

Beyond the principles

Evidence of whether locally led adaptation works in practice

Anaa Hassan, Paul Mitchell and Aaron Acuda



About the authors

Anaa Hassan is a climate finance associate at IIED (www.iied.org/people/anaa-hassan)

Paul Mitchell is an adaptation and resilience senior associate at IIED (www.iied.org/people/paul-mitchell)

Aaron Acuda is a researcher at IIED (www.iied.org/people/aaron-acuda)

Corresponding author: anaa.hassan@iied.org

Acknowledgements

This report was made possible through the generous support of the Netherlands Ministry of Foreign Affairs. We are also grateful to the Generating Ambition for Locally Led Adaptation (GA-LLA) programme partners for their valuable comments and review throughout the course of this work.

Furthermore, huge thanks to Dr Mohsen Gul (senior researcher at IIED) for his contributions in fine-tuning the framework. We further extend our thanks to Ros Cook for the report's layout, and to Annette McGill for editorial support.

Published by IIED, October 2026

Hassan, A, Mitchell, P and Acuda, A (2026) Beyond the principles: evidence of whether locally led adaptation works in practice. IIED, London.

iied.org/22778iied

ISBN: 978-1-83759-240-1

This publication has been reviewed according to IIED's peer review policy, which sets out a rigorous, documented and accountable process (see www.iied.org/research-excellence-impact for more information). The reviewers were Dr. Katherine Browne from Stockholm Environment Institute and Dr. Pia Treichel from RMIT University.

International Institute for Environment and Development
44 Southampton Buildings, London WC2A 1AP, UK
Tel: +44 (0)20 3463 7399
www.iied.org

 www.linkedin.com/company/iied

 www.facebook.com/thelIED

Download more publications at iied.org/publications

IIED publications may be shared and republished in accordance with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License (CC BY-NC-ND 4.0). Under the terms of this licence, anyone can copy, distribute and display the material, providing that they credit the original source and don't use it for commercial purposes or make derivatives. Different licences may apply to some illustrative elements, in which instance the licence will be displayed alongside. IIED is happy to discuss any aspect of further usage. Get more information via www.iied.org/about-publications

Cover photo: representatives of partners in the 'Generating Ambition for Locally Led Adaptation' programme assessing evidence of impact during an annual learning reflection workshop. Credit: Paul Mitchell/IIED.

IIED is a charity registered in England, Charity No.800066 and in Scotland, OSCR Reg No.SC039864 and a company limited by guarantee registered in England No.2188452.



Beyond the principles

Evidence of whether locally led adaptation works in practice

Anaa Hassan, Paul Mitchell and Aaron Acuda

Contents

Abbreviations	5
Summary	6
1 Introduction	8
1.1 Purpose: why we built the framework	8
1.2 What the literature tells us, and the gap this framework fills	9
2 The LLA Evidence and Impact Framework	12
3 Methodology	14
3.1 Literature identification and selection	14
3.2 Application	16
3.3 Testing and refinement	17
3.4 Limitations	21
4 Results and analysis	22
4.1 Composition of the corpus	22
4.2 Overall classification	23
4.3 Evidence strength across categories	23
4.4 What robust LLA looks like in the literature	24
4.5 Interpreting the No Evidence of LLA and Not Classifiable clusters	25
4.6 The state of play: where robust evidence is strongest and where the gaps lie	25
4.7 Evidence gaps and implications for future research	29
4.8 Discussion: key findings	30
4.9 Implications for the framework	30
4.10 Implications of the limitations for the findings	30
5 Conclusions and way forward	32
Annex 1: The LLA Evidence and Impact Framework	34
References	40

Abbreviations

AI	Artificial intelligence
GA-LLA	Generating Ambition for Locally Led Adaptation (programme)
GGA	Global Goal on Adaptation
IIED	International Institute for Environment and Development
LDC	Least Developed Country
LLA	Locally led adaptation
NRM	Natural resource management
P1–P8	The eight principles for Locally Led Adaptation
SIDS	Small Island Developing States
USAID	United States Agency for International Development

Summary

Locally led adaptation (LLA) has moved rapidly from the margins to the mainstream of climate policy. The eight principles for LLA,¹ endorsed by more than 150 institutions, including governments, since 2021, are increasingly shaping how adaptation finance is discussed and designed. What has not kept pace is the evidence base. A review of the literature, summarised in this report, found that while conceptual and policy interest in LLA is growing — particularly among donors, multilateral climate funds and development agencies — very few instruments exist to systematically document how LLA is implemented and what kinds of evidence demonstrate that adaptation is genuinely both locally led and impactful. LLA-specific tools are limited, and those available simply assess whether a project succeeded; almost none interrogate the quality and robustness of the evidence on the decision making, agency and participation that make adaptation ‘locally led’ in the first place.

To address that gap, IIED, as a part of the Generating Ambition for Locally Led Adaptation (GA-LLA) programme,² has developed an LLA Evidence and Impact Framework. The framework assesses the eight LLA principles across two complementary dimensions: Process (how an intervention was designed and led) and Impact (what changed as a result). Each question is scored on a 0–3 scale (or marked ‘not applicable’), and together, these ratings place the strength of the available evidence into one of four categories: No Evidence, Limited Evidence, Substantive Evidence, or Robust/ Triangulated Evidence. Importantly, the framework assesses the strength of the evidence for LLA rather than passing judgement on the underlying project; established development tools already exist for evaluating project quality itself.

This report introduces the framework and then describes its first application at scale. Using a structured, mixed manual and artificial intelligence (AI)-assisted review, the framework was applied to 106 unique papers spanning 2005 to 2026 and covering a wide geography (the Arctic, the Caribbean, the Pacific, South and Southeast Asia, sub-Saharan Africa and parts of Latin America and the Middle East) and all sectors covered by the Global Goal on Adaptation (GGA). The exercise had three aims: to map where LLA-aligned evidence exists in the literature; to stress-test whether the framework can be operationalised consistently across diverse document types; and to establish the current status and quality of the LLA evidence base, identifying where robust evidence is strongest, where it is thin and what biases it carries.

Reaching this corpus, and this version of the framework, required iteration. An initial application to a larger candidate pool of 299 papers was checked against independent human review and found to be systematically too lenient. This was used to drive two successive revisions of the scoring codebook and a careful manual re-filtering of the working corpus. We then re-scored every paper in the final 106-paper corpus from the source text under the current codebook. Section 3.3 sets out this process in full, since understanding it is essential to trusting what follows.

Three findings stand out. First, LLA-aligned practice is real, but the aggregate picture is sobering: 21.7% of the corpus (23 of 106 papers) meets the Robust LLA threshold of high evidence on both Process and Impact, while 45.3% (48 papers) falls into No Evidence of LLA. Document type matters enormously: 49% of project evaluations are categorised as Robust LLA, while just 15% are categorised as No Evidence of LLA — close to a mirror image of the aggregate pattern. Case studies, policy analyses and conceptual framing papers score far lower, in part because many cannot structurally document the funding and legacy evidence the framework assesses. Stratifying by document type is therefore essential to reading the results most usefully.

Second, two further outcomes — ‘Not Classifiable’ (11.3%, 12 papers) and ‘Excluded from corpus’ (2.8%, 3 papers) — capture sources the framework is not designed to evaluate: rich research on gender, Indigenous knowledge or adaptive capacity that studies a phenomenon rather than an implemented project. Distinguishing these from genuine No Evidence of LLA is one of the framework’s more consequential design features.

1 www.iied.org/principles-for-locally-led-adaptation

2 To learn more about the GA-LLA programme see: www.iied.org/generating-ambition-for-locally-led-adaptation

Third, there is clear evidence that geographic and sectoral patterns are, in places, counter-intuitive. Robust evidence is not concentrated in Small Island Developing States (SIDS) and Least Developed Countries (LDCs). SIDS/LDC papers score marginally lower on average than the rest of the corpus. Health-related LLA evidence remains close to absent outside North America and the Caribbean, and Actions and Outcomes is the weakest evidence category across the corpus, reflecting how hard it is to evidence tangible, equitably distributed, community-led outcomes specifically.

These findings come with four important caveats. First, because the corpus is English-language only, francophone, lusophone and other non-English literatures are under-represented, which may (at least partially) explain the geographic gaps. Second, the framework can only score what is documented, so silences — especially on funding and legacy — may reflect reporting conventions as much as practice. A lack of evidence should not be misconstrued as implying a lack of action; rather, this is a review of a set corpus of literature and certain dimensions cannot surface in document types not designed to address them.

Third, the process of scoring and categorisation inevitably introduces 'edge-cases': papers that sit immediately above or below the threshold for inclusion in any one category. In this research, 23 of the 106 papers were categorised as Robust LLA. A number of borderline papers sit just below the threshold for inclusion in this category; including them would raise the 'Robust' count and shift some shares at the margin, even though the overall patterns would hold. And finally, it's important to acknowledge that AI-assisted scoring carries scope for bias. Several steps were taken to minimise this risk: the framework was piloted manually first and the AI-assisted scoring was validated against manual scoring on around 15% of the corpus and cross-checked and resolved by the research team.

Fourth, this project is itself a documented example of what it takes to use AI responsibly for evidence synthesis at scale. AI-assisted scoring made it feasible to apply a detailed framework consistently across the full corpus and produced a text-anchored justification for every score — work that would otherwise have required many months of manual review. That efficiency came with a condition though: the results could only be trusted once they had been checked against independent human judgement. They were found wanting on specific, identifiable grounds and so were iteratively corrected until the AI and human scoring converged. Section 3.3 documents this process in full, including where it initially fell short, in the belief that an honest account of what AI-assisted review requires to be trustworthy is as useful a contribution as the evidence findings themselves.

Taken together, this application of the framework shows that the framework is fit for purpose, with refinements. It produces interpretable, internally consistent readings that align with expert understanding of the field, and it surfaces both explicit LLA and a large body of 'incidental' LLA in adjacent literatures.

1

Introduction

1.1 Purpose: why we built the framework

Over the past decade, LLA has shifted from a niche idea championed by a handful of practitioners to a central organising concept in international climate policy. The eight principles for LLA, developed in 2020 and with increasing signatories since 2021, crystallised a broad consensus: that the people and communities most exposed to climate risk should have genuine agency over how adaptation is designed, financed and delivered, and that doing so produces more effective, more equitable and more durable results. The principles are now widely endorsed,³ and the language of 'local ownership', 'direct-to-local finance' and 'downward accountability' appears routinely in donor climate funding strategies and adaptation funding mechanisms.

Endorsement, however, is not the same as evidence. As LLA has gained traction, a practical question has become harder to avoid: how do we actually know when adaptation is locally led, and how do we know whether LLA approaches are effective? Answering these questions requires more than a definition. It requires a shared, structured way of recognising and weighing the evidence, a means of distinguishing an intervention that is genuinely shaped and governed by local actors from one that merely consults them, and a way to tell the difference between a strong claim and a well-documented one.

