ESI Preprints

Not Peer-reviewed

Policy Pathways for Responsible AI-Assisted Surveys: Institutional Adoption of H-VALIDA in Small Developing States - Translating Methodological Innovation into Governance Practice

Dr. Florence Yasmine Andrews, PhD
Assistant Professor of Economics & MBA Programme Coordinator
University of Belize, San Ignacio, Cayo District, Belize

Approved: 19 August 2026
Posted: 21 August 2026

Cite As:
Andrews, F.Y. (2026). *Policy Pathways for Responsible AI-Assisted Surveys: Institutional Adoption of H-VALIDA in Small Developing States - Translating Methodological Innovation into Governance Practice*. ESI Preprints. https://doi.org/10.19044/esipreprint.8.2026.p528

Abstract

Methodological innovation counts for little when it remains confined to academic conversation. For AI-assisted survey systems to strengthen evidence-based decision-making, the principles of transparency, human validation, contextual adaptation, and ethical accountability must be translated into operational policies, institutional procedures, and everyday research practice. This article examines the policy and practice implications of the H-VALIDA (Hybrid Validation and Auditing for Localized, Interpretable, Documented, and Accountable AI-Assisted Surveys) framework across national planning, government agencies, statistical institutions, universities, non-governmental organisations, digital governance, community engagement, and Sustainable Development Goal monitoring in Belize and comparable small developing states. It advances a phased institutional adoption pathway and identifies the risks that surface when AI-assisted surveys are introduced without corresponding safeguards. Drawing on critical data studies, human-centred AI scholarship, and the documented constraints of statistical systems in Small Island Developing States, the analysis maintains that technology must serve people, that data systems must represent communities fairly, and that artificial intelligence must remain accountable to human judgment and public purpose. The contribution is a context-sensitive, capacity-aware model of responsible AI

governance that privileges incremental institutional learning over premature technological scale-up.

Keywords: H-VALIDA; Policy; Responsible AI; Digital Governance; Institutional Adoption; Ethical Accountability; Small Developing States; Survey Methodology; Sustainable Development Goals; Data Equity

1. Introduction

1.1. Background

Transparent and validated methodologies for AI-assisted surveys carry real consequences for policy design, institutional practice, digital governance, and sustainable development planning. National development planning depends on reliable data about poverty, health, education, employment, climate vulnerability, and institutional performance. AI-assisted surveys can accelerate data production, reduce the costs of coding and translation, and support more frequent monitoring cycles. Those gains materialise only when systems are governed by standards that protect quality, inclusion, and public trust.

In small developing states the stakes rise still further. Limited fiscal space, thin institutional capacity, high exposure to climate and economic shocks, and historically uneven statistical coverage mean that errors in data systems can distort the allocation of scarce public resources, obscure the needs of marginalised groups, and erode institutional legitimacy. The rapid diffusion of generative AI tools into survey design, instrument translation, coding of open-ended responses, and sampling-frame construction therefore presents both an opportunity and a governance challenge of unusual intensity.

Many national statistical offices in small states were established or expanded in the post-independence period with external technical assistance oriented toward census and basic administrative statistics. Survey capacity developed more unevenly, often project by project. The arrival of AI tools lands on institutional foundations that are already stretched. Without deliberate governance, automation risks being layered onto fragile processes rather than integrated into strengthened ones. The political salience of survey data has also increased as governments and international partners rely more heavily on indicators for budget allocations and performance reporting, creating structural incentives to prioritise speed over transparent validation.

A related consideration is the rapid commercialisation of generative AI services. Vendors compete on ease of use and claimed performance, often with limited disclosure of training data composition or suitability for low-resource language varieties. Public institutions that lack independent evaluation capacity are poorly positioned to assess these claims. Without corresponding safeguards, AI adoption risks producing findings that appear efficient yet remain insufficiently validated, systematically under-representing rural, Indigenous, older, or digitally excluded populations, and leaving statistical offices unable to document or defend their methods.

