

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

Dr. Florence Yasmine Andrews, PhD

Assistant Professor of Economics and MBA Programme Coordinator
University of Belize, Belmopan, Belize

[Doi:10.19044/esj.2026.v22n25p32](https://doi.org/10.19044/esj.2026.v22n25p32)

Submitted: 13 August 2026

Accepted: 25 September 2026

Published: 30 September 2026

Copyright 2026 Author(s)

Under Creative Commons CC-BY 4.0

OPEN ACCESS

Cite As:

Andrews, F.Y. (2026). *H-VALIDA: A Hybrid Methodological Framework for Transparent AI-Assisted Surveys in Data-Scarce Contexts: A Conceptual Framework for Responsible AI Integration*. European Scientific Journal, ESJ, 22 (25), 32.

<https://doi.org/10.19044/esj.2026.v22n25p32>

Abstract

Survey research has taken up artificial intelligence quickly, largely because of speed, scale, and lower cost. Those same gains bring methodological risks that become especially serious where data are thin, several languages are in use, or cultural meaning is tightly layered. This article puts forward H-VALIDA (Hybrid Validation and Auditing for Localized, Interpretable, Documented, and Accountable AI-Assisted Surveys) as both a conceptual and an operational framework for AI-assisted surveys that are more reliable, more sensitive to context, and more firmly grounded in ethics. The framework draws together critical data studies, human-centred AI, the total survey error tradition, classical survey methodology, and decolonial principles of data governance. Seven interlocking pillars organise the work: Hybrid Human-AI Collaboration, Contextual Sampling, Algorithmic Auditing, Human Validation, Source Triangulation, Open Documentation, and Ethical Accountability. Those pillars are mapped onto an expanded total survey error taxonomy, and the CARE Principles for Indigenous data governance run through the design as a whole. The article is a methodological design study rather than a field experiment. Development of the framework followed a five-stage protocol: a diagnostic review of documented weaknesses in AI-assisted surveys; a structured synthesis of the literature; integration of principles; construction of the pillars; and operationalisation. Prospective appraisal rested

on a specified heuristic rubric, a worked multilingual scenario, and a comparison with a purely automated baseline. No expert-panel scores, statistical tests, or national field implementation are reported here. Claims of improvement are therefore stated as theoretically grounded and operationally specified propositions that still require empirical confirmation. The central argument is that AI systems function as sociotechnical arrangements rather than as neutral instruments, and that validity depends on continuous human oversight, an open record of process, and deliberate adjustment to local conditions.

Keywords: Artificial Intelligence; AI-Assisted Surveys; Hybrid Methodology; Algorithmic Auditing; Human-Centred AI; Ethical Governance; Total Survey Error; CARE Principles

1. Introduction

1.1 Background and Rationale

Survey systems still underwrite evidence-based governance, public-health planning, educational assessment, climate-adaptation work, and the tracking of sustainable development goals. Governments, international agencies, and civil-society organisations continue to rely on survey evidence when they allocate resources, design interventions, or judge the effects of policy. Large language models, automated classification tools, and predictive analytics now commonly assist questionnaire design, multilingual translation, coding, anomaly detection, and interpretive reporting (Argyle et al., 2023; Jansen et al., 2023; von der Heyde et al., 2024).

The attractions are genuine. Computational systems can process large volumes of text quickly, support concurrent translation across several languages, surface patterns that human coders may miss, and cut both the cost and the calendar time of conventional operations. In settings where resources are tight, those efficiencies are particularly tempting. Accumulating evidence, however, shows that AI systems are never neutral instruments. They carry the imprint of the training data on which they rest, the institutional assumptions built into their design, the linguistic hierarchies that organise global digital corpora, and the historical power relations that decide whose knowledge is treated as authoritative (Crawford, 2021; Noble, 2018; Bender et al., 2021).

In data-scarce and multicultural settings the risks become sharper. Models trained mainly on material from high-resource, English-dominant, urban environments can systematically misrepresent populations in small developing states, multilingual societies, rural communities, and Indigenous groups (Barocas & Selbst, 2016; Mohamed et al., 2020). Outputs may look scientifically authoritative while still lacking contextual validity, linguistic precision, or cultural legitimacy. When such outputs feed into policy, the

practical consequences can include the exclusion of vulnerable populations, the misallocation of resources, and a gradual erosion of public trust in official statistics.

This article introduces H-VALIDA as a hybrid methodological framework that places human judgement, algorithmic auditing, and ethical accountability inside everyday AI-assisted survey practice. The framework is intended for environments marked by informational asymmetry, institutional constraint, and sociocultural complexity. Methodological validity is treated here not as a purely technical accomplishment but as a sociotechnical and ethical achievement that requires continuous human oversight, transparent documentation, and deliberate adaptation to local conditions.