A review of the existing literature, undertaken as part of this work and summarised in Section 1.2 below, confirmed that this is where the field is weakest. While conceptual and policy interest in locally led approaches is abundant, relatively few frameworks exist to systematically document how LLA is implemented in practice and what type of evidence demonstrates that adaptation processes are genuinely both locally led and impactful. The instruments that do exist, such as CARE's community-based adaptation framework (CARE International, 2010), which measures success against elements such as participation and community ownership, the evaluation framework that Ford (Ford et al., 2018) applied to determine what makes a 'successful project', and IIED's Tracking Adaptation and Measuring Development approach and the common community-based adaptation success features set out by Rashid and colleagues in 2013 (Rashid et al., 2013), tend to assess project success or outcomes. They rarely examine the nature of the evidence generated on decision making, local agency and participation — the very things that make adaptation 'local'.

The LLA Evidence and Impact Framework was developed by IIED, in consultation with partners in the GA-LLA programme drawing on experience supporting adaptation initiatives across multiple countries, to help close that gap. It is deliberately not an attempt to define what LLA 'is' in absolute terms, nor a replacement for project- or programme-level monitoring and evaluation. It is a practical lens for compiling, assessing and communicating the strength of evidence on how locally led approaches contribute to resilience, forming a common set of criteria against which the robustness of that evidence can be measured and compared across contexts.

This report documents the first time the framework has been applied at scale. As such, the application was designed around three intertwined objectives. The first was to map the evidence base: to identify within the existing literature where adaptation interventions were explicitly designed and labelled as locally led adaptation, and where interventions

³ LLA Alliance: join the endorsers of the locally led adaptation principles — www.iied.org/lla-alliance-join-endorsers-locally-led-adaptation-principles

aligned with the LLA principles without ever being framed as such (what we term 'incidental' LLA). The second was to test and refine the framework: to verify whether it can be operationalised as a consistent analytical lens across diverse document types, and to surface where its scoring guidance, document-type handling and AI prompt structure need adjustment. The third was to establish a state of play of the robust LLA evidence base: to interrogate the subset of papers that meet the robust evidence threshold and assess where the evidence is strongest, where the gaps lie and what geographic, thematic and methodological biases the published record carries.

Throughout, one distinction is fundamental and easy to overlook: the framework scores the strength of available evidence, not the inherent quality or success of the underlying projects. A low score signals that a paper does not document a given dimension of LLA, which may reflect the conventions of academic writing as much as anything about practice on the ground. Keeping the distinction clear is critical to ensure the results of this analysis inform the scaling of LLA approaches from the evidence base and do not imply that other adaptation approaches do not have validity.

Getting this right has practical implications, not merely academic ones. Adaptation finance is scaling, and a growing share of it is being justified on locally led grounds (in other words, direct-access modalities, enhanced direct access, community resilience funds and the like). If the field cannot distinguish robust evidence of local leadership from the confident assertion of LLA, then the case for shifting money and decision-making power towards local actors rests on advocacy alone, and is correspondingly fragile when challenged. A shared, structured way of weighing evidence does two things at the same time: it strengthens the legitimate case where the evidence is strong, and it makes the gaps transparent and visible where it is not. Both are necessary if LLA is to move from a concept that institutions endorse to a practice that finance consistently supports.

A further reason for a structured lens is that much LLA-aligned practice is never labelled as such. A great deal of the most rigorous community-led adaptation appears in literatures that predate or sit alongside the LLA framing: community-based adaptation, Indigenous-led adaptation, participatory action research, community-driven development, and so on. Treating only self-identified 'LLA' work as relevant would discard most of the available evidence. Applying the framework retrospectively can surface this 'incidental' LLA: practice that meets the substance of the principles without ever using the vocabulary, and which would otherwise remain invisible to anyone searching by label alone. We anticipate that this will shift over time as the language of LLA becomes more prevalent in the literature.

Five questions guided the work:

1. What proportion of the local-level adaptation literature provides evidence of LLA-aligned Process and Impact?
2. Which evidence categories are most consistently documented, and which are systematically under-reported?
3. How does evidence quality vary by document type?
4. Within the subset of papers with robust evidence, how is that evidence distributed across regions and sectors, and where do the most consequential gaps lie?
5. And, threaded through all of these, what does the corpus reveal about the practical applicability of the evidence framework itself?

1.2 What the literature tells us, and the gap this framework fills

The framework was built on a structured review of how academic and grey literature define and present valid evidence of LLA. The review was concerned with how evidence is generated and what makes it credible. The review began with an initial Google Scholar search using the keywords 'locally led adaptation', 'locally led climate action' and 'local climate adaptation' in titles and abstracts, yielding a total of 49 papers for analysis. To expand the scope and strengthen the evidence base, a second search was conducted on Google Scholar incorporating broader terms related to community-based approaches. This included the keywords 'community-based AND adaptation' and 'community-driven AND adaptation'. The results were then filtered to include only studies that explicitly referenced climate change adaptation within the social sciences field and also excluded blog posts. The search focused on literature published from 2010 onwards, leading to an additional 63 papers being reviewed. In total, 112 papers were evaluated for this analysis. (This 112-paper review is distinct from, and should not be confused with, the working corpus described in Section 3.1: it was conducted separately to inform how the framework itself should be designed, in other words, what counts as credible

evidence of LLA, and which existing evidence frameworks were worth drawing on, rather than being the set of papers the finished framework was subsequently applied to.) Four lessons shaped the design of the framework, and a fifth — the absence of any existing instrument fit for the task — explains why a new framework was needed.

First, on method: the most robust evidence on LLA comes from approaches rooted in local realities, treating communities as primary knowledge-holders and triangulating multiple sources. Most empirical papers triangulated their findings, and decolonising, participatory and grounded methods were widespread. These include research rooted in Indigenous Fijian values using the Vanua methodology and Talanoa approach (Yee et al., 2025), grounded research across Bangladesh, the Philippines and Vanuatu (Westoby et al., 2021) and storytelling-based Tok Stori methods in the Solomon Islands (Global Center on Adaptation, 2023). Community-centred and anticipatory action research (Bennett et al., 2016) and case study methods assessing appropriateness, effectiveness and equity (Clissold and McNamara, 2019) featured prominently, while survey-only approaches were found to risk positive response bias (McNamara, 2013). The framework's scoring scale reflects this, reserving its top score for multi-source, triangulated evidence.

Second, on definitions: only about one paper in eight defined key parameters such as effectiveness, sustainability, scalability, appropriateness and equity. Where they did, the definitions were inconsistent. Effectiveness was framed as meeting objectives efficiently (Ford et al., 2018), as inevitably shaped by systems of power (Bedelian et al., 2024), or as a donor-driven, top-down construct (Mills-Novoa, 2023); sustainability as institutionalising local entities for lasting legacy (Tye and Suarez, 2021; Bedelian et al., 2024); scalability as embedding approaches in enabling policy frameworks (Reich, 2025); appropriateness as cultural compatibility (Clissold and McNamara, 2019); and equity as inclusive, justice-oriented participation (Omukuti, 2020; Fisher, 2012; Rajabali et al., 2024; McNamara et al., 2020). Because two papers can both claim 'effective' adaptation while meaning quite different things, the framework anchors its categories to concrete, observable questions rather than to contested labels.

Third, on inclusion: fewer than a quarter of papers explicitly documented how gender and intersectionality shaped the evidence they generated, a core concern of the GA-LLA programme and of LLA approaches more broadly. This is better understood as a reporting gap than as a practice gap. As noted above, decolonising and participatory methods were widespread, and researchers using them are often already attentive to gender-sensitive basics such as holding separate focus groups; what was comparatively rare was making that engagement visible as documented evidence. The distinction matters, because inclusion that is practised but never recorded cannot be assessed, compared or learned from. It matters for quality too: conventional engagement can entrench the very inequalities it claims to address, since women may be reluctant to speak in male-dominated settings, opening the door to elite capture (Tye and Suarez, 2021). Where papers made their approach explicit, responses included separate focus group discussions for women in patriarchal contexts (Beauchamp et al., 2021) and conducting research in settings women prefer (Yee et al., 2025; Rajabali et al., 2024). The framework reflects this by treating the mapping of structural inequalities and the documented influence of marginalised groups as scored evidence in their own right, rather than as optional context.

Fourth, on attributes: across the reviewed work, valid evidence was consistently marked as flexible and context-specific (Tye, 2021), multi-scalar and systematic (Westoby et al., 2019; Bryan and Behrman, 2013; Gero, 2011), critically attentive to what 'local' means and to power dynamics between actors (Yee et al., 2025; Akampurira, 2024), co-produced, plural and justice-aware (Global Center on Adaptation, 2023; Roche et al., 2020; Mannke, 2011; Tye et al., 2023; Rahman et al., 2023), and culturally sensitive, grounded in community values (McNamara, 2013; Reich, 2015). These qualities directly informed the framework's design — its emphasis on flexibility, local agency and power, and collaboration and co-production — and underpin its insistence that Indigenous, traditional and locally situated knowledge counts as evidence so long as it is credible and verifiable within its context.

Crucially, the review found no readily applicable evidence framework for LLA: existing tools assess project success rather than the nature of the evidence itself. Several LLA-focused publications and evaluative tools, however, offered valuable elements to inform our evidence framework. A justice-and-power framing integrating Fraser's three dimensions of injustice — redistribution, recognition and representation (Fraser, 2006) — with Lukes' three faces of power — decision making, non-decision making and hegemonic power (Lukes, 1974) — shows how inequalities between elite organisational actors and marginalised communities shape the adaptation solutions that emerge, and argues that LLA should actively shift decision-making autonomy towards local actors (Rahman et al., 2023). A continuum model developed by the United States Agency for International Development (USAID) classifies interventions along a spectrum of local leadership. People-centric framings added further substance: a widely used community-based

adaptation framework measuring success against participation, informed and inclusive decision making, and community ownership (CARE, 2010); the common success features distilled by Rashid et al. (2013) — people-centred, process-driven, community-led, knowledge-oriented, empowerment-focused and accountability-driven; the four optimisation levers identified by McNamara et al. (2020); and an evaluation framework applied in Indigenous communities (Ford et al., 2018). The framework draws on these complementary insights but repurposes them: rather than judging whether adaptation is effective or how 'local' it is, it consolidates them into a single, standardised structure, built around the eight LLA principles, for assessing something the existing literature does not — the robustness, consistency and comparability of the evidence that adaptation is genuinely locally led.

2

The LLA Evidence and Impact Framework

The framework operationalises the eight LLA principles across two complementary parts comprising six evidence categories. Part 1, Process (Context, Design, Funding) concerns how interventions were conceived, covering the use of local and Indigenous knowledge, decision-making structures, and funding attributes; it comprises Categories 1 to 3. Part 2, Impact (Outcomes, Capacity, Learning) concerns what changed as a result, and covers tangible outcomes, institutional legacy, sustainability and learning systems; it comprises Categories 4 to 6.

Category 1: Context and Risk captures whether local and Indigenous knowledge of hazards, vulnerabilities and structural inequalities informed the design baseline.

Category 2: Design and Decision Making covers whether local stakeholders, especially marginalised groups, had genuine influence or veto over decisions, and whether power-sharing arrangements were documented.

