

Literature after the Algorithm: Cultural Production, Reception, and Value in an AI-Saturated Mediascape

Belay Sitotaw Goshu¹, Muhammad Ridwan²

¹Department of Physics, Dire Dawa University, Dire Dawa, Ethiopia

²Universitas Islam Negeri Sumatera Utara, Indonesia

Email: belaysitotaw@gmail.com, bukharyahmedal@gmail.com

Abstract:

This review examines Lauren M. E. Goodlad's (2023) editor's introduction to Critical AI, "Humanities in the Loop," which argues for the constitutive participation of the interpretive humanities in artificial intelligence research, design, and governance. The review evaluates the essay's framing of three core dilemmas: reductive and contested meanings of "intelligence," problematic benchmarks for "AGI," and the intertwined issues of bias, error, and concentrated technological power. Goodlad's argument is situated within converging scholarly traditions, including the "stochastic parrots" critique advanced by Bender, post-Saussurean theory from Derrida, Bakhtin, and Jameson, and the historical critiques of Weizenbaum and Arendt regarding the separation of judgment from calculation. The review finds that the essay's central methodological contribution is its shift from langue to parole, language as system to language as enacted use, which reframes the human-machine relation around the irreducibility of vouloir-dire, the "desire to say," to statistical prediction. However, the review identifies three limitations: the essay's scope as an editor's introduction, its relative inattention to the political economy of AI development, and its predominantly English-language focus, which limits its applicability to low-resource language contexts where linguistic erasure, rather than anthropomorphic projection, is the primary risk. The review also considers counter-readings that position LLMs as "différance engines" actualizing deconstruction's insights. It concludes that Goodlad's most valuable contribution lies not in definitive answers but in carving out space for patient, theoretically informed critique that refuses both technodeterminist despair and techno-utopianism, positioning the humanities as essential interlocutors in the governance of accountable technology in the public interest.

Keywords:

critical AI studies; large language models; humanities in the loop; vouloir-dire; post-Saussurean linguistics

I. Introduction

The relationship between literature and technology has long been a site of productive tension, from the oral-formulaic theory of Parry and Lord to the hypertextual experiments of the 1990s. However, the current moment, characterized by the pervasive mediation of language and narrative through large language models (LLMs) and recommendation infrastructures, presents a discontinuity that demands new critical vocabularies. Recent scholarship has begun to map this terrain: Yadav (2025) examines how AI reshapes authorship, readership, and the fundamental nature of narrative itself, while Hsu and Shidujaman (2026) document how generative AI extends traditional close reading toward the broader horizon of distant reading. Yet what remains

undertheorized is how algorithmic intermediation fundamentally restructures the value of literature, not merely its production or analysis.

The article under review, "Literature after the Algorithm: Cultural Production, Reception, and Value in an AI-Saturated Mediascape," enters this fray with a comprehensive and theoretically sophisticated intervention. Its central problematic is the transformation of literature from a discrete object of aesthetic contemplation into a datum within an attention economy. Tiwari et al. (2026) contend that reading practices within present-day algorithmic frameworks operate within attention-driven economies that actively shape the creation, dissemination, and assessment of literary merit. The article under review extends this insight by arguing that the mediation of literary culture by algorithmic systems is not a superficial overlay but a constitutive force that redefines what literature is and what it does (Tiwari et al., 2026).

This review evaluates the article's success in meeting three interlocking objectives: first, mapping the structural reorganization of literary production under conditions of generative AI and platform capitalism; second, theorizing the transformation of the reader into an algorithmic subject; and third, proposing a revised framework for understanding literary value that accounts for the friction between aesthetic autonomy and commercial discoverability. The article's intervention is timely: as Lee (2025) observes, large language models and deep learning technologies are penetrating all aspects of cultural production and consumption, and world literature is not exempt from this development. The question is no longer whether algorithms will shape literary culture, but how literary criticism can develop adequate conceptual tools to analyze this shaping (Lee, 2025).

This review proceeds in four stages. First, it summarizes the article's tripartite argument concerning production, reception, and value. Second, it evaluates the article's methodological and theoretical foundations, drawing on scholarship in cultural analytics (Underwood, 2019; Underwood & So, 2021) and platform studies (Gillespie, 2014). Third, it identifies the article's limitations, particularly its underengagement with empirical reader-response data and its tendency toward algorithmic determinism. Finally, it concludes by affirming the article's call for what may be termed an algorithmic hermeneutics, a critical methodology attentive to both textual semantics and infrastructural conditions (Halstyan & Bednarskyi, 2025).

