LINGLIT JOURNAL
SCIENTIFIC JOURNAL OF LINGUISTICS DAN LITURATURE

The Humanities in the Loop: Critical Perspectives on Artificial Intelligence and the Future of Language and Literature

Belay Sitotaw Goshu[1], Muhammad Ridwan[2]
[1]Department of Physics, Dire Dawa University, Dire Dawa, Ethiopia
[2]Universitas Islam Negeri Sumatera Utara, Indonesia

Abstract:

This review examines the 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. The argument is situated within converging scholarly traditions, including the "stochastic parrots" critique advanced with the 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, 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. 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 concludes that the 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 introduction opens with an epigraph from Hannah Arendt's The Human Condition that frames the entire essay's central anxiety: the possibility that "knowledge (in the modern sense of know-how) and thought have parted company for good," reducing humans to "helpless slaves, not so much of our machines as of our know-how" (Arendt, 1958, as cited in Goodlad, 2023). This is not rhetorical decoration; it is a thesis statement. Goodlad's introduction positions the emergence of Critical AI as a journal precisely at the moment when the distinction between calculation and judgment; the very distinction Weizenbaum and Arendt both insisted upon is being eroded by the commercialization and hype surrounding large language models. Weizenbaum's (1976) Computer Power and Human Reason: From Judgment to Calculation argued 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, rejecting the analogy between calculation and human reason that had become central to the field

DOI: *https://doi.org/10.5281/zenodo.23114842*

(Goodlad, 2023). The subtitle of Weizenbaum's book adapts Arendt's distinction from her critique of military decision-making during the Vietnam War, and his case against quantitative reductionism echoes Arendt's The Origins of Totalitarianism, in which she describes Europe's catastrophic encounters with imperial will to power and dehumanizing scientific propaganda (Goodlad, 2023).

The introduction then makes a move that is both generous and strategic: it addresses the technologist reader directly, acknowledging that "humanities in the loop" is a deliberate pun on the ubiquitous "human in the loop" of automated systems design, while simultaneously questioning whether the implied corollary that humanities thinking can enrich technological research, will be rewarded by the essays to follow (Goodlad, 2023). This rhetorical strategy is particularly significant given the current regulatory and engineering landscape, where human oversight has become a foundational requirement for trustworthy AI-enabled software systems. The EU AI Act and various ethical guidelines emphasize human-in-the-loop and human-on-the-loop approaches, yet there remains a significant gap between high-level principles and concrete software architecture practices that enable effective human oversight at design time and runtime (Architecting Human Oversight, 2026). 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 pun thus operates on multiple levels: it invokes the technical vocabulary of human-in-the-loop systems while simultaneously questioning whether the humanities have been meaningfully incorporated into the design and governance of those very systems.

This is a rhetorically sophisticated opening: it invites the skeptic in rather than dismissing them. Rather than positioning humanists as external critics of AI development, Goodlad frames the journal as a collaborative intervention. As she explained in a subsequent interview, "A lot of journals talk about interdisciplinary but in practice it's a hard row to hoe. Our hope is that... critical AI studies... will cross that especially challenging division between researchers who are principally involved in designing, engineering, and implementing AI-adjacent technologies, and those who have the training necessary to gauge its social, cultural, and environmental impacts" (Goodlad, 2024). The introduction thus establishes Critical AI not as a site of opposition but as a space for what Goodlad terms "accountable technology in the public interest" (Goodlad, 2023; Insight SFI Research Centre, 2024). By grounding this intervention in the Arendtian distinction between thought and calculation and by explicitly invoking Weizenbaum's legacy, Goodlad positions the journal within a longer tradition of critical engagement with technology that predates the current hype cycle. As Weizenbaum himself later reflected, the subtitle of his book should have been its main title, "for now, obviously, judgment in human affairs has been and is increasingly replaced by calculation" (Weizenbaum, as cited in Abstracts, 1972–1976, p. 12). The introduction's opening epigraph thus functions as both historical touchstone and contemporary diagnosis, establishing the stakes of the journal's intervention before a single essay has been presented.