Category 3: Funding Attributes addresses the financial mechanics of the intervention: whether funds flowed directly to local actors, whether there was flexibility and contingency, what the overhead structure looked like, and whether follow-on funding was secured. This category speaks most directly to the common critique that adaptation funding flows are opaque.

Category 4: Actions and Outcomes covers what was implemented and what changed for whom, including evidence on equity of outcomes.

Category 5: Capacity and Sustainability covers post-project sustainability, transfer of capacity to local institutions and whether the intervention left durable structures in place.

Category 6: Learning and Transparency covers the documentation practice of the intervention itself — whether methods, assumptions and decisions were recorded, whether learning was captured in feedback loops and whether reporting was accessible to non-specialist audiences. Importantly, Category 6 captures documentation and learning transparency, not funding transparency. This distinction matters when reading the results, since Category 3, Funding Attributes, consistently scores low even where Learning and Transparency (Category 6) scores high. The two are conceptually and empirically separate.

Each category contains three to five evaluation questions (20 in total), scored on a 0–3 scale with a not applicable (N/A) option for papers not suited to a given question. The scoring scale, set out in Table 1, is the analytical heart of the method, and it is important to clearly state its logic. It measures how well something is evidenced, not how good it was. The four points on the scale correspond to four bands of evidence strength — No Evidence (0), Limited Evidence (1), Substantive Evidence (2) and Robust/Triangulated Evidence (3) — and each score carries a brief justification referencing the specific source. Because the questions sit within the framework's two dimensions, Process and Impact are scored separately and then read together, so that a well-designed process with little demonstrated impact, or a strong outcome with weak evidence of local leadership, is surfaced rather than averaged away.

The full framework, including all questions and scoring guidance, is reproduced in Annex 1.

Table 1. Scoring scale to assess the level of evidence of LLA in a given document

Score	Level	Evidence threshold
0	No Evidence	No verifiable documentation; claims unsupported by records or traceable sources.
1	Limited Evidence	Some indication of LLA principles, but partial or descriptive; without triangulation.
2	Substantive Evidence	Clear, documented, verifiable evidence, including in-depth anecdotal testimony.
3	Robust/ Triangulated Evidence	Multiple sources, including records, observations and local testimonies, demonstrating consistent practice and clear outcomes.
N/A	Not Applicable	The question's underlying mechanism cannot structurally apply to this source — not merely unmentioned, but inapplicable.

The three category scores within each part are averaged to give a single score for Process and a single score for Impact, with 1.5 acting as the threshold between 'low' and 'high'. The two part-scores are then combined, following Table 5, to place each study in one of four evidence-based classifications: Robust LLA (high Process, high Impact); Strong process LLA (high Process, low Impact); Top-Down Impact (low Process, high Impact); or No Evidence of LLA (low Process, low Impact). Because every score rests on documented evidence, two further outcomes arise where that evidence is too thin to support a number. Where two or more of the three categories on one side are marked N/A, that side is treated as Not Classifiable and only the classifiable side is reported — common for research that studies a phenomenon rather than an implemented project. Where two or more categories are N/A on both sides at once, the study is excluded from the corpus and left out of the denominator in any classification-percentage calculation.

3

Methodology

The study used a structured, mixed manual-and-AI-assisted literature review, organised in three sequential stages: identification and selection of literature; testing and refinement of the framework; and full-corpus application and synthesis. The design was chosen to balance the breadth that automated tools make possible with the judgement that interpretive analysis of adaptation evidence demands. Each stage is described below.

3.1 Literature identification and selection

We identified literature to review using the framework through a two-stage search. We first sourced candidate papers through parallel keyword searches, using both AI-assisted tools (Elicit AI and Research Rabbit) and manual searches on Scopus and Web of Science. Our search combined two concept clusters — an adaptation cluster (terms such as 'adapt*', 'resilien*', 'risk*', 'climat*', 'environment*') and a locally led cluster ('locally-led', 'locally led', 'community-led', 'community-driven', 'Indigenous-led', 'grassroots adaptation', 'participatory adaptation' and related variants). We included documentation from 2005 onwards, anchoring the review to the year of the first international Community-based Adaptation conference, after which the concept gained substantial traction in both implementation and documentation.

Screening then narrowed the field in stages. The manual Scopus and Web of Science search alone returned roughly 2,500 papers, while Elicit AI and Research Rabbit AI tools returned 100 peer-reviewed articles and 179 grey and academic literature results, respectively. The combined results were then exported to Rayyan AI, which identified 135 duplicates, resulting in 2,644 unique records. Manual light title screening reduced this to around 1,029, and further manual abstract screening produced 340 candidate papers for full-text review. Full-text PDFs were uploaded to Elicit AI. Applying the full inclusion and exclusion criteria in Elicit AI, clearly checking the outputs and applying a further de-duplication check resulted in a final corpus of 299 unique papers for analysis (Refer to Tables 2 and 3 and Figure 1 for a summary).

Table 2. Literature identification

Key literature source	Results
Scopus and Web of Science	2,500 items of academic literature
Elicit AI	100 peer-reviewed articles
Research Rabbit AI	81 peer-reviewed articles and 98 items of grey literature

Table 3. Screening process

Operation and tool	Results
Initial duplicate identification	Removal of 135 duplications, resulting in 2,644 unique records
Manual title screening	1,029 unique records
Manual abstract screening	340 candidate papers for full review
Application of full inclusion and exclusion criteria in Elicit AI	299 articles for final analysis

This 299-paper corpus was the starting point for the validation and revision process described in Section 3.3. Following that process, the corpus was manually re-filtered — removing duplicates missed in the first pass, papers that, on closer reading, did not describe a real, identifiable adaptation effort, and papers found to be near-duplicates under different reference identifiers — down to a working set of 108 papers considered suitable for fresh, from-scratch re-scoring. During that re-scoring exercise, a small number of additional duplicate uploads were identified and resolved, yielding a final, verified, de-duplicated working corpus of 106 unique papers. Every figure and table in Section 4 of this report is based on this 106-paper corpus, re-scored under the current codebook, unless explicitly stated otherwise. The inclusion and exclusion criteria are set out below.

Figure 1. Literature search, screening and validation flow

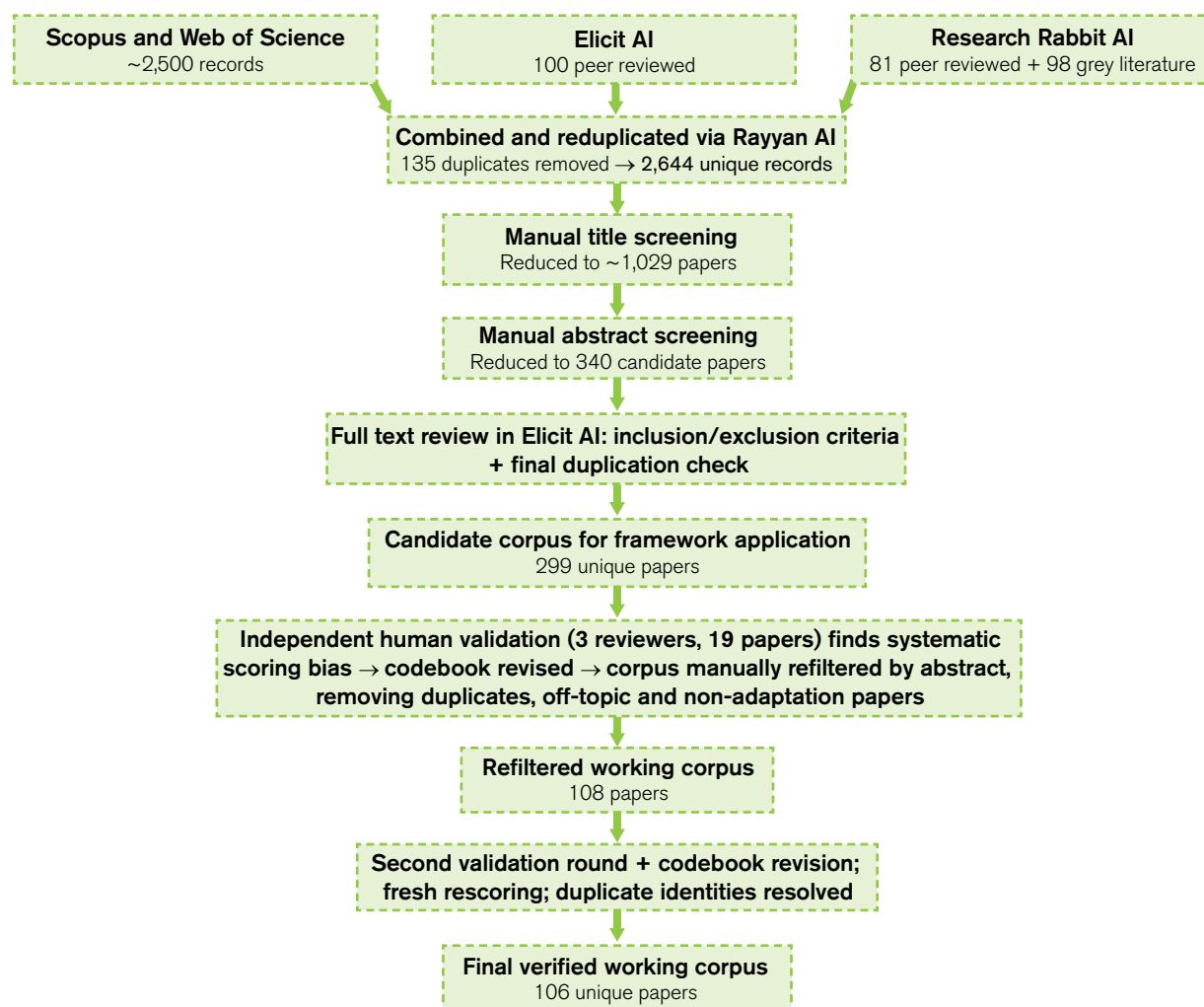


Table 4. Inclusion and exclusion criteria for literature search

Dimension	Inclusion	Exclusion
Time	Published 2005–2026	Published before 2005
Language	Full text in English	Abstract only in English; full text in another language
Document type	Peer-reviewed articles; grey literature (reports, policy briefs, working papers)	Books, conference proceedings, pure project documents without analytical content
Scope	Adaptation focus with a clear local/ community-level component	Mitigation-only; or national/regional analysis without a local sub-component
Evidence	Empirical data, implemented interventions, or systematic reviews of implementation	Purely theoretical papers; insufficient methodological detail

3.2 Application

As described above, before scaling we manually applied the framework to a small sample of papers, to confirm that the evidence categories generated interpretable results, that we could score consistently, and that any ambiguities surfaced. This pre-test led us to refine the scoring guidance, the document-type classifier we use to tag each paper, and the structure of the analytical prompt.

Once validated, the structured prompt was applied to the full corpus using Claude AI. For each paper, the AI extracted a score from 0 to 3 for each question with a text-anchored justification, the type of evidence observed, the category averages, the Process and Impact averages, a final classification, a confidence rating, and explicit 'assumptions avoided' and 'key evidence gaps' entries designed to keep scoring conservative. We then compared the manual and AI outputs on the pilot subset to test inter-method agreement. Crucially, the research team retained final analytical interpretation throughout, treating AI outputs as structured analytical input rather than as definitive conclusions.