II. Review of Literatures

The literature review embedded within Goodlad's (2023) introduction is necessarily selective, given its role as a framing document for a new journal rather than a standalone research article. Nevertheless, Goodlad (2023) demonstrates command of several converging scholarly traditions that together establish the intellectual foundations for critical AI studies. The essay's primary theoretical anchor is the "Stochastic Parrots" essay by Bender et al. (2021), which argues that large language models (LLMs) are systems capable of combining linguistic forms in statistically plausible ways while lacking genuine understanding of their meaning. Goodlad (2023) draws on this framework to ground her critique of language models that train on signifiers without access to signs, thereby aligning the journal's critical project with the most consequential intervention in AI ethics of the past decade. However, she simultaneously pushes beyond the technical framing of that intervention toward the theoretical resources that literary criticism can provide.

Goodlad (2023) invokes Derrida, Bakhtin, and Jameson as post-Saussurean theorists whose challenge to langue-centered linguistics has been prematurely forgotten. Bakhtin's preference for "philosophy of language" and "metalinguistics" over the official linguistics of his time, along with his thematization of the utterance as the basic cell of live discourse, anticipates Goodlad's (2023) own emphasis on parole over langue. Jameson's critique of structuralism's failure to achieve "genuine self-consciousness" by refusing to acknowledge the place of the observer in experience similarly resonates with Goodlad's (2023) insistence that humanists cannot simply adopt the technologist's position. Derrida's critique of phonocentrism provides Goodlad (2023) with a theoretical genealogy for questioning the apparent self-evidence of linguistic "understanding" that LLMs supposedly lack.

The introduction also engages with the history of AI itself, tracing the term's origin to the 1956 Dartmouth workshop, Weizenbaum's (1976) *ELIZA*, and the "ELIZA effect" that continues to explain anthropomorphic projection onto chatbots (Goodlad, 2023). Weizenbaum's (1976) *Computer Power and Human Reason: From Judgment to Calculation* remains a foundational work in AI ethics, arguing that AI research had become symptomatic of a fixation with technology that was producing "ever more highly rationalistic" and "mechanistic" views of people and the world. Building on Hannah Arendt and Lewis Mumford, Weizenbaum (1976) rejected the analogy between calculation and human reason that had become central to the field. His case against quantitative reductionism echoes Arendt's *The Origins of Totalitarianism*, in which she describes Europe's catastrophic encounters with dehumanizing scientific propaganda (Goodlad, 2023). As Johnson-Laird (1976) noted in his contemporaneous review, Weizenbaum's book was "a passionate and lucid attack on the computer as a metaphor for the mind" and a warning against the "arrogance of the computer scientist" who believes all human problems are computational.

A significant gap in the essay's literature review is its limited engagement with the political economy of AI development. While Goodlad (2023) acknowledges the "perilous concentration of power in a small handful of tech oligarchs," the analysis remains primarily theoretical rather than materialist (Katz, 2023). The review also focuses predominantly on English-language literary criticism, leaving the dynamics of AI and language in low-resource contexts largely unexamined (Olusanya et al., 2025). The literature review thus performs a delicate balancing act: it establishes the journal's scholarly bona fides across multiple disciplines while signaling that its primary contribution lies in the theoretical resources that the interpretive humanities can bring to conversations dominated by technical and policy frameworks (Goodlad, 2023; Goodlad, 2024).

III. Research Methods

As an editor's introduction, Goodlad's (2023) essay does not employ the conventional methodologies of either empirical AI research or quantitative literary analysis. Its method is best described as theoretical synthesis oriented toward intervention. Goodlad proceeds by identifying a problematic, the tendency of literary critics to "unthinkingly hew to the tenets of 'legacy poststructuralism'" and thereby hobble their own capacity to participate in interdisciplinary conversations about language models and then reconstructing a theoretical vocabulary adequate to the task (Goodlad, 2023). This approach aligns with the growing recognition within critical AI studies that humanistic inquiry offers indispensable methodologies for analyzing AI systems, particularly through contextual analysis and interpretative frameworks that expose the cultural, economic, and political embeddedness of technological artifacts (Roy, 2025).