1.2. Problem Statement

The principal problem is the gap between methodological innovation and institutional uptake. Even a well-specified framework remains ineffective if government agencies, statistical offices, universities, and development partners lack practical pathways for adoption, clear ethical protocols, and realistic phased implementation strategies. Existing ethical AI guidelines often assume organisational infrastructures, legal frameworks, and technical literacies that are unevenly distributed. Guidance produced in high-resource environments may prescribe comprehensive algorithmic audits or specialised AI ethics boards that small statistical offices cannot staff or fund.

The problem is compounded by the asynchronous nature of technological and institutional change. AI tools can be adopted within weeks; institutional protocols, staff skills, and community trust accumulate over years. When the former outpaces the latter, a governance vacuum opens. Furthermore, the distribution of AI literacy within small-state institutions is typically highly uneven, and many surveys are commissioned or co-funded by international development partners whose reporting timelines can favour rapid digital tools over slower, more inclusive designs. Without domestic policy anchors that treat documentation, human oversight, and inclusion as non-negotiable, external incentives may systematically undermine methodological quality.

1.3. Objectives

The objectives of this article are fourfold. First, to examine the policy and practice implications of the H-VALIDA framework across eight institutional domains: national policy and planning; government agencies and statistical institutions; universities and research institutions; non-governmental organisations and development partners; digital governance and responsible AI adoption; community engagement and data equity; Sustainable Development Goal monitoring; and practical pathways for institutional adoption. Second, to propose a concrete, phased institutional adoption pathway that respects capacity constraints while establishing non-negotiable protections. Third, to identify the risks of ignoring policy and practice implications when AI-assisted surveys are introduced. Fourth, to contribute to emerging international discussions on ethical AI governance in small states by offering a context-sensitive model adaptable to Small Island Developing States, low-resource environments, and multicultural societies.

1.4. Significance and Contribution

This article contributes to emerging international discussions concerning ethical AI governance in small states and developing nations by proposing a context-sensitive model adaptable to Small Island Developing States and regions affected by informational scarcity. Practically, the phased pathway and domain-specific recommendations offer statistical offices, planning ministries, universities, and development partners a usable template rather than another set of abstract principles. Theoretically, the analysis challenges the assumption that ethical AI governance must begin with comprehensive technical infrastructure; instead, it argues that institutional commitment to documentation, human oversight, and inclusion can form a robust foundation even under severe resource constraints. In doing so, the article positions small developing states not merely as recipients of externally designed AI systems but as sites of methodological innovation and governance experimentation.

2. Literature and Policy Context

2.1. Digital Governance and Responsible AI

AI adoption in the public sector raises questions of transparency, accountability, dependency, and sovereignty. High-level ethical principles alone cannot guarantee ethical AI; institutional accountability mechanisms are required (Mittelstadt, 2019; Floridi, 2019). Preference should be given to tools that allow documentation of prompts, model versions, and processing steps. Proprietary systems that function as complete black boxes create structural barriers to accountability.

Scholarship on algorithmic governance has repeatedly demonstrated that opacity in automated systems can undermine democratic accountability and public trust (Crawford, 2021; Noble, 2018). When statistical offices cannot explain how classifications were generated or how translations were produced, they lose the capacity to defend their products against legitimate critique. In small states, where technical expertise is concentrated in a limited number of individuals, this opacity is especially risky.

Empirical studies of public-sector AI adoption have identified documentation gaps and accountability diffusion as recurring failure modes. Explicit assignment of human responsibility for final classifications, translations, and methodological disclosures is therefore essential. Uncritical reliance on large commercial language models trained predominantly on Global North data can reproduce epistemic asymmetries. Local linguistic variants, Creole forms, Indigenous knowledge categories, and context-specific social classifications may be poorly represented or systematically distorted (Abebe, 2020; Shneiderman, 2022). International instruments such as the UNESCO Recommendation on the Ethics of Artificial Intelligence and the OECD AI Principles provide high-level normative guidance (UNESCO, 2021; OECD, 2019), yet their translation into operational procedures for survey production in small statistical offices remains underdeveloped.

