

Policy Pathways for Responsible AI-Assisted Surveys in Small Developing States: Institutional Adoption of H-VALIDA: A Conceptual Policy Analysis

Dr. Florence Yasmine Andrews, PhD

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

[Doi:10.19044/esj.2026.v22n28p25](https://doi.org/10.19044/esj.2026.v22n28p25)

Submitted: 13 August 2026

Accepted: 30 September 2026

Published: 31 October 2026

Copyright 2026 Author(s)

Under Creative Commons CC-BY 4.0

OPEN ACCESS

Cite As:

Andrews, F. Y. (2026). *Policy Pathways for Responsible AI-Assisted Surveys in Small Developing States: Institutional Adoption of H-VALIDA: A Conceptual Policy Analysis*. European Scientific Journal, ESJ, 22 (28), 25. <https://doi.org/10.19044/esj.2026.v22n28p25>

Abstract

Methodological novelty seldom moves public decisions while it stays locked inside academic conversation. Survey systems that lean on artificial intelligence can support planning that is actually evidence-led, but only once transparency, human checking, local fit, and ethical answerability have been written into everyday procedure rather than left as slogans. This paper is a conceptual and normative policy reading of the Hybrid Validation and Auditing for Localized, Interpretable, Documented, and Accountable AI-Assisted Surveys framework, abbreviated here as H-VALIDA. It reports neither a field trial nor original survey microdata. Documents were gathered through a purposive search organised in four clusters fixed in advance: responsible and human-centred AI; survey methods and official-statistics quality; Indigenous data governance and data justice; and small-state, Caribbean, and Belize statistical and digital-policy texts. Widely cited international instruments were kept where they function as governance reference points rather than as local manuals. The reading itself combined heuristic use of published criteria, structured scenario work, comparative conceptual placement, documentary triangulation, and internal consistency checks across eight institutional domains. Five H-VALIDA pieces: localisation, interpretability, documentation, accountability, and hybrid validation, are set out as a minimum production protocol for offices working

with thin statistical capacity, multilingual measurement, and donor-driven calendars. A five-phase adoption sequence is offered only as a planning sketch, together with operational floors for human review, documentation artefacts, role assignment, and implementation metrics. Predictable breaking points include budget shortfalls, staff turnover, political discontinuity, and external reporting pressure. The argument ends where it must: responsible uptake is institutionally thinkable if Phase 1 safeguards are treated as non-negotiable. Effectiveness claims stay propositions until someone tests them.

Keywords: H-VALIDA; Responsible AI; Digital Governance; Institutional Adoption; Survey Methodology; Small Developing States; Small Island Developing States; Data Equity; Conceptual Framework

1. Introduction

1.1. Background

National development planning still depends on information that is at least reasonably trustworthy: poverty, health, schooling, work, climate exposure, and how institutions themselves perform. AI tools now sit inside survey design, instrument translation, coding of open answers, and the construction of sampling frames. Speed and cost gains are not imaginary. They arrive, however, only if quality, inclusion, and public trust are kept intact.

In small developing states the cost of getting this wrong sits higher than a textbook discussion of statistical error usually admits. Tight fiscal space, thin offices, climate and economic exposure, and historically patchy statistical coverage mean that mistakes can send scarce public money in the wrong direction and leave rural, Indigenous, older, or digitally excluded people out of view (World Bank, 2022; United Nations, 2023). In Belize those pressures meet a multilingual social field in which English, Belizean Kriol, Spanish, Maya languages, and Garifuna sit side by side, and a geography that pulls urban centres away from rural and cay communities (Statistical Institute of Belize [SIB], 2022; Government of Belize, 2022).

The 2022 Population and Housing Census counted a household population of 397,483 persons. Of that total, 57.8 percent lived in rural households. Ethnic self-identification included Mestizo/Hispanic/Latino (51.7 percent), Creole (25.2 percent), Maya (9.8 percent), and Garifuna (4.0 percent) (SIB, 2024). Census forms already record ethnicity, language, disability, internet access, and economic activity. An AI-assisted coding or translation protocol that is tone-deaf to those linguistic and geographic facts is more likely to produce systematic measurement error than a merely ethical lapse.

Many national statistical offices in small states grew after independence with outside help aimed at censuses and basic administrative series. Survey capacity often arrived project by project. Generative AI

therefore lands on foundations that were already stretched. Vendor performance claims commonly omit training-data composition or fitness for low-resource language varieties. Offices without independent evaluation capacity are poorly placed to test those claims (Jobin et al., 2019; Mittelstadt, 2019).