Two features of the AI-assisted stage deserve emphasis because they shape how the results should be trusted. First, the requirement for a text-anchored justification on every score means that no score floats free of the source: a paper could not be credited with, say, robust funding evidence unless the model could point to specific passages documenting fund flows. This is what makes the conservative bias on funding and capacity a feature rather than a flaw. The AI was explicitly instructed not to infer beyond what the text supports, so a low funding score reliably indicates silence in the source rather than a generous or stingy reading. Second, the 'assumptions avoided' field forced the analysis to record where it declined to fill gaps, creating an audit trail of the corpus's blind spots that itself became analytically useful, surfacing the recurring funding, follow-on and post-project silences reported in Section 4. The document-type classifier, refined during the pre-test, then allowed every result to be read against the document type that produced it, which, as the results make plain, is indispensable to interpreting the headline figures correctly.

Each paper was finally mapped to one of five quadrants based on its Process and Impact averages, with 1.5 acting as the threshold between low and high evidence.

Table 5. Classification quadrants

Quadrant	Process/Impact	Interpretation
Robust LLA	High LLA Process (≥ 1.5)/ High adaptation Impact (≥ 1.5)	Strong local ownership with verifiable, equitable outcomes.
Strong Process LLA	High LLA Process/Low adaptation Impact	Strong local leadership; outcomes still maturing.
Top-Down Impact	Low LLA Process/High adaptation Impact	Results achieved but with weak local legitimacy; sustainability at risk.
No Evidence of LLA	Low LLA Process/Low adaptation Impact	Neither approach nor outcomes align with LLA; fundamental redesign required.
Not Classifiable	More than two out of three categories N/A on one side (Process or Impact)	Insufficient evidence to trust a number on that side; commonly diagnostic or social-science research rather than project evaluation.

Two features of this design carry through to the results and should be kept in view. First, the framework was built to assess the quality and strength of evidence of LLA, not to evaluate project or programme performance. Even so, several of its evidence categories, particularly those concerning funding flows and post-project sustainability, depend on the kind of documentation that project evaluations and monitoring and evaluation systems generate, and are simply not addressable in document types such as national policy analyses or conceptual reviews. Tagging every paper with a document-type classifier was therefore essential, so that results could be read in a stratified way rather than as a single undifferentiated total. Second, the 1.5 cut-off between low and high evidence is a deliberate analytical convention rather than a natural boundary. It was adopted to enable the two-by-two classification: sorting each of the Process and Impact axes into 'high' (≥ 1.5) or 'low' (< 1.5) is what yields the four quadrants, whereas a graded low/medium/high scale on each axis would produce a nine-cell matrix that is harder to interpret and to act on. The rule was applied uniformly and mechanically — Process ≥ 1.5 is high and < 1.5 is low, and likewise for Impact — with no manual override

or case-by-case judgement, including for papers sitting directly on the threshold. The trade-off is that a binary split discards information near the boundary and makes a paper's quadrant sensitive to small scoring differences around 1.5. Rather than adjust these cases, we keep them visible: the underlying 0–3 question scores and the continuous Process and Impact averages are reported alongside each quadrant, and papers falling close to the cut-off are flagged as borderline, so that a paper averaging 1.4 is not read as equivalent to one averaging 0.4. This preserves a single, reproducible classification rule while alerting readers to the cases where a high/low label rests on a narrow margin.

3.3 Testing and refinement

Before scaling, the framework was piloted manually on a small sample of around 30 papers (approximately 10% of the candidate corpus) to surface ambiguities, confirm that the evidence categories generated interpretable results and confirm that scoring could be applied consistently. After this pilot, a structured prompt implementing the framework's original 23-question codebook (referred to below as v1.1) was applied to the full 299-paper candidate corpus using an AI system (Claude AI). For each paper, the AI produced a 0–3 score for every question with a text-anchored justification, along with category averages, Process and Impact averages, a final classification, and a confidence rating. Applied at this stage, the framework classified 79 of the 299 papers (26.4%) as Robust LLA and 176 (58.9%) as No Evidence of LLA.

Validation round 1: three independent human reviewers, 19 papers

To test whether this AI-assisted scoring could be trusted, three independent human reviewers — referred to here as Reviewer A, B and C — separately re-scored a sample of papers by hand, using the same 23-question codebook, without seeing the AI's scores first. The logic was simple: if the AI's scoring were sound, its classifications should broadly agree with a careful human reading of the same papers.

They did not agree. Reviewer A, who scored strictly and required an explicit citation or statement before crediting any evidence, found the AI scored on average 1.64 points higher than their own scoring, on both Process and Impact, on a 0–3 scale. Reviewers B and C scored more generously, willing to credit a paper's general discussion even without a direct citation — and even against this more generous standard, the AI still scored 0.9 to 1.1 points higher on average. Reviewers A and C, two humans independently reading the same papers, differed from each other by about 1.1 points, establishing that some disagreement between careful readers is normal. Set against that baseline, the AI's classification matched Reviewer A's on only 1 of 19 papers checked — about 5% agreement.

Three specific, repeatable causes were identified. First, a reviewer-stance problem: the AI tended to credit a paper's general tone and context rather than requiring a specific, citable statement, crediting what a paper implied rather than what it actually said. Second, a framework-fit problem: many papers in the corpus were literature reviews, opinion pieces or policy discussions rather than evaluations of one or more specific projects, programmes or interventions, yet the AI was still attempting to answer all 23 questions for these papers, scoring some based on things the document type was never designed to address. Third, a category-specific problem in Learning and Transparency: the AI frequently mistook an author's own academic reflection on their research methods for evidence of a genuine, project-level community learning practice.

Further inspection also confirmed a more basic pattern: of the 79 papers originally classified as Robust LLA under v1.1, only about a third held up under a fresh, careful re-read once the codebook issues above were corrected — the remainder mostly dropped to No Evidence of LLA, and none moved in the opposite direction. This one-directional pattern of correction — almost always downward, never upward — is itself strong evidence that the original scoring carried a systematic leniency bias rather than random noise; a genuinely random scoring error would be expected to correct in both directions roughly equally.

Codebook revision 1, and re-filtering the corpus

These findings drove a full rewrite of the scoring codebook, targeting the three causes identified above: a clearer evidence standard requiring a paper to specifically and substantively discuss a topic before it can be credited, rather than scoring what seemed generally plausible for that type of document; a formal framework-fit check applied before scoring, so that questions inapplicable to a review, opinion piece or policy discussion are marked 'not applicable' rather than scored and forced; and a rule explicitly distinguishing an author's own methodological reflection from a genuine project-level learning practice.

Separately from the codebook, the 299-paper corpus itself was manually reviewed and re-filtered — removing duplicates, off-topic papers, and papers that did not actually describe a real, identifiable adaptation effort — down to a working set of 108 papers considered suitable for a full, fresh re-scoring exercise. A calibration check was then run: a small batch of previously human-validated papers was re-scored under the revised codebook and compared against the original human scores, confirming the revised framework was landing closer to human judgement before it was applied to the full 108-paper set.

Full re-scoring under the revised codebook

The 108 filtered papers were then re-scored from scratch, in six batches, reading each paper fresh rather than adjusting the old AI scores. Compared against the original 299-paper corpus, the shift was dramatic: the Robust LLA share fell from 56.4% (measured on the subset of the 299-paper corpus corresponding to these same papers) to 23.8%, while No Evidence of LLA rose from 29.1% to 65.3%. As with the first validation round, the correction was strongly one-directional — of papers originally No Evidence of LLA, 80% were confirmed as No Evidence of LLA again, and none moved up to Robust LLA — reinforcing that the original scoring bias was systematic rather than random.

Validation round 2: two further reviewers, ten papers, and specific remaining issues

To test the revised codebook, two reviewers — referred to here as Reviewer C and Reviewer D — each independently and manually re-scored five papers (ten papers total), comparing them question-by-question against the AI's scores on the same papers. This round showed that the revision had resolved the systemic leniency bias: the AI was no longer consistently too harsh or too lenient on average. But four specific, repeatable problems remained.

- The AI sometimes scored a genuinely inapplicable question as 0 (no evidence) rather than marking it N/A — for example, scoring a project 0 on international-funding transparency when the project was entirely funded locally, rather than recognising the question simply did not apply.
- The AI under-credited real community decision making power when the evidence took the form of an informal but specific quote (for example, a community member describing having declined a proposal) rather than a formally named approval process.
- The AI was too strict on 'learning loop' questions, requiring a whole formal, ongoing feedback system before giving credit, when a single well-documented instance of a plan changing in response to community input should have counted.
- Two reviewers, on two different papers, both independently flagged that the paper was not really evaluating an implemented adaptation project at all — one was a monitoring/research programme, one a general study of social dynamics — yet the AI had scored both fully, as if they were standard project evaluations.

A further, more mundane but important finding emerged directly from building the comparison: two real data-handling errors were found in how the validation spreadsheets themselves had been assembled, including one paper's scores being misaligned by one question, and two papers wrongly treated as 'declined by the reviewer', when in fact they had been scored in full under a mislabelled paper title. Once corrected, one paper (a Solomon Islands community-development case study) matched the AI's score exactly, highlighting that some of what looks like reviewer disagreement is really data-quality noise that a careful cross-check can and should catch.

Codebook revision 2

This second round drove a further, more structural revision. Not Applicable was formalised as a genuine fourth scoring option, distinct from a score of 0: N/A questions are now excluded from category averages rather than counted as a zero, resolving the first issue. Community refusal or veto evidenced by a specific, corroborated quote is now credited at the same level as a formally named approval process, resolving the second issue. A single, well-evidenced instance of a plan changing in response to community feedback is now sufficient for the relevant Learning and Transparency and Actions and Outcomes questions, rather than requiring a documented, ongoing, formal system, resolving the third.

The fourth issue — papers not really evaluating an implemented project at all — required a new rule rather than a re-wording. A proactive exclusion clause was added, letting the framework recognise when a source's Process and Impact sides both lack any real evidential basis (formally, when two or more of three categories are N/A on both sides simultaneously) and recommend the source be excluded from the corpus entirely, rather than forced into a classification

that would misrepresent it. A related, less severe safeguard was added for the more common case where only one side (Process or Impact) lacks a basis: such sources are now labelled Not Classifiable on that side rather than assigned a misleadingly precise number computed from one or two remaining data points.

Alongside these rule changes, the question set itself was directly edited. In Funding Attributes, one question (on flexible mid-course funding adjustments) was removed as it substantially overlapped with a remaining question, and the international-overhead question was reworded to apply, and be scoreable, only where external funding exists at all. In Actions and Outcomes, three questions were reworded to lower the evidentiary bar in the ways that validation round 2 indicated were reasonable: crediting a change-mechanism that existed and was used even once, rather than requiring proof a specific outcome is attributable to it; dropping the requirement that a marginalised group's power specifically 'increased' or be tied to a named project committee; and allowing evolved, not only originally stated, community priorities to count for alignment. In Capacity and Sustainability, two of four questions (on independent post-intervention management, and on securing follow-on funding) were removed, and the remaining capacity-transfer question was reworded to no longer require that the transfer be 'formal and legal'. In Learning and Transparency, the iterative feedback question was reworded to credit either the existence of a mechanism or evidence it was actually used, rather than requiring both.

