INNOVATION IN LANGUAGE LEARNING International Conference

Generative AI as a Mediated Discursive Operator: A Reporting Framework for Critical Discourse Studies

Ruqayyah N. Moafa

Department of Foreign Languages, College of Arts and Humanities, Jazan University, Saudi Arabia

Abstract

Critical discourse studies read texts as sites of power, and its apparatus presupposes an author whose position can be located. Generative language models remove that presupposition while leaving the political stakes intact. Two responses dominate the emerging literature, and both fall short: reading the model as a speaker mistakes distributional regularity for intention, while reading it as a tool cannot register training-data curation, alignment, safety filtering, ownership, or scale as discursive acts. Neither yields a rule for what an analyst may infer from output alone. This paper develops the Mediated Discursive Operator (MDO) as an analytical construct naming a relational position within a production chain: a statistical engine whose outputs enter and reshape discourse at scale on terms set upstream, with causal efficacy but without intention, belief, or authorship. Three corollaries are derived, each with scope conditions and a disconfirmation condition: agency is distributed and structurally unequal; bias claims are scope-bound but can support structural critique through convergence; and critique must follow the value chain. These are operationalized as the CDA-GenAI Critical Agenda, a reporting framework rather than a validated methodology. Arabic is proposed as a critical test case. The paper is conceptual, and a program for empirical validation is specified.

Keywords: *algorithmic mediation; generative AI; critical discourse studies; discursive agency; large language models; political economy; research transparency*

1. Introduction

Critical discourse studies has a well-developed apparatus for reading texts as sites of power, and that apparatus presupposes an author whose position can be located [1], [2]. Generative language models remove that presupposition while leaving the political stakes intact: a growing share of professional, educational, and public text now has no author, no locatable intentionality, and no disclosed provenance. The field's two available responses are both inadequate. Reading the model as a speaker mistakes distributional regularity for intention and grants autonomous speech to a corporate product; reading it as a tool cannot register training-data curation, alignment, safety filtering, ownership, or scale as discursive acts. Neither yields a rule for what an analyst may infer from output alone. Gillings et al. [3] set out the resulting challenges for critical discourse studies across three fronts: the status of machine-generated text as data, the dissolution of traceable authorship, and linguistic homogenization. That diagnosis is the starting point here, and what it leaves open is the constructive step: a named position for the system and a rule governing what may be inferred from its output. This paper proposes the Mediated Discursive Operator (MDO) as an analytical construct naming a relational position within a production chain, derives three corollaries, and operationalizes them as the CDA-GenAI Critical Agenda (CGCA), a reporting framework for authors, reviewers, and supervisors. What it makes possible is specific: locating a bias finding at a named stage of production, separating output-level evidence from upstream attribution by an explicit rule of evidence, and stating which form of agency a claim concerns. The paper is conceptual, and conceptual plausibility is not empirical validation.

The paper is addressed to three groups: researchers reporting analyses of generative-AI discourse, reviewers and editors assessing such work, and supervisors advising on design. The MDO is not original at the level of its component insights. Distributed agency comes from actor-network theory and agential realism [4], [5]; value-chain analysis from platform and infrastructure studies [6], [7]; scope-bound bias from critical AI ethics [8], [9]; and the extraction critique from data colonialism [10]; the account of AI as a material and political formation comes from critical AI studies [11], and the catalog of language-model harms from Weidinger et al. [12]. What is offered is their integration into a construct with derived propositions, an explicit rule of evidence, and stated conditions of failure.

2. Conceptual Approach and Framework Construction

This is a theoretical paper, and its reasoning should be auditable. The problem was identified from a mismatch: the emerging critical literature on generative AI produces bias findings and anthropomorphism critiques without a shared account of what an analyst may infer from output alone. Competing perspectives were selected by their standing as the two positions taken in that literature (the model as speaker and the model as tool), and compared on a single criterion: whether each can register the production chain as discursive. The construct was developed by asking what a third position must contain to satisfy both the anti-anthropomorphism commitment and the political-economic one. The corollaries were derived by asking what follows for agency, for bias, and for the scope of critique if the thesis holds; each is stated with a disconfirmation condition (Table 2). Boundaries were established by seeking cases the construct should exclude, and empirical implications by asking, for each corollary, what evidence a study could collect. The procedure is reconstructed rather than followed prospectively, as is normal for conceptual work. Conceptual plausibility is not empirical validation.

3. Two Misframings of the Discursive Object

3.1 The Speaking-Subject Misframing

A first family of approaches treats the model as a speaker. On this view the system has values, a voice, and in one recent study an emergent self-identity [13]; this literature asks what values a model's outputs are consistent with, a claim about output regularities that the register then describes in belief vocabulary [14], [15]. The framing has a genuine analytical advantage: it takes seriously that outputs carry evaluative stance and that users respond to them as they would to speech. The anthropomorphic inference occurs at a specific point, however. Grammatical first-person performance is a property of the text; intention, belief, and authorship are properties of agents, and no inference runs from the first to the second. The political consequence of eliding this is substantial: treating output as autonomous speech grants a corporate product the standing of a participant and removes from view the parties whose decisions shaped what it says [16], [17]. Pinell's [18] Gadamerian assessment reaches a compatible conclusion by a different route, arguing that generative output lacks the interpretive horizon that would make it language in the full sense.