1.2 Problem Statement

AI-assisted surveys can intensify existing methodological weaknesses when they are introduced without systematic validation, contextual adaptation, and transparent governance. Systems trained chiefly on data from high-resource contexts may systematically misrepresent populations in small developing states, multilingual societies, and rural or Indigenous communities (Barocas & Selbst, 2016). Opacity in algorithmic decision processes makes the problem worse. When researchers cannot reconstruct how a classification was produced, how a translation was generated, or why a particular weighting scheme was chosen, the capacity to interrogate or contest the outputs is sharply reduced (Raji et al., 2020; Mittelstadt, 2019). In small societies, where individuals and communities are more readily identifiable, the privacy and ethical stakes of that opacity rise still further.

The difficulty is therefore not merely technical. It is epistemological and ethical. It concerns whose realities are accepted as legitimate knowledge, whose voices automated systems amplify or silence, and who carries the costs when those systems fail. Addressing the problem requires a methodological response that treats human judgement, contextual knowledge, and ethical accountability as constitutive elements of survey quality rather than as optional extras.

1.3 Research Objectives

The primary aim is to develop and set out a transparent, hybrid, context-sensitive, and ethically accountable conceptual framework for AI-assisted surveys. The specific objectives are as follows:

- To identify the principal methodological weaknesses that affect AI-assisted surveys in data-scarce contexts, including urban bias, linguistic distortion, cultural misclassification, and opacity;

- To synthesise principles drawn from critical data studies, human-centred AI, the total survey error framework, survey methodology, and decolonial thought;
- To construct seven interdependent methodological pillars and map them onto an expanded total survey error taxonomy, incorporating the CARE Principles;
- To specify practical procedures for validation, auditing, documentation, and ethical oversight that institutions with limited resources can implement in stages;
- To illustrate, through a specified heuristic rubric and a worked scenario, the prospective advantages of hybrid approaches relative to purely automated systems, while keeping conceptual propositions distinct from empirically demonstrated results.

1.4 Central Research Question

How can a transparent, hybrid, human-centred, and context-sensitive methodology be specified so as to improve the reliability, validity, auditability, and ethical legitimacy of AI-assisted surveys in data-scarce environments, pending empirical confirmation in the field?

1.5 Significance and Contribution

H-VALIDA contributes to methodological theory by enlarging conventional notions of survey quality to include algorithmic validity, contextual validity, linguistic validity, cultural validity, ethical validity, interpretive validity, and documentation validity. These dimensions are not peripheral to statistical quality; in the age of artificial intelligence they form part of its substance (Groves & Lyberg, 2010; Biemer, 2010).

The framework contributes to practice by offering operational procedures that institutions with limited resources can introduce in stages. It does not presuppose large computational infrastructure, proprietary audit tools, or extensive specialist capacity. External documentation, structured human review, mixed-mode sampling, and institutional accountability are emphasised instead.

H-VALIDA contributes to ethical governance by treating transparency and human accountability as integral to methodological legitimacy. By engaging the total survey error tradition and the CARE Principles explicitly, the framework answers the demand for responsible AI practices that remain grounded in local realities rather than in abstract universal principles (Mittelstadt, 2019; Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019). The present article positions the framework as conceptually specified and operationally usable, not as a field-validated instrument.

2. Literature Review

2.1 Evolution of Survey Methodology and the Total Survey Error Framework

Survey research has moved through successive technological regimes: paper instruments and face-to-face interviewing; computer-assisted telephone and personal interviewing; internet panels, mobile platforms, and real-time digital systems. Each phase expanded capacity and, at the same time, introduced new error structures. The total survey error (TSE) paradigm remains the central organising framework of the field (Groves & Lyberg, 2010; Biemer, 2010). Total survey error is the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. The classical decomposition distinguishes sampling error from nonsampling error, the latter including coverage error, nonresponse error, measurement error, and processing error. Specification error, the mismatch between the intended concept and the operational definition, is sometimes treated as a distinct component.

One lasting value of the TSE framework is its usefulness at the design stage. By forcing attention onto multiple sources of error and onto the costs of reducing each, the framework encourages a balanced reduction of error rather than the minimisation of any single component. Yet the classical TSE framework was developed mainly in the context of probability sample surveys conducted under relatively controlled conditions in high-resource environments. Artificial intelligence introduces layers of complexity that the classical framework only partly anticipates: opacity, automation bias, and dependence on training data whose provenance, representativeness, linguistic orientation, and cultural assumptions are often poorly understood by the survey team (Bender et al., 2021; Raji et al., 2020).