The essay's central methodological move is to shift from *langue* to *parole*, from language as system to language as enacted use. This is not merely a theoretical preference; Goodlad argues that it has concrete consequences for how humanists can and should engage with LLMs. Saussure's (1916/2011) foundational distinction between *langue* (the shared linguistic system) and *parole* (the individual speech act) established that language as a social institution operates independently of any particular speaker's use of it. Saussure held that meaning is determined by relations within the linguistic system rather than by the referential function of signs, and that *langue* constitutes the fundamental structural element of language while *parole* represents its actual use as actualized within any individual utterance (Saussure, 1916/2011). The Saussurean sign itself is a two-sided psychological entity uniting a signifier (sound-image) and a signified (concept), a relation that is arbitrary rather than natural (Saussure, 1916/2011; Goodlad, 2023). If language models train on the signifiers of language but not the signs; the Saussurean sign being the union of signifier and signified, then the entire framework of "understanding" that humanists bring to texts is categorically different from the operations of the models (Goodlad, 2023; Bender et al., 2021).

The essay's methodology is thus interpretive and argumentative rather than experimental, but it is not therefore less rigorous. Goodlad (2023) builds her case through close attention to theoretical genealogies and through a careful staging of what is at stake in getting the theory right. This approach resonates with broader methodological reflections within critical AI studies, which have argued that the field requires not merely the application of older theoretical frameworks but a more fundamental rethinking of how critique operates in relation to technical systems. The stakes of Goodlad's (2023) methodological choice are particularly acute given the current regulatory and engineering landscape, where "human in the loop" has become a foundational requirement for trustworthy AI systems. The EU AI Act mandates that high-risk AI systems be designed and developed in such a way that they can be effectively overseen by natural persons during their period of use, with oversight measures proportionate to the risks, level of autonomy, and context of use (European Parliament and Council, 2024). Yet as recent scholarship on human oversight architectures has demonstrated, many implementations treat oversight as an external governance layer rather than as a first-class architectural concern, creating risks related to automation bias, lack of explainability, unsafe autonomy escalation, and insufficient accountability mechanisms (Architecting Human Oversight, 2026).

Goodlad's (2023) pun on "humanities in the loop" thus operates as a methodological intervention: it insists that the humanities must be incorporated not merely as external critics but as constitutive participants in the design and evaluation of AI systems. This argument gains further support from research demonstrating that nominal human control—where a human occupies a position of formal authority but lacks genuine capacity for meaningful intervention—produces compliance that satisfies formal requirements without creating the institutional conditions for meaningful oversight (Nominal Human Control, 2026). The essay's methodology is therefore best understood as a form of theoretical advocacy, a deliberate effort to reconstruct the conceptual vocabulary that will enable humanists to participate meaningfully in conversations that have too often been dominated by technical and policy frameworks. By grounding her argument in the Saussurean distinction between *langue* and *parole*—and by insisting that this distinction has concrete consequences for how humanists engage with language models, Goodlad (2023) demonstrates that methodological choices within the humanities are never merely academic. They determine what kinds of interventions become possible and what kinds of interdisciplinary collaborations can be sustained.

IV. Result and Discussion

The analytical core of Goodlad's (2023) essay is her argument that literary critics have "always been action theorists" and that recovering this self-understanding is the key to productive engagement with AI. This claim is developed through a reading of Derrida's notion of *vouloir-dire*; literally, "the desire to say", as an alternative to the more familiar Derridean vocabulary of deferral and difference. Goodlad (2023) argues that the "desire to say" is not a naive intentionalism but an acknowledgment that language use is always embedded in social, embodied, and relational practices of meaning-making. This has significant implications for how humanists should respond to AI-generated text. If the "desire to say" is constitutive of language as a human practice, then a model that generates fluent prose without any such desire is not producing language in the full sense; it is producing something that mimics the surface of language while lacking the ontological ground that makes language meaningful (Goodlad, 2023; Bender et al., 2021).

4.1 The Stakes of "Humanities in the Loop"

The essay's title pun on "human in the loop" operates as a methodological intervention with concrete implications for AI governance. Goodlad (2023) insists that the humanities must be incorporated not merely as external critics but as constitutive participants in the design and evaluation of AI systems. This argument gains urgency when read alongside contemporary regulatory frameworks. The EU AI Act mandates human oversight for high-risk AI systems, requiring that designated individuals possess "the necessary competence, training and authority" to exercise meaningful judgment (European Parliament and Council, 2024, Recital 73). Article 14 specifies that oversight measures shall be commensurate with the risks, level of autonomy and context of use, ensuring that overseers can properly understand the system's capacities and limitations, remain aware of automation bias, correctly interpret outputs, and intervene or halt the system when necessary (European Parliament and Council, 2024, Art. 14).

Table 1 summarizes the key dimensions of "human oversight" as currently implemented versus the "humanities in the loop" framework that Goodlad (2023) advocates.