Full re-scoring under the final codebook and validation round 3

Every paper across all six batches was re-scored fresh from source text under this final, 20-question codebook — not adjusted from its prior score. During this pass, a careful check for duplicate paper identities across batches (by both reference ID and title/content similarity) resolved the working corpus to a final, verified 106 unique papers; this is the corpus underlying every table in Section 4.

A third validation round then re-tested 10 papers — the same 10 papers used in round 2, re-reviewed by a human reviewer under the final codebook and compared against the AI's fresh re-scoring of the same papers. This round found 87% of comparable question-level scores within one point of the independent reviewer's, with no net directional bias (a mean difference of 0.0 across the sample) — evidence that the iterative correction process had converged. A small number of specific disagreements remain open on individual papers and are not claimed to be resolved here; they are documented in the underlying validation records rather than papered over.

This validation history means the classification percentages reported throughout this document should not be read as unusually harsh; they are the result of removing a documented, one-directional leniency bias that the original scoring carried, confirmed independently across three separate rounds of human review. Furthermore, the framework's proactive exclusion clause and Not Classifiable safeguard mean that a substantial minority of the corpus — 15 of 106 papers, or 14.2% — now sits outside the four standard classifications altogether, which is itself a genuine methodological finding: a meaningful share of the published literature on locally led and community-based adaptation is not, in substance, evaluating an implemented adaptation effort, and no framework of this kind should force a score onto it regardless.

Why 1.5 and what would the alternative look like? 1.5 is the exact midpoint of the 0–3 scoring scale, which is why it was chosen: it requires no separate justification beyond being half of the maximum possible evidence strength, whereas any other cut-off would need its own argument for that specific value. To test how much this choice matters in practice, Robust LLA counts were recalculated at thresholds from 1.3 to 1.7 using the same 91 papers with both a Process and an Impact score (the 12 Not Classifiable and three Excluded papers are unaffected by this test either way) (see Table 6, below).

Table 6. Robust LLA count at alternative threshold values

Threshold	Number of papers with Robust LLA	% of 91 scoreable papers	% of full 106	Change compared to 1.5 threshold	Number of papers with No Evidence of LLA
1.3	32	35.2%	30.2%	+9	38
1.4	30	33.0%	28.3%	+7	42
1.5 (current)	23	25.3%	21.7%	—	48
1.6	20	22.0%	18.9%	–3	52
1.7	15	16.5%	14.2%	–8	56

Two things support 1.5 as the right compromise rather than an arbitrary pick. First, the change across this range is gradual rather than a step function — finer-grained testing (every 0.02 between 1.30 and 1.60) found no natural break point or cliff anywhere in the distribution. A gap in the data — very few papers scoring between, say, 1.4 and 1.6 — would have given us a non-arbitrary place to draw the line, because the data itself would be separating two groups of papers. We found no such gap, which is itself evidence that no other specific threshold is more ‘correct’ than the midpoint. Second, the size of the effect is real but bounded: moving the threshold by ± 0.2 shifts the Robust LLA share by roughly 8–9 percentage points of the full corpus in either direction — enough to report transparently (as we do here), but not so large that it undermines the aggregate finding that a genuine minority of the corpus demonstrates robust evidence of LLA on any reasonable threshold in this range.

The narrative above describes an iterative process with several moving parts: corpus size, reviewer and codebook version all changed together at different points. Table 7 sets out every step in sequence, making explicit which papers were involved, who reviewed them (human or AI), and how many papers were in scope at each stage, from the original 299-paper candidate pool through to the final, verified 106-paper working corpus.

Table 7. Sequence of validation steps, reviewers, and corpus size

Step	What happened	Reviewed by	Papers in scope	Codebook version
1	Candidate literature search and screening (keyword search, database screening, inclusion/exclusion criteria applied)	Manual + AI-assisted search	~2,500 initial hits → 299 candidates	—
2	Initial AI-assisted scoring of the full candidate corpus	AI	299 papers	v1.1 (23 questions)
3	Validation round 1: independent hand-scoring of a subsample, compared against the AI scores	Three human reviewers (A, B, C)	19 papers	v1.1 (23 questions)
4	Codebook revised in response to Round 1 findings (evidence standard, framework-fit gate, learning-category fix)	Research team	—	v1.1 → v2.0
5	Manual re-filtering of the full candidate corpus by abstract (removing duplicates, off-topic and non-adaptation papers)	Human (manual review)	299 reviewed → 108 retained	—
6	Calibration check: a small batch of previously human-validated papers re-scored under the revised codebook	AI, checked against prior human scores	Small calibration batch	v2.0 (23 questions)
7	Full re-scoring of the filtered corpus, from source text, under the revised codebook	AI	108 papers	v2.0 (23 questions)
8	Validation round 2: independent hand-scoring of a further subsample, compared against the AI scores	Two human reviewers (C, D)	Ten papers	v2.0 (23 questions)
9	Codebook revised a second time (N/A option formalised, three questions removed, six reworded, Not Classifiable/Excluded outcomes added)	Research team	—	v2.0 → v3.0
10	Full re-scoring under the final codebook; duplicate paper identities resolved during this pass	AI	106 papers (final)	v3.0 (20 questions)
11	Validation round 3: the same 10 papers from Round 2, re-scored by hand under the final codebook and compared again	One human reviewer	Ten papers	v3.0 (20 questions)

Three numbers recur throughout this report: 299 is the candidate corpus assembled at Step 1, before any human validation or re-filtering; 108 is the corpus remaining after manual re-filtering at Step 5, before the final round of AI re-scoring and de-duplication; and 106 is the final, verified working corpus underlying every table and figure in Section 4. A separate, unrelated figure — 112 — appears only in Section 1.3, describing a distinct literature review conducted to inform the framework's design, not the corpus it was applied to.

3.4 Limitations

Six limitations are worth noting. Section 3.4 returns to how several of them shape specific findings.

Language. The corpus is restricted to English-language documents, which structurally under-represents francophone, lusophone and other non-English adaptation literatures. This is a structural factor behind the geographic gaps reported in Section 4.7 and means those gaps should be read as gaps in the searchable evidence rather than proof of absent practice.

Documentary silence. The framework can only assess what is documented. An absence of evidence on, say, funding flows or post-project legacy may reflect the reporting conventions of the literature as much as weak practice on the ground.

Document-type ceilings. Some document types cannot, by their nature, document project-level detail: a national policy analysis or conceptual review will not report fund flows, capacity transfer or follow-on funding. The framework's N/A, Not Classifiable and Excluded outcomes are designed specifically to stop this structural gap being misread as evidence of weak practice. Because the mix of document types in the corpus is itself a product of the search method (see below), aggregate figures should always be read alongside the document-type breakdown in Table 10 rather than as a single undifferentiated total.

Search coverage and the grey literature constraint. Candidate literature was identified primarily through established academic databases (Scopus and Web of Science). These index peer-reviewed research well but capture grey literature poorly — and project and programme evaluations, which are the single richest source of exactly the funding-flow and capacity-transfer evidence the framework looks for, are overwhelmingly grey literature. Because no consolidated, searchable database of adaptation project evaluations exists across funders, this body of evidence cannot be retrieved systematically even though it is known to be substantial. The corpus is therefore weighted towards peer-reviewed research and is relatively thin in project evaluations. This is a structural constraint on any systematic review in this field rather than a flaw specific to this study, but it does mean the evidence base assessed here is narrower than the full universe of LLA evidence, and that low scores on evaluation-dependent categories partly reflect what the search could reach.

Threshold sensitivity. The 1.5 cut-off between low and high evidence is a deliberate analytical convention rather than a natural boundary. A number of papers sit close to this line on one or both dimensions, so a small shift in the threshold would move some between classifications; the broad patterns reported in Section 4 are not sensitive to precisely where the line is drawn.

The use of AI. AI-assisted scoring carries demonstrated scope for systematic bias: the validation history in Section 3.3 is not a hypothetical caveat but a documented account of a real leniency bias found and corrected. The residual risk was actively managed: a text-anchored justification was required for every score, scoring was cross-validated against manual human scoring across successive rounds and final interpretive authority rested with the research team throughout — but it is not claimed to be eliminated. Future users of this framework should budget for further validation as a normal part of applying it, not a one-off exercise.

4

Results and analysis

The framework was applied to the full, verified working corpus of 106 unique papers, spanning 2005 to 2026 and covering a broad geographic and sectoral sweep. This section reports the main results, then turns to a deeper ‘state of play’ examination of the Robust subset, before drawing out the implications for both the evidence base and the framework itself.

4.1 Composition of the corpus

Because document type strongly shapes which of the framework’s evidence categories can be documented at all, every paper was first tagged with a document-type category, allowing the analysis to be stratified. The composition that emerged is one of the most consequential findings of the exercise, but it must be read in light of the search method (see Section 3.1). Project evaluations, which are the document type most able to evidence funding flows and capacity transfer, make up 39% of the corpus. This share reflects the limited indexing of grey literature in academic databases at least as much as the actual balance of activity in the field.

Table 8. Corpus composition by document type.

Framework fit	Number of papers	Share	Typical content
Case study	49	46%	Site-level or multi-site ethnographic/empirical studies of adaptation practice
Project evaluation	41	39%	Evaluations of named, funded adaptation projects — the document type richest in the verifiable evidence the framework assesses
Policy analysis	10	9%	National/sub-national policy or institutional analyses with a local lens
Conceptual framing	6	6%	Reviews, scoping papers, frameworks and multi-country syntheses
Total	106	100%	

Project evaluations make up a substantial share of this corpus (39%), reflecting a corpus deliberately assembled to prioritise papers that focus on an identifiable, evaluable adaptation effort. Even so, the majority of the corpus (61%) consists of case studies, policy analyses and conceptual framing, for which many framework dimensions — particularly funding flows and post-project sustainability — may be structurally not addressable. This is precisely the pattern the framework’s Not Applicable, Not Classifiable and Excluded outcomes exist to handle correctly.

4.2 Overall classification

Applied across the full corpus, the framework produced the distribution outlined below, according to the classification system described in Table 9.

Just over one fifth of the corpus (22.3% of scoreable papers) demonstrates Robust LLA — clear evidence of high-quality local leadership coupled with verifiable, equitable outcomes. No Evidence of LLA, at 46.6% of scoreable papers, is the single largest category. Read at face value, this looks stark, but document type explains a large share of this figure, as Table 10 shows.

