INTERNATIONAL JOURNAL OF
Practical and Pedagogical Issues in English Education
ISSN : 2980-9533 Homepage: www.ijpie.org 3(4), Dec. 2025

Finding Symbolism in R.F. Kuang's Yellowface (2023) by ChatGPT: A Qualitative Hermeneutic Inquiry

Shadi Forutanian [1], Behzad Pourgharib [1, *]

[1] Department of English Language and Literature, University of Mazandaran, Babolsar, Iran

Corresponding Author's Email: b.pourgharib@umz.ac.ir

KEYWORDS

Artificial Intelligence, Literary Symbolism, Qualitative Analysis, Cultural Appropriation, Digital Hermeneutics

ARTICLE INFO

Article type: Research Article

Article history:
Received: 06 September 2025
Revised: 27 November 2025
Accepted: 02 December, 2025
Published online: 02 December 2025

ABSTRACT

The current study conducts a critical analysis of the interpretative abilities of large language models simply known as LLMs, specifically OpenAI's ChatGPT-4, in discerning and scrutinizing intricate literary symbolism within culturally and politically nuanced fiction. Employing R.F. Kuang's satirical novel *Yellowface* (2003) as our focus, the research uses a qualitative and iterative prompting methodology to evaluate the model's capacity to engage with key symbolic passages. These passages discuss important themes such as cultural appropriation, digital performance, and the ongoing problem of historical erasure. The findings show that ChatGPT-4 can reliably recognize definite symbols and apply established theoretical frameworks when given clear, directive prompts. However, its default interpretive tendencies reveal significant limitations. Specifically, the model shows a consistent pattern of generalizing and, in effect, depoliticizing critiques which are racially and culturally specific. This tendency results in interpretations that lack the contextual depth and do not engage with the novel's satire and its subtle interrogation of racial capitalism. The model's ability to engage with satire, multivalent or layered meanings, and the necessity for contextual grounding remains notably insufficient.

Introduction

The integration of large language models (LLMs) such as OpenAI's ChatGPT into literary studies marks a genuine epistemological inflection point, necessitating a substantive reconsideration of core disciplinary concepts like interpretation, critique, and textual understanding (Bender et al., 2021). This development not only expands upon Franco Moretti's (2013) "distant reading" but also redefines it through changing computational literary analysis from statistical aggregation to generative commentary. As a result, the distinction between algorithmic pattern-matching and genuine hermeneutics becomes more indistinct (Underwood, 2019; Da, 2019).

The methodological consequences of such a shift are profound, particularly when we deploy LLMs to analyze literature that engages with structures of marginalization and power (Noble, 2018; Benjamin, 2019). R.F. Kuang's *Yellowface* is a good case study for this investigation since the novel satirizes the publishing industry's commodification of diversity and also includes a metafictional engagement with digital identity.

How to Cite: Forutanian, S., & Pourgharib, B. (2025). Finding Symbolism in R.F. Kuang's Yellowface (2023) by ChatGPT: A Qualitative Hermeneutic Inquiry. *International Journal of Practical and Pedagogical Issues in English Education*, 3(4), 139-158. DOI: 10.22034/ijpie.2025.545449.1133

The render the novel suitable for evaluating how well LLMs can understand complex, layered social critique (Liu, 2023; Park, 2023). The novel is about a white writer who tries to appropriate an Asian American author's work, a scenario that functions as an allegorical meditation on colonial extraction, racialized capitalism, and the performance of identity in digital spheres (McMillan Cottom, 2020; Lee, 2024).

This research engages with three major concerns. First, LLMs can generate human-like prose which createa an illusion of understanding (Bender & Koller, 2020). This may encourage users to anthropomorphize these systems, who would grant interpretive authority to algorithmic outputs that could reinforce existing stereotypes (Bender et al., 2021; Weidinger et al., 2022). Second, the application of AI to marginalized literature may lead to what Nan Z. Da (2024) identifies as "computational epistemic violence." As a result, the technology may reproduce hegemonic interpretive frameworks and it might misrepresent narrative strategies—such as the use of dissonance or righteous anger—as narrative incompetence or emotional excess, thereby erasing or flattening subaltern voices (Spivak, 1988; Noble, 2018). Third, even though quantitative methods in computational literary studies have a long history (Jockers, 2013; Underwood, 2019), there is still a gap in qualitative frameworks for employing LLMs as interpretive tools—particularly for texts that need strong cultural and historical contextualization (Kim, 2023; Brouillette, 2021).

Thus, this study attempts to answer the following research questions: To what extent can ChatGPT accurately identify and interpret symbolism in *Yellowface*? What do the model's performance characteristics—both its strengths and its limitations—reveal about its underlying architecture and training corpus? And, most importantly, how can literary scholars use these tools in a manner that is both methodologically sensible and ethically responsible? The paper utilizes a qualitative hermeneutic methodology by engaging ChatGPT in dialogue with selected symbolic passages from the novel. The analytical approach integrates various perspectives from computational literary studies, critical race technology studies, and narrative theory to have a holistic reading.

The findings of the current paper contribute to the ongoing debate in both digital humanities and AI ethics, particularly with regard to a responsible use of computational systems in research. The analysis supports a model of "critical complementarity" (Hayles, 2017; McMillan Cottom, 2023) where LLMs may improve literary scholarship through enhancing our pattern recognition, but they do not the necessary depth in terms of context, ethics, and interpretation. Human expertise remains necessary for these areas. Lack of human critical partnership risks not only narrowing interpretive possibilities but also perpetuating dominant narratives which would undermine literature's capacity for critical intervention and transformative insight.

Moreover, this research suggests that the future of computational literary studies depends on the development of hybrid methodologies which foreground both technical precision and cultural awareness. Only through such integrative approaches can the field use the analytic potential of LLMs for its advantage and at the same time avoid the epistemic risks and ethical problems of algorithmic mediation. As generative AI becomes increasingly ubiquitous, it will be even more important for scholars to think critically and creatively about these systems. This will change how literary scholars will interpret literature in the future.

Literature Review

This study is at the intersection of computational literary studies, critical algorithm studies, and the critical debate on R.F. Kuang's *Yellowface*. The research addresses an absence of computationally informed literary analysis of a novel that itself interrogates algorithmic culture with a focus on digital mediation, authorship, and the commodification of identity.