These new error sources are social and political because they shape whose knowledge is treated as authoritative. When an AI system trained predominantly on data from high-resource, English-dominant, urban contexts is used to code open-ended responses from rural multilingual communities, the resulting measurement error is structured by historical patterns of linguistic and epistemic exclusion. H-VALIDA builds on the TSE tradition by expanding the error taxonomy to include algorithmic error, linguistic error, cultural error, and documentation error. In data-scarce and multicultural environments these additional components can dominate total error.

2.2 Artificial Intelligence as a Sociotechnical System

Critical data studies show that AI systems are shaped by training data, institutional assumptions, linguistic structures, and historical power relations (Crawford, 2021; Latour, 2005). Data are never raw; they are constructed through social processes of collection, categorisation, and valuation. Large

language models trained on uncensored internet text risk generating plausible but unfounded content (Bender et al., 2021). Such models can reproduce stereotypes, amplify majority perspectives, and produce text that appears authoritative while lacking grounding in the specific social realities under investigation. High-level ethical principles alone cannot guarantee ethical AI; institutional accountability mechanisms are required (Mittelstadt, 2019; Floridi, 2019; UNESCO, 2021). Methodological evaluation must therefore reach beyond accuracy, precision, recall, and F1 scores to include ethical, political, social, and epistemological dimensions.

2.3 Algorithmic Bias, Explainability, and Auditing in Survey Contexts

Bias may arise from non-representative training data, flawed feature selection, inappropriate model architectures, or deployment in contexts for which systems were not designed (Barocas & Selbst, 2016; Abebe, 2020). In survey research the same problem appears as translation bias, classification bias, sampling bias, weighting bias, prediction bias, and interpretive bias. Urban, linguistic, gender, and socioeconomic biases are particularly relevant in developing-country settings.

Recent empirical work on large language models in survey and public-opinion research reports both promise and instability. Models can assist item drafting, classification, and synthetic response generation, yet performance varies sharply across languages, dialects, and culturally specific constructs (Argyle et al., 2023; Jansen et al., 2023; Ziems et al., 2024; von der Heyde et al., 2024). Automated translation frequently misses idiomatic expressions, code-switching, and locally embedded concepts (Hovy & Prabhumoye, 2021). Explainable AI and algorithmic auditing have developed as responses to opacity, but much of that literature assumes high-resource environments and access to internal model parameters (Raji et al., 2020). Proprietary models often function as black boxes. Under these conditions, external audit trails that document inputs, prompts, outputs, human interventions, and validation decisions offer a realistic path to accountability.

2.4 Human-Centred AI and Hybrid Methodologies

Human-centred AI emphasises systems that support rather than replace human decision-making (Shneiderman, 2022). Cultural meaning, linguistic nuance, community context, and ethical judgement often cannot be captured adequately by automated systems alone. Hybrid methodologies therefore position computational tools as assistants within broader systems of human oversight. The strongest research systems combine the computational strengths of AI with the interpretive, ethical, cultural, and contextual strengths of human judgement. That orientation matters most where the social and

political consequences of misrepresentation are high and where the capacity to correct errors after the fact is limited.

2.5 Data Scarcity, Multilingualism, and Decolonial Perspectives

Data scarcity means that available data may be incomplete, fragmented, outdated, or insufficiently representative. AI systems often require large, high-quality datasets if they are to perform reliably. When data are limited or biased, models produce unreliable outputs. Language is not merely a technical issue of translation accuracy; it bears directly on meaning, interpretation, trust, and validity. Decolonial approaches show that many contemporary systems reproduce colonial patterns: extraction of data from populations with little control over its use; imposition of categories developed in the Global North; concentration of technological capacity in a small number of actors; and marginalisation of non-Western knowledge systems (Mohamed et al., 2020).

Frameworks such as the CARE Principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) and OCAP (Ownership, Control, Access, and Possession) set out concrete requirements for ethical engagement with Indigenous data (Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019). H-VALIDA is designed in explicit dialogue with these concerns. Its emphasis on contextual sampling, local and bilingual review, source triangulation, open documentation, and ethical accountability is intended to resist the uncritical transfer of tools and categories from high-resource to data-scarce environments.

2.6 Research Gap and Positioning of H-VALIDA

Despite growing attention to ethical AI and human-centred design, limited operational guidance exists for validating AI-assisted survey systems in small developing states, multicultural societies, and contexts of data scarcity. Existing responsible-AI and survey-quality frameworks often assume high-resource capacities, relatively homogeneous linguistic environments, and access to internal model parameters or large audit datasets (UNESCO, 2021; Mittelstadt, 2019). H-VALIDA addresses this gap by offering an operational model tailored to constrained conditions. It does not claim to replace TSE, CARE, or human-centred AI. It integrates those traditions into a survey-lifecycle protocol that can be adopted in phases.

