GLM_REVIEW_v8

Peer Review: Governance or Competition? Divergent Frames in AI Policy Discourse Across the US and Global South

Review of v8 (Revised Manuscript)

Overall Assessment

This revised manuscript presents a comparative analysis of AI policy framing between the United States and Global South nations (South Africa and Brazil). The research question—how nations construct AI as a policy problem—remains theoretically significant and timely. The authors have made substantial improvements in response to previous reviewers' concerns, particularly in transparency, statistical reporting, and acknowledgment of limitations. The use of LLMs as content coders following CommDAAF methodology is innovative and methodologically interesting.

However, several methodological concerns from previous reviews remain unaddressed, particularly regarding document type comparability, causal inference, and coding validity. The manuscript is improved but still has significant limitations that constrain the strength of its claims.

Recommendation: Minor Revision

---

1. Were Prior Methodological Concerns Addressed?

1.1. Substantially Addressed

Statistical Analysis and Effect Sizes (Partially Addressed) The authors have added Cramér's V effect size measures with conventional benchmarks (V = .10 small, .30 medium, .50 large), and have implemented Bonferroni correction for multiple comparisons (α = .05/8 = .006). This addresses Reviewer #1's concern about underdeveloped statistics and inflated Type I error rates. The reporting now distinguishes between findings that survive correction (Governance, Sovereignty, Safety) and those that do not (Rights, Innovation), providing appropriate caution.

Rights Frame Reliability (Acknowledged) The previous reviews correctly identified that rights framing reliability (κ = .52) fell below acceptable thresholds. The authors now explicitly acknowledge this limitation and provide appropriate caveats: "Findings involving rights framing should be interpreted with caution; the apparent difference between US (9%) and Global South (16%) may partly reflect measurement error rather than substantive divergence." This is an appropriate handling of a methodological weakness.

Document Type Limitation (Extended Discussion) The manuscript now contains an extended discussion of the document type comparability problem (lines 216-222), explicitly acknowledging that congressional hearings and mixed policy documents have different rhetorical conventions that may affect framing patterns. This limitation is also highlighted in the final Limitations section. While the authors acknowledge the problem without methodological solution (e.g., multivariate controls), the transparency is improved.

India Sample (Flagged as Exploratory) The manuscript now properly treats India as exploratory: “The Indian sample (N=7) is too small for reliable analysis and is reported only as exploratory.” It is excluded from primary analysis and presented separately in a dedicated section. This addresses Reviewer #2’s concern about misleading “Global South” claims based on inadequate representation.

Transparency on Coding Process The authors now provide more detail on the LLM coding process, including the initial κ = .21 failure mode due to “document-type bias” and the specific prompt revision that addressed it. The explanation is plausible and transparent about what went wrong and how it was fixed.

1.2. Partially Addressed

Confidence Intervals (Mentioned but Not Fully Provided) Table 1 now includes a column for “Difference” but does not display actual confidence intervals around frame proportions or differences. The text mentions confidence intervals (e.g., “95% CI [18.5%, 38.5%]” in Response to Reviewers), but these are not systematically displayed in the Results tables as Reviewer #1 requested.

Adjudication Transparency (Improved but Still Concerning) The authors now describe the adjudication process more clearly: “Two coders reviewed 15 disagreements and assigned final codes based on the predominant emphasis.” This provides more transparency than v7. However, the concern about κ = 1.0 after adjudication—which suggests the adjudicators simply imposed their own coding rather than serving as an independent third perspective—is not fully resolved.

1.3. Not Addressed

Document Type Comparability (Acknowledged but Not Controlled) The authors acknowledge the document type problem extensively but do not address it methodologically. Reviewer #1 requested either: (a) statistical controls for document type; (b) restriction to comparable document types; or (c) explicit modeling of document type as a confounder. None of these approaches are implemented. The manuscript remains vulnerable to the critique that observed framing differences may reflect genre effects rather than national context effects.

Multivariate Analysis (Not Implemented) Reviewer #1’s recommendation for multivariate analysis (logit or multinomial logit) controlling for document type, year, and other confounders is not implemented. The analysis remains bivariate (chi-square tests of frame distribution differences) despite clear recognition that document type, year, and sample size differences could confound the relationship between national context and framing.

Human-Validated Coding (Not Added) Both reviewers requested a human-coded validation subset (20-30 documents) to establish LLM coding accuracy against ground truth. This was not added. The manuscript still relies entirely on LLM-LLM cross-validation, which—as Reviewer #1 correctly noted—demonstrates consistency but not necessarily validity. Both models could share biases that produce agreement on incorrect codings.

Complete Prompts (Not Provided) Reviewer #1 and #2 both requested complete prompts in appendices for replicability. The authors describe the key intervention (“code the dominant MESSAGE about AI in the hearing, NOT the document type”) but do not provide the full prompts. This remains a replicability gap.

Search Queries and Document Registry (Not Provided) The manuscript still lacks specific search queries, API parameters, and a document registry with IDs that would enable replication. Reviewer #1’s concern about insufficient detail for replicability is not addressed.

Causal Inference (Speculation Remains) The Discussion section still advances speculative explanations for the observed divergence (adopter vs. developer positioning, geopolitical positioning) without empirical testing. The authors acknowledge that these are hypotheses, but the discussion still reads as if these are plausible explanations rather than untested speculation. Reviewer #2's concern about circular reasoning ("US is a developer because it frames AI competitively; it frames AI competitively because it's a developer") is not fully addressed.