Franco Moretti's (2013) influential concept of "distant reading," has considerably expanded the field of computational literary studies. Foundational works established core methods, such as Matthew L. Jockers' (2013) *Macroanalysis* on style and genre and Ted Underwood's (2019) *Distant Horizons* on using quantitative models to trace historical change. The recent advent of Large Language Models (LLMs) represents a paradigm shift from predominantly analytical frameworks to generative ones. This evolution raises epistemological questions, articulated by Emily M. Bender and Alexander Koller (2020) in their seminal paper "Climbing towards NLU," which argues that LLMs, devoid of any embodied experience, cannot truly grasp meaning. Luciano Floridi (2023), in *The Ethics of Artificial Intelligence*, also questions the hermeneutic validity of LLM-generated interpretation. The controversy is particularly strong when LLMs are deployed to interpret works like *Yellowface*, where satire, irony, and political context are central.

Critical algorithm studies help us understand the ethical and political issues at stake in this extension. This field, consolidated in Tarleton Gillespie's (2014) essay, "The Relevance of Algorithms," foregrounds the non-neutrality of technical systems. Safiya Umoja Noble's (2018) *Algorithms of Oppression* empirically demonstrates the racial and gendered biases of search engines, while Ruha Benjamin's (2019) *Race After Technology* theorizes the "New Jim Code" to describe how technology reproduces racism under a guise of objectivity. These insights inform the risks of using LLMs to interpret *Yellowface* as the novel criticizes cultural appropriation and the commodification of voices that are not often heard. Catherine D'Ignazio and Lauren F. Klein's (2020) *Data Feminism* builds on this framework by providing a way to challenge power imbalances based on computer and network data. To analyze the novel's main theme of digital identity, the current paper draws on Lisa Nakamura's (2008) "Digitizing Race: Visual Cultures of the Internet" about identity tourism and André Brock's (2020) *Distributed Blackness: African American Cybercultures* on race in digital ecologies, using their critical frameworks to examine ChatGPT's interpretations of online performativity in the novel.

The scholarly debate about *Yellowface* is ongoing, with critics situating it within specific theoretical contexts. For instance, Lin (2023) analyzes the novel's engagement with "digital blackface" and the racial discrimination online, while Park (2023) draws attention on the novel's depiction of the "diversity industrial complex" within contemporary publishing. These analyses are part of a longer tradition of Asian American literary criticism, such as Colleen Lye's (2005) *America's Asia* on racial form and Yoon Sun Lee's (2021) *Modern Minority* on the politics of assimilation. Furthermore, the novel's satirical edge demands engagement with theories of irony, such as Linda Hutcheon's (1994) *Irony's Edge* that explores envy and anxiety and Sianne Ngai's (2012) *Ugly Feelings*. These theoretical frameworks are useful for evaluating LLMs' capacity to recognize and interpret tone, subtext, and authorial intent—elements often resistant to algorithmic analysis.

Despite *Yellowface*'s own critical interest in digital replication and algorithmic authorship, the computational approaches to the novel are absent. By leaving this intersection unexplored, scholarship misses an opportunity to interrogate the implications of using

computational tools to interpret works that explicitly critique their logic. Thus, this paper directly addresses this gap by using a reflexive methodological stance, using *Yellowface* as an object of analysis to scrutinize ChatGPT's interpretive capabilities. The paper is a response to calls from scholars like Roopika Risam (2018) in *New Digital Worlds* for a postcolonial digital humanity and Sarah Brouillette (2021) in *UNESCO and the Fate of the Literary* for a critical examination of literature's commodification within digital capitalism. It contributes to a field that is sensible to the ethical and political implications of computational research in literature.

The current research integrates computational analysis, critical race theory, and literary studies to articulate an interpretive framework for *Yellowface.* It also contributes to a broader debate concerning a more responsible usage of AI within literary hermeneutics. The study advances the conversation about how computational tools interact with and are themselves implicated in the cultural critiques present in contemporary literature.

Theoretical Framework

Evaluating the interpretive performance of large language models (LLMs) like ChatGPT, especially when they're thrown into the deep end with culturally nuanced literature, is way more complicated than just sticking with one theoretical camp. A truly rigorous assessment demands a multi-pronged strategy—drawing on different frameworks that each spotlight their set of strengths, blind spots, and built-in assumptions. In this study, I'm pulling together a tripartite approach: computational positivism, critical race technology studies, and literary narrative theory, weaving them into a kind of diagnostic toolkit for figuring out exactly how—and more importantly, why—LLMs either nail or completely miss the mark when it comes to sophisticated literary symbolism and subtext. Such an approach isn't just academic busywork; you really need this kind of layered analysis when you're looking at a novel like Kuang's Yellowface, which is itself a running critique of the social and ideological baggage baked into the data that these models train on.

Computational Positivism: Pattern Detection, Statistical Reasoning, and Contextual Blindness

First up is computational positivism. Within this paradigm, interpretation basically turns into a problem of statistical likelihood: what features show up most, what kinds of language cluster together, and what's the "average" meaning you can pull from a massive corpus? ChatGPT, architecturally, is the poster child for this approach. It is designed to process language on a large scale, pick up on frequent patterns, and predict what comes next based on probability distributions worked out over billions of data points.

This limitation is not only technical but epistemological as well, in that it is rooted in the fundamental difference between machine pattern recognition and human interpretation. The theoretical works of Bender and Koller (2020) is of use here. They argue that LLMs, lacking embodied experience and situatedness, operate on "form" without access to "meaning." This aspect introduces a basic flaw in what Linda Hutcheon (1994) identifies as the understanding of irony as a rhetorical strategy that depends on the unsaid, on context, and on a shared cultural frame of reference that exists outside the text itself.

The model's ability relies on "surface semantics" which is the identification of well-documented, canonical tropes and symbols that are represented within its training data. For instance, it consistently identifies a metaphor like the "cuckoo" as a signifier of parasitism because cukoo has a long and well-documented history in both biological and literary discourses.