3.2 The Passive-Tool Misframing

A second family treats the model as an instrument and asks whether it performs a task as well as a human analyst would [19], [20], [21], a question also pursued in annotation research where model and human labels are compared directly [22]. Its valid insight is the refusal of anthropomorphism: the tool framing correctly denies the system intention and belief. What instrumentality cannot register is everything that makes the output what it is. Training-data selection, architecture and optimization, alignment and safety filtering, deployment interfaces, ownership, and scale are not features of a tool's use but decisions taken elsewhere, and each shape what can be said. The tool metaphor conceals institutional power precisely by locating the analytically interesting question at the interface, where the user is the only agent in view [23], [24].

3.3 What the Two Framings Share

Both treat the system as an analytical endpoint. The speaking-subject framing makes that endpoint a will; the tool framing makes it an instrument. Neither asks what conditions produced the output, which is the question a critical account requires. The MDO is not a compromise between them: it rejects the shared premise, and its analytical payoff lies in what follows from doing so.

4. The Mediated Discursive Operator

Consider a newsreader on a 24-hour channel reporting that markets opened lower. The audience takes the sentence neither as her personal view nor as the output of a neutral instrument, but as a corporately

mediated speech act: a human voice carries it, while wording, framing, and timing have been shaped by the channel's institutional position. She is not a conduit, since she selects emphasis and controls pacing, yet her room for autonomous discursive action is bounded by editorial protocol. The analogy has an obvious limit, in that she has intentions and a generative model does not; what it isolates is a discursive event whose production conditions are neither subjectivist nor instrumentalist. Suchman's [25] account of human-machine reconfiguration makes the general point that the boundary between agent and apparatus is an outcome of arrangements rather than a given.

The MDO thesis. A generative language model, considered in deployment, occupies the position of a Mediated Discursive Operator: a statistical engine whose outputs enter and reshape human discourse at scale on terms set upstream, by training-data selection, architectural and optimization choices, alignment and safety regimes, and deployment policy. 'Operator' denotes selection, transformation, and scaling, not volition; the system has causal efficacy without intention, belief, or authorship. The MDO is an analytical construct, not a class of software: what it names is a relation, and the same weights occupy different MDO positions under different deployment conditions. Its unit of analysis is the situated output event taken together with its production conditions; its level is the relation between a deployed system, an internally differentiated corporate node, and an uptake context. Three features are necessary: outputs are generated rather than retrieved, the production chain is controlled by parties absent from the interaction, and the system operates at a scale that makes its regularities consequential for discourse. Where any is absent, as in a locally run open-weights model with documented provenance or a system whose outputs do not circulate, the construct does not apply, and saying so is what makes it falsifiable.

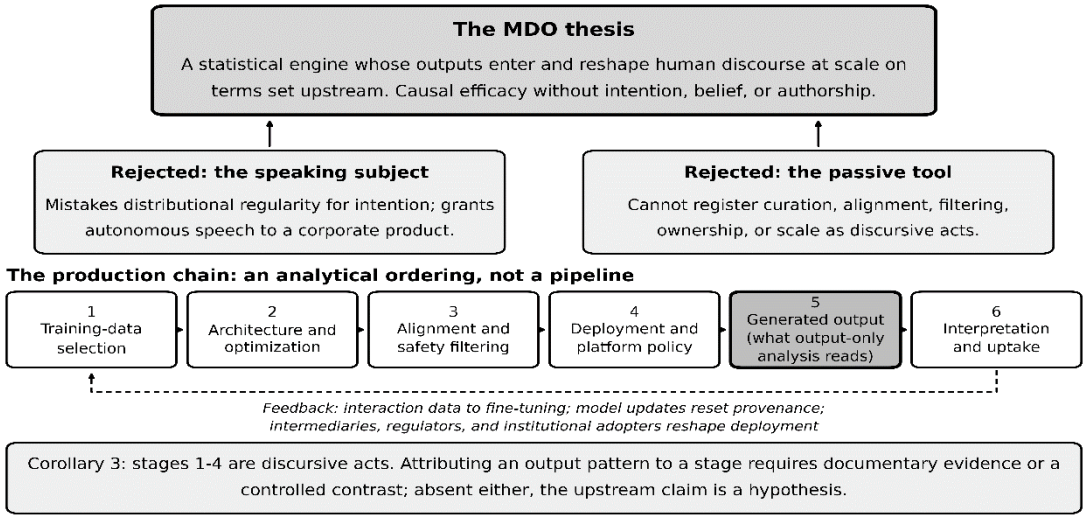


Fig. 1. The MDO position and the production chain. The construct is defined against two rejected framings (upper panels). The chain below is an analytical ordering rather than a causal pipeline: the dashed return path marks feedback from uptake to fine-tuning, model updates that reset provenance, and reshaping by intermediaries, regulators, and institutional adopters. Stage 5 is what output-only analysis reads. Conceptual schematic; no empirical findings are represented.

Because 'agency' is the term most often asked to carry the whole argument, the paper distinguishes seven senses and uses them separately (Table 1). The distinction that matters most is between the system's causal efficacy, in that outputs have effects no participant authored, and the institutional power exercised through design decisions users cannot negotiate. Moral responsibility and legal accountability attach to human and institutional actors, never to a statistical process, and the two can diverge. Corollary 1 is a claim about relations among these senses, not about agency in general.

Table 1. Seven senses of agency distinguished in the MDO framework.