1.2 Problem, Scope, and Article Type

The problem at the centre of the paper is the distance between methodological proposals and what offices actually take up. Ethical AI guidance often presupposes organisational machinery, legal frameworks, and technical literacy that small statistical offices cannot staff or pay for (Jobin et al., 2019; Mittelstadt, 2019). Tools can be adopted in weeks. Protocols, skills, and community trust take years. External partners, for their part, frequently commission surveys against reporting calendars that reward rapid digital shortcuts.

What follows is a design-oriented conceptual and normative policy analysis. It is not a systematic review, not a field evaluation, and not an assertion that H-VALIDA has already been stress-tested in live survey shops. The paper builds a capacity-aware prototype pathway rather than a proven implementation model. Limits of capacity, political discontinuity, and technological change are treated as part of the problem, not as afterthoughts. H-VALIDA was first specified as a methodological packaging of localisation, interpretability, documentation, accountability, and hybrid validation for AI-assisted survey production. This article does not invent those principles. What it adds is the translation of that packaging into an eight-domain institutional pathway, an indicative phased sequence, operational floors for validation and documentation, and a set of failure points particular to small developing states and Small Island Developing States (SIDS). Whether the packaging actually reduces error remains an empirical question for later work.

1.3 Objectives and Contribution

Four objectives organise the paper. First, the policy implications of the proposed framework are examined across eight institutional domains. Second, a sequenced adoption pathway is constructed as a hypothetical prototype that institutions may adapt. Third, risks of ungoverned adoption and anticipated failure points are identified, and minimum operational controls are specified. Fourth, the paper seeks to contribute a context-sensitive model for small-state discussions of ethical AI governance.

The intended contribution is both practical and theoretical. Practically, the pathway and domain recommendations offer a template rather than another list of abstract principles. Theoretically, the analysis questions the assumption that ethical AI governance must begin with comprehensive technical

infrastructure. Institutional commitment to documentation, human oversight, and inclusion can provide a foundation even under severe resource constraints. Small developing states are positioned not only as recipients of externally designed systems but as possible sites of modest governance experimentation. That claim is offered as an analytically derived proposition, not as an empirically demonstrated effect.

2. Literature and Policy Context

2.1 Digital Governance and Responsible AI

High-level ethical principles cannot, by themselves, guarantee ethical AI; institutional accountability mechanisms are required (Mittelstadt, 2019; Floridi, 2019). Work on algorithmic governance has shown how opacity can undercut democratic answerability (Crawford, 2021; Noble, 2018). In small states, where technical expertise is concentrated in a handful of officers, undocumented model versions and unlogged prompts become especially hazardous.

International instruments supply normative orientation rather than shop-floor instructions. The UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021) and the OECD AI Principles, revised in 2024 to address generative and general-purpose systems, information integrity, and downstream reuse (OECD, 2019/2024), remain high-level. The National Institute of Standards and Technology (NIST) AI Risk Management Framework offers a process vocabulary of map, measure, manage, and govern (NIST, 2023). The European Union Artificial Intelligence Act introduces risk-tiered obligations that will shape vendor behaviour even outside the Union (European Union, 2024). UNESCO's Caribbean Artificial Intelligence Policy Roadmap translates similar concerns into a regional agenda organised around culture and creativity, governance and transformation, education and upskilling, and resiliency and sustainability (UNESCO, 2024). None of these texts tells a small statistical office how to document a translation prompt, sample AI-coded livelihood answers, or protect a local evaluation dataset through a change of government.

2.2 Survey Quality, Measurement Error, and Official Statistics

Because the object of concern is AI-assisted surveys, the analysis also sits inside survey methodology. The total survey error framework treats coverage, sampling, nonresponse, measurement, and processing error as jointly relevant to quality under budget constraints (Biemer, 2010; Groves et al., 2009). Mixed-mode designs can cut coverage and nonresponse error while introducing mode-related measurement differences (de Leeuw, 2005; Biemer, 2010). Official-statistics quality frameworks emphasise relevance, accuracy,

timeliness, coherence, and accessibility, together with process documentation (United Nations, 2014).