The present contribution is a conceptual and operational framework. It is not a completed empirical validation study. That distinction is held throughout the remainder of the article.

3. Research Methodology

3.1 Research Design and Positioning

This article presents a methodological design study. It does not report primary field data collected from human respondents. The research adopts a pragmatic orientation that combines qualitative inquiry, comparative methodological analysis, heuristic systems design, documentary analysis, and a structured appraisal protocol. The design is consistent with the tradition of methodological innovation research, in which the primary contribution is a transferable framework whose subsequent empirical testing constitutes a recommended agenda for future work (Lincoln & Guba, 1985).

The absence of primary human-subject data is a deliberate design choice at this stage of the work. It is also a limitation. Heuristic appraisal and a worked scenario cannot stand in for field implementation. The article therefore does not claim that H-VALIDA has been empirically validated at national scale. What it does claim is that the framework is conceptually coherent, operationally specified, and prospectively advantageous relative to ungoverned automation.

3.2 Framework Development Process

H-VALIDA was developed through an iterative five-stage protocol.

- Stage 1 consisted of a diagnostic analysis of documented weaknesses in traditional and AI-assisted surveys in data-scarce contexts, drawing on published evidence of bias, exclusion, opacity, translation failure, and automation bias.
- Stage 2 was a critical review of international models and standards for AI validation, algorithmic auditing, ethical data governance including the CARE Principles, and survey quality assurance under the TSE paradigm. Literature was identified through targeted searches of survey methodology, AI ethics, human-centred AI, multilingual NLP, and Indigenous data governance, with snowballing from core texts (Groves & Lyberg, 2010; Bender et al., 2021; Shneiderman, 2022; Mohamed et al., 2020). Inclusion favoured sources that addressed error control, accountability, multilingual practice, or data sovereignty. The review was purposive rather than a registered systematic review; that limitation is acknowledged.
- Stage 3 synthesised principles drawn from critical data studies, human-centred AI, systems theory, survey methodology, constructivist epistemology, and decolonial approaches.
- Stage 4 constructed seven interdependent pillars, each addressing a specific class of vulnerability and each mapped onto an expanded error taxonomy.

- Stage 5 operationalised each pillar into procedures applicable across the survey lifecycle, with attention to feasibility under resource constraints. The validation matrix in Section 4.4 is the principal operational artefact of this stage.

3.3 Appraisal Strategy: What Was and Was Not Done

No formal expert panel was convened. No proprietary or open large language model was systematically benchmarked on a labelled survey corpus. No inter-rater reliability coefficients, confidence intervals, or tests of statistical significance are reported. Terminology in earlier drafts that implied completed expert scoring or statistical demonstration has been withdrawn.

What the study does provide is a transparent appraisal protocol that future empirical work can apply. The protocol has four components.

First, a heuristic rubric with six criteria: reliability, transparency, ethical robustness, contextual sensitivity, applicability under resource constraint, and reproducibility. Each criterion is defined, linked to one or more pillars, and assigned an evidence requirement (Table 1).

Second, comparative logic contrasting a hybrid workflow with a purely automated baseline on the same hypothetical survey task. The baseline is defined as end-to-end use of a general-purpose large language model for item drafting, translation, open-ended coding, and interpretive summary, without structured human review, mixed-mode sampling, or an external audit log.

Third, a worked multilingual scenario (Section 4.6) that traces questionnaire support, translation, sampling, coding, review, auditing, documentation, and ethical checks through the seven pillars. The scenario is illustrative; it is not an experimental result.

Fourth, documentary triangulation against published evidence of AI failure modes in translation, classification, and representation. Internal consistency checks examined whether the seven pillars cohere as an interdependent system rather than as a loose list of recommendations.

This strategy establishes conceptual coherence, operational clarity, and a testable agenda. It does not establish statistical generalisation.

Table 1: Heuristic appraisal rubric for H-VALIDA (protocol for future empirical use)

Criterion	Definition	Primary pillars	Evidence required in field tests
Reliability	Stability of estimates and classifications under documented procedures	1, 4, 5	Agreement between AI draft codes and human final codes; test–retest on a subsample
Transparency	Capacity to reconstruct tools, prompts, outputs, and decisions	3, 6	Completeness of the living methodological log; independent auditor reconstruction
Ethical robustness	Consent, privacy, sovereignty, and CARE compliance	7	Consent language review; privacy risk assessment; community authority records

Contextual sensitivity	Fit to local language, geography, and knowledge systems	2, 4	Bilingual review notes; rural and Indigenous coverage rates
Applicability	Feasibility under limited staff, budget, and compute	All	Staff hours, tools used, phased adoption record
Reproducibility	Another team can follow the recorded pathway	3, 6	Independent reconstruction from the archive without undocumented steps

3.4 Ethical Considerations and Limitations

No primary human-subject data were collected. Principles of respect, transparency, fairness, accountability, privacy, and inclusion guided the conceptual work. The principal limitation is the absence of full national field implementation. A second limitation is contextual heterogeneity: data-scarce environments are not homogeneous, and transferability must be tested rather than assumed. A third limitation is the purposive literature base. Additional empirical studies of LLM-assisted item writing, coding, and translation should continue to update the framework.

