Operationalising the Ethical Use of AI in Development Research

A Framework for FCDO-Funded Research Partners

Lucas Sempe

2026-02-01

Abstract

Artificial intelligence tools — large language models, automated coding, machine-assisted evidence synthesis — are transforming development research practice. Yet frameworks for their ethical use remain aspirational. Drawing on a cross-national analysis of 2,216 AI policy documents across 193 countries (Sempé, 2026), this brief identifies the operationalisation gap as the binding constraint: organisations endorse ethical values but rarely translate them into enforceable research protocols. Following the UNESCO Recommendation's three-tier architecture — values, principles, and policy actions — and drawing on convergent standards from the OECD, EU AI Act, and the broader AI ethics literature, we propose an operational framework and checklist for FCDO-funded research partners seeking to move beyond principle statements toward accountable, transparent, and inclusive AI-assisted research.

1 Introduction

Development research organisations increasingly rely on AI tools. Large language models assist with systematic review screening, qualitative coding, survey translation, and data analysis. Automated pipelines extract and synthesise evidence at scales that were prohibitively expensive five years ago. The FCDO's Research Commissioning Centre (RCC), led by 3ie and the University of Birmingham, commissions research across thematic areas where AI-assisted methods are becoming standard practice.

These tools create genuine efficiency gains — but they also create ethical risks that existing research ethics frameworks were not designed to address. Algorithmic bias can distort evidence synthesis. Privacy-sensitive data may flow through third-party APIs. Automated outputs may lack the interpretive nuance that development contexts require. And the opacity of large language models complicates the transparency and reproducibility norms that underpin credible research.

The operationalisation gap. A recent cross-national analysis of 2,216 AI policy documents catalogued in the OECD.AI Policy Observatory (Sempé, 2026) reveals a consistent finding: ethical principles are widely endorsed but rarely operationalised. Across three companion studies examining implementation capacity, ethics governance depth, and UNESCO alignment:

- 97% of policies score below the midpoint on ethical operationalisation.
- Accountability mechanisms (monitoring, evaluation, independent oversight) are the most systematically omitted governance features globally — an *accountability paradox.*
- The gap between *articulating* principles and *implementing* them characterises AI governance regardless of income level or region.

The same pattern risks emerging in research practice. Organisations adopt responsible AI statements; fewer establish the review processes, documentation standards, and audit mechanisms that make those statements operational. This brief addresses that gap.

2 What the Evidence Tells Us

Why analyse 2,216 policy documents? Because multiple international frameworks — developed independently — converge on the same ethical architecture, yet none had been tested at scale.

The UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021), adopted by 193 member states in 2021, provides the most comprehensive structure: **4 values** (the *why*), **10 principles** (the *what*), and **11 policy action areas** (the *how*). But UNESCO is not alone. Jobin et al. (2019) mapped 84 AI ethics guidelines and found convergence around transparency, fairness, non-maleficence, responsibility, and privacy. Floridi et al. (2018) proposed the AI4People framework — beneficence, non-maleficence, autonomy, justice, and explicability — bridging bioethics into AI governance. The OECD AI Principles (OECD, 2019) established five principles and five policy recommendations for trustworthy AI, adopted by 46 countries. The EU AI Act (European Parliament and Council, 2024) shifted from aspirational principles to binding, risk-based regulation with enforcement mechanisms.

These frameworks agree on *what matters*. What the cross-national analysis (Sempé, 2026) reveals is that they all fail in the same way: **values are articulated, principles are endorsed, but policy actions are not operationalised.** The UNESCO study documented this directly — coverage rates are 55% for values, 53% for principles, but just 41% for policy action areas. Gender (9.6%), culture (9.1%), and communication (2.0%) are virtually absent. Human rights — the stated foundation of the entire framework — appears in only 22.9% of policies.