AI tools touch several error sources at once. Machine translation can inject measurement error that clusters in particular language communities. Automated coding of open-ended responses is a processing-error problem that resembles earlier computer-assisted coding, except that commercial language models are less inspectable and their mistakes may travel together across prompts and model families. Digital-first modes shift coverage toward younger and urban respondents. These are familiar survey-science problems wearing new technical clothes. Responsible AI guidance that ignores total survey error is therefore incomplete for statistical production.

2.3 Data Equity, Language, and Indigenous Data Governance

Survey systems help decide who is counted. Critical data studies have shown that categories, frames, and modes of contact are not neutral (Barocas & Selbst, 2016; Eubanks, 2018; Bender et al., 2021). Indigenous data sovereignty scholarship treats control over collection, ownership, and interpretation as a question of self-determination (Kukutai & Taylor, 2016; Carroll et al., 2020). In Belize, census instruments already record ethnicity and language, and rural–urban differences in living conditions remain large (SIB, 2022; SIB, 2025). Any AI-assisted coding or translation protocol that is insensitive to Kriol, Spanish, Maya languages, or Garifuna risks systematic measurement error.

2.4 Small-State Statistical Systems and the Belize Reference Case

World Bank and United Nations writing on digital development in small states emphasises thin institutional capacity, dependence on external technical assistance, and limited local-language resources (World Bank, 2022; United Nations, 2023). A 2024 statistical needs assessment of the Statistical Institute of Belize reported progress in human resources and regular survey administration, alongside remaining constraints in coordination, administrative-data integration, and resource stability (World Bank, 2024). Recent household operations, including the 2022 census and the Multiple Indicator Cluster Survey, illustrate both domestic production capacity and high partner co-financing (SIB, 2022; SIB & UNICEF, 2025). That mix is typical of SIDS statistical systems: competence exists, but continuity is vulnerable to funding cycles and staff movement.

Belize also supplies a legal and digital-governance setting that should be distinguished from generic small-state inference. The Statistical Institute of Belize Act (2006) mandates impartial collection, analysis, and dissemination of official statistics. The Data Protection Act (2021) establishes rules for personal-data processing, consent, and limited research and statistics

exemptions (Government of Belize, 2021). The Digital Government Act (2022) creates an institutional basis for e-government services, interoperability, and privacy-consistent information access (Government of Belize, 2022b). These instruments do not yet specify AI-assisted survey protocols. They do, however, provide legal hooks to which documentation, human validation, and limitation reporting could be attached. Evidence about Belize is therefore treated as a reference case, not as a proxy for all SIDS.

2.5 Research Gap and Positioning of H-VALIDA

Practical guidance for translating responsible-AI principles into survey protocols in small developing states remains thin. Most frameworks either assume high-resource organisations or remain at the level of principle lists (Jobin et al., 2019). H-VALIDA is not presented as a replacement for the OECD principles, UNESCO guidance, NIST processes, model cards, or datasheets for datasets (Mitchell et al., 2019; Gebru et al., 2021). Those artefacts already exist. The proposed contribution is narrower: a hybrid survey-production protocol that sequences minimum operational controls for localisation, interpretability, documentation, accountability, and human-machine validation under SIDS constraints.

3. Research Methodology

3.1 Article Type

The study is a design-oriented conceptual and policy analysis. No primary field data were collected. Widespread institutional adoption of AI-assisted survey methods under explicit ethical frameworks remains emergent in small developing states. A prospective analysis can therefore supply a reference point against which later empirical studies may be compared. The article is labelled and structured as a conceptual policy analysis rather than as an empirical results paper.

3.2 Identification of Literature and Policy Documents

Sources were identified through purposive, not systematic, search conducted between 2023 and 2026. Four clusters were specified in advance: (a) responsible and human-centred AI and critical data studies; (b) survey methodology and official-statistics quality; (c) Indigenous data governance and data justice; and (d) small-state, Caribbean, and Belize statistical and digital-policy documents. Search venues included Google Scholar, publisher platforms for Nature Machine Intelligence, Public Opinion Quarterly, and related journals, the UNESCO Digital Library, OECD legal instruments, NIST publications, United Nations Statistical Commission texts, and official Belize government and SIB repositories.