---

2. Remaining Issues

2.1. Document Type Confounding (Unresolved)

The most significant remaining issue is the document type comparability problem. The authors acknowledge it extensively but do not address it methodologically. The core empirical finding—US discourse emphasizes competition framing (Sovereignty + Innovation = 43%) while Global South discourse emphasizes Governance framing (41%)—may be driven by: - Genre differences (hearings vs. reports) rather than national context - Venue effects (congressional committees vs. policy submissions) - Rhetorical conventions (adversarial testimony vs. institutional analysis)

Without statistical controls, matched samples, or explicit modeling, the causal claim that national context drives framing patterns cannot be defended. This is a fundamental validity problem.

Recommended Action: Conduct multivariate analysis (logit or multinomial logit) controlling for document type, year, and other potential confounders. Alternatively, conduct a robustness check analyzing only the most comparable document types across contexts.

2.2. Reliability Thresholds for Primary Claims

The primary claims—Governance divergence (V = .24) and Sovereignty divergence (V = .31)—are based on frames with κ > .70 ("substantial" agreement). However, the Rights frame (κ = .52) is still reported with comparative claims (US 9% vs. Global South 17%). While the authors acknowledge the limitation, including this comparison in Table 1 may mislead readers who do not read the fine print in the Method section.

Recommended Action: Either: (a) remove Rights from primary analysis and report only as exploratory; or (b) flag Rights clearly in Table 1 with an asterisk indicating reliability concerns.

2.3. Missing Human Validation

The lack of human-coded validation remains a gap. LLM-LLM cross-validation demonstrates consistency but not accuracy. Both models could miscode in the same direction if they share training data biases or interpretive assumptions. A modest human validation subset (n=20-30) would substantially strengthen confidence in the coding scheme's validity.

Recommended Action: Have human coders code a validation subset (minimum 20 documents distributed across countries and frames) and report human-LLM agreement statistics. This is a modest investment for substantial validity gains.

2.4. Incomplete Confidence Interval Reporting

The manuscript mentions confidence intervals in the text but does not systematically display them in tables. Readers cannot visually assess precision around frame proportions or differences. This is a basic standard for statistical reporting in empirical social science.

Recommended Action: Add 95% confidence intervals to Table 1 for all frame proportions (US and Global South columns) and for the difference column.

2.5. Search Strategy Transparency

The manuscript describes the search process but does not provide the specific search queries or API parameters used. This limits replicability. A researcher attempting to reproduce the study would need to know: what exactly was searched in GovInfo? What date ranges? What specific API endpoints?

Recommended Action: Provide complete search query strings, API parameters, and a document registry with IDs, titles, URLs, and metadata in an appendix.

2.6. Speculative Causal Explanations

The Discussion section advances explanations for the observed divergence (adopter vs. developer positioning; geopolitical positioning) without empirical testing. The authors acknowledge that these are hypotheses, but the structure of the Discussion suggests these are plausible explanations rather than unsubstantiated speculation. The circularity concern (identifying countries as “developers” based on their framing, then explaining framing by developer status) is not fully addressed.

Recommended Action: Restructure the Discussion to separate: (1) empirical findings (frame distributions and differences); (2) descriptive patterns (committees emphasizing different frames, temporal patterns if any); (3) hypotheses for future testing (developer/adopter, geopolitical positioning). Remove language suggesting these explanations have empirical support.

---

3. Positive Developments

The authors should be commended for several improvements:

1. **Bonferroni correction** appropriately applied to address multiple comparisons.
2. **Cramér’s V** effect sizes reported with interpretive benchmarks.
3. **Explicit acknowledgment** of Rights frame reliability limitations.
4. **Extended discussion** of document type comparability problems.
5. **Transparency** about initial coding failures and prompt revisions.
6. **Proper treatment** of India as exploratory rather than representative.
7. **More cautious claims** throughout, avoiding causal overreach compared to v7.

These changes show serious engagement with reviewer feedback and improved methodological transparency.

---

4. Recommendation: Minor Revision

This manuscript has improved substantially from v7. The authors have addressed several critical concerns (statistical underdevelopment, rights frame reliability, India sample) and demonstrated transparency about remaining limitations. The conceptual contribution—documenting divergent AI framing patterns between US and Global South contexts—remains valuable and understudied in the literature.

However, fundamental methodological issues remain that constrain the validity of claims:

1. **Document type confounding** is acknowledged but not controlled—this is the most significant threat to validity.
2. **Human validation** is absent despite both reviewers requesting it.
3. **Multivariate analysis** is not implemented despite clear confounding variables.
4. **Search strategy and prompts** lack transparency for replicability.

These issues prevent me from recommending acceptance at this stage. However, I believe these can be addressed through relatively straightforward revisions that do not require complete reanalysis or fundamental redesign:

- Add human-coded validation subset (n=20-30)
- Provide complete search queries and prompts in appendix
- Add confidence intervals to all tables
- Flag Rights frame in Table 1 with reliability warning
- Conduct basic multivariate analysis (logit with document type control)
- Separate hypotheses from empirical claims in Discussion

If the authors address these four revisions—particularly the human validation and multivariate analysis—I would recommend acceptance. The manuscript makes a valuable contribution to understanding cross-national AI policy discourse, and with these methodological refinements, the findings would be presented with appropriate rigor.

The paper’s strengths—novel comparative focus, timely research question, innovative LLM-coding methodology—outweigh the remaining limitations if addressed through minor revision.

Reviewer’s Note: This review evaluates the manuscript’s progress from v7 to v8, focusing on whether prior concerns were addressed and what substantive issues remain. The recommendation reflects balanced consideration of conceptual contribution versus methodological rigor.