2.2. Data Equity and Community Representation

Survey systems influence who is counted, whose experiences become visible, and whose needs are prioritised in policy. Inclusion is a condition of validity. Institutions must design sampling and data collection strategies that actively reach rural communities, Indigenous populations, coastal and island settlements, low-income households, older adults, persons with disabilities, and digitally excluded groups. Community participation in validation strengthens both accuracy and legitimacy (UNESCO, 2021; Taylor, 2017).

Critical data studies have long shown that data systems are never neutral mirrors of social reality; they are shaped by decisions about categories, sampling frames, modes of contact, and the relative weight given to different forms of knowledge (Barocas & Selbst, 2016; Eubanks, 2018; Bender et al., 2021). In multicultural small states such as Belize, where Creole, Spanish, Mayan languages, Garifuna, and English coexist, these design decisions carry heightened consequences. Indigenous data sovereignty scholarship further underscores that control over the collection, ownership, and interpretation of data about Indigenous peoples is itself a matter of self-determination (Kukutai & Taylor, 2016).

2.3. Sustainable Development Goal Monitoring

Reliable survey data are foundational to SDG monitoring. When AI-assisted surveys contribute to indicator measurement, documentation of methods, validation procedures, and residual limitations should accompany reported figures. The 2030 Agenda's emphasis on leaving no one behind places particular responsibility on national statistical systems to disaggregate data and surface patterns of exclusion (United Nations, 2015; United Nations, 2023). Integrating H-VALIDA principles into national SDG data protocols offers a practical means of aligning technological adoption with the equity commitments of the Agenda.

2.4. Research Gap and Positioning

While literature on ethical AI and algorithmic auditing has expanded rapidly, practical guidance for translating methodological frameworks into national guidelines, agency protocols, university training, NGO standards, and community-engaged processes remains underdeveloped for small developing states. Most existing frameworks either assume high-resource institutional environments or remain at the level of abstract principle (Jobin et al., 2019). This article addresses that gap by offering a sequenced, domain-specific, capacity-aware pathway grounded in the realities of small-state statistical systems.

2.5. Extended Literature Synthesis

The literature on ethical AI has matured from high-level principle lists toward more operational concerns with auditing, documentation, and organisational accountability (Mittelstadt, 2019; Floridi, 2019). Human-centred AI scholarship provides a complementary foundation: Shneiderman's framework insists on human control, reliability, and safety as design requirements rather than after-the-fact corrections (Shneiderman, 2022). Critical data studies illuminate the distributional stakes of survey design (Crawford, 2021; Noble, 2018; Eubanks, 2018). Documentation practices developed in the machine-learning community, datasheets for datasets and model cards for model reporting, supply useful templates that can be adapted to survey contexts (Gebru et al., 2021; Mitchell et al., 2019). The CARE Principles for Indigenous data governance articulate expectations that data about Indigenous peoples should serve community goals and remain under community authority where appropriate (Carroll et al., 2020). The literature on digital development in small states underscores thin institutional capacity, high dependence on external technical assistance, and limited local language resources as recurring features (World Bank, 2022).

3. Research Methodology

3.1. Approach

This article is a policy-oriented analytical study grounded in the H-VALIDA methodological framework. It synthesises insights from critical data studies, human-centred AI, algorithmic governance theory, and sustainable development policy analysis. No primary field data were collected. The approach is deliberately prospective and design-oriented: rather than evaluating an already-implemented national policy, it constructs a feasible pathway that institutions can adapt. Widespread institutional adoption of AI-assisted survey methods under explicit ethical frameworks is still emergent in small developing states; a design-oriented analysis creates a reference point against which future empirical studies can measure progress.

3.2. Analytical Framework

Implications are examined across eight institutional domains. For each domain, the analysis identifies concrete policy and practice implications and the conditions required for responsible implementation. Within each domain the analysis proceeds from diagnosis to prescription: identifying specific risks of ungoverned AI use, the minimum safeguards required, and the organisational or policy instruments through which those safeguards can be institutionalised.

3.3. Limitations

The analysis is prospective rather than evaluative of existing national policies. Institutional capacity, resource constraints, AI literacy, and rapid technological change will affect the ease of implementation. Full empirical assessment of adoption pathways remains a task for future research. A further limitation concerns the rapid evolution of generative AI capabilities; the article therefore focuses on institutional and procedural principles, documentation, human oversight, inclusion, accountability, that remain relevant across technological generations.

