



EUROPEAN SCIENTIFIC JOURNAL
by European Scientific Institute



Paper: “H-VALIDA: A Hybrid Methodological Framework for Transparent AI-Assisted Surveys in Data-Scarce Contexts: A Conceptual Framework for Responsible AI Integration”

Submitted: 13 August 2026

Accepted: 25 September 2026

Published: 30 September 2026

Corresponding Author: Florence Yasmine Andrews

Doi: 10.19044/esj.2026.v22n25p32

Peer review:

Reviewer 1: Xi Chen

University of California Santa Cruz, USA

Reviewer 2: Elda Xhumari

University of Tirana, Albania

Reviewer B:
Recommendation: Revisions Required

The TITLE is clear and it is adequate to the content of the article.

The title is clear, informative, and consistent with the manuscript's focus on a hybrid methodology for AI-assisted surveys in data-scarce contexts. However, it is relatively long, and the subtitle partly repeats the methodological emphasis already expressed in the main title. The authors may consider shortening it for greater precision and impact.

The ABSTRACT clearly presents objectives, methods, results, and conclusions.

The abstract clearly identifies the problem, purpose, theoretical foundations, seven pillars, and principal argument of the manuscript. However, the methods and results are presented only in broad terms. It should briefly explain how the heuristic evaluation and structured simulation were conducted, including the evaluation criteria and evidence used. Statements regarding the framework's reliability and practical value should also be qualified because no empirical field validation is reported.

There are a few grammatical errors and spelling mistakes in this article.

The manuscript is generally well written and uses an appropriate academic style. Nevertheless, several sentences are overly long or repetitive, and some grammatical and punctuation problems require correction. For example, the sentence beginning "The design is consistent with the tradition of methodological innovation research; in which..." uses a semicolon incorrectly. Some sections also repeat similar claims concerning transparency, contextual validity, and human oversight. Careful proofreading and moderate language editing are recommended to improve concision, consistency, and readability.

The study METHODS are explained clearly.

The manuscript identifies five stages of framework development and describes heuristic evaluation, comparative analysis, structured simulation logic, literature triangulation, and internal consistency checking. However, the methodology is not sufficiently detailed to permit replication or independent assessment. The paper should specify how the literature was identified and selected, who participated in the "expert-informed" evaluation, the experts' selection criteria and number, the evaluation instrument, the simulation scenarios and inputs, the AI systems or models tested, the comparison procedure, and how the findings were analyzed. If no formal experts or actual simulations were used, the terminology should be revised accordingly. A more transparent framework-development protocol and a completed validation matrix would considerably strengthen the study.

The results are clear and do not contain errors.

The seven pillars are presented clearly and are logically mapped onto an expanded total survey error taxonomy. This is a useful conceptual contribution. Nevertheless, the Results section primarily describes the proposed framework rather than reporting verifiable evaluation results. The statement that the evaluation “demonstrated” significant improvements is not supported by quantitative results, comparative data, documented expert assessments, test cases, or clearly reported simulation outputs. The authors should provide concrete evidence, such as scenarios, performance criteria, reviewer ratings, examples of AI errors detected through human validation, or a comparative results table. Otherwise, these statements should be reframed as theoretical propositions or expected benefits.

The CONCLUSION or summary is accurate and supported by the content.

The conclusion accurately summarizes the framework and is generally consistent with the manuscript’s theoretical discussion. However, several conclusions are expressed too definitively given the absence of empirical field testing and fully documented validation results. Claims that H-VALIDA improves reliability, interpretability, or ethical robustness should be presented as proposed or prospective benefits requiring empirical confirmation. The conclusion is also lengthy and repeats material from the Discussion and Recommendations sections. It would benefit from condensation and a sharper distinction between established findings, conceptual contributions, and future research expectations.

The list of REFERENCES is comprehensive and appropriate.

The references are relevant to the manuscript’s main themes, including total survey error, human-centred AI, algorithmic bias, decolonial AI, and Indigenous data governance. However, the literature base is relatively limited for a methodological framework covering several substantial research fields. Additional recent research on AI-assisted survey design, automated coding, multilingual survey translation, algorithmic auditing, mixed-mode sampling, validation frameworks, and applicable AI governance standards should be incorporated.

The citation–reference correspondence also needs correction. Floridi (2019), Lincoln and Guba (1985), and UNESCO (2021) appear in the reference list but are not clearly cited in the main text. All references should be checked systematically to ensure that each listed source is cited in the manuscript and that every in-text citation has a corresponding reference. Reference formatting and DOI inclusion should also be standardized.

Please rate the TITLE of this paper.

[Poor] 1-5 [Excellent]

Please rate the ABSTRACT of this paper.

[Poor] 1-5 [Excellent]

3

Please rate the LANGUAGE of this paper.

[Poor] 1-5 [Excellent]

3

Please rate the METHODS of this paper.

[Poor] 1-5 [Excellent]

2

Please rate the RESULTS of this paper.