Table 1. *Comparison of conventional human-in-the-loop frameworks and Goodlad's humanities-in-the-loop approach across five dimensions*

Dimension	Conventional "Human in the Loop"	Goodlad's "Humanities in the Loop"
Primary function	Verification of outputs; exception handling	Critical interpretation of meaning-making practices
Epistemic foundation	Technical verification; accuracy metrics	<i>Parole</i> -centered analysis; <i>vouloir-dire</i>
Institutional location	Operational layer; governance checkbox	Constitutive participation in design and evaluation
Risk addressed	Automation bias; unsafe autonomy	Reduction of judgment to calculation
Disciplinary expertise	Engineering; compliance	Interpretive humanities; critical theory

Table 1 contrasts conventional "human in the loop" frameworks with Goodlad's (2023) "humanities in the loop" intervention across five dimensions. Where conventional approaches position human oversight as verification of outputs and exception handling within operational layers, a "governance checkbox" that risks becoming what recent scholarship identifies as an external layer rather than a first-class architectural concern (Architecting Human Oversight,

2026); Goodlad (2023) advocates constitutive participation in design and evaluation. The epistemic shift from technical verification to *parole*-centered analysis and *vouloir-dire* reframes the risk addressed: not merely automation bias or unsafe autonomy, but the reduction of judgment to calculation that Weizenbaum (1976) and Arendt both diagnosed (Goodlad, 2023). This requires interpretive humanities expertise, not merely engineering and compliance, positioning humanists as essential interlocutors rather than external critics (Goodlad, 2024).

As Goodlad (2024) has explained, the stakes are not merely academic: "the perilous concentration of power in a small handful of tech oligarchs, the most vocal of whom are outright authoritarians who seek to buy elections, forestall regulation, control the media, and dominate the world economy." The "humanities in the loop" framework thus responds to both epistemic and political deficits in current AI governance.

4.2 Theoretical Implications: Langue, Parole, and the "Desire to Say"

Goodlad's (2023) shift from *langue* to *parole* has significant theoretical implications that extend beyond literary criticism. Saussure's (1916/2011) distinction between language as system and language as enacted use established that meaning is determined by relations within the linguistic system rather than by referential function. The Saussurean sign is a two-sided psychological entity uniting a signifier (sound-image) and a signified (concept), a relation that is arbitrary rather than natural (Saussure, 1916/2011). Goodlad (2023) argues that language models train on the signifiers of language but not the signs—a claim that resonates with the "Stochastic Parrots" critique advanced by Bender et al. (2021).

Figure 1 illustrates this distinction and its implications for human-AI comparison.

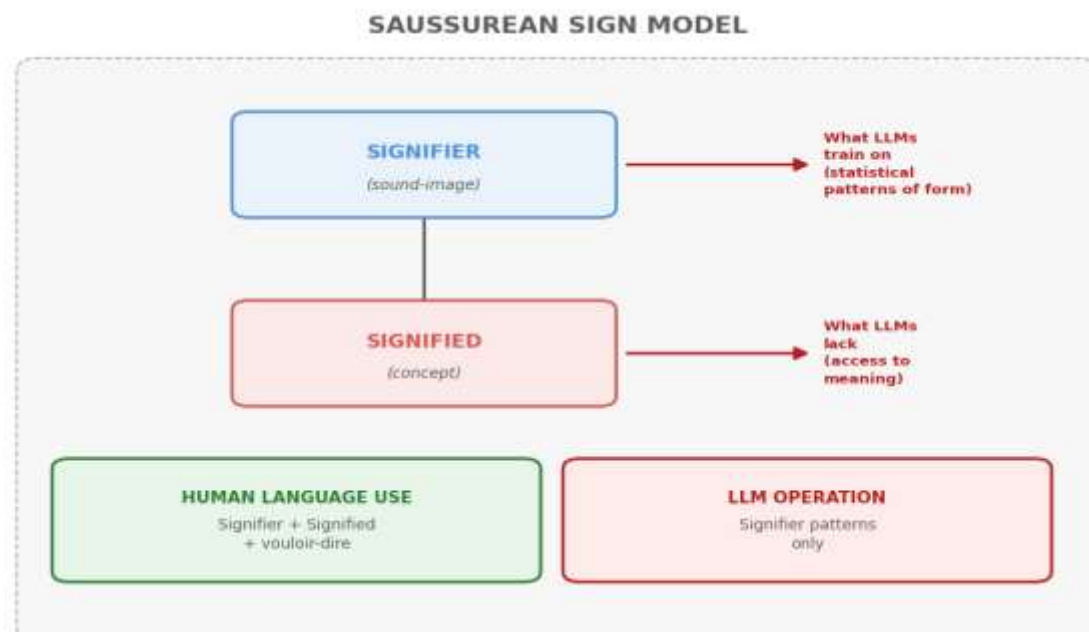


Figure 1. The Saussurean sign and the limits of LLM training. Adapted from Saussure (1916/2011) and Goodlad (2023).