4. The H-VALIDA Framework

4.1 Overview

H-VALIDA is a transparent heuristic methodology designed to support reliability, contextual sensitivity, traceability, reproducibility, and ethical legitimacy in AI-assisted surveys in developing and data-constrained contexts. Artificial intelligence functions not as a replacement for human judgement but as a collaborative tool operating within accountable systems of human oversight. The seven pillars are interdependent: weakness in one undermines the others. The framework is modular so that institutions with limited resources can adopt pillars sequentially.

The name encodes the core commitments: Hybrid human-AI collaboration; Validation and Auditing; Localized adaptation; Interpretable outputs; Documented process; and Accountable responsibility. Those commitments structure every stage of the survey lifecycle.

4.2 Mapping the Seven Pillars to an Expanded Total Survey Error Taxonomy

A distinctive contribution of H-VALIDA is the explicit mapping of its pillars onto an expanded TSE taxonomy. The mapping shows that the framework is a systematic strategy for controlling major error sources that arise when artificial intelligence enters survey practice under conditions of data scarcity and cultural complexity (Table 2).

Table 2: Mapping of H-VALIDA pillars to an expanded total survey error taxonomy

Pillar	Primary error targets	Principal control mechanism
1. Hybrid Human-AI Collaboration	Automation bias; processing error	Pre-specified division of labour; humans retain final authority
2. Contextual Sampling	Coverage error; nonresponse error	Mixed-mode design; inclusion of digitally excluded groups
3. Algorithmic Auditing	Algorithmic error; opacity	External audit trail of tools, versions, prompts, and interventions
4. Human Validation	Measurement, linguistic, and cultural error	Structured review by local and bilingual reviewers
5. Source Triangulation	Data scarcity effects; specification error	Multiple sources; discrepancies treated as analytic signals
6. Open Documentation	Documentation error; irreproducibility	Living methodological log archived with the dataset
7. Ethical Accountability	Ethical and governance harms beyond classical TSE	Consent including AI disclosure; CARE; institutional responsibility

The pillars do not replace classical error control.
They operationalise it under new technological and social conditions.

4.3 The Seven Pillars in Detail

Pillar 1: Hybrid Human-AI Collaboration

AI supports rather than replaces human judgement. Computational tools may assist with data organisation, preliminary coding, pattern detection, translation support, anomaly identification, and statistical interpretation. Critical decision points remain under human authority: questionnaire finalisation, translation approval, coding category definition, interpretation of ambiguous responses, and final analytical conclusions. The working principle is that AI may propose and humans dispose (Shneiderman, 2022).

The division of labour must be specified in advance and documented. Before a survey begins, the research team should define which tasks will be assigned to AI systems, which will remain exclusively human, and which will involve sequential human-AI interaction. That specification is intended to prevent the gradual, unexamined expansion of automation into domains where human judgement is essential.

Pillar 2: Contextual Sampling

Sampling strategies must take account of geographic diversity, rural-urban differences, linguistic communities, Indigenous populations, and digitally excluded groups. Mixed-mode data collection, combining face-to-face, telephone, and digital modes, is recommended where digital inequality is significant. Representativeness cannot be achieved by digital convenience alone. In data-scarce environments, sample design is an ethical act, because exclusion from the sample is equivalent to exclusion from knowledge and, ultimately, from policy attention.

Pillar 3: Algorithmic Auditing

Researchers must document the tool used, the version, the prompts entered, the outputs generated, the human review status, the errors corrected, and the conditions of reproducibility. Even when proprietary systems limit internal explainability, external documentation of inputs, outputs, and human interventions creates an audit trail (Raji et al., 2020). Most commercially available large language models are proprietary. Under those conditions, the only feasible form of auditability is systematic recording of what was done. Auditing under H-VALIDA is a continuous practice integrated into every stage of the survey lifecycle.

Pillar 4: Human Validation

Human validation is indispensable for correcting culturally insensitive categorisations, identifying hidden biases, validating multilingual interpretations, and strengthening the legitimacy of analytical outputs. Validation may involve researchers, subject experts, community representatives, bilingual reviewers, and local stakeholders. Structured review protocols specify what is to be reviewed, by whom, against what criteria, and with what authority to revise. Human validation is a constitutive element of methodological validity, not a residual quality check performed after automation has done the real work.

