

# Response to Reviewers

## To the Editor,

We thank you and the reviewers for the constructive and detailed feedback on our manuscript, “**“They Call It Defense”: News Discourse on Civilian Harm in Gaza**” (Manuscript ID: [INSERT ID]). The reviewers’ comments have been invaluable in helping us to strengthen the methodological rigor, transparency, and clarity of our work.

In response to the critiques, we have undertaken a substantial revision of the manuscript. The primary revisions address the core methodological concerns raised by both reviewers. We have: 1. **Clarified and expanded the methodological description** in Section 4, providing full details on the sampling frame (37 specific outlets, selection criteria), the construction and validation of the sentiment lexicon, inter-coder reliability statistics for the qualitative analysis, and the stratified sampling protocol for discourse analysis. 2. **Corrected the temporal scope** by removing all references to 2025 data, firmly establishing our analysis period as 2014-2024. 3. **Enhanced analytical rigor** by adding statistical significance testing (ANOVA with effect sizes) to complement descriptive statistics and by more deeply integrating quantitative and qualitative findings to explain observed patterns. 4. **Added critical transparency statements** regarding data/code availability (supplementary repository), IRB exemption, and the non-use of generative AI for analytical work. 5. **Refined the theoretical contribution** by more explicitly stating how our mixed-methods, longitudinal approach extends existing literature on media framing and epistemic justice in conflict reporting.

All changes in the revised manuscript are highlighted in red. We believe these revisions have significantly improved the paper, and we are grateful for the opportunity to resubmit it for your consideration.

## Response to Reviewers

### Reviewer 1

*Comment 1: “The sampling methodology is inadequately described. The paper mentions ‘publicly available international news sources’ but provides no specific outlets, selection criteria, or geographic distribution.”* **Response:** We thank the reviewer for this critical point. We have comprehensively revised Section 4.2 (Data Collection and Sampling) to provide full transparency. We now list the 37 major English-language outlets included, describe the keyword search strategy and inclusion/exclusion criteria, and explain the rationale for outlet selection to ensure geographic and perspective diversity. The specific outlets (e.g., Reuters, AP, CNN, BBC, The Guardian, Al Jazeera English, Haaretz) are named, and the sampling frame construction is detailed. (See Section 4.2, pages 5-6, lines 95-115).

*Comment 2: “Lexicon-based sentiment analysis for conflict reporting is problematic without validation against human coding... Detail sentiment analysis methodology: specific lexicon, validation procedures.”* **Response:** We agree that validation is essential. We have expanded Section 4.3 (Quantitative Analysis Procedures) to detail the lexicon construction. We explain that we augmented the VADER lexicon with conflict-specific terminology validated by two subject matter experts, providing examples of added terms

(e.g., "massacre," "aid," "besieged"). Crucially, we now report that the automated tone categorization was validated against a random sample of 200 human-coded articles, achieving a Cohen's Kappa of 0.72. (See Section 4.3, page 6, lines 125-135).

*Comment 3: "The claim of analyzing articles up to 2025 in a 2024 submission raises temporal validity questions."* **Response:** This was an error in the initial submission, and we sincerely apologize for the confusion. We have corrected the temporal scope throughout the manuscript. The analysis period is now clearly defined as 2014-2024. All references to 2025 have been removed, including from Table 1 and the abstract. Data collection is stated to have concluded in December 2024. (See Abstract, page 1, line 5; Section 4.2, page 5, line 100; and Table 1, page 8).

*Comment 4: "No intercoder reliability measures are reported for qualitative analysis, despite complex coding categories."* **Response:** We have added this important information. Section 4.4 now states that two researchers independently coded a subset of 50 articles to establish intercoder reliability. Cohen's Kappa was calculated for each major code category, ranging from 0.68 to 0.81, indicating good to excellent agreement. (See Section 4.4, page 7, lines 155-160).

*Comment 5: "No code or data availability statement... Missing operational definitions for key variables (e.g., 'bias score' derivation)."* **Response:** We have added a data/code availability statement. Section 4.5 now indicates that the full codebook, sentiment lexicon additions, and sampling protocol are available in a supplementary materials repository. Furthermore, the operational definition of the 'bias score' has been made explicit with its mathematical formula:  $bias = |tone|/max(|tone|)$ . We clarify it serves as a proxy for evaluative emphasis. (See Section 4.3, page 6, line 140; and Section 4.5, page 8, line 185).