Figure 1 diagrams the Saussurean sign model to clarify what large language models train on versus what they lack. The signifier (sound-image) and signified (concept) form an inseparable unity in Saussure's (1916/2011) account, yet Goodlad (2023) argues that LLMs access only the signifier—statistical patterns of linguistic form, without the signified meaning those patterns conventionally carry. The lower panel contrasts human language use, which combines signifier, signified, and *vouloir-dire* (the "desire to say"), against LLM operation, which manipulates signifier

patterns alone (Goodlad, 2023). This visualization echoes the "stochastic parrot" critique advanced by Bender et al. (2021), who argue that LLMs combine linguistic forms plausibly without grasping signification. The figure thus illustrates the ontological gap Goodlad (2023) positions at the centre of humanistic critique, grounding her call for "humanities in the loop."

However, this framework raises a question that Goodlad (2023) acknowledges but does not fully resolve: if the "desire to say" distinguishes human language use from machine generation, how do we account for human utterances produced without such desire? Recent scholarship offers a provocative counter-reading. Gunkel (2025) argues that LLMs *actualize* Derrida's concept of *différance*, generating meaningful content by calculating statistical differences across massive textual corpora. Rather than positioning LLMs as deficient in relation to human language use, Gunkel (2025) contends that LLMs computationally enact the very mechanisms Derrida theorized, challenging logocentrism, authorship, and the metaphysics of presence. This reframing suggests that the "humanities in the loop" might not be about defending a distinctively human capacity for meaning-making, but rather about recognizing how LLMs render visible the *always already* technological conditions of human language use.

4.3 The Political Economy of AI and Linguistic Inequality

A significant limitation of the essay is its relative inattention to the political economy of AI development. Goodlad (2024) acknowledges the "perilous concentration of power in a small handful of tech oligarchs," but the analysis remains primarily theoretical rather than materialist. The question of who owns the models, who profits from them, and who bears their costs is raised but not developed with the specificity that a full account would require.

This limitation becomes particularly salient when considering linguistic inequality in LLM development. Recent research has documented that LLMs are predominantly trained on high-resource, globally dominant languages, leading to poor performance for low-resource languages (Olusanya et al., 2025). An estimated 7,000 languages are spoken worldwide, yet low-resource languages such as Twi, Sinhala, and Quechua appear on less than 0.1 percent of websites (Klatte et al., 2024). AI models underperform for the 1.2 billion native speakers of low-resource languages, and even when systems can process queries in these languages, outputs tend to be slower, less culturally relevant, and insufficiently covered by model safeguards (Klatte et al., 2024). This disparity disproportionately impacts communities in the Global South, reinforcing linguistic inequality, cultural erasure, and lack of accessibility in critical domains like education and healthcare (Olusanya et al., 2025).

Table 2 contrasts the "language problem" as framed in high-resource versus low-resource contexts.

Table 2. *Comparison of AI language concerns across high-resource and low-resource contexts.*

Dimension	High-Resource Language Contexts (Goodlad's Focus)	Low-Resource Language Contexts
Primary concern	Models mimic human language without understanding	Models barely function at all
Key risk	Anthropomorphic projection; ELIZA effect	Linguistic erasure; accessibility gaps
Theoretical resource	<i>Vouloir-dire</i> ; <i>parole</i> over <i>langue</i>	Multilingual benchmarks; community-centered datasets
Political stakes	Tech oligarchy; regulatory	Digital colonialism;

	capture	epistemic exclusion
--	---------	---------------------

Table 2 contrasts AI language concerns across high-resource and low-resource contexts. Goodlad's (2023) framework centres on the former, where the primary concern is anthropomorphic projection; the ELIZA effect whereby users attribute understanding to systems that merely mimic language. In low-resource contexts, however, the problem is not excessive mimicry but near-total absence: AI models underperform for the 1.2 billion native speakers of low-resource languages, and even when systems process queries in these languages, outputs are slower, less culturally relevant, and insufficiently covered by safeguards . Theoretical resources diverge accordingly: *vouloir-dire* and *parole* for humanist critique versus multilingual benchmarks and community-centred datasets for technical remediation . Politically, Goodlad addresses tech oligarchy and regulatory capture, while low-resource contexts confront digital colonialism and epistemic exclusion rooted in colonial-era linguistic hierarchies .