Three specific findings matter for research organisations:

- **Operationalisation is the binding constraint.** Across all frameworks, the gap is not in *what principles to endorse* but in *how to enforce them.* 97% of policies score below the midpoint on operationalisation. What Floridi et al. (2018) called "translating principles into practice" remains unachieved — for countries and, by extension, for the organisations that use AI.
- **Accountability is the most needed and most omitted feature.** The capacity study found Accountability (C4) averaging just 0.48/4.0 globally — the weakest dimension. Implementation science predicts this: accountability creates transparency that exposes failures (Mazmanian & Sabatier, 1983; Pressman & Wildavsky, 1973). Governments avoid it; research organisations risk the same.
- **Resources are not the barrier.** GDP has *zero effect* on ethics governance quality. Kenya, Rwanda, Brazil, and Colombia all demonstrate sophisticated frameworks despite modest budgets. Ethical operationalisation requires institutional design and political will — not money. FCDO research partners in low- and middle-income countries are not at a disadvantage.

3 From Values to Policy Actions: A Framework for Research

The UNESCO Recommendation provides a three-tier architecture that organises ethical AI governance from the abstract to the concrete. The cross-national analysis confirmed this structure empirically: values are easiest to adopt, principles require more effort, and policy actions — where operationalisation happens — are where governance collapses. We adopt what we call a **UNESCO+ architecture**: the UNESCO structure as organising skeleton, enriched by two additional analytical traditions. The convergent AI ethics literature — synthesised through five ethics governance dimensions (E1–E5) — provides the institutional specificity that UNESCO's principles name but do not operationalise: governance mechanisms, rights protection architecture, and participation structures. Implementation science — synthesised through five capacity conditions (C1–C5) — ensures that policy actions are clear, resourced, enforceable, monitorable,

and coordinated. The "+" ensures that the framework is not only normatively comprehensive but institutionally specific and implementable.

3.1 Tier 1 — Values: Why ethical AI matters for research

UNESCO identifies four foundational values: **human rights and dignity**, **peaceful and just societies**, **diversity and inclusiveness**, and **environment and ecosystem flourishing**. These map onto convergent values across frameworks — Floridi's beneficence and autonomy, the OECD's human-centred values, and the EU AI Act's fundamental rights grounding.

For research organisations, values answer the question: *What are we protecting when we govern AI use?*

- **Human rights and dignity** — research participants' privacy, autonomy, and protection from algorithmic harm. The cross-national analysis found this value present in only 22.9% of AI policies — a striking gap for a rights-based framework. Research organisations must not replicate this omission.
- **Diversity and inclusiveness** — ensuring AI-mediated research does not systematically marginalise populations it claims to serve. As Gwagwa et al. (2020) document, AI tools deployed in African contexts often reflect Western training data rather than local realities.
- **Environment** — recognising the computational costs of AI tools and their carbon footprint, relevant to research programmes operating under sustainability commitments.

Values should be stated explicitly in organisational AI policies — not as aspirational language but as the criteria against which all subsequent principles and actions are assessed.

3.2 Tier 2 — Principles: What ethical standards govern AI use

UNESCO specifies ten principles; the convergence analysis shows these map onto standards found across all major frameworks:

Table 1: Convergent ethical principles for AI in research

Principle (UNESCO)	Convergent sources	Meaning for research
Transparency & explicability	OECD, Floridi (explicability), Jobin	Disclose AI use, document methods, make prompts reproducible
Fairness & non-discrimination	OECD, EU AI Act (prohibited practices), Jobin	Test for differential AI performance across languages, populations, contexts
Responsibility & accountability	OECD, Floridi (justice), Jobin	Designate who is responsible for AI-assisted outputs; establish oversight
Human oversight & determination	EU AI Act (human-in-the-loop), OECD	Meaningful human review of AI outputs — not rubber-stamping
Privacy & data governance	EU AI Act, OECD, Jobin	Verify API providers' data policies; protect respondent information
Safety & security	OECD, Floridi (non-maleficence), EU AI Act	Validate AI outputs before they inform research conclusions

Principle (UNESCO)	Convergent sources	Meaning for research
Proportionality & do no harm	Floridi (non-maleficence)	Assess whether AI tools are appropriate for the specific research context
Sustainability	UNESCO, OECD	Consider computational costs and environmental impact
Awareness & literacy	UNESCO	Train researchers in AI capabilities *and* limitations
Multi-stakeholder governance	UNESCO, OECD	Include affected communities in decisions about AI-mediated research