Sense	Held by	Analytical status in the MDO
Linguistic agency	Users	Choice among forms in composing prompts and editing output

Sense	Held by	Analytical status in the MDO
Interactional agency	Users	Selecting, accepting, refusing, recontextualizing output
Causal efficacy	The system	Effects without intention, belief, or authorship
Institutional power	Corporate and deploying actors	Design and deployment decisions users cannot negotiate
Moral responsibility	Human and institutional actors	Never the system; requires separate normative argument
Legal accountability	Human and institutional actors	May diverge from moral responsibility across jurisdictions
Attributed agency	Ascribed by users and texts	An object of analysis, not a property of the system

4.1 Position in the Conceptual Landscape

The MDO stands in a specific relation to several adjacent traditions, and the differences are not merely terminological. Actor-network theory supplies distributed agency but treats actors as analytically equal pending investigation; the MDO weights the corporate node in advance and states the ground for doing so. Agential realism supplies material-discursive entanglement but dissolves the apparatus-phenomenon boundary the MDO needs to locate accountability at named stages. Platform studies supply corporate infrastructural power but analyzes platforms as intermediaries of user-generated content, whereas an MDO generates the text at issue, making provenance rather than circulation the analytical problem. Data colonialism supplies the extraction critique; the MDO adds the return leg, in which extracted distributions are re-imposed as output on the communities they were drawn from. Kera [26] reaches a convergent conclusion from a different direction, describing large language models as deixis machines that perform the structure of enunciation with no speaker to inhabit it; that account works at the level of address, whereas the MDO works at the level of production, and the two are complementary. Within critical discourse studies, Kopf [27] differentiates four ways of studying social media critically, and the MDO names a fifth analytical situation: content whose author-function is corporate and statistical. Ledin and Machin [28] argue that critical discourse studies must move from metafunctional description to materiality; the MDO extends that move to the conditions of production, which is the warrant Corollary 3 requires. Sociomaterial and agential accounts of responsibility in computing reach adjacent conclusions from an organizational starting point [29], [30]. A comparison across twenty adjacent frameworks, stating overlap, difference, and contribution for each, is available from the author on request.

4.2 The 'Social Actor' Framing

One adjacent proposal warrants sustained engagement. Hamdi [31] reads generative AI as a social actor through van Leeuwen's [32] framework, and Lakel [33] proposes the terms Large Discourse Model and Artificial Discursive Agent. These are not rivals to the MDO so much as its complement: they ask how generative AI is represented in discourse, whereas the MDO asks what kind of operator it is given its production conditions. A representational analysis shows how institutional texts position the system; a production-side analysis shows which upstream decisions the positioning conceals. KhosraviNik's [34] distinction between content and practice supplies the vocabulary for keeping the two levels apart.

5. Three Corollaries

Three propositions follow from the thesis. Each is stated with its derivation, scope conditions, observable implications, and a disconfirmation condition, so that the framework can fail (Table 2).

Table 2. The three corollaries as derived propositions.

	Corollary 1: asymmetric agency	Corollary 2: scope-bound bias	Corollary 3: value-chain critique
Proposition	Agency is real, distributed, and structurally unequal across the senses in Table 1.	Bias claims are warranted only under stated scope conditions; structural claims require convergence across them.	Output-only analysis does not by itself warrant claims about ideological production.
Derivation	If terms are set upstream by absent parties, agency cannot be located solely in the interaction.	If output is distribution-weighted, a finding indexes the conditions that produced it.	If upstream stages are discursive acts, the analytical object exceeds the output.
Scope condition	Deployments where users cannot negotiate configuration.	Claims naming a model, version, prompt regime, and language community.	Cases where production decisions are documented or contrastable.
Observable implication	Output regularities persist across varied user prompting within a fixed configuration.	Measured effects vary systematically across models, versions, and language varieties.	Documented upstream changes correspond to measurable shifts in output regularity.
Disconfirmation	Varying prompts within a configuration accounts for more of the variation in output regularity than varying the configuration does.	Bias findings show no systematic variation across models, prompt regimes, or language communities.	Documented version or policy changes are followed by no systematic shift in output regularity.
Research design	Prompt-controlled comparison across configurations, with prompts held constant.	Multilingual, multi-model comparison with effect sizes.	Longitudinal snapshots paired with policy and documentation analysis.
CDS contribution	Replaces a yes/no agency dispute with a mapping task.	Lets local evidence support structural critique without overgeneralizing.	Extends critique from text to the conditions of its production.

Two points require emphasis. First, asymmetry is not powerlessness: users retain linguistic and interactional agency, and Corollary 1 claims only that this agency operates within terms it cannot set. Second, scope-binding does not deny structural power. A single finding indexed to one model and one prompt regime cannot support a claim about 'AI bias' in general, but convergent findings across models, prompt regimes, and language communities do support structural claims, and historically accumulated inequality is what makes such convergence expected rather than surprising [35], [36], [37].

Corollary 3 raises the hardest methodological problem in the paper: how to relate a textual feature to an upstream decision without speculation. The framework's answer is a rule of evidence. A textual regularity, adequately sampled, is evidence of an output pattern and nothing more. Attributing that pattern to a stage of production requires either documentary evidence (a data statement, model card, alignment description, or content policy) or a controlled contrast in which the same prompt is run across versions, models, or language varieties. Where neither is available, the upstream claim is a hypothesis and must be labeled one. This is a constraint rather than a solution: proprietary opacity means many upstream claims will remain hypotheses, which is precisely the condition Axis A asks analysts to record rather than write around. Work on corporate safety discourse shows what documentary analysis at this level looks like [38], and Ahmed and Mahmood's [39] reading of ChatGPT indicates what an output-level analysis can and cannot reach.

Table 3. The production chain: actors, powers, and discursive consequences.