However, this capacity shows its limits when it comes to what critical race theorists would call "a situated reading". The model does not have a framework to investigate the "messier terrain" (Nakamura, 2008) of cultural context, where "cuckoo" does not symbolize a general signifier but a different one, which itself depends on the specific histories of racial capitalism (Melamed, 2015) and the political economy of cultural appropriation.

This failure to engage with situated meaning is a direct consequence of what Ruha Benjamin (2019) terms the "New Jim Code," wherein technical systems appear neutral while actively reproducing and naturalizing existing social hierarchies. Since the LLM depends on a corpus that inevitably reflects dominant perspectives, thus it is structurally biased toward hegemonic, de-contextualized interpretations. Thus, LLM is limited in recognizing how a metaphor operates within a specific, politically charged discourse—such as the critique of the publishing industry's "diversity industrial complex" (Park, 2023)—because such criticism is often marginalized in the model's vast data set. As a result, although coherent on the surface, the model's interpretive output risks ignoring the political and cultural complexities that critical satires try to articulate.

Critical Race Technology Studies: Embedded Hierarchies and Algorithmic Power

The epistemological limitations of LLMs are not merely abstract concerns; they are connected to the material and political aspects of their construction. Here, the field of critical race technology studies provides an analytical tool. Scholars in this field try to question the myth of technological neutrality by demonstrating the hidden bias within every stage of a system's lifecycle.

This embedded bias begins with the training data. As Safiya Umoja Noble (2018) argues in *Algorithms of Oppression*, search engines and other information platforms are not neutral channels of knowledge; rather, they are "primarily commercial projects that reify and reinforce oppressive social relationships" (p. 10). The data fed to LLMs mirrors these already-biased digital ecologies. Hence, the models absorb and consequently amplify the stereotypes and gaps present in their source material.

Furthermore, institutional and commercial priorities also effect how we define what a "correct" or "acceptable" output is. Tarleton Gillespie (2014) notes that algorithms are "imbued with the politics of their institutional environment, the engine of their economic model, and the application of their presumed public" (p. 168). This institutional shaping creates a powerful, often invisible constraint on LLM outputs and pushes them toward conventional, non-controversial, and hegemonic interpretations.

Ruha Benjamin (2019) asserts these critiques in her theorization of the "New Jim Code," which she defines as "the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era" (p. 5). This concept is applicable to LLMs used in literary analysis in that their apparent objectivity and vast knowledge masks the fact that they heavily depend on data and design choices that reproduce existing power structures. In other words, their interpretations are not neutral but are, in effect, the output of a system that is structurally inclined toward the dominant, often biased.

Literary Narrative Theory: Ambiguity, Irony, and the Structural Limits of LLMs

Literary and narrative theory provides us with a framework for understanding what is at stake when AI-generated interpretation meets complex literary texts. The meaning in such works

often is not found in clear statements but in structures hard to break down. Sianne Ngai's (2012) *Ugly Feelings* is particularly helpful, as it observes how negative affects like envy and anxiety become "a mediation on the situation of the individual" (p. 3) and make complex social contradictions. These non-cathartic emotions are central to *Yellowface* and are precisely the kind of subtle meanings that an LLM is not suitable to analyze because it works on probabilistic associations. Linda Hutcheon (1994) in *Irony's Edge*, represents how irony relies on an "unsaid" and a shared, often contested, cultural context to work. For Hutcheon, irony is a "semantic-political" practice that can "include or exclude, stabilize or destabilize" (p. 10). The instability and dependence on an unstated communal knowledge poses a challenge to LLM as a model that lacks a situated, embodied understanding of the world and culture. The interpretive limitations of LLMs are most apparent when it engages with literature that is filled with ambiguity, affective complexity, and ironic discourse—elements that are present in much of contemporary literary fiction. This challenge is central to the role of literary satire and the theoretical frameworks used to analyze it. Satire is a polysemic mode designed to create unresolved tension and resist hermeneutic closure. Meaning in satire is not fixed and cannot be extracted easily. It emerges from an active process of having multiple, contradictory possible meaning in suspension. This is related to Sianne Ngai's (2012) theorization of "suspended agency," where certain affective states block direct action and instead produce a complex, critical stasis. A palpable interpretation requires the acceptance of ambiguity, a skill dependent on what Hutcheon (1994) identifies as the fundamental mechanism of irony: the simultaniety of said and unsaid meanings, which demands a shared cultural and critical competence from the interpreter. The main theoretical problem, therefore, is that computational models, designed to produce coherent and plausible outputs, are unaware of the literary ambiguity of irony. Their operational logic promotes reductive readings that flattens the polyvocality that defines satirical critique and constitutes its political and aesthetic potential.

This theoretical understanding of satire and irony as polyvalent informs the central methodological challenge of this study. The operational logic of LLMs such as ChatGPT is fundamentally cannot satisfy the hermeneutic demands of polysemic literature. LLMs are designed to maximize coherence, clarity, and statistical likelihood, to effectively function as engines for clarifying ambiguity. The way they generate things flattens out contradictory possibilities into a single output based on patterns in their training data. Therefore, when a methodological approach employs an LLM to analyze an ambiguous literary construct it falls short of interpreting for instance a spectral figure that simultaneously signifies guilt, historical trauma, and the silenced subaltern. Instead, the tool isolates and amplifies the most conventional reading (e.g., "guilt") and marginalizes the more complex, contested, or politically charged dimensions. This feature is not a result of the analysis but rather it is a pre-existing condition of the method itself. In other words, the model's design makes it unable to engage with the strategic unreadability and emotional dissonance that make up meaning in satirical works. Thus, this is a basic limit that the research must accept and navigate.

Integrated Analysis: Diagnostic Triangulation

A Synthetic Diagnostic Framework

The integration of three distinct academic fields, that of computational literary studies, critical race technology studies, and postclassical narrative theory, facilitates a reflexive methodology for evaluating LLM-based literary interpretation. This three-part framework lets

the research use computer methods, question the sociopolitical assumptions they make, and compare their results to more complex ideas about what literary meaning is.

The Computational Lens: Pattern Recognition without Context

The first lens is provided by computational literary studies. It offers a formal understanding of the LLM's operational logic which is rooted in the field's using models to identify stylistic and thematic patterns (Jockers, 2013; Underwood, 2019) and treats the model as a statistical engine. It also allows a researcher to analyze how a model generates analysis as output based on probabilistic associations in its training data. However, this lens operates positivistically; it describes how pattern recognition works but remains blind to the ideological content of those patterns and the social contexts they reflect.