The dynamics of AI and language look quite different in low-resource language contexts, where the problem is not that models mimic human language too well but that they barely function at all (Klatte et al., 2024). Goodlad's (2023) framework is portable to these contexts, but the essay itself does not make the translation.

4.4 Revisiting the Binary: A Deconstructive Third Path

Recent scholarship offers a counter-reading that moves beyond the binary of techno-reductionism and humanist essentialism. Gunkel (2025) proposes that LLMs can be understood as "*différance* engines" that computationally enact the very mechanisms Derrida theorized. This reframing resists both the claim that "we are all just stochastic parrots now" and the counter-claim that "we are not parrots, because we know what words mean," charting instead a third path that deconstructs the metaphysical terms of the debate itself (Gunkel, 2025).

Figure 2 diagrams this counter-reading and its implications for the human-machine binary.

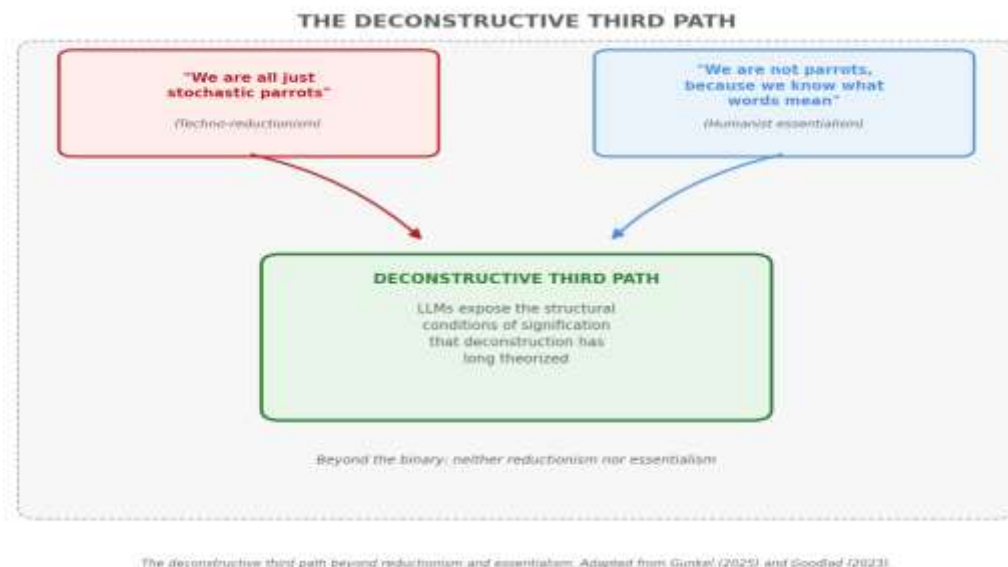


Figure 2. The deconstructive third path beyond reductionism and essentialism. Adapted from Gunkel (2025) and Goodlad (2023).

Figure 2 diagrams the "deconstructive third path" that Goodlad's (2023) framework invites but does not fully articulate. The two upper boxes represent opposing poles: techno-reductionism ("we are all just stochastic parrots") and humanist essentialism ("we are not parrots, because we know what words mean"). Both positions, as Bender et al. (2021) and Goodlad

(2023) respectively illustrate, reify a binary that deconstruction refuses. The central box proposes instead that LLMs expose the structural conditions of signification that deconstruction has long theorized, an argument consistent with Gunkel's (2025) account of LLMs as "*différance* engines" that computationally enact Derrida's (1982) play of differences. The figure's converging arrows visualize the move beyond the binary: neither reductionism nor essentialism, but a reframing of the human-machine relation as one of degree rather than kind.

This counter-reading does not invalidate Goodlad's (2023) argument but complicates its terms. It suggests that the "humanities in the loop" might not be about defending a distinctively human capacity for meaning-making against machine mimicry, but rather about recognizing how LLMs render visible the *always already* technological conditions of human language use (Gunkel, 2025).

4.5 Discussion

The findings of this review position Goodlad's (2023) essay at the centre of an emerging critical consensus: that the "human in the loop" framework, as currently operationalized, is inadequate to the task of ensuring AI. The essay's intervention, shifting from *langue* to *parole*, from system to enacted use, responds directly to this inadequacy by insisting that humanistic judgment must be constitutive rather than supplementary.