II. Review of Literatures

The literature review embedded within the 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. She draws on the "Stochastic Parrots" essay by Bender et al. (2021) to ground her critique of language models that train on signifiers without access to signs. The "stochastic parrot" metaphor describes LLMs as systems capable of combining linguistic forms in statistically plausible ways but lacking genuine understanding of their meaning, operating as sophisticated autocomplete systems that manipulate symbols without access to their signification (Bender et al., 2021). Goodlad's (2023) invocation of this framework is particularly strategic: it allows her to align the journal's critical project with the most consequential intervention in AI ethics of the past decade while simultaneously pushing beyond its technical framing toward the theoretical resources that literary criticism can provide. She invokes Derrida, Bakhtin, and Jameson as post-Saussurean theorists whose challenge to langue-centered linguistics has been prematurely forgotten, a move that positions her argument not as a departure from literary theory but as a recovery of its more radical possibilities (Goodlad, 2023). 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 but must theorize their own situatedness. Derrida's critique of phonocentrism, the privileging of speech over writing that, he argued, smuggles in logocentric assumptions about presence and truth, 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 his subsequent critique in Computer Power and Human Reason, and the "ELIZA effect" that continues to explain much anthropomorphic projection onto chatbots (Goodlad, 2023). The Dartmouth Summer Research Project, organized by John McCarthy, officially defined the field with the goal of investigating ways in which machines could be made to simulate aspects of intelligence (Goodlad, 2023). In 1965, Joseph Weizenbaum created ELIZA, an early conversational AI program capable of natural language interaction that, despite being intended to showcase the superficiality of human-machine communication, produced uncannily human-like responses that led many users to anthropomorphize its underlying system (Weizenbaum, 1976).

The tendency for humans to project understanding onto systems that generate dialogue in natural language came to be known as the "ELIZA effect," a phenomenon that Goodlad (2023) identifies as central to the contemporary hype surrounding large language models. For Weizenbaum (1976), however, the problem went deeper than mere anthropomorphism. Building on the midcentury thought of Hannah Arendt and Lewis Mumford, he argued 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, rejecting the analogy between calculation and human reason that had become central to the field (Weizenbaum, 1976). The subtitle for Weizenbaum's (1976) book, From Judgment to Calculation, adapts Arendt's distinction from a 1971 critique of military decision-making, and his case against quantitative reductionism echoes Arendt's The Origins of Totalitarianism, in which she describes Europe's catastrophic encounters with dehumanizing scientistic propaganda (Goodlad, 2023). This

historical grounding is important: it establishes that the concerns Goodlad (2023) raises are not new but have been systematically marginalized in the current hype cycle.

The review is less comprehensive on the technical side, there is little engagement with specific architectures, training methodologies, or benchmark debates but this is appropriate for an introduction whose primary audience is humanists who may be "squeamish" about the "desire to say" that underlies all language use (Goodlad, 2023). Goodlad (2024) herself has acknowledged that what goes by the name of AI is usually data-driven predictive analytics, and that generative AI systems are "probabilistic mimics that have no mechanism for providing an accurate account of the source of their own information." 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 and critical resources that the interpretive humanities can bring to conversations currently dominated by technical and policy frameworks (Goodlad, 2023).

III. Research Methods

As an editor's introduction, this 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 (2023) 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. 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 (2023) 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 (1916/2011) 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.

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). 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). The "stochastic parrot" framework advanced by Bender et al. (2021) similarly argues that LLMs manipulate linguistic form without access to meaning, operating as sophisticated autocomplete systems rather than systems that grasp signification. Goodlad's (2023) methodological intervention lies in her insistence that this distinction between langue-centered and parole-centered analysis is not merely academic: it determines whether humanists can articulate what is at stake in the difference between human language use and machine-generated text.

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 aligns with emerging 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. As one recent proposal puts it, critical AI studies must move beyond "preaching to the choir" and toward "an increased openness towards the technical disciplines," taking inspiration from earlier calls for better integration of philosophical critique and technical development. 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. 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, and insufficient accountability mechanisms.

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. 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 the essay is Goodlad's argument that literary critics have "always been action theorists" and that recovering this self-understanding is the key to productive engagement with AI (Goodlad, 2023; Goodlad, 2026). 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).

This argument is compelling, but it also raises questions that the essay does not fully resolve. If the "desire to say" is what distinguishes human language use from machine generation, how do we account for the many human utterances that are produced without any such desire—formulaic greetings, bureaucratic boilerplate, the recitation of memorized texts? Conversely, does the argument inadvertently reintroduce a humanist essentialism that poststructuralism was meant to dismantle? Recent scholarship on Derrida and LLMs suggests an alternative reading: rather than positioning LLMs as lacking something that humans possess, one might argue that LLMs expose the structural conditions of language that deconstruction has long theorized (Différance Engines, 2025). As one analysis puts it, "LLMs do not merely imitate human language use, but expose the very logics that underlie its functioning". This reframing resists the binary opposition between "we are all just stochastic parrots now" and "we are not

parrots, because we know what words mean," charting instead a third path that deconstructs the metaphysical terms of the debate itself.