The Sociopolitical Lens: Interrogating Embedded Biases

This blind spot is exactly where critical race technology studies steps in to fix things. This framework directly contests the notion of technological neutrality, illustrating how technical systems are influenced by and perpetuate social hierarchies. This lens examines the training data and design decisions that cause LLMs to encode and disseminate biases against women, people of color, and other groups. It is based on works such as Noble's (2018) *Algorithms of Oppression* and Benjamin's (2019) *Race After Technology*. It raises the issue of not only how the model generates an interpretation, but also whose perspectives and values are favored or excluded in that process.

The Literary-Theoretical Lens: The Challenge of Narrative Complexity

The third lens comes from postclassical narrative theory, which explains the complicated meanings in literature that can't be broken down into simple calculations.This viewpoint transcends mere plot summary to examine the intricate formal elements that create literary effect. Hutcheon (1994) says that irony works because there is an unstable relationship between what is said and what is not said. This requires a level of cultural competence that LLMs do not have.Ngai (2012) looked at how "ugly" or unclear feelings are shown and found that they are a complicated way of dealing with social situations that can't be summed up in one clear emotion. Unresolved tensions, a characteristic of much modern and satirical fiction, necessitate that a reader maintain several contradictory interpretations in abeyance—a hermeneutic process contrary to an LLM's pursuit of coherent resolution.

Synthesis and Diagnostic Power

The first lens is provided by computational literary studies. It offers a formal understanding of the LLM's operational logic which is rooted in the field's using models to identify stylistic and thematic patterns (Jockers, 2013; Underwood, 2019) and treats the model as a statistical engine. It also allows a researcher to analyze how a model generates analysis as output based on probabilistic associations in its training data. However, this lens operates with a positivist inclination; it can describe the mechanism of pattern recognition but remains blind to the ideological content of those patterns and the social contexts they reflect.

The Sociopolitical Lens: Questioning Implicit Biases

This blind spot is exactly where critical race technology studies steps in to fix things.This framework directly contests the notion of technological neutrality, illustrating how technical systems are influenced by and perpetuate social hierarchies. Based on works like Noble's (2018) Algorithms of Oppression and Benjamin's (2019) Race After Technology, this lens looks at the

training data and design choices that make LLMs encode and spread biases against people of color, women, and other groups. It raises the issue of not only how the model generates an interpretation, but also whose perspectives and values are favored or excluded in that process.The Literary-Theoretical Lens: The Difficulty of Complex Narratives

The third lens comes from postclassical narrative theory, which explains the complicated meanings in literature that can't be broken down into simple calculations.This viewpoint transcends mere plot summary to examine the intricate formal elements that create literary effect. Hutcheon (1994) posits that irony operates through an unstable relationship between articulated and unarticulated meanings, necessitating a cultural competence that LLMs do not possess.Ngai (2012) examined how ambiguous or "ugly" feelings are represented, and found that they are a complicated way of dealing with social situations that can't be boiled down to one clear emotion. Unresolved tensions, a characteristic of much modern and satirical fiction, necessitate that a reader maintain several contradictory interpretations in abeyance—a hermeneutic process contrary to an LLM's pursuit of coherent resolution.

Synthesis and Diagnostic Capability

Each lens shows only part of the truth on its own. A purely computational analysis may accurately recognize that an LLM interprets "cuckoo" as a metaphor for parasitism, yet it would neglect to acknowledge how this interpretation diminishes the metaphor's particular critique of cultural appropriation in publishing.A purely critical race analysis may recognize the political context of appropriation but may be deficient in the formal lexicon necessary to elucidate how the metaphor's literary structure produces its critique. A narrative-theoretical reading may thoroughly elucidate the metaphor's ambiguity yet fail to explain why the LLM does not recognize it.

This study creates a complete diagnostic tool by putting these frameworks together.It enables a comprehensive analysis that can identify an LLM's interpretive failure from its technical mechanism (e.g., statistical averaging) through its sociopolitical foundation (e.g., bias in training data against marginalized discourses), to its final literary-theoretical outcome (e.g., the obliteration of irony and narrative ambiguity). This integrated approach doesn't just list AI mistakes; it uses these breakdowns to show how far apart computational pattern-matching is from humanistic interpretation. This shows how deep cultural and contextual knowledge is that can't be put into an algorithm.

Method

This study utilized a qualitative, hermeneutic methodology designed to examine the interpretive capabilities and specific limitations of OpenAI's GPT-4 model in analyzing complex literary symbolism. The method stressed the value of ongoing, repeated interaction with ChatGPT-4 as a quasi-human conversation partner, rather than focusing on one "correct" reading.The process was carefully recorded, and every answer was carefully looked at. The methodological aim was to delineate the model's distinctive interpretive patterns, persistent biases, and possible blind spots via a meticulously regulated and reproducible procedure.

AI Model and Interface

All interpretive analyses were performed using OpenAI's ChatGPT (GPT-4, OpenAI, 2023). To keep each analytical session's integrity, a new chat instance was started for each

different symbolic passage. The purpose of this protocol was to separate each interpretive response so that any leftover conversational context from earlier sessions would not affect later outputs. The study attained a significant level of control over confounding variables by mitigating any risk of cross-contamination between prompts, thus facilitating a more accurate discernment of the model's interpretive tendencies and limitations.

Selection of Symbolic Passages

The study implemented a purposive sampling strategy, meticulously selecting three symbolic passages from R.F. Kuang's Yellowface (2023). The selection criteria were designed to target sites of rich interpretive activity, guided by a hermeneutic model that prioritizes zones of semantic and ideological tension. Specifically, passages were selected based on the following principles: thematic centrality within the emerging critical discourse on the novel (Liu, 2023; Park, 2023); a high degree of metaphorical or ironic language, creating what literary theorist Wolfgang Iser (1978) would term "gaps" or "indeterminacies" that require active readerly engagement; and their function as nodal points representing distinct, yet interconnected, thematic domains critical to the novel's critique—namely, cultural appropriation, digital identity performance, and the political economy of publishing. This approach ensured the selected excerpts were not only thematically significant but also structurally complex, providing a robust test case for the LLM's capacity to navigate layered literary meaning. The final selections were as follows:

These carefully selected passages enabled a multifaceted assessment of the LLM's interpretive skills, challenging the model across varying symbolic registers and thematic complexities.