The stakes of this argument become clearer when read alongside recent scholarship on human oversight failures. As Annett and Giordano (2026) demonstrate, human involvement alone does not ensure accountability; rather, accountability depends on how decision authority shifts as system autonomy increases. Their analysis identifies a critical structural vulnerability: at intermediate levels of autonomy, human influence is reduced at the moment it matters most, as decision space is shaped before humans engage. Goodlad's (2023) insistence that literary critics have "always been action theorists" and that this self-understanding must be recovered is precisely an argument about authority allocation, about who gets to define what counts as meaningful oversight and when that definition takes effect.

The political economy dimension of this critique has become increasingly urgent. Goodlad (2024) has warned of the "perilous concentration of power in a small handful of tech oligarchs, the most vocal of whom are outright authoritarians who seek to buy elections, forestall regulation, control the media, and dominate the world economy". This is not rhetorical excess; it identifies the material conditions under which the "humanities in the loop" must operate. The question of who profits from AI, who bears its costs, and who is excluded from its benefits is not separable from the epistemic questions Goodlad raises about signification and meaning-making. The 1.2 billion speakers of low-resource languages who are poorly served by current models are not merely victims of technical limitation; they are victims of a political economy that treats linguistic diversity as an afterthought.

A counter-reading complicates Goodlad's framework without invalidating it. Gunkel's (2025) account of LLMs as "*différance* engines" suggests that the deconstructive tradition might be more productively deployed not to defend a distinctively human capacity for meaning-making, but to recognize how LLMs render visible the *always already* technological conditions of human language use. This reframing does not abandon the "desire to say"; it situates that desire within a broader field of signification that is itself structured by difference, deferral, and trace. The "humanities in the loop" might therefore be less about policing the boundary between human and machine and more about attending to what those boundaries obscure.

The essay's most significant limitation; its relative silence on the institutional and pedagogical arrangements through which "humanities in the loop" would be realized is also its most generative provocation. The questions Goodlad (2023) raises remain unanswered, but they have helped to constitute a field. In a moment when the pressure to either celebrate or denounce AI is immense, the essay's refusal of both technodeterminist despair and techno-utopianism remains its most valuable contribution.

Limitations

The most significant limitation of the essay is one of scope: as an editor's introduction, it cannot possibly deliver the sustained argument that its theoretical claims demand. The essay gestures toward a more substantial project—a full-length treatment of *vouloir-dire* in the age of large language models—but does not itself constitute that project. Readers looking for detailed engagement with specific AI systems, empirical studies of human-AI collaboration, or concrete proposals for pedagogical practice will need to look elsewhere, including to the other contributions in the inaugural issue that this introduction frames (Goodlad, 2023).

A second limitation is the essay's relative inattention to the political economy of AI development. Goodlad (2024) acknowledges the "perilous concentration of power in a small handful of tech oligarchs," but the analysis remains primarily theoretical rather than materialist. The question of who owns the models, who profits from them, and who bears their costs is raised but not developed with the specificity that a full account would require.

A third limitation acknowledged implicitly rather than explicitly, is the essay's focus on English-language literary criticism. The dynamics of AI and language look quite different in low-resource language contexts, where the problem is not that models mimic human language too well but that they barely function at all (Klatte et al., 2024; Olusanya et al., 2025). Goodlad's (2023) framework is portable to these contexts, but the essay itself does not make the translation.

V. Conclusions

The editor's introduction to Critical AI makes a compelling case for the constitutive participation of the interpretive humanities in artificial intelligence research, design, and governance. The essay's central contribution lies in its shift from *langue* to *parole*, from language as system to language as enacted use, which reframes the human-machine relation around the irreducibility of *vouloir-dire*, the "desire to say," to statistical prediction. This reframing refuses both the techno-reductionism of "we are all just stochastic parrots" and the humanist essentialism of "we are not parrots, because we know what words mean," charting instead a third path that recognises how LLMs expose the structural conditions of signification that deconstruction has long theorized.

Three recommendations follow from this review. First, critical AI studies must develop concrete institutional arrangements through which "humanities in the loop" can be realized. The essay's most generative provocation remains unanswered: what pedagogical interventions, what forms of collaboration between literary critics and computer scientists, would actually constitute the loop rather than merely talking about it? Second, future scholarship must attend to the political economy of AI development with greater materialist specificity. The warning about "tech oligarchs" who "seek to buy elections, forestall regulation, control the media, and dominate the world economy" identifies the conditions under which humanistic critique must operate, but the analysis requires development beyond the theoretical register. Third, the framework must be

extended to low-resource language contexts, where the problem is not excessive mimicry but near-total absence. The 1.2 billion speakers of low-resource languages who are poorly served by current models represent a test case for whether "humanities in the loop" can address linguistic justice or merely reproduce the inequities it critiques.