Stage	Principal actors	Type of power	Discursive consequence	Evidence available
1 Training-data selection	Data vendors, curators, licensors	Definitional: what counts as language	Sets the distribution from which output is sampled, including which varieties are thin or absent	Data statements, licensing disclosures, audits of open counterparts
2 Architecture and optimization	Researchers, engineers, compute providers	Formal: what counts as a good continuation	Tokenization and objective favor the modal register	Technical reports, tokenizer vocabularies, open-weights comparators

Stage	Principal actors	Type of power	Discursive consequence	Evidence available
3 Alignment and safety filtering	Annotation contractors, policy and red teams	Normative: what counts as acceptable speech	An evaluative regime sediments into stance, hedging, and refusals	Model cards, content policies, refusal probes across topics and languages
4 Deployment and platform policy	Product teams, API resellers, downstream developers, institutional adopters; regulators as constraint	Access and framing: who meets which system on what defaults	Same weights, different discourse; adoption converts a tool into a condition of participation	Interface and version documentation, procurement and institutional policy
5 Generated output	No author; a distribution-weighted composite	Effect, not decision	The textual regularities analysis can read	The texts, with prompts, version, date, and parameters recorded
6 Interpretation and uptake	Users, professional and educational communities	Situated: real, consequential, unequally resourced	Whether a pattern becomes a discursive fact	Reception study, interviews, institutional policy, community interpretation

6. The CDA-GenAI Critical Agenda

The MDO thesis is theoretical; to do analytical work it requires an operational form. The CGCA is a reporting framework and critical workflow, not a methodology and not a validated instrument. Each of its five axes operationalizes one corollary as a reporting requirement (Table 4, Figure 2; the axes are restated as an author-facing checklist in Appendix A). Remling's [40] operationalization of poststructuralist policy analysis is the in-journal precedent: a theoretical apparatus rendered auditable without a claim to empirical validation. Any analogy with consensus-developed reporting standards in clinical research concerns function, not epistemic status; those instruments were built through extensive consensus exercises, and the CGCA has undergone nothing equivalent.

Table 4. The five CGCA axes: research question, unit, admissible claim, and reproducibility requirement.

Axis	Research question	Unit of analysis	Admissible claim	Reproducibility requirement	Corollary operationalized
A Provenance disclosure	What is knowable about this system's production, and what is not?	Deployed version at a stated date	Descriptive; non-disclosure is itself a finding	Deposit the provenance record with the analysis	3
B Disaggregated agreement	Does agreement hold for interpretive judgments or only surface features?	The coded item, stratified by level	Interpretive claims only as far as interpretive-stratum agreement allows	Deposit codebook and per-stratum statistics	1
C Scope-bound bias	Under which conditions does this pattern hold?	Prompt-output pair indexed to model, version, variety	Bounded to conditions varied; structural attribution requires cited convergence	Deposit prompts, outputs, and the variation design	2
D Centering the periphery	Whose scholarship and varieties are treated as the default case?	The citation base and varieties analyzed	Claims about a community require engagement with its scholarship	Report the citation-base audit result	3
E Replicable materials	Could another researcher see what was seen?	The deposited materials set	Findings hold for deposited materials; snapshot-specific findings labeled	Deposit at submission; name withheld items and reasons	3

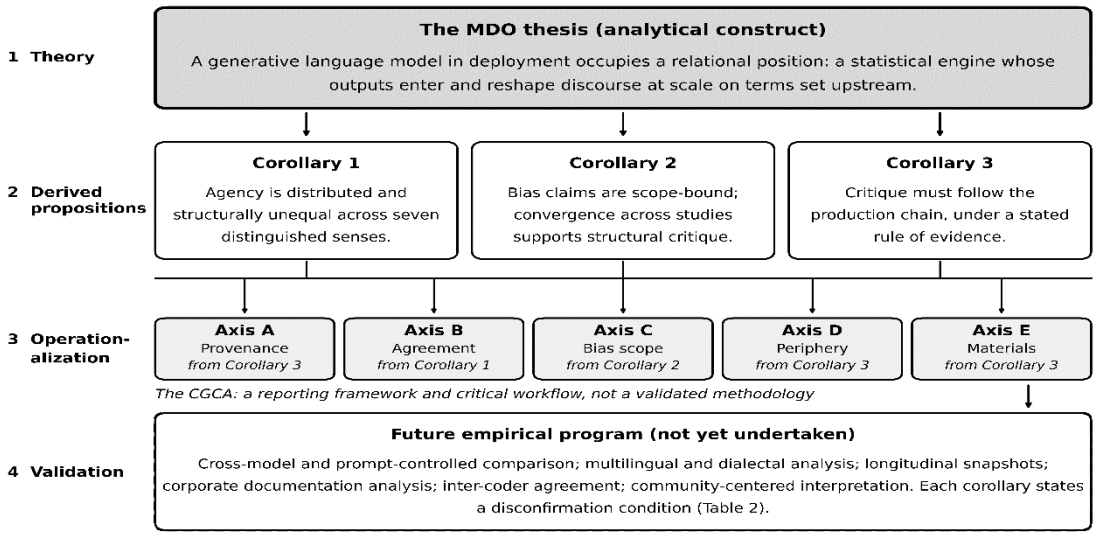


Fig. 2. From theory to validation. Tier 1 states the MDO thesis as an analytical construct. Tier 2 gives the three corollaries derived from it. Tier 3 gives the five CGCA axes, each labeled with the corollary it operationalizes. Tier 4 marks the empirical program as not yet undertaken (dashed border). The tiers distinguish conceptual propositions from operationalization and from validation; nothing in the figure reports an empirical finding.