3.4. Analytical Stance

The analysis adopts an explicitly normative orientation: it treats the production of reliable, inclusive, and accountable knowledge for public decision-making as a non-negotiable responsibility of the institutions that commission and use survey data. This orientation is consistent with the public-purpose mandate of official statistics and with the equity commitments of the 2030 Agenda. The recommendations prioritise sequenced, capacity-aware steps over comprehensive blueprints that exceed available resources.

4. Results: Policy and Practice Implications

4.1 Implications for National Policy and Planning

National development planning depends on reliable data. Policy implications include the need for national guidelines on the responsible use of AI in survey research and official statistics. Such guidelines should require documentation of AI tools and model versions, human validation of automated outputs, bias auditing, multilingual review, and clear disclosure of methodological limitations. A national AI-in-statistics guideline need not be lengthy; its core function is to establish non-negotiable expectations and embed them in statistical legislation or quality assurance frameworks.

The content of a national guideline should be developed through inclusive consultation involving line ministries, civil society organisations, and community representatives. Once adopted, monitoring mechanisms, periodic self-assessment, occasional independent review by university partners, and public reporting of compliance, can maintain pressure for adherence. In small states, national policy also has a coordinating function: a common standard reduces the risk that individual projects invent ad-hoc approaches under time pressure. A practical mechanism is a small technical working group convened by the national statistical office or planning ministry to draft the guideline, develop supporting templates, and monitor early adoption.

4.2. Implications for Government Agencies and Statistical Institutions

Agencies should establish internal protocols for AI-assisted survey design, documentation, and validation, beginning with mandatory documentation of AI tools and human review of automated classifications, then expanding toward more comprehensive auditing as capacity grows. Statistical institutions should invest in AI literacy and algorithmic auditing skills among existing staff through short, practice-oriented training modules focused on prompt documentation, sampling of AI-coded responses for human comparison, and recognition of linguistic bias. Partnerships with universities can supply curriculum content.

Agencies should prioritise mixed-mode data collection strategies. Heavy reliance on online or mobile surveys risks systematic exclusion of rural households, older adults, and digitally disconnected communities. Data protection and privacy procedures must be strengthened in light of AI processing; in small communities the risk of indirect identification is elevated. Statistical offices should maintain an internal register of AI tools and versions used, and develop local evaluation datasets, samples of previously coded responses, translated instruments, or classified codes, against which new AI tools can be tested before operational deployment. Such datasets serve as a practical instrument of methodological sovereignty.

4.3. Implications for Universities and Research Institutions

Universities play a critical role in methodological innovation, training, ethical review, and independent validation. They can incorporate H-VALIDA principles into research methods curricula, postgraduate training, and research ethics procedures, and serve as independent validators of AI-assisted survey outputs. In small developing states, universities often serve as one of the few concentrations of research capacity outside government. By embedding H-VALIDA principles in methods courses and requiring explicit statements of AI use in ethics review, universities can cultivate a generation of researchers who treat human validation and documentation as standard practice. Collaborative projects between university research centres and the national statistical office can produce shared audit templates and local-language evaluation datasets.

4.4. Implications for Non-Governmental Organisations and Development Partners

NGOs and international development partners frequently commission surveys under time pressure that favour rapid digital tools. Efficiency must be balanced with validation. Development partners should require documentation of AI use, human review of automated analysis, and explicit attention to inclusion of marginalised groups as conditions of funding and reporting. When partners treat AI documentation and human validation as non-negotiable elements of survey quality, they create incentives for adoption that purely domestic policy may lack. A concrete step is for major partners operating in a given country to adopt a common minimum standard for AI-assisted surveys, coordinated with the national statistical office.