Table 9. Final classification across all 106 papers

Classification	Papers	Share	Mean Process/Impact
Robust LLA	23	21.7%	22.3%
Strong Process LLA	18	17.0%	17.5%
Top-Down Impact	2	1.9%	1.9%
No Evidence of LLA	48	45.3%	46.6%
Not Classifiable	12	11.3%	11.7%

Table 10. Classification by document type (out of 103 scoreable papers, excluding the three Excluded papers)

Framework fit	Robust LLA	Strong Process	Top Down	No Evidence	Not Classifiable	Total
Project evaluation	20 (49%)	14 (34%)	1 (2%)	6 (15%)	0 (0%)	41
Case study	3 (7%)	2 (4%)	0 (0%)	31 (67%)	10 (22%)	46
Policy analysis	0 (0%)	2 (20%)	0 (0%)	6 (60%)	2 (20%)	10
Conceptual framing	0 (0%)	0 (0%)	1 (17%)	5 (83%)	0 (0%)	6

Among project evaluations specifically, 49% are Robust LLA and only 15% are No Evidence of LLA — close to a mirror image of the aggregate pattern. Mean Process/Impact scores confirm the same pattern: project evaluations average 1.85/1.39, against 1.07/0.72 for case studies, 1.05/0.65 for policy analyses and 0.94/1.06 for conceptual framing. The outcome Not Classifiable is concentrated almost entirely in case studies and policy analyses (12 of 12 instances) — exactly the document types most likely to study a phenomenon (gender roles, Indigenous knowledge, adaptive capacity) without evaluating an implemented project. Document-type stratification is essential to reading the figures usefully.

4.3 Evidence strength across categories

Disaggregating scores across the six evidence categories reveals a consistent and instructive pattern. Some dimensions of LLA are well-documented across the literature; others are systematically under-evidenced, regardless of region or document type.

Table 11. Mean scores by evidence category (out of 103 scoreable papers, organised by strength of documentation)

Evidence category	Mean (0–3)	Interpretation
1. Context and Risk	1.49	Well-documented
2. Design and Decision Making	1.56	Strongest dimension
3. Funding Attributes	1.13	Mixed
4. Actions and Outcomes	0.96	Weakest dimension
5. Capacity and Sustainability	1.09	Under-documented
6. Learning and Transparency	1.01	Under-documented

Design and Decision Making and Context and Risk are comparatively well-evidenced across the corpus, reflecting how often local knowledge and participatory processes are documented in some form. Actions and Outcomes is the weakest category by a clear margin. This reflects the literature itself: tangible, equitably distributed, community-led outcomes are rarely documented in the corpus, leaving little evidence for the framework to register. Capacity and Sustainability and Learning and Transparency are the second- and third-weakest, consistent with ex-post evaluation and institutional follow-through being rarely documented in the published record. Funding Attributes sits in the middle of the pack, in part because the framework's conditional funding question (scored N/A rather than 0 for projects with no external funding component at all — see Annex 1.6) avoids penalising fully self-funded or locally funded projects for a question that simply does not apply to them. Future refinement of this framework should prioritise clearer evidentiary guidance for Actions and Outcomes, the category most in need of it.

Within the Robust LLA subset specifically (23 papers), all six categories score well above the 1.5 threshold (range 1.75–2.33), confirming that papers meeting the Robust LLA bar do so by scoring consistently across categories, rather than compensating for one weak dimension with an exceptionally strong one.

4.4 What robust LLA looks like in the literature

Among the 23 Robust LLA papers, the attributes that earn a Robust classification are, by construction, present: local or Indigenous knowledge driving risk framing and design, documented decision-making structures with community approval or veto power, and, where funding is described, resources channelled directly to local entities. These are what a high Process score requires, so their recurrence in this subset is definitional rather than an independent finding; they are reported here to show what the classification looks like in concrete terms, not as a discovery that robust papers happen to share them.

Among these robust papers, Indigenous or community knowledge shapes the project from the outset: it determines which risks the project treats as important and what the response should look like. This contrasts with the more common pattern, in which an external organisation has already defined the problem and the intervention, and consults communities only to refine a design that is largely settled. Research products are frequently co-authored or community-led. This points to local actors holding roles in knowledge production — authorship, framing, analysis — and not only in the adaptation projects being studied. Furthermore, decision-making structures are documented, often including community approval or veto power exercised through a specific, quoted instance rather than only a formally named protocol. Where funding is described, it channels resources directly to local entities. Overall, within the scored dimensions, what distinguishes the strongest papers is less the presence of any single attribute than the way evidence is corroborated across sources: quantitative data, direct community testimony and a specific, well-documented instance of a plan or programme changing in response to community input tend to appear together rather than singly.

Table 12. The five highest-scoring Robust LLA papers, ranked by combined Process + Impact score

Paper	Country	Process	Impact
Indigenous stewardship and co-management of a Great Barrier Reef mangrove ecosystem	Australia	2.47	2.46
Indigenous seed conservation, Andhra Pradesh	India	2.36	2.41
Community-built bridge, East Kwaio	Solomon Islands	2.69	2.06
Inuit-led integrated environment and health monitoring	Canada	2.50	2.16
Transformative Climate Communities programme	United States	2.56	2.04

The Great Barrier Reef case involves Yirrganydji Traditional Custodians directly co-authoring the underlying paper and co-managing the site with a corporate partner, backed by Australia's first biodiversity sustainability loan. The Andhra Pradesh case documents a community advocacy campaign that won a policy shift from Joint Forest Management to Community Forest Management. The East Kwaio bridge was entirely self-funded by the community, with an explicit, quoted account of the community declining external funding offers. These five cases span three continents and a range of funding models — from fully self-financed (Solomon Islands) to substantially state- or corporate-co-financed (Australia, United States) — suggesting Robust LLA is achievable under more than one funding arrangement, and does not require a single model.

4.5 Interpreting the No Evidence of LLA and Not Classifiable clusters

The 48 papers classified No Evidence of LLA, and the 12 classified Not Classifiable, require careful and distinct interpretation — the label 'No Evidence of LLA' conceals more than one underlying story.

- **Genuine No Evidence of LLA:** interventions that were top-down by design and produced limited, inequitable or unsustainable outcomes. Notably, many of these papers report those shortcomings candidly themselves. The framework is doing exactly what it should here: correctly identifying misalignment with LLA principles rather than papering over a critical or self-reflective source.
- **Document-type artefacts, correctly redirected to Not Classifiable rather than No Evidence of LLA:** papers for which the framework structurally cannot evidence certain categories — as Section 4.2 shows, 22% of case studies and 20% of policy analyses land in Not Classifiable rather than being forced into a low No Evidence of LLA score.
- **Borderline cases:** papers whose Process or Impact averages fall just below the 1.5 cut-off, where a small variation in evidence richness would tip them into a higher category. Using a band of ± 0.1 around the threshold (in other words, a score between 1.40 and 1.60) as a reasonably conservative definition of 'borderline', 20 of the 106 papers (19%) have at least one of their two averages in this range: 11 on Process, 10 on Impact (some papers are borderline on both). This count is sensitive to how 'borderline' is defined: widening the band to ± 0.2 roughly doubles it, to 32 papers (30%). No single band is inherently correct, so this range is reported to make that sensitivity visible rather than to imply false precision. These are not separately enumerated here but are a normal feature of any threshold-based classification and are discussed in Section 3.4. Table 5 in Section 3.3 quantifies how much the choice of threshold itself matters: Robust LLA ranges from 32 papers (30.2%) at a threshold of 1.3 to 15 papers (14.2%) at 1.7, with the current 1.5 threshold sitting at the midpoint of this range by design.

This clean separation of 'genuinely weak practice' from 'wrong instrument for this source' is one of the framework's more consequential design features, and it means the 45.3% aggregate No Evidence of LLA figure can be read as a genuine reading of weak LLA practice, rather than one inflated by sources the framework was never designed to evaluate.

4.6 The state of play: where robust evidence is strongest and where the gaps lie

This section interrogates the geographic and sectoral distribution of the 106-paper corpus. Region, country and sector were assigned to every paper to allow this analysis. The aim is not simply to count where robust evidence exists, but to read what its distribution reveals: which regions and sectors are well-evidenced, where the gaps lie and what geographic, thematic and methodological patterns the literature carries. The five subsections below build on one another: 4.6.1 maps Robust LLA share by region; 4.6.2 overlays SIDS and LDCs to test whether these structurally prioritised contexts show a correspondingly strong evidence base; 4.6.3 turns from geography to sector; 4.6.4 combines the two, cross-tabulating where robust evidence actually pools by region and sector together, with the SIDS/LDC overlay repeated as an additional lens; and 4.6.5 asks a different question — not where robust evidence sits, but which of the six evidence categories are best and worst documented within the Robust LLA subset itself.

4.6.1 Geographic distribution

Table 13. Robust LLA share by region (out of 103 scoreable papers)

Region	Number of papers	Number with Robust LLA	Robust LLA share	Mean Process/ Impact
South Asia	22	1	5%	1.12/0.67
Pacific and Oceania	20	6	30%	1.52/1.14
Sub-Saharan Africa	20	6	30%	1.51/1.24
Southeast Asia	16	3	19%	1.22/0.81
North America	10	3	30%	1.56/1.27
Latin America	5	2	40%	1.56/1.30
Caribbean	4	1	25%	1.45/1.09
Europe	2	0	0%	1.32/0.54
Global/Multi-country	2	0	0%	0.79/0.78
East Asia	1	0	0%	0.81/0.58
Arctic	1	1	100%	2.50/2.16

South Asia is the single largest region in the corpus (22 papers) but has the lowest Robust LLA share of any region with five or more papers, at just 5%. This is driven heavily by a large cluster of Bangladesh case studies and diagnostic research papers rather than project evaluations — a document-type effect layered on top of the geographic one, and a reminder that regional patterns in this corpus cannot be read independently of Table 9. Latin America (40%), North America, sub-Saharan Africa, and Pacific and Oceania (each 30%) form the strongest tier, though several of these regions have small sample sizes, and it is important to avoid over-generalising from this limited set. The single Arctic paper (Inuit-led monitoring, Canada) is Robust LLA, but one paper cannot establish a regional pattern on its own.

4.6.2 Small Island Developing States and Least Developed Countries

The standard regional breakdown obscures a cross-cutting question that is very relevant for policy: how well does the robust evidence base cover the countries that are simultaneously the most climate-vulnerable and the most capacity-constrained? Two internationally recognised groupings, SIDS and LDCs, capture this. Treated as overlay tags applied on top of the regional classification, they allow the analysis to ask whether the structural attention these groupings receive in climate finance is matched by a robust evidence base on LLA.

Table 14. SIDS/LDC overlay (out of 103 scoreable papers)

Group	Number of papers	Number with Robust LLA	Robust LLA share	Mean Process/ Impact
SIDS and/or LDCs	36	7	19%	1.36/0.90
Neither	67	16	24%	1.42/1.08

Papers from the Solomon Islands and Vanuatu⁴ — two countries that are both SIDS and LDCs — might be expected to score especially strongly, given the structural attention these contexts receive in climate finance discussions. However, much of the adaptation finance reaching these countries flows through multilateral and bilateral mechanisms that are frequently designed and delivered in a top-down way, with donor conditionalities that constrain local control over design and funding, resulting in a pattern that would predict weaker Process scores, not stronger ones. The data here cannot adjudicate between these two readings (see Section 3.9 for further discussion); it shows only that the ten Solomon Islands and Vanuatu papers form a mixed pattern rather than a uniformly strong one — four Robust LLA, three Strong Process LLA and three No Evidence of LLA — and that SIDS/LDC papers as a combined group score marginally lower on average than the rest of the corpus on both Process and Impact.