6.1 A Conceptual Vignette

The following is a hypothetical illustration, not a study. No prompts were issued, no outputs were collected, and no coding was performed; the vignette exists to show what analytical decisions the axes require. Suppose a researcher asked a model to write short professional biographies for two fictional engineers differing only in name, and observed that one text foregrounded technical authority while the other foregrounded interpersonal warmth. A traditional reading would identify gendered professional discourse and stop. A CGCA-compliant reading would instead record the model, version, interface, and date, and note what the vendor does not disclose (Axis A); stratify any coding into surface-level and interpretive-level items before computing agreement (Axis B); state the claim as bounded to this model, version, prompt frame, and language, with an effect measure across repeated generations (Axis C); ask whose scholarship on gendered professional discourse the analysis draws on (Axis D); and deposit prompts and outputs so a later reader can re-run the contrast on a later snapshot (Axis E). Whether these requirements improve analytical quality is untested; what the vignette shows is which decisions the framework forces into the open.

6.2 Status and a Falsifiable Validation Pathway

The CGCA has not undergone empirical validation through prospective application, journal-level uptake, or measured effects on reporting quality, and the vignette above is illustrative rather than evidential. A three-stage pathway is proposed: pilot adoption in a small number of studies; comparison of reporting completeness between compliant and non-compliant analyses of the same material; and assessment of whether scope-bound claims survive replication better than unbounded ones. To keep the pathway falsifiable rather than aspirational, two kinds of criteria are distinguished in advance. The disconfirmation conditions in Table 2 are what would show the corollaries to be false: if scope-bound findings prove no more replicable than unbounded ones, or if varying prompts within a configuration explains more output variation than varying the configuration, the framework fails on its own terms. Two further criteria concern usefulness rather than truth: whether independent studies adopt the axes within 24 months of publication, and whether adopters report improved specificity in their claims. Failure on these would show the framework to be unused, not wrong. The full validation program spans cross-model

comparison, dialectal analysis, longitudinal snapshots, documentation analysis, and community-centered interpretation; Table 2 states the disconfirmation condition on which each corollary would fail.

6.3 Relation to Adjacent Reporting Frameworks

Model Cards, Data Statements, and Datasheets for Datasets share the CGCA's commitment to structured disclosure but operate on the producer side and address developers [41], [42], [43]. The CGCA operates on the analyst side: it disciplines what a discourse analyst may claim about a system. Where a model card exists, Axis A can draw on it directly; where none exists, Axis A requires the analyst to record what is knowable and name what is not.

7. Anchoring the Proposal in Applied Linguistic Traditions

Systemic functional linguistics offers a vocabulary for specifying which output regularities to document, since a distributional process should be expected to produce systematic tendencies in transitivity, modality, and thematic structure; this is a hypothesis for testing rather than an established finding [44]. That such tendencies should exist at all follows from the few-shot generalization behavior documented since Brown et al. [45]. Axis B can accordingly specify which metafunctional dimensions a codebook tracks, and the stylistic convergence reported by Sourati et al. [46] can be examined as metafunctional convergence rather than undifferentiated sameness. The discourse-historical approach supplies strategies of nomination and predication that support systematic comparison across prompts and snapshots, letting Axis C scope conditions be specified in terms of which topoi appear under which conditions [47], [48]. Multimodal critical discourse studies becomes directly relevant as generative systems extend into image and audio; Hart's [49] analysis of AI-generated images of refugees illustrates the analytical purchase of treating generated images as multimodal discursive constructions, and the axes extend to visual provenance and to schemata separating compositional features from interpretive ones [50], [28].

7.1 Arabic as a Critical Test Case

Arabic is not an appended illustration but the framework's most informative test case, because what a well-resourced English deployment conceals, an Arabic one exposes. Ahmed [51] argues that the apparatus of discourse analysis itself carries Anglophone assumptions requiring decolonial revision; the MDO gives that argument a production-side form. Arabic NLP resources are documented as thin relative to English, and dialect coverage thinner still [52], [53], [54]. There is therefore reason to expect model output in Arabic to favor Modern Standard Arabic over regional varieties and to flatten register accordingly. That expectation is a hypothesis, not a finding: the sampling distribution of any proprietary deployment is undisclosed, which is what makes the case diagnostic. Provenance opacity is a defect in every language. In English, dense coverage masks its analytical cost, so the defect is easier to overlook; in Arabic the same opacity is immediately disabling, because a corporate coverage decision and a property of the language cannot be told apart without disclosure. The Arabic case therefore exposes a limitation that applies to English deployments no less. Nothing here treats Arabic, its varieties, or their speakers as homogeneous. Applied to this article's own citation base (Axis D), the audit is unflattering but stated: of 54 works cited, one is in a language other than English [33], and the scholarship engaged substantively from outside North America and Western Europe is limited to work published from the Arab region [39], [31], [54] and Brazil [36].

A programmatic agenda follows: corpus-level audits of Arabic representation in documented training sets, dialect-comparative prompt studies, and register analysis of Arabic output against human-authored professional corpora. Each is a test of Corollary 2's scope conditions rather than an illustration of them.

8. Objections, Limits, and Adoption

Several objections can reasonably be raised against this proposal; the four most consequential are answered in Table 5, and the full set of ten, with responses, is available from the author on request. Where an objection lands, the claim is narrowed rather than defended.

Table 5. Anticipated objections and responses.