4.5 Implications for Digital Governance and Responsible AI Adoption

H-VALIDA contributes to digital governance by insisting that AI systems used in survey research remain explainable, auditable, and subject to human oversight. Digital transformation strategies should explicitly address the risk that AI-assisted systems may deepen existing inequalities if they privilege digitally connected populations. Procurement policy is a practical lever: when government agencies procure AI tools, tender specifications can require documentation of model versions, support for prompt logging, contractual prohibitions on secondary data use, and demonstration of performance on local-language evaluation samples.

4.6. Implications for Community Engagement and Data Equity

Institutions should design sampling and data collection strategies that actively reach rural communities, Indigenous populations, Garifuna communities, coastal and island settlements, low-income households, older adults, persons with disabilities, and digitally excluded groups. Multilingual instruments and human linguistic validation are essential. Community participation in validation strengthens both accuracy and legitimacy. Feedback mechanisms that allow communities to review findings and correct misrepresentations support epistemic justice and public trust. Particular care is required when reporting findings about small communities, where the combination of geographic and demographic descriptors can render groups identifiable.

4.7. Implications for Sustainable Development Goal Monitoring

When AI-assisted surveys contribute to indicator measurement, documentation of methods, validation procedures, and residual limitations should accompany the reported figures. A practical response is to attach a standardised methodological annex, covering tools used, validation steps taken, and known residual risks, to any SDG-related survey report. International organisations that compile global SDG databases can support responsible practice by treating methodological transparency as a quality criterion rather than an optional extra.

4.8. Practical Pathways for Institutional Adoption

Full implementation may appear demanding for institutions with limited resources. A phased adoption pathway is therefore recommended. Each phase builds on the previous one, allowing institutions to demonstrate progress and accumulate experience without requiring simultaneous transformation of every procedure.

Phase 1: Core safeguards (0-6 months)

Mandatory documentation of AI tools, model versions, and prompts; human review of automated coding and translation; basic ethical disclosure of AI use to respondents. These measures require institutional commitment and procedural discipline more than specialised technical expertise. They can be introduced through internal circulars and checklist templates within a single planning cycle.

Phase 2: Auditing and bias review (6-18 months)

Structured checks for linguistic, urban, gender, and socioeconomic bias; comparison of AI and human coding samples; clear reporting of limitations. Sample sizes for comparison need not be large; even modest double-coding exercises can reveal systematic discrepancies.

Phase 3: Contextual and community validation (12-30 months)

Mixed-mode sampling strategies; local expert and bilingual review; pilot community feedback on findings where appropriate. External partners can usefully target support to this phase.

Phase 4: Institutional integration (24-48 months)

Incorporation of H-VALIDA principles into standard operating procedures, ethics review processes, training curricula, and SDG data protocols. When new officers encounter these principles as part of their induction, the practices are more likely to persist.

Phase 5: Continuous improvement (Ongoing)

Longitudinal review of framework performance, updating of protocols as AI technologies evolve, and sharing of lessons across institutions. Periodic review every two to three years allows institutions to retire outdated tools and adjust the balance between automation and human oversight.

Table 1: Phased Institutional Adoption Pathway for H-VALIDA

Phase	Core Activities	Primary Responsibility	Indicative Timeline
1	Documentation; human review; ethical disclosure	Statistical office / 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	SOPs; ethics integration; training curricula; SDG protocols	National statistical system + universities	24–48 months
5	Longitudinal review; protocol updates; inter-institutional learning	All stakeholders	Ongoing

Note: Timelines are indicative and should be adjusted to local capacity and resource cycles. Phases may overlap.

4.9. Risks of Ignoring Policy and Practice Implications

If AI-assisted surveys are adopted without corresponding safeguards, several interrelated risks materialise:

- Policy decisions may be based on biased or incomplete data, leading to misallocation of scarce public resources.
- Vulnerable populations may remain invisible in official statistics, undermining the commitment to leave no one behind.
- Public trust in data systems and institutions may erode when communities recognise systematic under-representation.
- Digital transformation may deepen rather than reduce inequality by privileging digitally connected populations.
- External technological dependency may increase as institutions become reliant on proprietary systems they cannot inspect.
- SDG monitoring may lose credibility if stakeholders question the validity of AI-assisted indicators.

These risks are particularly acute in small developing states where data systems are already constrained and the consequences of misallocated resources are severe.