⁴ Vanuatu graduated from LDC status in December 2020. The three Vanuatu papers in this corpus — Clarke et al. (2019), Carodenuto et al., and Buggy and McNamara — were either published before graduation or drew on fieldwork conducted while Vanuatu was still an LDC, so are treated here as LDC-context papers

This does not mean SIDS/LDC contexts produce weaker LLA practice in reality; strong individual papers (such as the entirely self-funded community-built bridge in East Kwaio, Solomon Islands) sit alongside weaker ones within the same small cluster, and the sample size within this overlay group (36 papers) is modest. This finding should be treated as provisional, pending a larger, document-type-matched sample in any future application of the framework.

4.6.3 Sectoral distribution

Table 15. Robust LLA share by sector (out of 103 scoreable papers)

Sector	Number of papers	Number with Robust LLA	Robust LLA share
Ecosystems and natural resource management	17	6	35%
Infrastructure and urban	15	5	33%
Health	4	1	25%
Water	24	5	21%
Food, agriculture and livelihoods	14	2	14%
Multi-sector	29	4	14%

Ecosystems and natural resource management (NRM), and infrastructure and urban are the strongest sectors, each with roughly a third of their papers meeting the Robust LLA threshold. Food, agriculture and livelihoods and the broad multi-sector bucket (papers not cleanly assignable to a single sector) are comparably weak. Health remains a small sample (4 papers) across the whole corpus, and the low absolute count means this sector's share should not be over-read, though the scarcity of health-focused LLA evidence itself is a genuine and consequential finding: climate-and-health LLA is the most under-evidenced sector in the corpus.

4.6.4 Geography by sector

Combining the regional and sectoral classifications shows that robust evidence is not just unevenly distributed but unevenly distributed in patterned ways. Some regions concentrate in particular sectors; others are more balanced. The matrix below maps where Robust LLA evidence specifically is pooling, with the SIDS and LDC rows added as an overlay that re-counts papers already included in the regional rows above, rather than as additional, separate papers.

Table 16. Number of papers per sector for each region, within the Robust LLA subset (23 papers)

Region	Ecosystems	Food, agriculture and livelihoods	Health	Infrastructure and urban	Multi-sector	Water	Total
Pacific and Oceania	1	0	0	1	3	1	6
Sub-Saharan Africa	2	1	0	0	1	2	6
North America	1	0	0	1	0	1	3
Southeast Asia	1	1	0	1	0	0	3
Latin America	0	0	0	1	0	1	2
South Asia	1	0	0	0	0	0	1
Caribbean	0	0	0	1	0	0	1
Arctic	0	0	1	0	0	0	1
SIDS (overlay)	1	0	0	2	3	1	7
LDC (overlay)	1	1	0	1	3	2	8

Sub-Saharan Africa and Pacific and Oceania each contribute six of the corpus's 23 Robust LLA papers, but in different sectoral profiles: sub-Saharan Africa concentrates in ecosystems/NRM and water, while Pacific and Oceania is unusually multi-sector heavy, reflecting several broad synthesis and comparative papers from that region. South Asia contributes only a single Robust LLA paper despite being the largest region in the corpus overall (see Table 12), and that paper falls in ecosystems/NRM rather than in the water or multi-sector categories that dominate the region's broader literature. Health appears in only one cell in the entire matrix — the single Arctic paper — reinforcing the near-total absence of health-focused Robust LLA evidence noted in Section 4.6.3. The SIDS and LDC overlay rows show that 7 and 8 of the corpus's 23 Robust LLA papers respectively touch these groupings, with multi-sector, infrastructure and water the most common sectors in both — but as Section 4.6.2 shows, these papers sit within a wider SIDS/LDC-tagged corpus of 36 papers whose average Process and Impact scores are not higher than the rest of the corpus, so their sectoral concentration here should not be read as evidence that SIDS/LDC status itself predicts strong practice.

4.6.5 Evidence-category strengths and weaknesses by cluster

A further layer cross-references the geographic picture with the six evidence categories, asking not just where robust evidence pools but which dimensions of LLA are best and worst evidenced within the Robust LLA subset specifically. Even within this high-scoring subset, category scores cluster well below the maximum of 3.0, and some regional sample sizes here are small (as few as one to three papers in several regions), so the patterns below should be read as indicative rather than definitive.

Table 17. Mean evidence-category scores by region, within the Robust LLA subset (23 papers)

Region	Number of papers	Context	Design	Funding	Outcomes	Capacity	Learning
Sub-Saharan Africa	6	1.78	1.96	1.89	1.73	1.92	1.61
Pacific and Oceania	6	2.00	2.46	1.94	1.93	2.08	2.00
North America	3	1.78	2.58	2.22	2.07	2.17	1.44
Southeast Asia	3	1.67	2.33	2.33	1.87	2.50	1.78
Latin America	2	2.00	2.12	1.50	1.80	1.50	1.83
Caribbean	1	2.00	2.75	2.33	2.20	2.50	1.33
Arctic	1	2.00	2.50	3.00	1.80	3.00	1.67
South Asia	1	2.33	2.75	2.00	2.40	2.50	2.33

Two patterns recur across regions with reasonably sized samples (sub-Saharan Africa and Pacific and Oceania, six papers each; North America and Southeast Asia, three papers each). First, Design and Decision Making is the strongest category in all four of these regions, consistent with the overall corpus finding (see Section 4.3) that this is the framework's strongest dimension generally — Robust LLA papers earn their classification substantially on the strength of well-documented community decision-making authority. Second, Learning and Transparency is the weakest category in both sub-Saharan Africa (1.61) and North America (1.44), the two regions where it can be assessed on a reasonable sample, suggesting that even among the strongest papers in the corpus, evidence of iterative feedback loops and multi-directional learning remains comparatively thin. Funding Attributes is not the weakest category in any of the four better-sampled regions — a marked difference from what might be expected, and consistent with the corpus-wide finding (see Section 4.3) that the framework's revised, conditional funding question has removed funding from its earlier position as the most binding evidentiary constraint.

Reading two of the larger regional clusters in more detail brings some texture to this picture, though the smaller clusters (Latin America, Caribbean, Arctic, South Asia, each with one or two papers) are not large enough to support this kind of detailed reading and are reported in Table 16 for completeness only. In sub-Saharan Africa, the six Robust LLA papers span water, ecosystems/NRM, and food, agriculture and livelihoods; Design and Decision Making and Funding Attributes are comparably strong (1.96 and 1.89), while Learning and Transparency and Actions and Outcomes are both comparatively weak (1.61 and 1.73) — an indication that even the region's strongest papers document community authority and resource access more richly than they document iterative adaptation or measurable outcomes. In the Pacific and Oceania, where the six Robust LLA papers are unusually multi-sector (see Table 15), Design and Decision Making is the standout strength (2.46),

and Actions and Outcomes is comparatively the weakest category (1.93) — though notably higher than the sub-Saharan Africa figure for the same category, suggesting the region's strength lies more in translating decision-making authority into documented action than several other regions manage.

Health-related Robust LLA evidence is limited to a single paper in the entire corpus (the Arctic-region Inuit-led monitoring case, Table 11), which scores at or near the maximum on Funding Attributes and Capacity and Sustainability (both 3.00) but more moderately on Learning and Transparency (1.67). A single paper cannot establish a regional or sectoral pattern. But its existence at all is notable, given how close to absent health-focused Robust LLA evidence is everywhere else in the corpus (see Section 4.6.3). This near-total absence of robust health-LLA work across sub-Saharan Africa, South Asia, the Pacific and Southeast Asia is, on this evidence, the most consequential thematic gap in the entire corpus.

4.7 Evidence gaps and implications for future research

Synthesising across the geographic, sectoral and category-level findings, four major evidence gaps emerge, and together they define a clear agenda for future LLA research and evidence generation.

- **Geographic:** The Middle East, North Africa and Central Asia produced zero papers of any classification — not just zero Robust ones — despite severe and growing climate risks and substantial donor programming in parts of this region. This is the single largest geographic gap in the corpus, and it is a complete absence rather than a matter of degree. Europe (2 papers), East Asia (1 paper) and the broader 'Global/multi-country' bucket (2 papers) are each too thinly covered to generalise from. The corpus is also heavily concentrated in a small number of countries — Bangladesh (11 papers), Indonesia (9 papers) and the Melanesian cluster (Solomon Islands/Vanuatu/Fiji) (14 papers combined) together account for over a third of the corpus — while francophone, lusophone, Arabic and Spanish literatures are structurally excluded by the English-language search. Repeating this search in additional languages, and targeted commissioning in partnership with regional institutions in the Middle East, North Africa, Central Asia and francophone/lusophone Africa, would likely close a genuine blind spot rather than simply adding volume to already-covered regions.
- **Sectoral:** Health is the most under-represented sector by a wide margin (four papers, of which only one is Robust LLA) — a clear priority given the rising profile of climate–health concerns internationally. The weakness is not confined to thinly covered sectors, however. Food, agriculture and livelihoods is well represented in the corpus, but only 11% of its papers reach Robust LLA, so the problem is not simply that some sectors have too few papers: even where coverage is good, documented LLA practice is often weak.
- **Evidence category:** Actions and Outcomes is the framework's weakest category overall (with a mean of 0.96), meaning the literature finds it hardest to evidence tangible, equitably distributed, community-led outcomes specifically — a different priority than funding transparency, which is no longer the binding constraint it once appeared to be once the framework correctly treats fully self-funded or locally funded projects as N/A (rather than penalising them). Capacity and Sustainability is consistently the second-weakest category, and ex-post evaluation of what happens to local institutions after projects end is rarely documented; patient, multi-year follow-up studies are needed to close this gap, not simply more project evaluations at the point of completion.
- **Methodological biases to acknowledge:** Indigenous-led framing is somewhat over-represented within the Robust LLA subset (3 of 23 papers are explicitly Indigenous-led in their title and framing) relative to its likely share of LLA practice worldwide — a strength of this literature, but one that may mean LLA practice outside Indigenous-led traditions is systematically under-recognised as Robust simply because it is documented differently. The same three-country cluster noted above (Bangladesh, Indonesia, the Solomon Islands/Vanuatu/Fiji) skews the regional picture in ways that make it hard to distinguish a genuine regional pattern from a publication-volume effect. And 14.2% of the corpus (15 of 106 papers) is Not Classifiable or Excluded, meaning a meaningful share of the 'locally led adaptation' literature is not, in substance, evaluating an implemented project — a pattern future literature searches should anticipate and scope for, rather than treat as a data-quality problem to be filtered out after the fact.

Combining these, three priority areas for future evidence generation stand out: (i) LLA research in the Middle East, North Africa, Central Asia, and in francophone/lusophone Africa, ideally after re-running this search in additional languages; (ii) health-focused LLA evidence in the global South specifically, given how close to absent it currently is; and (iii) ex-post legacy and capacity research — the most consistently under-studied dimension of Impact, and the one most directly tied to whether LLA delivers durable, sustainable change, rather than merely an initial improvement.

