning in terms of AI would, if true, simply be a new iteration at an unprecedented magnitude and accessibility of the same epistemic stance found in the Western critical tradition for a long time now and as associated with postcolonial critical practice. A student who uses AI to guide their interpretation of literary works would be given readings that embody a certain cultural point of view as if it were a given, without the institutional structures (footnotes, framing, engaging with the text in a classroom) that make it more visible that it is a human point of view. The finding would provide a valuable qualification and counterweight to the broader critiques of cultural bias and postcolonial training of AI, and hint that a reflexivity or awareness of the "locatedness of interpretation" has emerged in the models.

Proposition 2: Theory and Culture

The second proposition holds that Euro-American critical frameworks will be applied analytically to generate readings of all six texts, while South Asian interpretive traditions will be absent from the responses or mentioned without being substantively applied—and that this asymmetry will persist even in responses to *Gitanjali* and *The God of Small Things*. The operational test is the applied–mentioned distinction established in the analytic procedure. A response that organises its reading of the Keats ode around Romantic aesthetic theory and adds that a South Asian reader "might also think of *rasa*" has named a tradition without deploying it; this is precisely the structure that Mignolo (2011) describes as the assignment of non-Western knowledge to the status of culture rather than theory, and that Devy (1992) characterised as the colonial amnesia induced by English literary education. Spivak's (1988) concept of epistemic violence is operative here: the disqualification of South Asian frameworks as legitimate analytical tools, even when the texts being interpreted emerged from South Asian contexts, enacts the asymmetric relation of knowledge that the hypothesis anticipates.

The proposition would be disconfirmed if the responses show substantive application of the South Asian frameworks — such as if a response analyses the emotional flow of "Ode to a Nightingale" by operations of a particular rasa, or if a response reads and interprets Gitanjali with the devotional frameworks of Bengali. All of these would be positive counter-evidence, and would need to be recorded and reported even if there was no general trend.

There are two qualifications of interest. The writings of Said, Bhabha and Spivak are reported and classified as part of the so called "metropolitan postcolonial theory." Although it does not refer to a colonial reading in the traditional sense, its presence is indicative that a response to the colonial reading is not just a reproduction of a colonial reading (in the traditional sense). Chambers and Kamal (2026) have indicated that the canon of literary theory will continue to accept the Global South thought in the form(s) it takes after it is appropriated by metropolitan institutions. In this, a South Asian voice would have been needed only by Said, Bhabha and Spivak and a corpus in which they would be seen only from this point of view would be in keeping with the observation; a corpus that is counted either way would be described under the decision rules. The second qualification relates to the baseline of the institutions. Should this proposition be true, the models would be reproducing what the English curricula in Bangladesh are already known to do (Bain, 2026; Hamid, 2021). So the correct ending would be that AI doesn't create a new institutional hierarchy per se, it just does at scale and without any of the partial correctives that an individual teacher might include.

Proposition 3: Positioning and Stereotype

The third proposition holds that when the reader is located in Bangladesh, responses will change in ways that attribute a fixed and limited set of concerns to that reader—concerns such as colonial grievance, religious feeling, or cultural difference from a Western norm—and that Bangladesh will frequently be subsumed within an undifferentiated South Asian or Indian category, effacing its historically specific identity. Because the positioned prompt specifies only a geographical location, any additional content that a response introduces represents the model's own inference about what a reader in that location needs or wants; it is in this inferential mechanism that stereotyping or Othering, as defined in Table 1, would manifest.