Inclusion criteria were: relevance to AI governance, survey quality, data equity, or small-state statistical systems; publication in English; and sufficient bibliographic traceability to support citation. Exclusion criteria were: purely commercial vendor white papers without methodological disclosure; sources that addressed neither public-sector surveys nor governance; and unverifiable secondary summaries. International instruments were included when they are widely cited as governance references (UNESCO, OECD, NIST, United Nations Statistical Commission, European Union). Belize and Caribbean sources were added to discipline generic ethics language with institutional facts. The search is not claimed to be exhaustive. Recent generative-AI governance texts were prioritised where they address public-sector risk, documentation, or information integrity.

3.3 Derivation of Eight Domains and Five Components

The eight institutional domains were derived by mapping the organisations that typically commission, produce, fund, teach, or use official and project surveys in small states: national planning authorities; statistical offices and line agencies; universities; non-governmental organisations and development partners; digital-governance and procurement units; community and data-equity actors; Sustainable Development Goal (SDG) monitoring arrangements; and the cross-cutting problem of sequenced adoption. The five H-VALIDA components were derived by condensing recurring failure modes in the synthesised literature, context mismatch, opacity, missing process records, diffused responsibility, and unreviewed automation, into operational requirements for survey production.

3.4 Analytical Procedures

Five procedures were used. First, heuristic evaluation drew on published expert scholarship rather than on a convened expert panel. The phrase expert-informed therefore refers to criteria extracted from published work in survey methodology, human-centred AI, critical data studies, and Indigenous data governance (Shneiderman, 2022; Mittelstadt, 2019; Carroll et al., 2020; Biemer, 2010). No Delphi process or interview programme was conducted. Second, structured scenario reasoning examined four stylised situations common in small states. The scenarios are conceptual illustrations, not case studies. Third, comparative conceptual analysis positioned hybrid human–AI procedures against ungoverned automation with respect to transparency, interpretability, and accountability. No experimental claim of improved accuracy is made. Fourth, documentary triangulation cross-checked recommendations against international instruments and published accounts of statistical capacity. Fifth, internal consistency checks asked whether each H-VALIDA component could be traced across the eight domains.

Table 3 later summarises the resulting analytical matrix. The procedures establish a literature-grounded prototype. They do not constitute empirical validation.

3.5 Analytical Stance and Remaining Limits

The stance is normative: reliable, inclusive, and accountable public knowledge is treated as a responsibility of commissioning institutions. Recommendations favour sequenced steps over comprehensive blueprints. Rapid change in generative systems is acknowledged by focusing on durable procedural principles rather than on particular model names. Timelines in the pathway are planning assumptions, not evidence-derived implementation periods. Claims about feasibility, under-representation, vendor dependency, and the likely effects of safeguards are framed as propositions unless supported by cited documentary evidence.

4. Proposed Policy Framework and Analytical Propositions

The following outputs are analytically derived. They are not findings from field implementation. Section 4 is therefore presented as a proposed policy framework rather than as empirical results.

4.1 The H-VALIDA Components

H-VALIDA names five operational requirements for AI-assisted survey production. Table 1 states each component, its conceptual source, its function in survey work, and its relationship to existing responsible-AI artefacts.

Table 1: H-VALIDA Components in Relation to Existing Responsible-AI Approaches

Component	Conceptual foundation	Function in AI-assisted surveys	Relation to existing approaches
Localisation	Critical data studies; Indigenous data governance; survey translation practice	Adapt categories, language, and mode design to Kriol, Spanish, Maya languages, Garifuna, and rural–cay geographies	Treats linguistic and geographic adaptation as measurement-error control, not only as ethics
Interpretability	Human-centred AI; official-statistics quality	Require that classifications, translations, and sampling suggestions can be explained to a statistical officer	Specified as officer-level intelligibility rather than public model interpretability
Documentation	Datasheets; model cards; statistical process documentation	Record tool, version, prompt, date, operator, and known limits for each AI-assisted step	Adapts Gebru et al. (2021) and Mitchell et al. (2019) to survey workflow artefacts
Accountability	Algorithmic governance; public-purpose mandate of official statistics	Assign a named human owner for final codes, translations, and published indicators	Applies Mittelstadt’s argument that principles need institutional owners to the commissioning office

Hybrid validation	Total survey error; human-in-the-loop design	Treat automation as a first pass subject to sampled human review and, where feasible, community or bilingual checks	Validation intensity is scaled to capacity, while a non-zero human sample remains mandatory
-------------------	--	---	---

Note: The table specifies a proposed prototype. It does not claim empirical superiority over existing frameworks.