Pillar 5: Source Triangulation

When large, high-quality datasets are unavailable, researchers must rely on multiple sources of evidence, expert review, and clear reporting of limitations. A single AI-generated classification based on incomplete training data is a weak foundation for inference. The same classification, once cross-checked against administrative records, expert judgement, community feedback, and alternative analytical methods, becomes more robust. Triangulation does not eliminate uncertainty, but it makes uncertainty visible and manageable.

Pillar 6: Open Documentation

Every methodological decision, AI process, data source, prompt, model version, and validation step must be recorded. A living methodological log, updated at each stage and archived with the final dataset and report, enables subsequent researchers and auditors to reconstruct the analytical pathway. Open documentation does not require that every internal parameter of a proprietary model be disclosed. It does require that the research team's interactions with those models, and the decisions made on the basis of their outputs, be fully recorded.

Pillar 7: Ethical Accountability and the CARE Principles

Ethical responsibility must remain clear throughout the survey process. Key concerns include informed consent, including disclosure of AI use; privacy; confidentiality; data ownership; algorithmic discrimination;

surveillance risks; and respondent autonomy. In small societies, privacy concerns are heightened because individuals and communities may be more easily identifiable even when formal anonymisation procedures are applied. Institutions cannot outsource ethical responsibility to AI vendors.

H-VALIDA incorporates the CARE Principles. Collective Benefit requires that data ecosystems enable communities to derive benefit from the data. Authority to Control recognises community rights in governing collection, ownership, and application. Responsibility requires those working with community data to demonstrate how use supports self-determination. Ethics requires that rights and wellbeing remain the primary concern at all stages (Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019).

4.4 The Validation Matrix as an Operational Instrument

Section 4.4 of the earlier manuscript named a validation matrix without presenting one. Table 3 supplies the missing instrument. Institutions may add rows for multilingual review quality, community feedback, data-protection compliance, and longitudinal consistency. The matrix is a practical instrument rather than a rigid checklist. Thresholds are starting points and should be calibrated locally.

Table 3: H-VALIDA validation matrix

Dimension	Indicator	Responsible party	Evidence required	Provisional threshold	Corrective action
Sampling	Rural, linguistic, and digitally excluded coverage versus frame	Sampling lead	Mode mix table; frame notes	No major group omitted without documented reason	Add face-to-face or community liaison mode
AI coding	Share of AI draft codes revised by human reviewers	Coding supervisor	Double-coded subsample; revision log	Review 100% of sensitive items; at least 20% of remainder	Retrain codebook; change prompts; increase review
Translation	Bilingual sign-off on meaning, not only wording	Bilingual reviewers	Back-translation notes; idiom log	Zero unresolved culturally loaded items	Revise item; add local example; drop item
Documentation	Completeness of tool, version, prompt, output, and decision fields	Project manager	Living log completeness check	No undocumented AI interaction in the production path	Halt analysis until log is complete
Interpretation	Triangulation of AI summary with alternative sources	Lead analyst	Discrepancy memo	All material discrepancies recorded and adjudicated	Qualify or withdraw the claim
Ethics and CARE	AI disclosure in consent; community authority where applicable	Ethics officer	Consent form; data-governance record	Explicit AI disclosure; documented authority pathway	Do not field until consent and authority are in place

4.5 Comparative Logic: Hybrid versus Purely Automated Workflows

Heuristic comparison of a hybrid workflow with a purely automated baseline suggests, as a theoretical proposition, that human oversight is likely to improve detection of culturally insensitive categorisations, hidden biases, and multilingual misinterpretation. That proposition is consistent with published evidence on automation bias, translation failure, and representation harm (Bender et al., 2021; Hovy & Prabhumoye, 2021; Barocas & Selbst, 2016). It is not a statistical finding of the present study.

Key risks of ungoverned AI-assisted surveys include algorithmic bias from non-representative training data; urban bias from digitally mediated sampling; linguistic bias against non-dominant languages or mixed-language responses; cultural misinterpretation of local expressions; false objectivity created by apparently scientific outputs based on incomplete data; opacity of decision processes; ethical harms related to consent, privacy, and autonomy; and institutional dependency on externally designed systems. These risks are ethical and political because they shape whose realities are recognised as legitimate knowledge.

4.6 Worked Scenario: A Multilingual Livelihoods Module in a Small State

The following scenario is illustrative. It is not an experimental result. Consider a national statistical office in a small multilingual state that wishes to field a short livelihoods module covering informal work, subsistence production, and climate-related disruption. Official languages include English and a widely spoken Creole; several Indigenous languages are used in rural districts. Internet access is uneven.