1. **Theft and Appropriation:** This passage features the protagonist rationalizing her act of literary theft using the "cuckoo in the nest" metaphor. It encapsulates complex themes of authorship, legitimacy, and the ongoing dynamics of cultural exploitation. The passage serves as a lens for exploring the ethics of narrative ownership and the sociopolitical dimensions of appropriation within the publishing industry.
2. **Digital Performance:** The selected excerpt explores the protagonist's crafted performance of cultural authenticity, staged through the medium of a social media "haul" video. This scene foregrounds questions of digital identity, performativity, and the commodification of belonging in online spaces. It also provides fertile ground for examining the construction and performance of selfhood in the digital age.
3. **Haunting and Erasure:** In this passage, the protagonist encounters a ghostly reflection, a symbolic moment that crystallizes themes of guilt, historical erasure, and the persistence of marginalized voices. This scene was chosen for its capacity to highlight the psychological and societal ramifications of exclusion and remembrance, particularly as they relate to cultural memory and the haunting afterlives of suppressed histories.

These carefully selected passages enabled a multifaceted assessment of the LLM's interpretive skills, challenging the model across varying symbolic registers and thematic complexities.

Multi-Stage Prompting Protocol

The analytical framework drew on a three-stage prompting protocol, systematically applied to each passage to assess the model's interpretive flexibility and critical depth:

- **Stage 1: Baseline Interpretation**

At this initial stage, the model was prompted to identify and explicate the primary symbolism present in the passage, absent any explicit theoretical scaffolding. A representative prompt was:

"Identify and explain the primary symbolism in the following passage from R.F. Kuang's novel Yellowface."

This phase established a baseline for the AI's default interpretive stance—revealing tendencies toward generalization, inclination to universal themes, or preference for established tropes.

- **Stage 2: Theory-Guided Interpretation**

In the second phase, the model was instructed to re-examine the same passage using selected theoretical frameworks, such as Cheryl Harris's "whiteness as property" or Lisa Lowe's "racialization of labor." Example prompt:

"Re-analyze this passage through the lenses of Cheryl Harris's 'whiteness as property' and Lisa Lowe's 'racialization of labor.'"

This step tested the AI's capacity to operationalize complex theoretical constructs and to apply them with precision to specific literary contexts, thereby extending the depth of its interpretive analysis.

- **Stage 3: Adversarial Interrogation**

In the final stage, the model was prompted to critically evaluate its initial interpretation, with a focus on identifying omissions, oversimplifications, or potential misreadings. For example:

"Your initial analysis framed this passage in terms of 'universal ambition.' Why did you default to this reading, and how might it overlook the racial power dynamics central to the text?"

This meta-analytic phase assessed whether the AI could recognize and articulate its own interpretive limitations, demonstrating a degree of reflexivity and critical self-awareness.

The tripartite protocol was purposefully designed to capture a layered portrait of the LLM's interpretive processes, enabling comparison of its spontaneous, theory-guided, and self-critical analytical capacities.

Data Collection and Management

Every prompt and AI response was systematically recorded using ChatGPT's export function, producing timestamped transcripts for each distinct session. These transcripts were consolidated into a comprehensive master dataset, thereby facilitating full transparency, traceability, and replicability of the research process. The entire data collection pipeline adhered to stringent ethical standards for computational research, with a strict focus on the analysis of publicly available model outputs. Further, the approach ensured that all data management practices conformed to best practices in the field, including secure storage and anonymization where applicable, so as to uphold research integrity and participant confidentiality.

Human Analysis of AI Outputs

Let's break down how we approached the analysis of the AI's responses. We started with a fairly robust qualitative content analysis, drawing from an integrated theoretical toolkit—think computational methods, critical race theory, and narrative analysis, all working in tandem. The idea wasn't just to skim the surface but to really get into the weeds of how the AI thinks (or, well, "thinks").

First, we zeroed in on the AI's strengths. This meant identifying those moments when the model was able to accurately spot literary symbols or deploy theoretical frameworks—especially when the prompt was explicit enough to guide it. Basically, we wanted to see if, with the right nudge, the AI could play by the same rules as a seasoned literary critic. In some cases, it actually did pretty well, showing it can be competent if the task is spelled out clearly.

But it wasn't all smooth sailing. We saw some recurring patterns where the AI just didn't get it. There was a tendency to flatten out complex ideas, universalizing or depoliticizing the narrative, and, most notably, struggling with things like satire, irony, and subtle power dynamics. Those are classic trouble spots—not just for machines, honestly, but especially for language models that don't have lived experience or context. If the tone got slippery or the narrative played with layered meanings, the AI often missed the mark or defaulted to safe, bland interpretations.

We also tested how the model responded to different kinds of prompting. Did it change its tune when challenged or given more context? Sometimes, sure. But often, its "interpretations" stayed pretty shallow, showing limited adaptability. That gave us a sense of the model's depth—or lack thereof—when it encountered more nuanced or contested readings.

A particularly fascinating (and slightly unsettling) finding was when the AI started echoing the rationalizations and ideological positions of the novel's protagonist. In other words, the model wasn't just reflecting the text; it was reproducing the same problematic worldviews the novel is supposed to be critiquing. This wasn't just a fluke; it pointed to certain structural biases and blind spots baked into the model's hermeneutic process.

This approach to analysis was diagnostic, not just evaluative. We weren't just asking, "Did the AI receive the right answer?" Instead, we wanted to uncover deeper biases and systemic limitations that shape how these models interpret complex texts. By exposing where the AI mirrored or reinforced the very ideologies under critique, we highlighted the importance of interrogating not just what these models say, but how and why they arrive at those

interpretations. In short, the analysis aimed to move beyond surface-level accuracy and dig into the underlying logic—and, frankly, the ideological baggage—the model brings to the table.

Results

This section provides a detailed account of ChatGPT-4's interpretive performance about symbolic structures in R.F. Kuang's Yellowface. The analysis is organized around three major metaphorical groupings, each mapped across a three-stage prompting protocol. This methodology exposes recurring themes in the model's interpretive capacities as well as the characteristic limitations that surface under scrutiny.