The distinctive claim is therefore modest. H-VALIDA does not invent model cards, human oversight, or Indigenous data principles. It packages those requirements as a minimum survey-production protocol for offices that cannot staff an AI ethics board, cannot train a local large language model, and cannot halt donor-funded fieldwork while comprehensive audits are designed. Whether the package reduces error is unknown until tested.

4.2 Operational Minimums

Reviewer concern that the framework remained underspecified is addressed here as a set of provisional operational minimums. These thresholds are planning devices, not empirically calibrated performance standards.

Human validation. For every AI-assisted coding or classification task that feeds published indicators, a stratified human sample should be independently double-coded. As a starting rule of thumb, offices may review at least 10 percent of records or 100 records, whichever is larger, stratified by language community, urban-rural location, and any category known to be conceptually unstable (for example, informal livelihoods). Where the universe is smaller than 100 records, full human review is preferred. Discrepancy rates should be tabulated by stratum. A discrepancy rate above 10 percent in any priority stratum is proposed as an escalation trigger requiring prompt revision, additional bilingual review, or withdrawal of automation for that task. These figures are illustrative and should be adjusted after local piloting.

Translation. Machine translation is treated as a draft only. Independent bilingual review of the full instrument, plus a cognitive pilot in each major language community, remains the validity-critical step. Residual conceptual mismatches should be disclosed in the methodological annex.

Documentation artefacts. Each AI-assisted step should generate a one-page process record stating: tool name and version; date; operator; prompt or parameter set; purpose of use; sample reviewed; discrepancy summary; named approving officer; and known limits. Records should be stored as institutional files, not personal folders.

Roles. The commissioning statistical office retains non-delegable responsibility for published codes, translations, and indicators. A named survey manager approves AI use. A second officer or university reviewer signs the human-sample check. Vendors may supply tools; they do not own official estimates.

Implementation metrics. Progress can be tracked by the share of AI-assisted surveys with complete tool and prompt records; the share with recorded human sample review; the existence and currency of local evaluation datasets; incorporation into standard operating procedures; and the frequency with which residual limitations appear beside published indicators. These metrics measure adoption of procedures, not improvement in estimate quality.

4.3 Implications across Eight Institutional Domains

National policy and planning. A short national AI-in-statistics guideline can establish non-negotiable expectations: documentation of tools and versions, human validation of automated outputs, bias and language review, and disclosure of residual limitations. In small states the guideline also coordinates otherwise fragmented project practice. Inclusive consultation with line ministries, civil society, and community representatives would reduce the risk that the document remains a planning-ministry artefact. Monitoring can be light: periodic self-assessment, occasional university review, and public reporting of compliance.

Government agencies and statistical institutions. Internal protocols should begin with mandatory documentation and human review, then expand toward structured auditing as capacity allows. Short, practice-oriented training on prompt logging, sample double-coding, and recognition of linguistic mismatch is more feasible than specialised degree programmes. Mixed-mode collection remains necessary because digital-only designs can exclude rural households, older adults, and disconnected communities (de Leeuw, 2005). An internal register of tools and modest local evaluation datasets, previously coded responses, translated items, contested livelihood terms, would give offices a limited form of methodological sovereignty. Those datasets are also fragile: staff turnover can erase tacit knowledge of how they were built.

Universities and research institutions. Universities can embed the five components in methods courses and ethics review, and can act as independent reviewers of documentation quality. In Belize and similar systems they are often one of the few research concentrations outside government. Collaborative construction of local-language evaluation sets would be a concrete joint product.

Non-governmental organisations and development partners. Partners frequently commission surveys under time pressure. If documentation, human review, and inclusion checks become contractual quality criteria, external finance can reinforce rather than erode domestic standards. Alignment among major partners and the national statistical office is more important than any single agency circular.

Digital governance and procurement. Tender specifications can require prompt logging, version disclosure, contractual limits on secondary

training use, and demonstration of performance on local-language samples. Preference for inspectable tools is a sovereignty measure, not a demand that every model be trained domestically. In Belize, procurement language can be aligned with the Data Protection Act (2021) and the Digital Government Act (2022) so that survey microdata are not reused for model training without separate informed consent.

Community engagement and data equity. Sampling and contact strategies should reach rural communities, Maya and Garifuna populations, cay settlements, low-income households, older adults, persons with disabilities, and digitally excluded groups. Machine translation, if used, is a draft. Bilingual human review and cognitive pilots remain the validity-critical step. Feedback mechanisms that allow communities to contest misrepresentation support legitimacy, particularly where small-area reporting can identify groups.