4.10 Illustrative Scenarios

To ground the implications, consider four illustrative scenarios drawn from the characteristic conditions of small developing states.

Illustration A: AI-Assisted Coding of Open-Ended Livelihood Responses. Under time pressure, a statistical office applies a commercial language model to code free-text livelihood responses. Without safeguards, informal and seasonal livelihoods common in coastal communities are misclassified, understating climate-sensitive livelihoods. Under an H-VALIDA approach, the model version and prompts are documented, a stratified sample is double-coded by human coders familiar with local terminology, discrepancies are analysed, and residual limitations are reported.

Illustration B: Multilingual Instrument Translation. Machine translation is proposed for instruments in English, Spanish, and Creole. Without safeguards, ambiguous or culturally mismatched terms produce measurement error concentrated in Creole-speaking communities. An H-VALIDA process treats translation as validation-critical: machine translation serves only as a first draft, independent bilingual reviewers flag conceptual mismatches, and a cognitive pilot tests comprehension.

Illustration C: Digital-First Mode Design. An employment survey planned with a predominantly online mode over-represents younger, urban respondents. H-VALIDA principles require that mode design be evaluated for representation consequences before efficiency gains are prioritised; a mixed-mode strategy with stratified human review of coding samples is adopted.

Illustration D: Partner-Commissioned Baseline Survey. A development partner prioritises rapid delivery. Without safeguards, documentation of tools and validation is minimal, making subsequent interpretation of change difficult. When partner requirements align with H-VALIDA guidelines, the contract requires documentation, human review, and a methodological annex, enabling clearer interpretation and cumulative capacity development.

4.11. Implementation Enablers and Common Obstacles

Successful institutionalisation depends on visible high-level sponsorship within the national statistical system, the availability of reusable templates and worked examples, and realistic sequencing. Common obstacles include staff turnover, concurrent organisational reforms, the perception that validation is "extra" work, and the absence of local evaluation datasets. Addressing turnover requires embedding principles in induction training and SOPs. Investing early in modest local evaluation datasets creates a lasting institutional asset.

4.12. Sequencing Decisions and Portfolio Approaches

Not every survey needs to implement the full pathway simultaneously. Institutions facing acute capacity constraints can adopt a portfolio approach: apply core Phase 1 safeguards to all AI-assisted surveys as a minimum standard, while selecting one or two high-priority surveys for deeper Phase 2 and Phase 3 treatment in any given planning cycle. High-priority candidates include surveys that feed into national development planning or SDG reporting, or that address highly vulnerable populations.

5. Discussion

5.1. Interpreting the Implications

The translation of H-VALIDA principles into policy and practice confirms that methodological legitimacy requires more than technical sophistication. It requires inclusive governance structures capable of ensuring transparency, fairness, representation, reproducibility, and public trust. Technology must serve people; data systems must represent communities fairly; and AI must remain accountable to human judgment and public purpose. Ethical accountability is an institutional practice that must be designed into survey workflows from the outset, not appended after technical decisions have already been made.

5.2. Institutional Feasibility

A recurring concern is whether resource-constrained institutions can realistically implement hybrid validation procedures. The phased adoption pathway addresses this by prioritising essential protections first. Core safeguards, documentation of tools and prompts, human review of automated outputs, and disclosure of AI use, require institutional commitment more than specialised technical expertise. Subsequent phases introduce more demanding activities only after institutions have accumulated experience. Partial implementation of core safeguards is already an improvement over ungoverned automation; progressive expansion as capacity grows is preferable to paralysis while waiting for ideal conditions.

5.3. Sovereignty and Dependency

Uncritical importation of AI tools risks reproducing forms of epistemic dependency and technological asymmetry. Policy implications therefore include preference for tools that support documentation and auditability, investment in local language resources and benchmark datasets, and collaboration among universities, statistical offices, and community organisations. Sovereignty does not require that every AI model be trained locally; it requires that national institutions retain the capacity to understand, validate, contest, and, where necessary, reject the outputs of systems they employ.