Under Pillar 1, the team assigns a general-purpose large language model a drafting role only. The model proposes item wordings and a first-pass codebook. Questionnaire finalisation, translation approval, and final codes remain human decisions.

Under Pillar 2, the sample combines a digital option for urban respondents with face-to-face enumeration in rural districts and community-facilitated sessions where connectivity is poor. Digital convenience sampling is rejected as a sole mode.

Under Pillar 3, the living log records model name, version date, every prompt, every model output used in production, and every human override.

Under Pillar 4, bilingual reviewers examine Creole and Indigenous versions for meaning rather than lexical equivalence. A model rendering of a Creole phrase about informal sharing of catch as wage labour would be flagged and recoded as reciprocal exchange.

Under Pillar 5, AI-generated summaries of open-ended responses are compared with agricultural extension notes and community meeting records.

Discrepancies are written into an analytic memo rather than silently resolved in favour of the model.

Under Pillar 6, the log, codebook, translation notes, and discrepancy memo are archived with the microdata. Under Pillar 7, the consent script states that some text may be processed with computational assistance under human review; community authorities are consulted where Indigenous respondents are involved; and small-area disclosure risk is assessed before any public release. The scenario shows how the pillars operate as a sequence of controls rather than as a list of aspirations.

5. Discussion

5.1 Interpreting the Framework

The development of H-VALIDA supports the argument that hybrid methodologies embedding AI within transparent, accountable, human-governed structures are a more defensible basis for AI-assisted surveys in data-scarce environments than ungoverned automation. Purely automated systems risk producing outputs that appear efficient and scientific while remaining insufficiently transparent, insufficiently reviewed, and insufficiently adapted to local realities.

Validity is treated as a relational and contextual achievement rather than as a property of the algorithm alone. The interdependence of the seven pillars is a central design feature. A survey that implements rigorous algorithmic auditing but neglects contextual sampling will still produce biased results. A survey that emphasises human validation but fails to document AI processes will lack reproducibility. A survey that documents everything but fails to secure meaningful informed consent or respect data sovereignty will violate ethical standards even if its statistical outputs are internally consistent.

5.2 Implications for Survey Methodology

H-VALIDA expands traditional concepts of validity, reliability, and representativeness. In addition to statistical criteria, the framework introduces algorithmic, contextual, linguistic, cultural, ethical, interpretive, and documentation validity as dimensions of applied quality. A technically consistent survey may still be unreliable if it excludes rural communities, misinterprets local language, overlooks Indigenous perspectives, or fails to account for digital inequality. Mapping the pillars onto an expanded TSE taxonomy allows survey methodologists already familiar with TSE concepts to locate the new requirements of AI-assisted practice within a familiar architecture (Groves & Lyberg, 2010; Biemer, 2010).

5.3 Implications for Human-Centred AI and Data Governance

The framework shows how computational systems can be specified to support human decision-making rather than replace it (Shneiderman, 2022). Documentation of prompts, model versions, outputs, validation decisions, and limitations is a practical mechanism for reproducibility even when proprietary systems limit internal explainability. The CARE Principles strengthen the ethical pillar by providing community-centred criteria for data governance. In resource-constrained settings, external documentation combined with CARE-informed governance is often the most feasible pathway to accountability.

5.4 Risks of Not Applying Hybrid Controls

If AI-assisted surveys are implemented without hybrid controls, vulnerable groups may be excluded from samples and from the knowledge those samples produce. AI outputs may embed biases invisible to researchers who lack contextual knowledge. Policy decisions may rest on inaccurate or incomplete data. Public trust in official statistics may decline. Privacy may be compromised, particularly in small societies. These risks are acute where the capacity to detect and correct errors after the fact is limited. In data-scarce settings, a single flawed survey may constitute the only available evidence on a given issue for years.

5.5 Relation to Existing Frameworks

H-VALIDA does not displace TSE, human-centred AI, algorithmic auditing, or CARE. TSE supplies the error-control architecture. Human-centred AI supplies the division of authority between tools and people. Algorithmic auditing supplies the reconstructability requirement. CARE supplies community-centred data governance. The contribution is to bind these traditions into a survey-lifecycle protocol that is usable where internal model access, large labelled corpora, and specialist audit units are unavailable.

5.6 Limitations

The principal limitation is the absence of full empirical field testing. Simulations and heuristic protocols cannot fully reproduce the complexity of real fieldwork, institutional decision-making, or respondent behaviour. Claims for H-VALIDA are therefore prospective. A second limitation is contextual diversity. Operational protocols will need local tailoring. A third limitation is the literature base, which, although expanded in this revision, cannot exhaust the rapidly growing empirical literature on LLM-assisted survey tasks.