Three distinct outcomes can be distinguished, each with different analytical implications. In the first, the positioned response is substantively unchanged from the

neutral one. Under the decision rules, this counts as evidence of hierarchy only if the neutral response is itself unlocated and universalist; if the neutral response is already reflexive about its standpoint, an unchanged positioned response may simply indicate that the model does not presume to know what a Bangladeshi reader thinks. In the second, the response introduces attributions—colonial grievance, religious consciousness, a presumed unfamiliarity with Western literary conventions—that follow from the stated location rather than from the text under discussion. This constitutes essentialism in Said's (1978) sense: the fixing of a group within predetermined characteristics. This outcome would align with findings on identity prompting (Cheng et al., 2023; Naous et al., 2024; Wang et al., 2025) and would exemplify what Spivak (1988) describes as speaking for the subaltern rather than creating the conditions for the subaltern to speak. The documented tendency of image models to view South Asian subjects through an outsider's gaze (Qadri et al., 2023) would here find its textual-interpretive counterpart. In the third outcome—which would refute the proposition—the response offers a differentiated, historically specific reading attentive to the particularity of Bangladesh: its linguistic nationalism, its Muslim-majority literary culture, its memory of 1947, 1952 and 1971, and the distinctive tradition of Bengali literary criticism, without reducing the reader to those contexts or denying their engagement with the universalising aspirations of the canonical texts themselves.

Proposition 4: Variation Across Models and Texts

The fourth proposition holds that the patterns described in Propositions 1 to 3 will be broadly consistent across all three AI models, and will be at least as pronounced for the South Asian-authored texts—*Gitanjali* and *The God of Small Things*—as for the British-authored ones. This proposition is designed to test a structural explanation against a design-specific one. If the patterns are consistent across ChatGPT, Claude and Gemini, this consistency would be compatible with an explanation grounded in the shared conditions of model production—training on similarly skewed corpora, fine-tuning practices oriented to similar evaluative standards, and deployment within a comparable commercial context—of the kind theorised by Kwet (2019) and Ricaurte (2019) under the rubric of digital colonialism. If patterns vary substantially by model, this would instead point towards design decisions by individual developers and would suggest a problem more tractable through targeted intervention.

The comparison across texts is to another question of a similar nature. The lack of South Asian frameworks of interpretation from even responses to Tagore and Roy, texts that have been created in the traditions that are the subject of such frameworks, does not seem to be enough to explain the epistemic hierarchy. It would rather signify the condition of structurality of literary interpretation in AI as a marginalisation of South Asian modes of reading which are not a result of texts chosen, but a result of the conditions under which the models were created. On the other hand, the more the South Asian authored texts produce South Asian frameworks than the British authored texts, the more it appears that the South Asian models are contextually sensitive and have to be interpreted with some care, as it would bring the structural thesis into question.

What the Protocol Can and Cannot Establish

The three levels of claim should be kept analytically separate in reporting the study. Empirical finding is the pattern of features of the discourse that occur in the primary corpus and are reoccurring in the stability corpus. The theoretical interpretation is that such patterns (if they exist) reproduce orientalist relations of knowledge, a claim which relies on the theoretical framework of the previous section and is thus put forward as an interpretation and not as a description. The key argument is of a third (evaluative) order.

One such argument becomes available when data reveal a systematic discrepancy between neutral and positioned conditions: a system that provides a culturally located reading only for readers who signal their own location makes the burden of visibility (of naming themselves 'Other') fall on the reader at the margins. In other words it would be a discovery that has implications directly for pedagogy and policy, apart from the more general theoretical assertions about Orientalism.

The protocol cannot establish why models respond as they do, cannot trace any pattern to specific features of training data or fine-tuning, and cannot assess how stable the responses are beyond the coarse check the stability corpus provides. It cannot speak to what students make of the interpretations they receive, or to how their reading, trust and resistance interact with the AI output. What it can determine is whether, at a given date, the readings that three widely available AI systems return to a student of English literature differ with that student's stated location, and how those differences are structured discursively. That is a narrower claim than many critics of AI cultural bias have made, but it is one that can be made precisely and that others can verify.

Implications for Literary Education and AI-Mediated Knowledge Production

Either outcome of the protocol carries significant implications for English studies in South Asia. If the propositions hold, the decolonising work described by Hamid (2021) and Al-Quaderi and Al Mahmud (2010) acquires a new and urgent object: a privately consultable interpreter that carries the existing hierarchy of critical frameworks into every answer, available to students at any hour and without the visible institutional framing that at least partially contextualises a human teacher's perspective. The scale of this effect may be substantial; Kasneci et al. (2023) note that AI uptake in education has outpaced the research infrastructure needed to evaluate it. If the propositions fail, this is equally significant, and in some respects more instructive: it would constitute evidence against a homogenising account of AI cultural influence that is often assumed rather than tested in the humanities, and would suggest that training on a corpus that includes substantial postcolonial scholarship has produced a partial reflexivity about interpretive standpoint that the critical literature has not anticipated.