Objection	Response
MDO is a new label for a sociotechnical system	Partly conceded: the category is not new, but the inferential rule attached to this position is what no existing account supplies.
The framework reproduces ANT with a corporate emphasis	ANT holds actors analytically equal pending investigation; the MDO weights the corporate node in advance and states what would disconfirm that weighting.
Corporate asymmetry is assumed, not demonstrated	Partly conceded: argued from documented opacity rather than measured and now stated as falsifiable.
CGCA is an agenda, not an operational method	Conceded: it is a reporting framework and critical workflow, not a validated methodology. Table 4 specifies it per axis.

Four limits deserve naming. The framework's value depends on uptake by institutional actors the author cannot command. It is calibrated for frontier-LLM discourse, and its extension to smaller, multimodal, or open-weights systems is untested. Axis B presumes a two-coder design that single-author studies cannot supply. And applied indiscriminately, the axes could function as a barrier rather than a discipline. Adoption in under-resourced contexts raises three further risks: Axis A presumes information developers may not publish in the relevant languages; Axis E presumes repository infrastructure that is unevenly distributed; and uneven uptake between well-resourced and resource-constrained venues could entrench rather than reduce Anglophone centrality. These are conditions on responsible adoption rather than reasons to defer it.

A more radical objection would ask whether critical discourse studies is the right tradition at all for an object with no locatable speaking subject. The argument here is that it is, but only once its apparatus is adapted: the field's core commitment is to reading language as a site of power, and an object whose production is concentrated in a handful of firms is a site of power in the field's own terms. What must be given up is the assumption that the text's producer is available for analysis; what replaces it is a rule for reasoning from text to production conditions under stated evidential constraints.

9. Conclusion

The payoff can be stated concretely. A researcher adopting the framework can locate a bias finding at a named stage of production rather than attributing it to 'the model' or to 'AI'; can distinguish what a text shows from what the production chain is hypothesized to have done, by a stated rule rather than authorial confidence; can say which sense of agency a claim concerns, so that a dispute about model agency becomes a tractable disagreement about causal efficacy or institutional power; and can report a finding in a form a later study on a later snapshot can contradict. Each of these capacities has partial precedents in adjacent literatures. What the paper offers is their integration into a single reporting framework, together with a specification of what would count as evidence against it.

Generative language models are becoming infrastructural for professional, educational, and public text production, and critical discourse studies has something to contribute to their analysis. That contribution will remain partial for as long as the field must choose between granting a corporate product the standing of a speaker and treating it as a neutral instrument. The MDO offers a third position, three propositions that follow from it, and a reporting framework that makes those propositions answerable to evidence.

REFERENCES

[1] Fairclough N., *Critical discourse analysis: The critical study of language*, 2nd ed., Routledge, 2013. https://doi.org/10.4324/9781315834368

[2] van Dijk T. A., "Critical discourse analysis", in D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis*, 2nd ed., Wiley-Blackwell, 2015, pp. 466–485. https://doi.org/10.1002/9781118584194.ch22

[3] Gillings M., Kohn T., Mautner G., "The rise of large language models: Challenges for Critical Discourse Studies", *Critical Discourse Studies*, 22(6), 2025, pp. 625–641. https://doi.org/10.1080/17405904.2024.2373733

[4] Barad K., *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*, Duke University Press, 2007.

[5] Latour B., *Reassembling the social: An introduction to actor-network-theory*, Oxford University Press, 2005.

[6] Plantin J.-C., Lagoze C., Edwards P. N., Sandvig C., "Infrastructure studies meet platform studies in the age of Google and Facebook", *New Media & Society*, 20(1), 2018, pp. 293–310. https://doi.org/10.1177/1461444816661553

[7] van Dijck J., *The culture of connectivity: A critical history of social media*, Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199970773.001.0001

[8] Bender E. M., Gebru T., McMillan-Major A., Shmitchell S., "On the dangers of stochastic parrots: Can language models be too big?", in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, ACM, 2021, pp. 610–623. https://doi.org/10.1145/3442188.3445922

[9] Birhane A., "The unseen Black faces of AI algorithms", *Nature*, 610(7932), 2022, pp. 451–452. https://doi.org/10.1038/d41586-022-03050-7

[10] Couldry N., Mejias U. A., *The costs of connection: How data is colonizing human life and appropriating it for capitalism*, Stanford University Press, 2019.

[11] Crawford K., *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*, Yale University Press, 2021.

[12] Weidinger L., Uesato J., Rauh M., Griffin C., Huang P.-S., Mellor J., Glaese A., Cheng M., Balle B., Kasirzadeh A., Biles C., Brown S., Kenton Z., Hawkins W., Stepleton T., Birhane A., Hendricks L. A., Rimell L., Isaac W., … Gabriel I., "Taxonomy of risks posed by language models", in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, ACM, 2022, pp. 214–229. https://doi.org/10.1145/3531146.3533088

[13] Lee M., "Emergence of self-identity in artificial intelligence: A mathematical framework and empirical study with generative large language models", *Axioms*, 14(1), 2025, Article 44. https://doi.org/10.3390/axioms14010044

[14] Sukiennik N., Gao C., Xu F., Li Y., *An evaluation of cultural value alignment in LLM*, arXiv (Preprint), 2025. https://arxiv.org/abs/2504.08863

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

[16] Heinsfeld B., Veletsianos G., "The language on GenAI: A critical exploration of personification metaphors in UNESCO's guidance for generative AI in education and research", *Journal of Interactive Media in Education*, 2025(1), 2025, Article 16. https://doi.org/10.5334/jime.983

[17] Ji J., "Demystify ChatGPT: Anthropomorphism around generative AI", *GRACE: Global Review of AI Community Ethics*, 2(1), 2024. https://ojs.stanford.edu/ojs/index.php/grace/article/view/3222