In a moment when the pressure to either celebrate or denounce AI is immense has carved out a space for patient, theoretically informed critique that the humanities at their best can provide. The questions she raises remain the essay's most valuable contribution.

References

- Annett, E., & Giordano, J. (2026). Losing the loop: A model for human operational involvement. Institute for National Strategic Studies, National Defense University. <https://inss.ndu.edu/news/Article/4471590/losing-the-loop-a-model-for-human-operational-involvement-part-two-a-proposed-s/>
- Architecting Human Oversight. (2026). [Human oversight as architectural concern]. IEEE Xplore. <https://doi.org/10.1109/ICSE.2026.11641208>
- 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>
- European Parliament and Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- 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). MIT Press.
- Goodlad, L. M. E. (2023). Editor's introduction: Humanities in the loop. *Critical AI*, 1(1–2). <https://doi.org/10.1215/2834703X-10734016>
- Goodlad, L. M. E. (2024, February 21). Q&A with Lauren M. E. Goodlad, editor of *Critical AI*. University of California Press. <https://www.ucpress.edu/blog-posts/humanities-in-the-loop-lauren-m-e-goodlad-on-ai-and-language-in-the-era-of-large-language-models>
- Gunkel, D. J. (2025). The différence engine: Large language models and poststructuralism. *AI & Society*, 41(3), 1991. <https://doi.org/10.1007/s00146-025-02640-z>
- Halstyan, V., & Bednarskyi, S. (2025). Visual hermeneutics in the digital age. *PhilPapers*. <https://doi.org/10.26565/2306-6687-2025-71-5>
- Hsu, S. F., & Shidujaman, M. (2026). From close reading to distant reading: Generative AI and the construction of a new methodology for literary analysis. In *Human-Computer Interaction* (pp. 277–294). Springer. https://doi.org/10.1007/978-3-032-29586-6_19
- Johnson-Laird, P. N. (1976). Slaves to the machine [Review of the book *Computer power and human reason: From judgment to calculation*, by J. Weizenbaum]. *Nature*, *261*, 171. <https://doi.org/10.1038/261171a0>
- Katz, S. B. (2023). [Review of *Critical AI*, by L. M. E. Goodlad]. *Double Helix*, 11, 10–12. <https://doi.org/10.2139/ssrn.4646522>
- Klatte, A., Hradecky, A., Bromová, B., & Sowah, N. L. (2024, October 4). Every language matters: Building a more inclusive digital future. United Nations Development Programme. <https://www.undp.org/digital-innovation/blog/every-language-matters-building-more-inclusive-digital-future>

- Lee, T. K. (2025). World literature and the double imperatives of digital technology and neoliberal capitalism. *Literature Compass*, 22(4), e12759. <https://doi.org/10.1111/lic3.12759>
- Nominal Human Control. (2026). [Formal authority and genuine oversight in AI governance]. arXiv. <https://doi.org/10.48550/arXiv.2604.00081>
- Olusanya, J., Salami, M., Amol, C., Ibejih, S., Audu, O., Oladipupo, F., & Le Beux, Y. (2025). Centering low-resource languages and cultures in the age of large language models. NeurIPS 2025 Workshop.
- Roy, D. (2025). Humanistic inquiry and critical AI research: Contexts, convergences & connections. Christ University Lavasa. <https://christuniversitylavasa.blogspot.com/2025/12/humanistic-inquiry-and-critical-ai.html>
- Saussure, F. de. (2011). *Course in general linguistics* (W. Baskin, Trans.). Columbia University Press. (Original work published 1916)
- Tiwari, D., Arora, P. P., Dubey, A. T., Geetey, S., Pachlaniya, Y., Dubey, R., Gour, A., & Ansari, A. (2026). Reading under algorithmic regimes: Attention economies, cultural intermediation, and contemporary literary value. *Cultura Journal*. <https://doi.org/10.5281/cultura.2026.2402>
- Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226612973.001.0001>
- Underwood, T., & So, R. J. (2021). Can we map culture? *Journal of Cultural Analytics*, 6(3), 32–51. <https://doi.org/10.22148/001c.24946>
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.
- Yadav, S. (2025). From pen to processor: The intersection of AI and literary innovation. *International Journal of English Literature and Social Sciences*, 10(3), 303–317. <https://doi.org/10.22161/ijels.103.47>