[Poor] 1-5 [Excellent]

2

Please rate the CONCLUSION of this paper.

[Poor] 1-5 [Excellent]

3

Please rate the REFERENCES of this paper.

[Poor] 1-5 [Excellent]

3

Overall Recommendation!!!

Return for major revision and resubmission

Comments and Suggestions to the Author(s):

The manuscript addresses a timely and important issue and proposes a potentially useful framework for responsible AI integration into survey research in data-scarce, multilingual, and culturally complex environments. The combination of human-centred AI, total survey error, algorithmic auditing, contextual validation, and the CARE Principles is conceptually valuable. The seven-pillar structure is clear and may provide a useful foundation for future empirical work.

The manuscript nevertheless requires major revision before its contribution can be properly evaluated. Its main weakness is the gap between the methodological claims and the evidence presented. The paper refers to “expert-informed heuristic evaluation,” “structured simulation logic,” and comparative findings, but does not provide sufficient information about the experts, evaluation protocol, simulation design, AI systems, datasets, scenarios, measures, or analysis. Consequently, the reported advantages of the hybrid framework cannot presently be verified or replicated.

The authors should:

Provide a transparent and reproducible account of the framework-development process, including the literature-selection procedure.

Describe the expert-informed evaluation in detail, including participant numbers, expertise, selection criteria, evaluation questions, scoring procedures, and synthesis of feedback.

Explain the structured simulations, including the tools or models tested, prompts or inputs, linguistic and cultural scenarios, comparison conditions, and evaluation metrics.

Present concrete findings in tables or worked examples rather than relying exclusively on narrative assertions.

Include the proposed validation matrix as a fully developed operational instrument with indicators, responsibilities, evidence requirements, decision thresholds, and corrective actions. Distinguish clearly between conceptual propositions, simulation findings, expert assessments, and empirically demonstrated results.

Remove or qualify unsupported expressions such as “demonstrated,” “significantly improve,” and “human oversight proved indispensable” unless appropriate evidence and statistical analysis are supplied.

Clarify how the framework differs from, integrates, or advances existing responsible-AI and survey-quality frameworks.

Reduce repetition across the Introduction, Results, Discussion, Conclusion, Recommendations, and Final Reflection sections.

Conduct careful language editing and a complete citation–reference consistency check.

The manuscript has potential, but its current form is closer to a well-developed conceptual or position paper than a validated methodological study. The article should either provide the missing validation evidence or explicitly reposition itself as a conceptual framework awaiting empirical validation.

Reviewer C:
Recommendation: Resubmit for Review

The TITLE is clear and it is adequate to the content of the article.

The title communicates the general idea, but it's overly long and makes misleading claim by using word "validated". The manuscript doesn't report any validation, quantitative evaluation, or reproducible experiment result.

The ABSTRACT clearly presents objectives, methods, results, and conclusions.

It clearly described the motivation and the seven proposed pillars, but it overstates the maturity and evidential status of the framework. It describes H-VALIDA as a “transparent heuristic approach” and presents it as a workable model, yet the study contains neither empirical field validation nor quantitative comparative evidence.

There are a few grammatical errors and spelling mistakes in this article.

The manuscript is generally readable, but it is excessively repetitive. The same claims—AI is sociotechnical, human oversight is necessary, transparency is important, and data-scarce contexts require contextual adaptation—are repeated across the Introduction, Literature Review, Results, Discussion, Conclusions, Recommendations, and Final Reflection. Considerable condensation is needed.

The study METHODS are explained clearly.

The methodology is the most serious weakness of the manuscript. Although the paper describes five stages of framework development, these stages remain high-level descriptions rather than reproducible methodological procedures. It makes the paper sounds like a motivation calling and lack concrete research or evidence. The manuscript states that validation involved “expert-informed heuristic evaluation,” “comparative analysis,” and “structured simulation logic.” However, essential methodological information is absent:

How many experts participated?
Who were the experts and what were their qualifications?
How were they recruited or selected?
What evaluation instrument or rubric was used?
What questions were experts asked?
What constituted a passing or successful evaluation?
Were evaluations conducted independently?
Was inter-rater agreement measured?
What AI systems or models were evaluated?
What prompts were used?
What datasets or synthetic inputs were used?
What simulation scenarios were constructed?
What baseline represented the “purely automated” condition?
What quantitative or qualitative metrics were used?
How were reliability, transparency, interpretability, and ethical robustness operationally measured?

Without this information, the validation process cannot be reproduced or independently assessed.

Furthermore, describing the absence of primary data as a “deliberate design choice rather than a limitation” is not fully convincing. It may be appropriate for an initial conceptual framework, but it substantially limits any claim that the framework has been validated.

The results are clear and do not contain errors.

The section labeled “Results” largely presents and explains the proposed framework rather than reporting results from an actual evaluation. The seven pillars are described clearly, but description of a framework should not be conflated with evidence that the framework works.