In both cases, the protocol operationalises what has hitherto been a theoretical suspicion. It turns a general concern about the cultural politics of AI into a specific, empirically testable question that any department of English in South Asia can examine with its own texts and its own students. It also points towards a pedagogical response: the standpoint of an AI interpretation, like that of any critic, is something that students can be taught to identify and interrogate—which critics it cites, which traditions it applies or omits, which reader it assumes, and what it takes for granted as universal. Critical AI literacy, in this sense, is not a technical skill but a humanistic one, and it may be precisely the kind of literacy that literary education is best equipped to develop.

Conclusion

This article asks whether generative AI, when it interprets canonical English literature, reproduces the relations of knowledge that Said called Orientalism. It does not answer that question empirically. It sets out what an answer requires.

Three things are offered. The first is a definition. Algorithmic Orientalism, a term already used for anti-Muslim bias in AI, is carried into literary interpretation and specified through four indicators: Othering, cultural stereotyping, epistemic hierarchy and universalism. The second is a set of safeguards against a framework that confirms itself. These are decision rules fixed in advance, explicit counter-evidence, a classification of

critical frameworks that goes beyond a Western–Eastern binary, and a baseline in documented curricular practice. The third is a protocol. Three models answer two prompts on six texts, with only the reader's location varied, and the 36 responses are analysed through Fairclough's three dimensions and checked against repeated generations and a second coder.

The limits are stated in the methodology. The protocol is small, confined to English and to three US-developed systems, and silent on reception. Carrying it out is the first task for further research. Extending it is the second: with many samples per condition, with systems developed outside the United States, with prompts in Bangla and other South Asian languages, and with comparison to the readings of students, teachers and published study guides in the region.

English literary study in South Asia began as a project in which one culture's readings were installed as the standard for another (Viswanathan, 1989). Whether generative AI continues that project or loosens it is an empirical question, and it can now be put precisely.

Recommendations

These recommendations follow from the theoretical framework and the scholarship reviewed above. They do not rest on empirical findings from the protocol, and should be revisited and revised once the protocol has been carried out and its results are available.

Departments of English in South Asia should incorporate critical AI literacy into their curricula, equipping students to analyse an AI interpretation as a discourse with a standpoint rather than as a neutral guide to meaning. This involves teaching students to identify which critics and theoretical traditions a given response names and applies, which interpretive frameworks it treats as universal, and which reader it implicitly constructs through its pronominal choices, its modality, and its presuppositions. The paired-prompt comparison central to this protocol—submitting both a neutral and a South Asian-positioned prompt to the same system for the same text—can serve as a productive classroom exercise, enabling students to observe directly whether and how their stated location changes the interpretation they receive, and to reflect critically on what those changes reveal about the system's assumptions.

Beyond classroom practice, departments should work towards giving South Asian interpretive traditions—Sanskritic, Bengali and Perso-Arabic—substantive analytical status in the curriculum. Students who have applied *rasa* theory, Bengali devotional poetics, or Islamic aesthetic criticism to specific literary texts are better positioned to notice when such frameworks are merely named in an AI response rather than used, and to understand what is at stake in that asymmetry. This is not a call for replacing existing theoretical frameworks but for expanding the repertoire of analytical tools that students regard as legitimate and usable, which is itself a condition of the critical literacy that the emergence of AI literary interpretation makes newly urgent.

AI developers should include literary and humanities interpretation in cultural standpoint evaluations, working with scholars from the regions concerned and attending to the applied–mentioned asymmetry between Western and non-Western critical frameworks. Existing evaluations concentrate on values inventories and social stereotypes and do not address the domain of literary meaning-making, where the stakes for educational equity are high and the patterns of cultural alignment are likely to be both systematic and consequential. Institutions and policymakers responsible for educational

guidance on AI should ensure that such guidance addresses interpretation in the humanities alongside academic integrity, recognising that the most significant questions raised by AI in literary education are not about plagiarism but about whose ways of reading are reproduced at scale.