The cross-national analysis operationalised these principles through five ethics governance dimensions (E1–E5) — framework depth, rights protection, governance mechanisms, operationalisation, and inclusion — providing the institutional architecture that UNESCO names but does not detail. The analysis found that principles are widely endorsed but shallowly implemented: coverage does not predict depth ($r = 0.02$, $p = 0.94$), and the most frequently mentioned principles appear as rhetorical gestures rather than substantive commitments. For research, stating "we are committed to transparency" is meaningless without specifying *what transparency requires* in practice — which is precisely what the E1–E5 dimensions provide.

3.3 Tier 3 — Policy actions: How to operationalise ethics

This is where governance fails. UNESCO defines eleven policy action areas; national coverage drops to 41%. For research organisations, policy actions are the concrete protocols, standards, and requirements that make values and principles enforceable. Without them, the first two tiers are performative. Implementation science (C1–C5) identifies what makes policy actions executable: they must have clear objectives (C1), dedicated resources (C2), enforcement authority (C3), accountability mechanisms (C4), and cross-institutional coordination (C5). The companion guidance document applies these conditions systematically to each standard.

Translating values and principles into research practice requires action in four areas:

Documentation and transparency. Every AI-assisted research output must document which tools were used (model identifiers, versions, access dates), which prompts or parameters were applied, what validation was performed, and how AI outputs were integrated into final results. This is the operational content of the transparency principle.

Quality assurance and validation. Minimum human verification rates for AI-generated outputs — for example, independent human coding of a random subsample (minimum 10–15%) in LLM-assisted systematic reviews. Pre-specified failure protocols: criteria for when AI outputs fail quality checks, including conditions for abandoning AI-assisted approaches in favour of manual methods.

Impact assessment and data governance. Before deploying AI tools, assess risks to participant privacy, data protection, and potential bias. Verify that AI tool providers' terms of service and data retention policies comply with research ethics requirements. Ensure informed consent covers AI-assisted data processing. Test for differential performance across languages, regions, and population groups.

Review and accountability. Establish a designated oversight function — a named individual or committee responsible for AI ethics review. Conduct post-project reviews of AI tool perfor-

mance and ethical issues. Create grievance mechanisms for researchers, participants, or partners to raise concerns. Report incidents to oversight bodies.

4 Operational Checklist

The following checklist maps concrete actions to the three-tier framework across the research lifecycle.

Table 2: Operational checklist for ethical AI use in research

Stage	Action	Tier
Design		
	State which values and principles govern AI use in this project	Values
	Assess whether AI tools are appropriate for the research question and context	Principles
	Conduct an AI impact assessment: identify risks to privacy, bias, and data protection	Policy actions
	Verify that AI tool providers' data policies comply with ethics requirements	Policy actions
	Register AI-assisted methods in the research protocol or pre-analysis plan	Policy actions
	Ensure informed consent covers AI-assisted data processing	Principles
Data & Analysis		
	Document all AI tools, versions, prompts, and parameters used	Policy actions
	Validate AI outputs against human-coded subsamples (minimum 10–15%)	Policy actions
	Test for differential performance across languages, regions, or population groups	Principles
	Maintain audit trails for all AI-assisted analytical steps	Policy actions
	Apply pre-specified failure protocols when quality thresholds are not met	Policy actions
Reporting		
	Disclose AI tool use, including model identifiers and access dates	Principles

Stage	Action	Tier
	Report validation results (human-AI agreement rates, error patterns)	Policy actions
	Acknowledge limitations of AI-assisted methods for the specific context	Values
	Make prompts and parameters available for reproducibility	Policy actions
Review & Learning	Conduct post-project review of AI tool performance and ethical issues	Policy actions
	Update organisational protocols based on lessons learned	Policy actions
	Share methodological insights with research partners	Values
	Report incidents or unexpected ethical issues to oversight body	Policy actions

Each item in Table 2 specifies a *verifiable* action, not an aspiration. The cross-national analysis showed that the distance between "we endorse transparency" and "we require disclosure of model identifiers in all publications" is precisely where ethical governance collapses.