Sustainable Development Goal monitoring. Methodological annexes covering tools, validation steps, and residual risks should accompany AI-assisted indicators. International compilers can treat such annexes as a quality criterion.

Practical adoption. Full simultaneous implementation will exceed many offices. Section 4.4 therefore sequences work. A portfolio approach is also possible: apply Phase 1 safeguards to all AI-assisted surveys, and reserve deeper validation for surveys that feed national planning or concern highly vulnerable populations.

4.4 Indicative Phased Pathway

The intervals below are planning assumptions. They are not derived from comparative implementation studies and should be adjusted to budget cycles, election calendars, and staffing. Phases may overlap. The pathway is a hypothetical prototype.

Phase 1, core safeguards (indicative 0-6 months): documentation of tools, versions, and prompts; human review of automated coding and translation; disclosure of AI use to respondents. These steps require procedural discipline more than specialised models and are treated as non-negotiable.

Phase 2, auditing and bias review (indicative 6-18 months): structured checks for linguistic, urban–rural, gender, and socioeconomic discrepancy; comparison of AI and human coding samples; reporting of limitations. Samples need not be large to reveal systematic mismatch.

Phase 3, contextual and community validation (indicative 12-30 months): mixed-mode strategies; local expert and bilingual review; limited community feedback pilots. External partners can usefully target this phase.

Phase 4, institutional integration (indicative 24-48 months): incorporation into standard operating procedures, ethics review, induction

training, and SDG protocols. Embedding in induction is one of the few realistic protections against turnover.

Phase 5, continuous review (ongoing): periodic updating as tools change; sharing of lessons. Review every two to three years is a working assumption.

Table 2: Indicative Phased Institutional Adoption Pathway

Phase	Core activities	Primary responsibility	Indicative timeline
1	Documentation; human review; ethical disclosure	Statistical office or survey unit	0-6 months
2	Bias checks; AI–human comparison; limitation reporting	Statistical office with university support	6-18 months
3	Mixed-mode design; local expert review; community feedback pilots	Multi-agency with community partners	12-30 months
4	Standard operating procedures; ethics integration; training; SDG protocols	National statistical system and universities	24-48 months
5	Longitudinal review; protocol updates; inter-institutional learning	All stakeholders	Ongoing

Note: Timelines are indicative planning assumptions and should be adjusted to local capacity, political calendars, and resource cycles. Phases may overlap.

4.5 Analytical Matrix

Table 3: Analytical Matrix Linking Risks, H-VALIDA Components, Sources, and Prototype Safeguards

Institutional risk	H-VALIDA component	Illustrative sources	Prototype safeguard
Linguistic measurement error in Kriol, Spanish, Maya, Garifuna instruments	Localisation; hybrid validation	Bender et al., 2021; Kukutai & Taylor, 2016; SIB, 2022	Machine translation as draft only; bilingual review; cognitive pilot
Opaque coding of open-ended livelihoods	Interpretability; documentation	Biemer, 2010; Mitchell et al., 2019; Mittelstadt, 2019	Logged prompts and versions; named officer signs final codes
Digital-first exclusion of rural and older respondents	Localisation; accountability	de Leeuw, 2005; Eubanks, 2018; World Bank, 2024	Mode design reviewed for coverage before efficiency claims
Donor timelines that skip validation	Accountability; documentation	World Bank, 2022; UNESCO, 2024; SIB & UNICEF, 2025	Minimum contractual clauses agreed with the national statistical office
Loss of evaluation datasets after staff exit	Documentation; hybrid validation	World Bank, 2024; Shneiderman, 2022	Datasets stored as institutional assets; procedures written into induction
Political discontinuity in a 48-month pathway	Accountability	United Nations, 2014; OECD, 2019/2024; Government of Belize, 2022b	Core Phase 1 rules placed in statistical legislation or quality frameworks

Note: The matrix is a conceptual synthesis, not an empirical risk register.