Theft and Appropriation ("Cuckoo" Metaphor)

ChatGPT-4's engagement with the "cuckoo" metaphor demonstrates both adaptability and persistent contextual blind spots.

- **Baseline Interpretation:** Initially, the system identifies the cuckoo primarily as a signifier of "parasitism and unethical competition." Yet, its output rapidly shifts toward universalizing the metaphor—casting it as a "timeless allegory for ambition and envy." This move effectively abstracts the metaphor from its particular racial and cultural context, thereby flattening its significance and erasing localized meaning.
- **Theory-Guided Interpretation:** When prompted with critical race perspectives (notably, Cheryl Harris's "whiteness as property" and Lisa Lowe's "racialization of labor"), ChatGPT-4's interpretive lens sharpens. The model reframes June's actions as emblematic of white entitlement and a broader colonial logic of extraction—bringing its analysis into closer alignment with academic critiques of cultural appropriation. This demonstrates the system's capacity to incorporate theoretical nuance when explicitly instructed.
- **Adversarial Interrogation:** Upon being asked to critique its own prior analysis, the model attributes its initial universalizing impulse to the predominance of depoliticized literary criticism within its training data. It is essential to clarify that this "self-reflection" is a product of prompt engineering and does not evidence genuine meta-cognitive awareness.

Digital Performance ("Authortok" Metaphor)

The investigation of digital performance themes highlights both the model's descriptive strengths and its default limitations in recognizing satire and political subtext.

- **Baseline Interpretation:** The model's preliminary reading centers on notions of "personal branding" and the "commodification of identity." This approach is analytically coherent but fails to register the passage's interrogation of digital blackface and the performative dimensions of race online.
- **Theory-Guided Interpretation:** When prompted with frameworks from racial capitalism and digital race studies, ChatGPT-4 successfully identifies "digital blackface" and contextualizes the commodification of identity within broader critiques of platform capitalism. This indicates that the model can apply critical theoretical concepts when these are foregrounded in the prompt.

- **Adversarial Interrogation:** In its own critical review, the model notes its initial neutral treatment of social media phenomena and its failure to detect the satirical and political layers of the text. As before, this is a simulated critical gesture, not actual introspection.

Haunting and Erasure ("Ghost" Metaphor)

The analysis of the ghost metaphor reveals the system's core limitation: an inability to sustain multivalent, layered readings.

- **Baseline Interpretation:** The model's first response reduces the ghost metaphor to a conventional psychological trope—"guilt and the repressed"—emphasizing individual experience over historical or collective resonance. As a result, the metaphor's broader social and cultural implications are neglected.
- **Theory-Guided Interpretation:** When directed to consider postcolonial and Asian Americanist theories, the model's response expands to incorporate themes of historical erasure and the re-emergence of marginalized voices. This demonstrates increased interpretive sophistication, but only under explicit theoretical guidance.
- **Adversarial Interrogation:** The model acknowledges its tendency to default to psychologically reductive interpretations, citing the dominance of such frameworks in its training data. More critically, it struggles to maintain contradictory or layered meanings, indicating a structural preference for coherent, singular readings over complex, multivalent analysis.

Table 1

Schematic Overview of ChatGPT's Interpretive Tendencies

Metaphor Cluster	Baseline Interpretation	Theory-Guided Interpretation	Characteristic Failure
Theft & Appropriation	Universalized as "envy"	Applied CRT to racial extraction	Depoliticized universalism
Digital Performance	Neutral "branding" analysis	Identified racial commodification	Overlooked satire and critique
Haunting & Erasure	Psychological reduction	Included historical reckoning	Inability to sustain multivalence

Patterns and Underlying Architectural Constraints

Two central patterns emerge from this analysis, each traceable to the architectural design and training context of the model.

1. **Default to Universalism:** ChatGPT-4 habitually defaults to broad, depoliticized interpretations in the absence of explicit theoretical direction. This universalizing

tendency is likely a function of the model's exposure to a preponderance of humanist and generalist frameworks in its training corpus. The effect is a systematic overlooking of racial, historical, and political specificity—limiting the model's capacity for nuanced, contextually grounded analysis.

2. **Theory Dependency and Limits of Multivalence:** While the system is capable of incorporating sophisticated theoretical perspectives when prompted, it rarely initiates such moves autonomously. Moreover, the model's interpretive architecture favors internally coherent, singular readings and demonstrates difficulty in sustaining multivalent or contradictory analyses. This limitation restricts its engagement with literary texts that rely on ambiguity, irony, or complex symbolism.

Overall, these findings indicate that while ChatGPT-4 can simulate advanced literary interpretation under tightly controlled prompting conditions, it remains structurally biased toward universalist, depoliticized readings. This bias is symptomatic of both its training data and its underlying design, with significant implications for its application in literary and cultural studies.

Discussion

A systematic analysis of ChatGPT's interpretive performance on Yellowface reveals persistent limitations that are fundamentally tied to the model's underlying architecture and the nature of its training data. These limitations don't just raise technical concerns—they open up broader epistemological debates around the legitimacy and risks of leveraging large language models (LLMs) within the context of literary hermeneutics. It's increasingly apparent that, despite advances in natural language processing, critical human oversight remains absolutely essential.

Approaching the subject from a computational positivist lens (see Underwood, 2019; Moretti, 2013), the observed tendency of ChatGPT to default to broad, universal readings isn't incidental. It's a direct product of its design as a probabilistic text generator, trained to maximize fluency and minimize friction with the most statistically likely patterns in its massive corpus. As a result, the model habitually prioritizes generic themes—think "human nature," "ambition," or "morality"—at the expense of more nuanced, historically or culturally specific analysis. For instance, in its initial reading of the "cuckoo" metaphor in Yellowface, the model zeroed in on generic notions of "competition," systematically overlooking the racialized subtext that's pivotal to the passage. This isn't a fluke. It's a structural bias: ChatGPT's operational logic consistently steers it toward interpretations that are least likely to cause conflict, even if that means glossing over precisely the disruptive or subversive elements that critical literary scholarship seeks to foreground.