5 Institutional Requirements

Moving from Tier 2 (principles) to Tier 3 (policy actions) requires institutional infrastructure — precisely the implementation capacity found missing in 96.5% of national policies.

Designated oversight. A named individual or committee responsible for AI ethics review within the organisation. This need not be a new body; integrating AI review into existing research ethics or quality assurance structures is sufficient, provided the mandate is explicit and resourced.

Documentation standards. Organisation-wide templates for documenting AI-assisted methods — standardised across projects to enable comparison, audit, and learning. The cross-national analysis found that documentation quality confounds governance assessment; the same applies to research: if AI methods are not documented consistently, quality cannot be assessed.

Training and capacity. Researchers using AI tools need practical training in both capabilities and limitations — not abstract ethics lectures but hands-on guidance on prompt engineering, output validation, bias detection, and failure recognition. Equally important is the deskilling risk: researchers who routinely delegate analysis, interpretation, and writing to AI may produce adequate outputs while gradually losing the intellectual capacities that make their judgment credible. Training must therefore address not only how to use AI tools but how to preserve the analytical and interpretive skills that AI threatens to atrophy — including periodic practice of key tasks without AI assistance. The evidence that operationalisation does not require large budgets applies here: training investments are modest relative to the risks of unvalidated AI

use.

Grievance and incident mechanisms. A clear pathway for researchers, participants, or partners to raise concerns about AI-assisted methods. The accountability paradox — that the governance feature most needed is the one most often omitted — can be consciously countered by building incident reporting into organisational culture before problems arise.

Periodic review. AI tools evolve rapidly. Protocols established in 2026 may be inadequate by 2027. Institutional arrangements must include scheduled reviews — at minimum annually — to assess whether ethical AI frameworks remain fit for purpose as tools, applications, and understanding develop.

6 Conclusion

The cross-national analysis of 2,216 AI policies (Sempé, 2026) demonstrates that the gap between ethical articulation and ethical operationalisation is the defining challenge of AI governance globally. Multiple international frameworks — UNESCO, OECD, EU AI Act, and the broader ethics literature — converge on the same values and principles. The problem is not knowing *what* to endorse; it is translating endorsement into enforceable practice. Research organisations are not exempt.

The UNESCO three-tier architecture — values, principles, policy actions — provides the structure. The convergence across frameworks confirms that these are not parochial standards but internationally validated norms. And the evidence that operationalisation does not depend on resources means every FCDO-funded partner, regardless of size or location, can build robust ethical AI practices now.

The checklist proposed here is deliberately minimal — a starting point, not an endpoint. It is designed to be achievable immediately and refined iteratively. What matters is crossing the threshold from aspiration to action.

> **i** This brief draws on a **cross-national analysis of 2,216 AI policies across 193 countries** (Sempé, 2026). The full analysis is available in three companion volumes: *AI Governance Implementation Capacity*, *AI Ethics Governance Depth*, and *UNESCO AI Ethics Recommendation Alignment*. Code, data, and methods: github.com/lsempe77/ai-governance-capacity.
> **Companion document.** This brief is accompanied by *Ethical Guidance for the Use of AI in FCDO-Funded Research, Evaluation, and Evidence Synthesis* (Sempé, 2026), which translates the framework presented here into stage-by-stage standards, decision prompts, checklists, and templates for research teams.

References

European Parliament and Council. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI act).*

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4Peopleâ an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*, 689–707.

Gwagwa, A., Kraemer-Mbula, E., Rizk, N., & Rutenberg, I. (2020). Artificial intelligence (AI) deployments in Africa: Benefits, challenges and policy dimensions. *The African Journal of*

Information and Communication, 26, 1–28. https://doi.org/10.23962/10539/30361

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Mazmanian, D. A., & Sabatier, P. A. (1983). *Implementation and public policy.* Scott Foresman.

OECD. (2019). *OECD principles on artificial intelligence.*

Pressman, J. L., & Wildavsky, A. (1973). *Implementation.* University of California Press.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.*