4.6 Anticipated Failure Points

Prescriptive language is easy to overstate. Several failure points are predictable. Severe budget shortfalls can reduce human review to a ritual signature. Political interference can pressure offices to publish faster than validation allows. Absolute shortages of bilingual or methods staff can make Phase 3 decorative. High turnover can destroy local evaluation datasets and the informal knowledge required to interpret them. Donor disbursement schedules can override domestic checklists. A change of minister can orphan a technical working group. The prototype therefore treats Phase 1 as the only element that should be described as non-negotiable. Later phases are optional intensifications, not a promise of smooth 48-month transition.

4.7 Risks of Ungoverned Adoption

If AI-assisted surveys proceed without corresponding safeguards, policy may rest on biased or incomplete data; vulnerable populations may remain statistically invisible; public trust may erode; digital transformation may deepen inequality; proprietary dependency may increase; and SDG indicators may lose credibility. These risks are acute where the margin for misallocated public expenditure is narrow. The statements in this subsection are risk hypotheses derived from the literature on opacity, exclusion, and official-statistics quality. They are not measured outcomes of AI use in Belize.

4.8 Illustrative Scenarios

Illustration A. A statistical office applies a commercial language model to code free-text livelihood responses. Informal and seasonal coastal livelihoods are misclassified. An H-VALIDA-consistent response would document the model and prompts, double-code a stratified sample with officers familiar with local terminology, analyse discrepancies, and report residual limits.

Illustration B. Machine translation is proposed for English, Spanish, and Kriol instruments. Conceptual mismatches concentrate error in Kriol-speaking communities. Translation is treated as validation-critical: the machine draft is reviewed by independent bilingual readers and tested in a cognitive pilot.

Illustration C. An employment survey planned as predominantly online over-represents younger urban respondents. Mode design is evaluated for representation before efficiency is prioritised, and a mixed-mode strategy is adopted.

Illustration D. A partner-commissioned baseline survey arrives with a short delivery window. Contractual alignment with national minimum standards requires documentation, human review, and a methodological annex, making later change estimates interpretable.

5. Discussion

5.1 Novelty and Humility

H-VALIDA is a proposed packaging of already circulating practices. Its possible value lies in specifying a minimum operational set for survey production in capacity-constrained, multilingual, donor-exposed systems. It does not uniquely solve SIDS constraints in the sense of a demonstrated technical fix. It proposes a different unit of design: the survey workflow of a small official-statistics office rather than the model card of a research laboratory or the principle list of an intergovernmental text.

5.2 Institutional Inertia and Political Economy

Section 4.6 listed failure points because high-level sponsorship in small states is often fleeting. Brain drain of technical staff is common. Partner reporting calendars frequently dictate tempo. Protecting protocols during political transitions therefore requires legal or quality-framework anchors, induction training that does not depend on a single champion, and shared custody of evaluation datasets with a university partner so that an election or a resignation does not erase institutional memory. Earmarking a modest share of every survey budget for validation is a more realistic defence than hoping that goodwill survives a fiscal shock.

Domestic political economy also matters. Offices with a clear legal mandate, some core funding, and visible professional identity are better placed to refuse undocumented automation. Offices that live project to project are not. The pathway cannot conjure the former from the latter; it can only state that Phase 1 remains worth attempting even when later phases stall.

5.3 Feasibility, Trade-offs, and Small-System Advantages

Documentation and human review consume staff time. Mixed-mode designs cost more than digital-only designs. Community validation adds stages. Those costs are real. The argument is that they are preferable to the costs of opaque or exclusionary indicators, not that they are free. Claims that smallness itself facilitates coordination, personal networks, or participatory validation are plausible but untested. They are stated here as hypotheses, not as institutional advantages already in hand.

5.4 Sovereignty, Trust, and Measurement of Progress

Sovereignty in this setting does not require local model training. It requires the capacity to understand, validate, contest, and reject outputs. Public trust is easier to lose than to rebuild. The process metrics listed in Section 4.2 can track whether the prototype is even being tried. Those metrics measure adoption of procedures, not improvement in estimate quality.

5.5 Transferability

Belize is the principal reference case because it combines multilingual measurement, rural–urban disparity, climate exposure, and a statistical office that already runs censuses and partner-supported household surveys. Similar configurations appear across Caribbean and other SIDS systems. Transfer still requires local legal, linguistic, and fiscal translation. Regional statistical organisations can lower the cost of that translation by sharing templates, not by imposing a single pathway.