[18] Pinell P., "Does artificial intelligence speak our language? A Gadamerian assessment of generative language models", *Political Research Quarterly*, 77(3), 2024, pp. 713–728. https://doi.org/10.1177/10659129241243038

[19] Garg R., Han J., Cheng Y., Fang Z., Swiecki Z., "Automated discourse analysis via generative artificial intelligence", in *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)*, ACM, 2024, pp. 814–820. https://doi.org/10.1145/3636555.3636879

[20] Hayes A. S., "'Conversing' with qualitative data: Enhancing qualitative research through large language models (LLMs)", *International Journal of Qualitative Methods*, 24, 2025, Article 16094069251322346. https://doi.org/10.1177/16094069251322346

[21] Nguyen D. C., Welch C., "Generative artificial intelligence in qualitative data analysis: Analyzing, or just chatting?", *Organizational Research Methods*, 29(1), 2026, pp. 3–39. https://doi.org/10.1177/10944281251377154

[22] Niu M., Jaiswal M., Mower Provost E., "From text to emotion: Unveiling the emotion annotation capabilities of LLMs", in *Proceedings of Interspeech 2024*, ISCA, 2024, pp. 2650–2654. https://doi.org/10.21437/Interspeech.2024-2282

[23] Baack S., "A critical analysis of the largest source for generative AI training data: Common Crawl", in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, ACM, 2024, pp. 2199–2208. https://doi.org/10.1145/3630106.3659033

[24] Perrotta C., Selwyn N., Ewin C., "Artificial intelligence and the affective labour of understanding: The intimate moderation of a language model", *New Media & Society*, 26(3), 2024, pp. 1585–1609. https://doi.org/10.1177/14614448221075296

[25] Suchman L., *Human-machine reconfigurations: Plans and situated actions*, 2nd ed., Cambridge University Press, 2007. https://doi.org/10.1017/CBO9780511808418
[26] Kera D. R., “Deixis machines and enunciation without a speaker in large language models”, *Discover Artificial Intelligence*, 6(1), 2026, Article 437. https://doi.org/10.1007/s44163-026-01157-8
[27] Kopf S., “Unravelling social media critical discourse studies (SM-CDS) – four approaches to studying social media through the critical lens”, *Critical Discourse Studies*, 23(2), 2026, pp. 215–232. https://doi.org/10.1080/17405904.2025.2463622
[28] Ledin P., Machin D., “Doing critical discourse studies with multimodality: From metafunctions to materiality”, *Critical Discourse Studies*, 16(5), 2019, pp. 497–513. https://doi.org/10.1080/17405904.2018.1468789
[29] Davison R. M., Chughtai H., Nielsen P., Marabelli M., Iannacci F., van Offenbeek M., Tarafdar M., Trenz M., Techatassanasoontorn A. A., Díaz Andrade A., Panteli N., “The ethics of using generative AI for qualitative data analysis”, *Information Systems Journal*, 34(5), 2024, pp. 1433–1439. https://doi.org/10.1111/isj.12504
[30] Kumar P. C., Cotter K., Cabrera L. Y., “Taking responsibility for meaning and mattering: An agential realist approach to generative AI and literacy”, *Reading Research Quarterly*, 59(4), 2024, pp. 570–578. https://doi.org/10.1002/rrq.570
[31] Hamdi S. A., “Generative AI as a new social actor in ESL/EFL context: A critical perspective”, *Arab World English Journal*, 17(1), 2026, pp. 371–378. https://doi.org/10.24093/awej/vol17no1.23
[32] van Leeuwen T., “The representation of social actors”, in C. R. Caldas-Coulthard & M. Coulthard (Eds.), *Texts and practices: Readings in critical discourse analysis*, Routledge, 1996, pp. 32–70.
[33] Lakel A., *Stop saying LLM: Large discourse models (LDM) and artificial discursive agent (ADA)?*, arXiv (Preprint in French), 2025. https://arxiv.org/abs/2512.19117
[34] KhosraviNik M., “Digital meaning-making across content and practice in social media critical discourse studies”, *Critical Discourse Studies*, 19(2), 2022, pp. 119–123. https://doi.org/10.1080/17405904.2020.1835683
[35] Blodgett S. L., Barocas S., Daumé III H., Wallach H., “Language (technology) is power: A critical survey of ‘bias’ in NLP”, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL, 2020, pp. 5454–5476. https://doi.org/10.18653/v1/2020.acl-main.485
[36] Bonil G., Hashiguti S., Silva J., Gondim J., Maia H., Silva N., Pedrini H., Avila S., *Yet another algorithmic bias: A discursive analysis of large language models reinforcing dominant discourses on gender and race*, arXiv (Preprint), 2025. https://arxiv.org/abs/2508.10304
[37] Palacios Barea M. A., Boeren D., Ferreira Goncalves J. F., “At the intersection of humanity and technology: A technofeminist intersectional critical discourse analysis of gender and race biases in the natural language processing model GPT-3”, *AI & Society*, 40(2), 2025, pp. 461–479. https://doi.org/10.1007/s00146-023-01804-z
[38] De A., Lima G., Zou Y., “What is safety? Corporate discourse, power, and the politics of generative AI safety”, in *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, ACM, 2026, pp. 1–18. https://doi.org/10.1145/3772318.3791632
[39] Ahmed T. N., Mahmood K. A., “A critical discourse analysis of ChatGPT’s role in knowledge and power production”, *Arab World English Journal*, 1(1), 2024, pp. 184–196. https://doi.org/10.24093/awej/ChatGPT.12
[40] Remling E., “Logics, assumptions and genre chains: A framework for poststructuralist policy analysis”, *Critical Discourse Studies*, 15(1), 2018, pp. 1–18. https://doi.org/10.1080/17405904.2017.1382382
[41] Bender E. M., Friedman B., “Data statements for natural language processing: Toward mitigating system bias and enabling better science”, *Transactions of the Association for Computational Linguistics*, 6, 2018, pp. 587–604. https://doi.org/10.1162/tacl_a_00041
[42] Gebru T., Morgenstern J., Vecchione B., Vaughan J. W., Wallach H., Daumé III H., Crawford K., “Datasheets for datasets”, *Communications of the ACM*, 64(12), 2021, pp. 86–92. https://doi.org/10.1145/3458723
[43] Mitchell M., Wu S., Zaldivar A., Barnes P., Vasserman L., Hutchinson B., Spitzer E., Raji I. D., Gebru T., “Model cards for model reporting”, in *Proceedings of the Conference on Fairness,*