The strongest concern occurs in Section 4.5, where the manuscript states that heuristic evaluation and simulation “demonstrated” that hybrid methodologies “significantly improve” reliability, interpretability, transparency, and ethical robustness. No evidence supporting the term “significantly” is presented. There are:

- no numerical results,
- no statistical tests,
- no sample sizes,
- no expert scores,
- no confidence intervals,
- no error rates,
- no baseline comparisons,
- no simulation outputs,
- no tables of results,
- and no worked examples.

Therefore, the claim of significant improvement is currently unsupported.

The CONCLUSION or summary is accurate and supported by the content.

The conclusion summarizes the manuscript consistently, but several statements remain stronger than the presented evidence allows. The manuscript concludes that human oversight improves algorithmic accountability and describes H-VALIDA as improving reliability, traceability, transparency, interpretability, and contextual validity.

Yet the manuscript later acknowledges that no complete field implementation has occurred and that the proposed advantages must still be confirmed empirically. The conclusion therefore needs to distinguish clearly between theoretical propositions and experimentally demonstrated findings.

The Conclusions, Recommendations for Practice, Recommendations for Future Research, and Final Reflection are also unnecessarily repetitive and could be substantially shortened.

The list of REFERENCES is comprehensive and appropriate.

The references include several important foundational sources, but the literature base is relatively limited for such a broad framework. Only 14 references are provided despite the manuscript attempting to integrate survey methodology, AI auditing, human-centered AI, algorithmic bias, multilingual AI, decolonial AI, Indigenous data governance, privacy, and ethical AI.

There are also citation consistency problems. The manuscript names Hansen, Hurwitz, Waksberg, Kish, Mahalanobis, Biemer and others when discussing the intellectual history of Total Survey Error, yet these works do not appear individually in the reference list. Conversely, some references listed at the end do not appear to be explicitly cited in the manuscript.

The literature review should also engage more directly with recent empirical research specifically evaluating AI/LLM-assisted survey design, coding, translation, and response analysis rather than relying predominantly on broad AI ethics literature.

Please rate the TITLE of this paper.

[Poor] 1-5 [Excellent]

3

Please rate the ABSTRACT of this paper.

[Poor] 1-5 [Excellent]

2

Please rate the LANGUAGE of this paper.

[Poor] 1-5 [Excellent]

3

Please rate the METHODS of this paper.

[Poor] 1-5 [Excellent]

1

Please rate the RESULTS of this paper.

[Poor] 1-5 [Excellent]

1

Please rate the CONCLUSION of this paper.

[Poor] 1-5 [Excellent]

2

Please rate the REFERENCES of this paper.

[Poor] 1-5 [Excellent]

2

Overall Recommendation!!!

Return for major revision and resubmission

Comments and Suggestions to the Author(s):

The manuscript addresses an important issue and brings together several relevant perspectives, including Total Survey Error, human-centered AI, algorithmic auditing, contextual validation, and the CARE Principles. However, in its current form, the manuscript reads substantially more like a conceptual position paper or framework proposal than a validated methodological study.

The central concern is the gap between the strength of the claims and the evidence presented. The manuscript repeatedly refers to validation, heuristic evaluation, comparative analysis, and simulation, but provides insufficient information to reproduce or independently assess these activities. No expert sample, evaluation instrument, scoring rubric, simulation design, model configuration, baseline, dataset, example output, quantitative metric, or comparative result is presented. Consequently, statements that H-VALIDA “demonstrated” or “significantly improve[s]” outcomes are not currently supported by the reported evidence.

A second major problem is the absence of figures, diagrams, tables, worked examples, and operational artifacts. This is particularly problematic for a methodological-framework paper. The seven pillars, their interactions, their relationship with the survey lifecycle, and their mapping to Total Survey Error should be visually represented. Most importantly, the section titled “The Validation Matrix” does not actually present a validation matrix. An actual matrix with dimensions, criteria, procedures, evidence requirements, responsible reviewers, and acceptance thresholds would make the framework considerably more useful.

Third, the distinction between framework development and framework validation needs to be corrected throughout the manuscript. Developing a theoretically grounded framework and subjecting it to informal or heuristic consideration does not establish empirical validity. The manuscript itself acknowledges that field testing has not occurred. The title, abstract, Results, Discussion, and Conclusion should therefore use more cautious terminology.

Fourth, the framework needs at least one detailed worked example. For instance, the authors could demonstrate an AI-assisted multilingual survey in which questionnaire generation, translation, sampling, AI coding, human review, auditing, documentation, and ethical checks are processed sequentially through the seven pillars. Without such an example, H-VALIDA remains difficult to distinguish operationally from a collection of well-established responsible-AI recommendations.

Finally, the manuscript is substantially longer than necessary because core arguments are repeatedly restated. The Results, Discussion, Conclusions, Recommendations, and Final Reflection contain considerable conceptual repetition. A shorter manuscript with concrete diagrams, tables, evaluation procedures, and examples would be substantially stronger.