Conclusions

Translating responsible-AI requirements into survey procedures is a governance problem rather than a software problem. The proposed H-VALIDA prototype specifies five operational components and an indicative phased pathway for small developing states. The article has not shown that the prototype works. It has argued that ungoverned automation is a poor default and that a minimum set of documentation, named human accountability, linguistic review, and sampled validation is institutionally conceivable even where comprehensive AI infrastructure is absent. Claims supported by the reviewed literature concern known survey-error sources, published capacity constraints, and the incompleteness of principle-only governance. Claims that remain the author’s normative proposals include the five-component packaging, the eight-domain mapping, the Phase 1 non-negotiable rule, and the illustrative validation thresholds. Matters requiring pilot testing include discrepancy rates, public trust, training effectiveness, and interaction with donor calendars.

Recommendations for National Policy

Governments may consider short national guidelines that treat documentation, human validation, language review, and limitation reporting as conditions of official statistical production; place those rules in statistical legislation or quality frameworks so that they survive political turnover; align partner requirements with the same minimum; and designate a small technical working group whose products are templates rather than speeches.

Recommendations for Institutional Practice

Agencies, universities, non-governmental organisations, and partners may consider beginning with Phase 1 safeguards on every AI-assisted survey; maintaining a register of tools; treating mixed-mode coverage as a validity condition; investing in bilingual review and modest local evaluation datasets held as institutional rather than personal files; inserting checklists into ethics review and partner contracts; retaining non-delegable responsibility in the

commissioning institution; and prohibiting secondary training use of survey microdata without separate informed consent.

Recommendations for Regional Actors and Future Research

Development partners can treat AI documentation and human validation as quality criteria. Regional statistical organisations can host shared repositories of audit templates and local-language resources. Future research should field-test the prototype; examine public trust; evaluate training modules; integrate Indigenous data-governance protocols more fully; compare Caribbean, Pacific, and other SIDS adaptations; and study how external funding calendars interact with validation time. Until those studies exist, claims about effectiveness should remain provisional.

Closing Remark

Technological sophistication does not by itself produce fair or legitimate official knowledge. In small developing states the margin for error in public decision-making is narrow. A cautious, sequenced, and empirically untested prototype is offered here as a starting design, not as a completed solution.

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 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), 1–3. <https://doi.org/10.1016/j.patter.2020.100036>
2. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
3. 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). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
4. 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>
5. Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., &

- Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
6. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
 7. de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.
 8. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
 9. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
 10. Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1, 261–262. <https://doi.org/10.1038/s42256-019-0055-y>
 11. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>
 12. Government of Belize. (2021). *Data Protection Act, 2021*. Government of Belize.
 13. Government of Belize. (2022). *Plan Belize Medium-Term Development Strategy*. Government of Belize.
 14. Government of Belize. (2022b). *Digital Government Act, 2022 (Act No. 24 of 2022)*. Government of Belize.
 15. Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.
 16. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
 17. Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press.
 18. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
 19. 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>

20. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
21. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.
22. Organisation for Economic Co-operation and Development. (2019/2024). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), as amended 3 May 2024. OECD.
23. Shneiderman, B. (2022). Human-centered AI. Oxford University Press.
24. Statistical Institute of Belize. (2022). Belize population and housing census 2022: Key findings. Statistical Institute of Belize. <https://sib.org.bz/census/2022-census/>
25. Statistical Institute of Belize. (2024). Statistical Institute of Belize presents key findings of the 2022 Population and Housing Census [Press release]. Statistical Institute of Belize.
26. Statistical Institute of Belize. (2025). Multidimensional poverty index: September 2025. Statistical Institute of Belize.
27. Statistical Institute of Belize Act. (2006). Laws of Belize.
28. Statistical Institute of Belize & United Nations Children's Fund. (2025). Launch of the Multiple Indicator Cluster Survey (MICS7) results [Press release]. Statistical Institute of Belize.
29. Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 1–14. <https://doi.org/10.1177/2053951717736335>
30. United Nations. (2014). Fundamental principles of official statistics. United Nations Statistical Commission.
31. United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. United Nations.
32. United Nations. (2023). The Sustainable Development Goals Report 2023. United Nations.
33. United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO.
34. United Nations Educational, Scientific and Cultural Organization. (2024). Caribbean artificial intelligence policy roadmap. UNESCO. <https://doi.org/10.58338/TELF3157>
35. World Bank. (2022). Digital development in small states. World Bank Group.
36. World Bank. (2024). Statistical needs assessment of the Statistical Institute of Belize. World Bank Group.