Researchers should replicate the protocol reported here and publish the findings in full, including negative results and partial disconfirmations of the propositions. The value of a pre-specified protocol lies partly in the commitment it creates to transparent reporting of whatever the data show, and the field requires studies that test the hypothesis against evidence rather than assuming it. Replication with repeated sampling per condition, with AI systems developed outside the United States, with prompts submitted in Bangla and other South Asian languages, and with reception studies that follow how students and teachers actually read and use AI interpretations would substantially extend the evidential basis from which the broader implications of this work could be drawn.

Declarations

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Ethical Considerations

No human subjects or personal information was used in this study. It is based exclusively on published books that are accessible to the public (and publisher descriptions and epitexts) cited in the references. Ethical approval was not necessary.

AI Use Statement

Generative AI (ChatGpt and Grammarly) was used for language editing and condensing text. It was not used for the data collection, coding the corpus and drawing findings. All text generated by AI is reviewed and edited by the author, and the author is responsible for the article's content.

Data Availability Statement

The data analysed are all published data, as mentioned in the references. The 39 corpus books are listed verifying, with the verification sources. The coding frame and coded dataset are available on reasonable request from the author.

References

Achebe, C. (1977). An image of Africa. *The Massachusetts Review, 18*(4), 782–794.

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews, 46*(1–2), 176–197. https://doi.org/10.1080/03080188.2020.1840225

Agarwal, D., Naaman, M., &Vashistha, A. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3713564

Ahmad, A. (1992). *In theory: Classes, nations, literatures*. Verso.

Al-Quaderi, G. G., & Al Mahmud, A. (2010). English literature in English medium schools in Bangladesh: The question of post-colonial pedagogy. *Asiatic: IIUM Journal of English Language and Literature, 4*(2), 121–154. https://doi.org/10.31436/asiatic.v4i2.526

Bain, K. (2026). Approach to poetry in curriculum: An exploratory study on the introductory poetry courses of different English departments in Bangladesh. *Bangladesh Journal of Educational Research, 12*(1), 32–38. https://doi.org/10.3329/bjer.v12i1.87254

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bhabha, H. K. (1994). *The location of culture*. Routledge.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), Article 100205. https://doi.org/10.1016/j.patter.2021.100205

Chakrabarty, D. (2000). *Provincializing Europe: Postcolonial thought and historical difference*. Princeton University Press.

Chambers, C., & Kamal, S. (2026). "Chip, chip, chip away": Revisiting the *Norton Anthology*'s canon essays in the MeToo era. *Bandung: Journal of the Global South*. Advance online publication. https://doi.org/10.1163/21983534-20262007

Cheng, M., Durmus, E., &Jurafsky, D. (2023). Marked personas: Using natural language prompts to measure stereotypes in language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 1504–1532). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.84

Dahman, A., &Saou, H. (2025). Algorithmic Orientalism: A systems analysis of anti-Muslim bias in artificial intelligence. *Journal of Islamic Sciences and Civilization, 10*(2). https://doi.org/10.55781/rsic.v10i2.541

Devy, G. N. (1992). *After amnesia: Tradition and change in Indian literary criticism*. Orient Longman.

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge.

Faruqi, S. R. (2004). A stranger in the city: The poetics of *sabk-e hindi*. *The Annual of Urdu Studies, 19*, 1–93.

Fish, S. (1980). *Is there a text in this class? The authority of interpretive communities*. Harvard University Press.