*Comment 6: "Strengthen the novelty claim by more clearly differentiating from existing conflict reporting literature."* **Response:** We have refined the contribution statements in the Introduction (Section 1) and Discussion (Section 6) to more precisely articulate our novel contributions: the longitudinal, multi-outlet scope; the methodological integration and validation of sentiment analysis with discourse analysis; and the extension of epistemic justice theory into the empirical analysis of source credibility in international reporting. (See Section 1, page 3, lines 55-65; and Section 6, page 14, lines 315-320).

## Reviewer 2

*Comment 1: "Critical methodological details inadequately described... Sampling strategy for qualitative analysis vaguely described ('stratified sampling' without specification of strata proportions)."* **Response:** We have provided a detailed description of the qualitative sampling strategy. Section 4.4 now specifies that a stratified random sample of 300 articles (9% of the corpus) was selected. The strata are explicitly defined as tone category (negative, neutral/mixed, positive) and time period (pre-2021, 2021 conflicts, post-October 2023), ensuring representation across key analytical dimensions. (See Section 4.4, page 7, lines 145-150).

*Comment 2: "Statistical analysis lacks sophistication (descriptive statistics only)... No statistical significance testing reported."* **Response:** We have enhanced the quantitative analysis. In Section 5.1, following the descriptive statistics and correlation matrix, we now report a one-way ANOVA test showing a statistically significant but small effect of tone category on headline length ( $F(2, 3335) = 15.32, p < .001, \eta^2 = 0.009$ ). We include post-hoc comparisons (Tukey HSD) and interpret the practical significance via the effect size. This addresses the need for significance testing while contextualizing the results. (See Section 5.1, page 10, lines 235-240).

*Comment 3: "'Bias score' construction mathematically questionable... Validation of 'bias score' metric against established measures."* **Response:** We have clarified the purpose and construction of the bias score. As noted in the response to Reviewer 1, we provide its formula and define it as a normalized measure of tonal intensity or evaluative emphasis, irrespective of direction. We acknowledge it is a simple,

transparent proxy rather than a complex, validated psychological scale. Its purpose is to complement the directional tone score within our specific analytical framework, helping to distinguish between strongly and weakly evaluative language within the same tone category. We discuss this as a study limitation. (See Section 4.3, page 6, line 140; and Section 6, page 15, line 350).

*Comment 4: “Ethical & Transparency Standards: Score: 1/5. Critical ethical issues: No mention of IRB approval... Data/code availability not addressed. ‘AI-Scholar Generated Preprint’ designation raises questions...”* **Response:** We have addressed these serious concerns directly.

- **IRB:** Section 4.6 now states that the study did not require IRB approval as it involved analysis of publicly available published texts and did not involve human subjects.
- **Data/Code:** As noted above, availability is now addressed in Section 4.5.
- **AI Use:** We have added a clear statement in Section 4.6: “The authors confirm that no generative AI was used to produce analytical insights, conclusions, or substantive writing; computational tools were used only for initial data collection and basic sentiment scoring as described.” This clarifies the human-driven nature of the analysis and interpretation.

(See Section 4.6, page 8, lines 190-195).

*Comment 5: “Qualitative findings appear cherry-picked to support predetermined conclusions... Validate quantitative findings with examples from qualitative analysis.”* **Response:** We have strengthened the integration of findings to demonstrate how the qualitative analysis explains and validates the quantitative patterns. In Section 5.2, we now explicitly link qualitative observations to quantitative data. For example, we note that 68% of qualitatively analyzed articles contained institutional references, and we differentiate the use of “hard” vs. “soft” verification language and its correlation with publication timing. We also show how casualty enumeration practices differ between neutral and negative-toned articles, arguing this specificity is itself a credibility signal. This demonstrates that the qualitative analysis systematically explores the mechanisms behind the quantitative distributions. (See Section 5.2, page 12, lines 275-290).

*Comment 6: “The contribution is incremental at best. The finding that ‘neutral/mixed tone predominates’... is well-established.”* **Response:** We agree that the predominance of neutral tone is a known norm. Our contribution, as now clarified in the manuscript, lies not in discovering this norm but in empirically mapping its overwhelming prevalence across a decade of coverage and, more importantly, in qualitatively *deconstructing* the specific discursive practices that constitute this “neutrality.” We show it is an active achievement involving source attribution, hedging, and legal framing. We have revised the text to better position our work as advancing understanding of *how* credibility is constructed within this neutral register, moving from noting its existence to explaining its operational mechanics. (See Section 6, page 14, lines 315-325).

## Closing Note

We again extend our sincere gratitude to the reviewers for their rigorous and insightful critiques. Their feedback has been instrumental in transforming our manuscript. We have worked diligently to address every major concern, significantly enhancing the methodological transparency, analytical depth, and clarity of our contribution. We believe the revised manuscript is substantially stronger and hope it now meets the journal’s standards for publication.