5.4. Ethical Accountability as Institutional Practice

Ethical responsibility cannot be outsourced to AI vendors. The organisation that commissions or conducts a survey remains accountable for the integrity of the knowledge it produces. When external firms supply AI-assisted survey services, contracts should explicitly allocate responsibility for validation, documentation, and ethical disclosure to the commissioning institution. Clauses that prohibit secondary use of data for model training without separate consent further protect respondents and institutional integrity.

5.5. Comparative Considerations for Small Island Developing States

Although the analysis is framed with particular reference to Belize, the institutional constraints and opportunities are shared across many Small Island Developing States and other small developing economies. Thin statistical capacity, multilingual populations, high climate vulnerability, and dependence on external technical assistance create a common structural context. The phased pathway is offered as a transferable template. Regional statistical organisations can play a useful convening role, facilitating the exchange of audit templates, training materials, and early lessons.

5.6. Tensions and Trade-offs

Responsible adoption inevitably involves trade-offs. Documentation and human review consume time and staff attention; mixed-mode designs are more expensive; community validation introduces additional stages. These costs are real. The argument is not that costs can be eliminated, but that they are preferable to the alternative costs of biased, opaque, or exclusionary data systems. In contexts where public resources are scarce and the consequences of misallocation severe, investing in the integrity of the knowledge base is itself a form of efficiency.

5.7. Broader Implications

Survey data interact with administrative data systems, geospatial information, and increasingly with machine-generated synthetic data. National data strategies need to extend H-VALIDA-aligned principles across the wider data ecosystem. Continuous professional development and communities of practice among statistical officers across neighbouring states are more likely to sustain capability than one-time training events. The gendered dimensions of data production and AI bias warrant explicit attention: bias auditing protocols should include gender as a standard dimension of review. Climate vulnerability measurement provides a particularly instructive domain, as the categories through which vulnerability is measured are contested and context-specific. The interaction between AI-assisted surveys and Indigenous data governance principles requires careful navigation; community validation and contextual adaptation pillars provide practical entry points for more respectful practice.

5.8. Political Economy of Survey Production

The feasibility of responsible AI adoption cannot be assessed solely in terms of domestic technical capacity. The political economy of survey production is heavily shaped by external funding cycles and partner reporting requirements. Aligning external incentives with national quality standards is a strategic priority. Practical mechanisms include joint statements of minimum standards between major partners and the national statistical office, incorporation of H-VALIDA-aligned checklists into partner monitoring guidelines, and earmarking a modest share of survey budgets for validation activities. Domestic political economy also matters: statistical offices with clear legal mandates, stable core funding, and visible high-level support are better positioned to insist on methodological standards.

5.9. Public Trust and Epistemic Legitimacy

Public trust in official statistics is a strategic asset that is costly to build and easy to erode. AI-assisted surveys that privilege digitally connected populations or apply classification systems poorly calibrated to local contexts risk such erosion. Transparency about methods, residual limitations, and validation steps is one mechanism for protecting trust; meaningful opportunities for community feedback is another. Epistemic legitimacy, the recognition that knowledge is not only technically competent but also fairly and accountably generated, matters especially in plural societies.

5.10. Measuring Progress

Institutions that adopt the phased pathway will benefit from simple metrics of progress: the proportion of AI-assisted surveys that maintain complete documentation of tools and prompts; the proportion that conduct and record sample-based human validation; the existence and currency of local evaluation datasets; the incorporation of H-VALIDA principles into SOPs and training curricula; and the frequency with which residual limitations are reported alongside published indicators.

Conclusions

H-VALIDA is not only a methodological framework; it is a call for responsible institutional practice and coherent policy. Translating its principles into national guidelines, agency protocols, university training, NGO standards, and community-engaged processes is essential if AI-assisted surveys are to strengthen rather than undermine evidence-based sustainable development. The ultimate implication is straightforward: technology must serve people, data systems must represent communities fairly, and artificial intelligence must remain accountable to human judgment and public purpose.