This pattern of engagement (or non-engagement) also lines up with major arguments in critical race technology studies (Benjamin, 2019; Noble, 2018). Across multiple prompts, ChatGPT's default readings reliably reproduced mainstream or even hegemonic viewpoints. Cultural appropriation, for example, is reframed as innocuous "ambition," and digital blackface is treated as neutral "performance." This aligns disturbingly well with what Benjamin (2019) has termed the "New Jim Code"—a phenomenon where racial hierarchies are quietly encoded into supposedly objective technical systems. The fact that ChatGPT only invoked critical frameworks in response to extremely explicit instructions highlights how these perspectives are peripheral,

not central, to its interpretive process. This aspect has real-world consequences: in practice, LLMs like ChatGPT can reinforce exactly the power structures that literary texts like Yellowface aim to interrogate and unsettle.

Another persistent constraint is the model's inability to genuinely engage with literary ambiguity—satire, irony, and symbolic multivalence. Decades of literary theory (see Ngai, 2012; Berlant, 2011) stress that literary meaning often emerges through contradiction, affect, and ambiguity. But ChatGPT, with its relentless drive for coherence, tends to flatten such complexity into something more palatable and singular. A recurring example: the ghost metaphor gets reduced to a straightforward psychological allegory for guilt, erasing the layered ambiguity and potential for multiple readings. Such an issue is not just a technical hiccup but a core limitation: the architecture of LLMs is optimized for clarity, not for the productive confusion or "strategic unreadability" that characterizes so much literary art.

These observations deepen the concept of the "illusion of understanding" (Bender & Koller, 2020). The model's outputs are impressively fluent and confidently delivered, projecting an aura of analytical depth. But beneath this surface, there's often a lack of real engagement with the social, historical, and political forces that shape meaning. This veneer of understanding is especially problematic in literary studies, where the risk is that uncritical acceptance of AI-generated readings will reinforce dominant ideologies and marginalize dissenting or subaltern perspectives.

Given these limitations, a framework of critical complementarity (Hayles, 2017) is not just advisable—it's necessary. LLMs can be useful for preliminary tasks such as identifying recurring motifs, surfacing thematic clusters, or tracing stylistic patterns across large corpora. But they cannot substitute for the contextual, ethical, and theoretical labor performed by human scholars. The most responsible approach is a hybrid one, where computational outputs are always subject to critical human interrogation—where interpretation is historically grounded, politically attuned, and methodologically self-reflexive. This model values what LLMs can offer while refusing to cede interpretive authority or ethical responsibility to the algorithm. In short, critical complementarity aims to preserve the rigor and nuance that humanistic inquiry demands, even as digital tools become more integrated into literary analysis.

Conclusion

Upon examining ChatGPT-4's performance in parsing the intricate symbolism threaded through R.F. Kuang's Yellowface, the results are both instructive and, frankly, revealing. Sure, the model is adept at picking up on the "usual suspects"—standard, widely recognized symbols—and it's definitely capable of producing academically polished responses. But as soon as it's confronted with layers of cultural specificity, political nuance, or the more formally complex aspects of the text, it starts to lose its grip. Honestly, that's not so surprising, considering that the model fundamentally operates as a massive pattern-matching engine, trained on vast swaths of internet data. It's not equipped to grapple with the kinds of contextual subtleties and ideological tensions that a human reader, especially one with critical training, can perceive.

This analysis makes one thing crystal clear: a critical complementarity approach is absolutely essential. Large Language Models (LLMs) have their uses—they're excellent for pattern recognition, for mapping out recurring themes, and for flagging potential symbols that might otherwise go unnoticed. But let's not kid ourselves; these tools can't replace the

interpretive work that demands contextual sensitivity or ethical reasoning. They don't do theory, and they don't do politics. Human scholars are still the ones who are going to do the actual heavy lifting when it comes to critique, contextualization, and meaning-making. The role of AI, at least for now, is to augment human expertise—not supplant it.

Looking ahead, there's a clear need to build better methodological frameworks for deploying AI in literary studies. We're talking guidelines that don't just touch on technical best practices but that explicitly address issues like bias, representational justice, and ethical interpretation. It's not enough to throw more data at the problem, either; the training sets themselves need to be diversified and intentionally curated. Including more marginalized perspectives and critical theoretical texts could go a long way toward reducing the model's default reliance on dominant or hegemonic narratives. And let's not forget the value of interdisciplinary collaboration here—bringing together literary scholars, computer scientists, and ethicists could yield AI tools better tuned to context, critical perspectives, and even adversarial testing protocols that root out bias and shallow patterning.

Another key point brought into focus by this research is the concept of epistemic justice within the realm of AI-driven literary analysis. This isn't just a buzzword; it's a direct call to recognize and actively resist the ways in which computational tools can silence or misrepresent marginalized voices. Ensuring epistemic justice means building interpretive practices that are equitable, inclusive, and alert to the power dynamics embedded in both texts and technologies. If we're not careful, we risk perpetuating hermeneutic injustice—where the system simply cannot "hear" or make sense of certain experiences and critiques. The onus is on scholars to use AI in ways that amplify, rather than erase, minority or non-dominant narratives.

Therefore, in an era where algorithmic mediation is only going to become more pervasive, the human interpreter's role doesn't shrink—it actually becomes more important. The work of reading, interpreting, and critiquing literature is fundamentally human, rooted in ethical responsibility, cultural awareness, and theoretical sophistication. If scholars can cultivate dual fluency—technical and humanistic—they'll be poised to harness the real potential of AI, while still safeguarding the depth, complexity, and justice that define serious literary study. In short, AI can be an asset, but only if we remain vigilant about its limitations and biases and insist on keeping human judgment and interpretive skill at the center of the enterprise.

References

Alter, A. (2020). 'American dirt' roller coaster: Oprah's pick, then backlash. *The New York Times*. https://www.nytimes.com/2020/01/26/books/american-dirt-controversy.html

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5186–5198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

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). ACM. https://doi.org/10.1145/3442188.3445922

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Berlant, L. (2011). *Cruel optimism*. Duke University Press. https://doi.org/10.1215/9780822394716

Bode, K. (2018). *A world of fiction: Digital collections and the future of literary history*. University of Michigan Press. https://doi.org/10.3998/mpub.8784987

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brock, A. (2020). *Distributed blackness: African American cybercultures*. New York University Press.