Accountability, and Transparency (FAT '19)*, ACM, 2019, pp. 220–229. https://doi.org/10.1145/3287560.3287596
[44] Halliday M. A. K., Matthiessen C. M. I. M., *Halliday's introduction to functional grammar*, 4th ed., Routledge, 2014. https://doi.org/10.4324/9780203431269
[45] Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J. D., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D. M., Wu J., Winter C., … Amodei D., "Language models are few-shot learners", *Advances in Neural Information Processing Systems*, 33, 2020, pp. 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
[46] Sourati Z., Ziabari A. S., Dehghani M., "The homogenizing effect of large language models on human expression and thought", *Trends in Cognitive Sciences*, 2026, Advance online publication. https://doi.org/10.1016/j.tics.2026.01.003
[47] Reisigl M., Wodak R., "The discourse-historical approach (DHA)", in R. Wodak & M. Meyer (Eds.), *Methods of critical discourse studies*, 3rd ed., SAGE, 2016, pp. 23–61. https://doi.org/10.4135/9781036235192.n2
[48] Wodak R., Meyer M. (Eds.), *Methods of critical discourse analysis*, 2nd ed., SAGE, 2009.
[49] Hart C., "Using generative AI to investigate the multimodal representation of refugees and migrants", *Discourse & Society*, 37(3), 2026, pp. 470–502. https://doi.org/10.1177/09579265251378665
[50] Kress G., van Leeuwen T., *Reading images: The grammar of visual design*, 2nd ed., Routledge, 2006. https://doi.org/10.4324/9780203619728
[51] Ahmed Y., "Political discourse analysis: A decolonial approach", *Critical Discourse Studies*, 18(1), 2021, pp. 139–155. https://doi.org/10.1080/17405904.2020.1755707
[52] Bergman A., Diab M., "Towards responsible natural language annotation for the varieties of Arabic", in *Findings of the Association for Computational Linguistics: ACL 2022*, ACL, 2022, pp. 364–371. https://doi.org/10.18653/v1/2022.findings-acl.31
[53] Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M., "The state and fate of linguistic diversity and inclusion in the NLP world", in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL, 2020, pp. 6282–6293. https://doi.org/10.18653/v1/2020.acl-main.560
[54] Shoufan A., Alameri S., "Natural language processing for dialectical Arabic: A survey", in *Proceedings of the Second Workshop on Arabic Natural Language Processing*, ACL, 2015, pp. 36–48. https://doi.org/10.18653/v1/w15-3205

Appendix A. The CGCA Author-Facing Checklist

The checklist restates the five axes as concrete questions, glossing each ambiguous term in parentheses, and may be reproduced in full or by reference in a methods section.

Axis A: Provenance Disclosure. A1. Model name and exact version (with snapshot or release date) specified? A2. Deployment channel (API, chat interface, embedded application) named? A3. Alignment regime (RLHF, DPO, other) disclosed to the extent publicly available? A4. Date range of analyzed outputs recorded? A5. Prompt template deposited or quoted in full?

Axis B: Disaggregated Agreement. B1. Codebook separating surface-level items (decidable from explicit lexical or structural features, such as a stated job title) from interpretive-level items (requiring inference about framing, stance, or agency)? B2. Agreement statistics (κ, F1, or accuracy) reported separately for each stratum? B3. Aggregate statistics, where reported, given in addition to rather than instead of the disaggregated figures? B4. At least one analytical claim resting on an interpretive-level statistic explicitly flagged as such? B5. Single-coder designs making B1–B4 inapplicable stated and justified?

Axis C: Scope-Bound Bias Claims. C1. Each bias claim specifying model (with version), prompt regime, and language community? C2. Effect size or comparable quantitative anchor reported for each claim? C3. Each claim written so that a reader could falsify it by replication? C4. Generic 'structural' attributions avoided unless explicitly defined, bounded, and distinguished from the study's own measured findings?

Axis D: Centering the Periphery. D1. Citation base audited for non-Anglophone and Global-South representation? D2. At least one substantive engagement (not citation alone) with scholarship from outside North America and Western Europe, where topic permits? D3. Analytical concepts checked against terms developed in relevant non-Anglophone traditions?



Axis E: Replicable Materials. E1. Prompts, outputs, and model-version metadata deposited in an open repository with a persistent identifier? E2. Analysis repeated on a second model or snapshot where feasible? E3. Non-depositable items (e.g. personally identifiable inputs) explicitly named and justified? The checklist is guidance, not a pass-fail instrument. Authors unable to address an item are invited to state and justify the omission, which reviewers should treat as adequate compliance; the reporting burden is discussed among the limits above.