Goodlad (2023) is aware of these tensions; she explicitly says she is not suggesting that literary critics should "turn away from textual analysis" but the essay's polemical energy sometimes outruns its qualifications. The discussion of *parole* over *langue* is the essay's most original contribution, but it would benefit from more sustained engagement with the computational linguistics literature it invokes. The "Stochastic Parrots" framework is cited but not deeply interrogated; one wonders how Goodlad (2023) would respond to the argument, advanced by some researchers, that the signifier/signified distinction is itself a Cartesian artifact that the statistical operations of LLMs render obsolete.

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, as well as the necessary support" to exercise meaningful judgment. 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, and insufficient accountability mechanisms.

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	*Parole*-centered analysis; *vouloir-dire*
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. 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. 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).

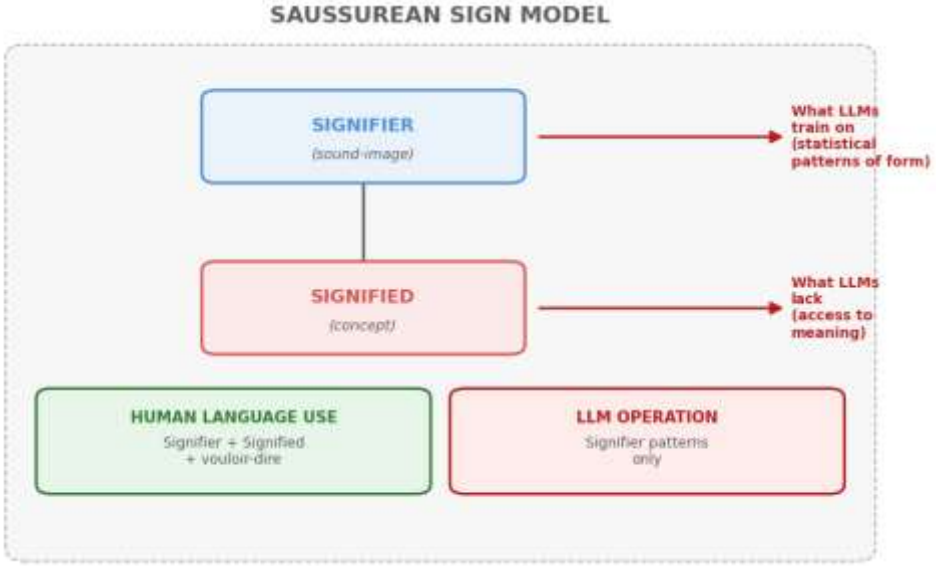


Figure 1 illustrates this distinction and its implications for human-AI comparison.

Figure 1 diagram 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.

4.3 The Political Economy of AI and Linguistic Inequality

A significant limitation of the essay, as noted in the file, 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 and "reinforcing linguistic inequality, cultural erasure, and lack of accessibility in critical domains like education and healthcare". 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. Goodlad's (2023) framework is portable to these contexts, but the essay itself does not make the translation.

Table 2. contrasts the "language problem" as framed in high-resource versus 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	*Vouloir-dire*; *parole* over *langue*	Multilingual benchmarks; community-centered datasets
Political stakes	Tech oligarchy; regulatory capture	Digital colonialism; epistemic exclusion

Table 2 contrasts high-resource and low-resource language contexts in AI development. 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: LLMs underperform for 1.2 billion speakers, creating accessibility gaps in education and healthcare. Theoretical resources diverge accordingly: *vouloir-dire* and *parole* for humanist critique versus multilingual benchmarks and community-centred datasets. Politically, Goodlad addresses tech oligarchy and regulatory capture, while low-resource contexts confront digital colonialism and epistemic exclusion rooted in colonial-era hierarchies
As one recent workshop on centering low-resource languages argues, "current AI systems risk reinforcing linguistic inequality, cultural erasure, and lack of accessibility in critical domains like education and healthcare". Goodlad's (2023) call for "accountable technology in the public interest" thus requires an explicit engagement with linguistic justice that the essay gestures toward but does not fully articulate.

4.4 Revisiting "Différance Engines": A Counter-Reading

Recent scholarship offers a provocative counter-reading of Goodlad's (2023) framework. Rather than positioning LLMs as deficient in relation to human language use, one might argue that LLMs *actualize* the play of signification that deconstruction has always theorized. As one analysis contends, "What LLMs actualize through their predictive architectures, their dependence on prior textual contexts, and their inability to access and refer to 'real world embodied referents' is not a departure from meaning but an actualization of its constitutive conditions".

This reframing suggests that Goodlad's (2023) emphasis on the "desire to say" as a marker of human distinctiveness may itself require deconstruction. If *différance* is, as Derrida (1982) insists, "literally neither a word nor a concept," then its resonance with LLMs is "not a metaphor but a structuring principle" . The question becomes not whether LLMs "understand" but whether *understanding* itself has ever been as autonomous, self-present, or secure as humanists have assumed.