This orientation is institutionally feasible even under resource constraints, provided that adoption is sequenced, that core safeguards are treated as non-negotiable, and that capacity is understood as something that can be built through deliberate practice. Small developing states need not wait for comprehensive technical infrastructure before beginning to govern AI-assisted surveys responsibly. The relative smallness of institutional systems can facilitate coordination; personal networks among statistical officers, university researchers, and civil society actors can support rapid learning; community-scale social relations can make participatory validation more feasible. These advantages can be mobilised provided that policy frameworks create the space and incentives for them to operate.

Recommendations for National Policy

Governments are recommended to:

- Develop national guidelines on the responsible use of AI in survey research and official statistics, establishing documentation, human validation, bias auditing, and multilingual review as conditions of official statistical production.
- Integrate AI-assisted data systems into broader digital governance and data strategy frameworks as instruments of public accountability.
- Support phased capacity-building programmes in AI literacy, algorithmic auditing, and data ethics for statistical and planning institutions.
- Align development partner requirements with national guidelines so that external funding reinforces rather than fragments responsible practice.
- Establish or designate a technical working group to draft supporting templates, monitor early adoption, and facilitate inter-agency learning.

Recommendations for Institutional Practice

Agencies, universities, NGOs, and development partners are recommended to:

- Adopt the phased implementation pathway, beginning with core safeguards of documentation, human review, and ethical disclosure.
- Establish internal protocols and maintain a simple register of AI tools and versions used in production processes.
- Prioritise mixed-mode sampling and active inclusion of marginalised populations, treating inclusion as a condition of validity.

- Invest in bilingual review capacity, local expert engagement, and the development of local evaluation datasets.
- Incorporate H-VALIDA checklists into project monitoring, evaluation, and ethics review processes.
- Treat ethical accountability as non-delegable: the commissioning institution retains responsibility even when technical execution is externalised.
- Include contractual clauses prohibiting secondary use of survey data for model training without separate, informed consent.

Recommendations for Future Research

Future research should evaluate the effectiveness of phased adoption pathways in real institutional settings; examine public trust in AI-assisted surveys; develop and test training programmes and audit templates tailored to small developing states; integrate Indigenous data governance principles into extensions of the framework; assess transferability across Caribbean, Pacific, and other SIDS contexts; and investigate the political economy of survey production under external funding pressures.

Recommendations for Regional and International Actors

Development partners are recommended to review their internal survey guidelines to ensure that AI documentation, human validation, and attention to inclusion are treated as quality criteria. Technical assistance programmes can be structured around the phased pathway. Regional statistical organisations can accelerate responsible adoption by convening communities of practice, hosting shared repositories of audit templates and local-language evaluation resources, and facilitating peer learning across small states facing similar constraints.

Final Reflection

The transition toward AI-assisted knowledge systems represents one of the defining transformations of contemporary research and governance. Yet technological sophistication alone cannot guarantee truth, fairness, or social legitimacy. Responsible AI requires methodological humility and institutional courage. Institutions that place human judgment, ethical accountability, and contextual sensitivity at the centre of AI-assisted survey practice will be better positioned to produce knowledge that is reliable, inclusive, transparent, and useful for the challenges of sustainable development. In small developing states, where the margin for error in public decision-making is narrow and the costs of exclusion are high, this orientation is not a luxury, it is a condition of democratic legitimacy and of genuine progress toward sustainable development that leaves no one behind.

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.
2. Barocas, S., & Selbst, A. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732.
3. Bender, E., 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).
4. Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., ... & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), 43.
5. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
6. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
7. Floridi, L. (2019). Establishing the rules for building trustworthy AI. Nature Machine Intelligence, 1, 261–262. https://doi.org/10.1038/s42256-019-0055-y
8. 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.
9. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399.
10. Kukutai, T., & Taylor, J. (Eds.). (2016). Indigenous data sovereignty: Toward an agenda. ANU Press.
11. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229).
12. 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
13. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

14. Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence. OECD.
15. Shneiderman, B. (2022). Human-centered AI. Oxford University Press.
16. Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), 1–14.
17. United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. United Nations.
18. United Nations. (2023). The Sustainable Development Goals Report 2023. United Nations.
19. United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO.
20. World Bank. (2022). Digital development in small states. World Bank Group.