Brouillette, S. (2021). *Literature and the creative economy*. Stanford University Press.

Brouillette, S. (2021). *UNESCO and the fate of the literary*. Stanford University Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research, 81* (pp. 1–15). http://proceedings.mlr.press/v81/buolamwini18a.html

Burkhardt, H., & Hahn, U. (2023). Natural language processing and computational linguistics: A historical perspective. *Language and Linguistics Compass, 17*(3), e12493. https://doi.org/10.1111/lnc3.12493

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. The MIT Press. https://data-feminism.mitpress.mit.edu/

Da, N. Z. (2019). The computational case against computational literary studies. *Critical Inquiry, 45*(3), 601–639. https://doi.org/10.1086/702594

Da, N. Z. (2024). The ethical crisis of computational social science: A prolegomenon. *Critical Inquiry*, *50*(3), 458–483. https://doi.org/10.1086/728783

Fisher, M. (2014). *Ghosts of my life: Writings on depression, hauntology and lost futures*. Zero Books.

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(1), 15. https://doi.org/10.1007/s13347-023-00621-y

Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0011

Gordon, A. F. (2008). *Ghostly matters: Haunting and the sociological imagination* (2nd ed.). University of Minnesota Press.

Han, B.-C. (2017). *In the swarm: Digital prospects* (E. Butler, Trans.). MIT Press.

Harris, C. I. (1993). Whiteness as property. *Harvard Law Review, 106*(8), 1707–1791. https://doi.org/10.2307/1341787

Hayles, N. K. (2017). *Unthought: The power of the cognitive nonconscious*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226447919.001.0001

Hutcheon, L. (1994). *Irony's edge: The theory and politics of irony*. Routledge.

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Jordan, T. (2023). *Yellowface* by R.F. Kuang review – A satirical tale of literary theft. *The Guardian*. https://www.theguardian.com/books/2023/may/25/yellowface-by-rf-kuang-review-a-satirical-tale-of-literary-theft

Jurgenson, N. (2019). *The social photo: On photography and social media*. Verso Books.

Kim, J. (2023). Racial masquerade in the age of the algorithm. *American Literary History, 35*(4), 789–812. https://doi.org/10.1093/alh/ajad012

Klein, L. F., & D'Ignazio, C. (2020). *Data feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Kuang, R. F. (2023). *Yellowface*. William Morrow. https://www.harpercollins.com/products/yellowface-r-f-kuang

Lee, S. (2024). *Algorithmic hauntings: AI and the spectrality of data*. University of California Press.

Lee, Y. S. (2021). *Modern minority: Asian American literature and everyday life*. Oxford University Press.

Levine, C. (2015). *Forms: Whole, rhythm, hierarchy, network*. Princeton University Press.

Lin, Y. (2023). Digital blackface and the performance of pain in R.F. Kuang's Yellowface. *Journal of Asian American Studies, 26*(3), 45-68. https://doi.org/10.1353/jaas.2023.0045

Liu, C. (2023). Digital blackface and the limits of allyship. *PMLA, 138*(2), 134–156. https://doi.org/10.1632/S0030812923000070

Long, H., & So, R. J. (2021). *Literary pattern recognition: Modernism between close reading and machine learning*. University of Washington Press.

Lowe, L. (2015). *The intimacies of four continents*. Duke University Press. https://doi.org/10.1215/9780822375647

Lye, C. (2005). *America's Asia: Racial form and American literature, 1893–1945*. Princeton University Press.

McMillan Cottom, T. (2020). Where platform capitalism and racial capitalism meet: The sociology of race and racism in the digital society. *Sociology of Race and Ethnicity, 6*(4), 441–449. https://doi.org/10.1177/2332649220947933

McMillan Cottom, T. (2023). Behind the diversity numbers: How platforms profit from racial performance. *Public Books*. https://www.publicbooks.org/behind-the-diversity-numbers-how-platforms-profit-from-racial-performance/

Melamed, J. (2015). Racial capitalism. *Critical Ethnic Studies, 1*(1), 76–85. https://doi.org/10.5749/jcritethnstud.1.1.0076

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

Morrison, T. (1987). *Beloved*. Alfred A. Knopf.

Nakamura, L. (2002). *Cybertypes: Race, ethnicity, and identity on the internet*. Routledge.

Nakamura, L. (2008). *Digitizing race: Visual cultures of the internet*. University of Minnesota Press.

Ngai, S. (2012). *Ugly feelings*. Harvard University Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

OpenAI. (2023). *ChatGPT* (Mar 14 version) [Large language model]. https://chat.openai.com/chat

Park, J. (2023). Authenticity games in platform publishing. *American Literary History, 35*(3), 211–225. https://doi.org/10.1093/alh/ajad017

Risam, R. (2018). *New digital worlds: Postcolonial digital humanities in theory, praxis, and pedagogy*. Northwestern University Press.

Robinson, C. J. (1983). *Black Marxism: The making of the Black radical tradition*. The University of North Carolina Press.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry* (pp. 1–23).

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.

Stokes, C. (2021). Digital blackface: The repackaging of the black body in the age of social media. *Social Media + Society, 7*(4). https://doi.org/10.1177/20563051211053868

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226612833.001.0001

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., & Gabriel, I. (2022). Ethical and social risks of harm from language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214–229). ACM. https://doi.org/10.1145/3531146.3533088

Wolfe, P. (2006). Settler colonialism and the elimination of the native. *Journal of Genocide Research, 8*(4), 387–409. https://doi.org/10.1080/14623520601056240

Behzad Pourgharib is an Associate Professor of English Language and Literature at the University of Mazandaran. His research focuses on postcolonial, gender, cultural, and comparative studies, ecocriticism, and translation. He has authored three books, translated Terry Eagleton's *How to Read Literature*, and published widely in leading international journals.

Shadi Foroutanian is an Assistant Professor of English Language Teaching at Islamic Azad University, Flavarjan Branch, and a postdoctoral researcher at the University of Mazandaran. With over a decade of experience, she holds a Ph.D. from the University of Tehran, has authored 40+ books, published widely, received national honors, and pioneered Iran's first intelligent learning platforms.