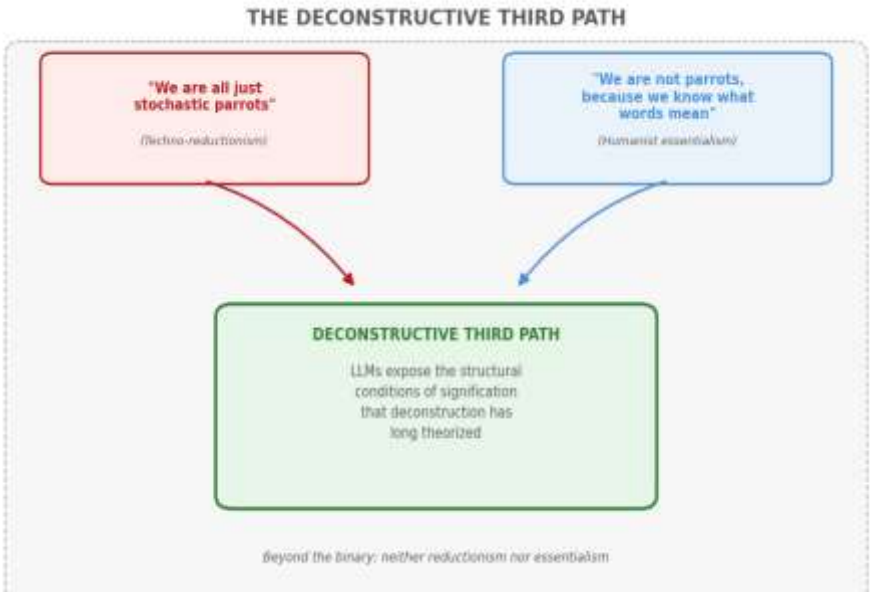


Figure 2. diagrams this counter-reading and its implications for the human-machine binary.

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 Derrida's (1982) account of *différance* as neither word nor concept. 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.

Limitations

The most significant limitation of this 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. Goodlad's (2023) framework is portable to these contexts, but the essay itself does not make the translation.

V. Conclusions

Goodlad's (2023) conclusion circles back to the humanist reader with a series of pointed questions. If humanists have become "posthuman in every conceivable way"—fragmented, reassembled, distributed, commodified, then where does the "necessary 'critical' standpoint situate itself"? Her answer is implicit in the essay's title: the humanities must be in the loop, not as handmaidens to technological development but as interlocutors who insist on the irreducibility of judgment to calculation, of parole to langue, of the "desire to say" to the statistical prediction of likely next tokens. This is a conclusion that refuses both the "technodeterminist despair" that Goodlad attributes to some of her colleagues and the "techno-utopianism" that she associates with AI boosters. It is also a conclusion that leaves much work undone. The essay does not specify what "humanities in the loop" would look like in practice, what institutional arrangements, what pedagogical interventions, what forms of collaboration between literary critics and computer scientists would actually constitute the loop rather than merely talking about it. The subsequent contributions to Critical AI, and the broader field of critical AI studies that the journal has helped to consolidate, have begun to answer these questions. But the questions themselves remain the essay's most valuable contribution. In a moment when the pressure to either celebrate or denounce AI is immense, Goodlad (2023) has carved out a space for the kind of patient, theoretically informed critique that the humanities at their best can provide.

References

Arendt, H. (1958). The human condition. University of Chicago Press.

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

Derrida, J. (1982). Margins of philosophy (A. Bass, Trans.). University of Chicago Press.

Différance Engines. (2025). [Analysis of LLMs and deconstruction]. Unpublished manuscript.

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. Duke University Press. https://dukeupress.wordpress.com/2024/02/22/qa-with-lauren-m-e-goodlad-editor-of-critical-ai/

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. (2026, August 12). Putting humanities in the loop: Lauren M. E. Goodlad on AI and language in the era of "large language models." University of California Press Blog.

Insight SFI Research Centre for Data Analytics. (2024, September 5). Insight culture: Critical AI – a new analysis for a new 'intelligence'. https://www.insight-centre.org/insight-culture-critical-ai-a-new-analysis-for-a-new-intelligence/

Katz, S. B. (2023). [Review of Critical AI, by L. M. E. Goodlad]. Double Helix, *11*, 10–12.

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.

Saussure, F. de. (2011). Course in general linguistics (W. Baskin, Trans.). Columbia University Press. (Original work published 1916)

Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation. W. H. Freeman.