Conclusions

Responsible use of artificial intelligence in survey research requires more than technological efficiency. AI-assisted surveys must be governed by

transparency, human validation, contextual sensitivity, ethical accountability, and reproducibility. The central contribution of this article is the articulation of H-VALIDA as a conceptually specified and operationally usable framework for data-scarce and culturally complex environments.

The following conclusions are offered as theoretically grounded propositions, not as experimentally demonstrated findings.

1. AI systems are sociotechnical and not inherently objective.
2. Contextual knowledge is essential for valid interpretation.
3. Transparency is a prerequisite for methodological legitimacy and public trust.
4. Human oversight is a necessary control for automation bias in multilingual and multicultural environments.
5. Survey quality is both statistical and socially constructed.
6. Ethical governance is inseparable from technological implementation; the CARE Principles provide concrete guidance.
7. Data scarcity requires methodological adaptation rather than resignation.
8. The TSE tradition remains a valuable foundation but must be expanded to address algorithmic, linguistic, cultural, and documentation error.

Recommendations for Practice

- Adopt hybrid human-AI workflows rather than fully automated systems, with a written division of labour.
- Maintain a living methodological log of tools, versions, prompts, translations, codes, and validation steps.
- Use mixed-mode sampling to reduce digital and geographic exclusion.
- Require structured human validation of AI-generated classifications, translations, and interpretations, preferably involving local and bilingual reviewers.
- Conduct external algorithmic audits focused on linguistic, cultural, urban, socioeconomic, and gender bias.
- Establish ethical protocols covering informed consent including disclosure of AI use, privacy, data ownership, and protection of vulnerable groups.
- Apply the CARE Principles where Indigenous or historically marginalised communities are involved.
- Treat the seven pillars as interdependent while allowing phased adoption according to institutional capacity.

Recommendations for Future Research

- Conduct full field tests of H-VALIDA across diverse communities, with pre-registered protocols and publicly archived logs.
- Measure agreement between AI draft codes and human final codes, revision rates, coverage by mode, and time costs.
- Compare named general-purpose and domain models against human-only and hybrid conditions on multilingual survey tasks.
- Develop sector-specific operational protocols and training materials for national statistical offices.
- Test transferability in other small developing states and multilingual societies, including community-based validation studies.

Closing Remark

Technological sophistication alone cannot guarantee truth, fairness, or social legitimacy. Responsible AI requires methodological humility: the recognition that computational power does not substitute for contextual understanding, ethical judgement, or democratic accountability. H-VALIDA proposes collaborative intelligence in which humans and machines coexist within accountable systems of knowledge production. In data-scarce, multilingual, and culturally complex environments, validity is a sociotechnical achievement, ethics is integral to methodology, and the populations whose lives are shaped by survey knowledge deserve systems that represent them accurately, respectfully, and accountably.

Author Contributions

The sole author conceived and designed the work, synthesised the literature, developed the framework, devised the appraisal strategy, drafted the manuscript, and revised the intellectual content. The author approved the final version and accepts accountability for every aspect of the study.

Conflict of Interest: The author reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The author is grateful to colleagues and practitioners who offered informal comments while the framework was taking shape. The author did not obtain any funding for this research.

References:

1. Abebe, R. (2020). From AI for good to AI for local good. *Patterns*, 1(2), 100012. <https://doi.org/10.1016/j.patter.2020.100012>

2. Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3), 337–351. <https://doi.org/10.1017/pan.2023.2>
3. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
4. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>
5. Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817–848. <https://doi.org/10.1093/poq/nfq058>
6. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
7. Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1, 261–262. <https://doi.org/10.1038/s42256-019-0064-x>
8. Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74(5), 849–879. <https://doi.org/10.1093/poq/nfq065>
9. Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, 15(8), e12432. <https://doi.org/10.1111/lnc3.12432>
10. Jansen, B. J., Jung, S., & Salminen, J. (2023). Employing large language models in survey research. *Natural Language Processing Journal*, 4, 100020. <https://doi.org/10.1016/j.nlp.2023.100020>
11. Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
12. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
13. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
14. Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
15. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
16. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020).

- Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). ACM. <https://doi.org/10.1145/3351095.3372873>
17. Research Data Alliance International Indigenous Data Sovereignty Interest Group. (2019). CARE Principles for Indigenous Data Governance. Global Indigenous Data Alliance. <https://www.gida-global.org/care>
 18. Shneiderman, B. (2022). Human-centered AI. Oxford University Press.
 19. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.
 20. von der Heyde, L., Haensch, A.-C., & Wenz, A. (2024). United in diversity? Contextual biases in LLM-based predictions of the 2024 European Parliament elections. arXiv. <https://doi.org/10.48550/arXiv.2409.09045>
 21. Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? Computational Linguistics, 50(1), 237–291. https://doi.org/10.1162/coli_a_00502