Frederking, V. (2025). Reliability of AI in the domain of literary literacy and literary text interpretation: An empirical study. *RISTAL: Research in Subject-matter Teaching and Learning, 8*(2), 70–87. https://doi.org/10.2478/ristal-2025-0007

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics, 50*(3), 1097–1179. https://doi.org/10.1162/coli_a_00524

Hamid, M. O. (2021). Interrogating the English of the English curriculum in postcolonial Bangladesh. *Asiatic: IIUM Journal of English Language and Literature, 15*(2), 11–29. https://doi.org/10.31436/asiatic.v15i2.2341

Hamid, M. O. (2023). English as a Southern language. *Language in Society, 52*(3), 409–432. https://doi.org/10.1017/S0047404522000069

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., …Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kaviraj, S. (2003). The two histories of literary culture in Bengal. In S. Pollock (Ed.), *Literary cultures in history: Reconstructions from South Asia* (pp. 503–566). University of California Press.

Khandelwal, K., Tonneau, M., Bean, A. M., Kirk, H. R., & Hale, S. A. (2024). Indian-BhED: A dataset for measuring India-centric biases in large language models. In *Proceedings of the 2024 International Conference on Information Technology for Social Good*. Association for Computing Machinery. https://doi.org/10.1145/3677525.3678666

King, G., Abbas, D., & Jegić, D. (2026). The political economy of AI: a content analysis of ChatGPT outputs and anti-Palestinian media bias. *Third World Quarterly*, *47*(8), 1501–1518. https://doi.org/10.1080/01436597.2026.2614728

Kwet, M. (2019). Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class, 60*(4), 3–26. https://doi.org/10.1177/0306396818823172

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Loomba, A. (2015). *Colonialism/postcolonialism* (3rd ed.). Routledge.

McKenna, S., &Tshuma, N. (2026). *Insidious by design: Implications of large language model algorithmic bias for the Global South*. https://arxiv.org/pdf/2606.28333

Mignolo, W. D. (2011). *The darker side of Western modernity: Global futures, decolonial options*. Duke University Press.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024). Having beer after prayer? Measuring cultural bias in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 16366–16393). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.862

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

O'Connor, C., &Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1–13. https://doi.org/10.1177/1609406919899220

Pollock, S. (Ed. & Trans.). (2016). *A rasa reader: Classical Indian aesthetics*. Columbia University Press.

Qadri, R., Shelby, R., Bennett, C. L., & Denton, E. (2023). AI's regimes of representation: A community-centered study of text-to-image models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 506–517). Association for Computing Machinery. https://doi.org/10.1145/3593013.3594016

Ricaurte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media, 20*(4), 350–365. https://doi.org/10.1177/1527476419831640

Roh, D. S., Huang, B., &Niu, G. A. (Eds.). (2015). *Techno-Orientalism: Imagining Asia in speculative fiction, history, and media*. Rutgers University Press.

Said, E. W. (1978). *Orientalism*. Pantheon Books.

Said, E. W. (1993). *Culture and imperialism*. Alfred A. Knopf.

Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., &Prabhakaran, V. (2021). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 315–328). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445896

Shahid, M. (2026). *Computational Orientalism: Measuring structural discourse bias in large language models using the Middle East Cultural Sensitivity Score (MECSS).* https://arxiv.org/pdf/2608.18100

Soraya, A. I., Ramadhani, R., & Bohang, M. B. A. (2026). When algorithms read folklore: AI, simulated reception, and cross-cultural studies of literary interpretation. *Utamax: Journal of Ultimate Research and Trends in Education, 8*(2), 162–181. https://www.researchgate.net/publication/415819329

Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). University of Illinois Press.

Tao, Y., Viberg, O., Baker, R. S., &Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), Article pgae346. https://doi.org/10.1093/pnasnexus/pgae346

Viswanathan, G. (1989). *Masks of conquest: Literary study and British rule in India*. Columbia University Press.

Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence, 7*(3), 400–411. https://doi.org/10.1038/s42256-025-00986-z

Wodak, R., & Meyer, M. (Eds.). (2016). *Methods of critical discourse studies* (3rd ed.). SAGE.

Zembylas, M. (2023). A decolonial approach to AI in higher education teaching and learning: Strategies for undoing the ethics of digital neocolonialism. *Learning, Media and Technology, 48*(1), 25–37. https://doi.org/10.1080/17439884.2021.2010094