4.8 Discussion: key findings

Read as a whole, the application supports several clear conclusions. LLA-aligned practice is real and visible in the literature, concentrated heavily in project evaluations (of which 49% are Robust LLA) relative to other document types (of which 0–7% are Robust LLA). That evidence is geographically uneven: Latin America, North America, sub-Saharan Africa and Pacific and Oceania form the strongest regional tier, while South Asia — the largest region in the corpus — has the lowest Robust LLA share, driven by document-type composition rather than an absence of strong practice. The SIDS/LDC finding — that papers from these contexts score no more strongly than the rest of the corpus, and marginally lower on average — is open to more than one interpretation and should be treated as a question for further research rather than a settled result.

Sectorally, health is a clear and consistent gap. On evidence categories, Design and Decision making and Context and Risk are strongest, and Actions and Outcomes is weakest. Finally, the framework's Not Classifiable and Excluded outcomes surface an important finding: a meaningful share of the LLA-adjacent literature is not evaluating an implemented project at all, and forcing it into a classification would misrepresent both those sources and the true failure rate among sources that were properly and fully assessed.

4.9 Implications for the framework

- Document-type stratification should remain a permanent, default feature of reporting: the gap between project evaluations (49% Robust LLA) and every other document type (0–7%) is stark, and reporting only aggregate totals would badly misrepresent the evidence.
- The N/A, Not Classifiable and Excluded outcomes work as intended and should be retained in any future version: they place 15 papers into a 'this source doesn't fit' category rather than a forced, misleading classification.
- The revised, lower evidentiary bar for Actions and Outcomes and Learning and Transparency questions (crediting single well-documented instances rather than requiring formal ongoing systems) meaningfully changed several papers' classifications in the expected direction and should be retained since they more accurately credit real evidence without lowering the bar for genuine engagement.
- Two items remain open for future work: the SIDS/LDC finding should be re-tested with a larger, more carefully matched sample; and Actions and Outcomes, the framework's weakest category, would benefit from clearer scoring examples and guidance.

4.10 Implications of the limitations for the findings

Section 3.4 sets out the framework's limitations in general terms; this section considers how they bear on the specific results reported above.

The language restriction is the clearest structural factor behind the geographic pattern in Section 4.7: the near-absence of robust papers from the Middle East and North Africa, Central Asia and francophone Africa should be read as a gap in the searchable, English-language evidence rather than proof of absent practice.

Because the framework scores only what is written, some of the No Evidence of LLA classifications are likely artefacts of reporting norms rather than genuinely weak interventions — particularly where funding flows or post-project legacy went undocumented.

The reliance on academic databases shaped the corpus composition directly: project and programme evaluations, the document types most likely to carry funding-flow and capacity-transfer evidence, are under-represented, which both depresses scores on those categories across the corpus and narrows the base from which the Robust LLA subset could be drawn. The weakness of funding evidence should therefore be read partly as a search coverage effect, not solely as a feature of practice.

The Robust LLA subset is the finding most exposed to the 1.5 threshold. A non-trivial number of papers fall within roughly a tenth of a point of the cut-off; had these borderline papers been classified as high, the Robust count would rise, and some regional and sectoral shares would shift at the margin. The headline patterns, however, are not sensitive to where the line is drawn: the weakness of funding evidence, the health and (outside Indigenous contexts) cultural-heritage gaps, and the geographic concentration of evidence all recur across the full corpus and within the Robust subset alike. The precise composition of the Robust subset should therefore be read as indicative rather than exact.

Finally, the scale of reclassification between the initial AI pass and the validated re-scoring (see Section 3.3) is itself relevant to interpreting the results: the one-directional movement — papers falling from Robust LLA towards No Evidence of LLA, rarely the reverse — is what identifies the original leniency as a systematic bias rather than random noise, and is why results are reported only from the validated pass.

5

Conclusions and way forward

Applied to a verified working corpus of 106 papers, the LLA Evidence and Impact Framework produces an interpretable, internally consistent reading of the local-adaptation literature. Reaching this framework and this corpus required a substantial process of validation and revision, described in full in Section 3.3: identifying a genuine, one-directional leniency bias in an early version of the scoring approach, rewriting the scoring codebook twice in direct response to specific, evidenced findings from three rounds of independent human review and re-scoring every paper from scratch under the final rules.

That validation process is itself a central finding of this report, not merely a methods footnote: a framework of this kind cannot be validated once and trusted indefinitely. It requires iterative, adversarial testing against independent human judgement, and a mechanism — the N/A, Not Classifiable and Excluded outcomes — for the framework to recognise the limits of its own applicability rather than forcing every source into a classification.

This research is, to our knowledge, one of relatively few documented cases where AI-assisted scoring was applied at scale to a qualitative evidence synthesis task, checked rigorously against independent human judgement, found wanting and iteratively corrected until it converged with that judgement — with the full process, including the failures, reported openly rather than smoothed over. AI-assisted scoring delivered clear efficiency gains: it made it feasible to apply a detailed, 20-question framework consistently across a corpus this size, a task that would have taken a small research team many months of full manual scoring, and it produced text-anchored justifications for every score, creating an audit trail that manual scoring alone rarely generates. But the efficiency gain was not free. Achieving results that could be trusted required three separate rounds of independent human validation, two substantial codebook rewrites and a full re-scoring of the corpus from scratch after each revision; human oversight was not a one-off check at the end but an integral, recurring part of the method throughout. The specific failure modes that emerged — a systematic leniency bias, difficulty recognising when a question genuinely did not apply, under-crediting informal-but-concrete evidence, and applying the framework to sources it was not designed to assess — are plausibly common risks for AI-assisted qualitative coding generally, not unique to this framework, and it may be useful for other teams considering a similar approach to budget realistic time and expertise for validation, rather than treating it as a lightweight add-on to an otherwise automated process.

On the substance, the results show real, if unevenly distributed, evidence of LLA: robust practice is concentrated in project evaluations and in a handful of strong regional and sectoral pockets. Document type, not just geography, drives much of the variation in what falls into No Evidence of LLA.

The SIDS/LDC finding (see Section 4.6.2) deserves unpacking rather than a single label, since it supports more than one plausible reading. It can be read as counter-intuitive, given how much structural attention SIDS and LDC contexts receive in climate finance discussions, on the assumption that this attention should translate into stronger local ownership. But it can equally be read as unsurprising, even expected: a large share of adaptation finance reaching SIDS and LDCs flows through multilateral and bilateral mechanisms that are often designed and delivered in a top-down way, with donor conditionalities and limited local control over design and funding decisions — precisely the pattern that would be expected to score poorly on Process. On this second reading, the finding is less a surprising reversal than a data point consistent with a broader, well-documented critique of paternalism in how adaptation finance is

directed towards the most climate-vulnerable countries. The analysis in this report cannot adjudicate between these two readings; distinguishing them would require examining the specific financing and governance arrangements behind each SIDS/LDC paper individually, rather than the category-level scores reported here. Both readings share the conclusion that the SIDS/LDC pattern merits further, more granular research before it informs funding or policy decisions.

Three next steps follow:

1. The SIDS/LDC and South Asia findings in Section 4.6 could be re-tested against a larger, document-type-matched sample (if it exists) before either is used for policy advocacy
2. New research should be commissioned specifically in health-related LLA, the most consistently under-evidenced sector in the corpus, and in project-evaluation-style research in South Asia specifically, given that region's low Robust LLA share is document-type-driven rather than a genuine absence of strong practice, and
3. This framework should continue to be validated against independent human review as a matter of course in any future application, not as a one-off exercise.

Annex 1: The LLA Evidence and Impact Framework

This annex presents the operational version of the LLA Evidence and Impact Framework, reflecting the rounds of codebook revision described in full in Section 3.3. It is a practical tool for documenting, assessing and learning from LLA in a way that is clear, consistent and comparable across contexts. It is not a substitute for project- or programme-level monitoring and evaluation, nor a mechanism for primary data collection; it is a structured means of compiling and weighing the strength of evidence that adaptation is genuinely locally led and impactful.

The eight Principles for Locally Led Adaptation

- P1: Devolving decision making to the lowest appropriate level
- P2: Addressing structural inequalities faced by marginalised groups
- P3: Providing patient and predictable funding that is easier to access
- P4: Investing in local capabilities to leave an institutional legacy
- P5: Building robust understanding of climate risk and uncertainty
- P6: Flexible programming and learning
- P7: Ensuring downward transparency and accountability
- P8: Collaborative action and investment

Structure: two dimensions, six categories, twenty questions

The framework organises the eight principles into two complementary dimensions — Process (how adaptation is designed and implemented) and Impact (what changes as a result) — operationalised through six evidence categories and 20 scored questions (reduced from 23 in the framework's initial version, following the revisions described in Section 3.3).

Part 1 (Process) covers Context and Risk (3 questions), Design and Decision Making (4 questions), and Funding Attributes (3 questions).

Part 2 (Impact) covers Actions and Outcomes (5 questions), Capacity and Sustainability (2 questions), and Learning and Transparency (3 questions).

Reducing the number of questions within a category does not change that category's weight in the Process or Impact average; each category still contributes equally to its domain regardless of how many questions feed it. Evidence may take many forms — quantitative data, qualitative accounts and Indigenous, traditional or locally situated knowledge — provided it is credible and verifiable within its context.

How to use the framework: a five-step process

The framework is applied to a single source of evidence about one adaptation intervention, producing one Process average, one Impact average and one classification for that source. In a research or evidence synthesis setting, the source is an existing document: a project evaluation, case study or peer-reviewed paper (as throughout this study). In a practitioner setting the 'source' is the assessor's own project documentation, used for self- or peer-assessment. The five steps below are the same in both settings; only the origin of the documentation differs.

Step 1: Define scope and context. Establish the specific adaptation activity, the timeframe being evaluated, and the key stakeholders involved.

Step 2: Systematic evidence collection. Use the standardised tables (see Part 1 and Part 2) to collect verifiable information, linking every piece of evidence to at least one LLA principle.

Step 3: Apply the scoring matrix. Assign a 0–3 score, or mark Not Applicable, to each question based on the strength and verifiability of documentation, scoring Process and Impact separately, and recording a brief justification for each score in the tables in Part 1 and Part 2.

Step 4: Synthesis and interpretation. (i) For each category, average its scored questions, excluding any marked Not Applicable. (ii) Calculate the Process average as the mean of the three Process category averages (Categories 1–3), and the Impact average as the mean of the three Impact category averages (Categories 4–6) — each category counts equally regardless of how many questions it holds. (iii) Apply the 1.5 threshold: a domain average of 1.5 or above is 'high', below 1.5 is 'low'. (iv) Read the Process/Impact combination against the interpretation and action guide to obtain the classification. (v) Apply the Not Applicable gating rules from the scoring scale table: if two or more of the three categories in a domain are entirely Not Applicable, record that domain as 'Not Classifiable'; if this holds for both domains, recommend the source for exclusion.

Step 5: Record, and (where applicable) share. Record the completed assessment transparently, with the justification for every score, so that it can be reviewed, reused or challenged. Where the framework is used by those implementing or funding the intervention, also apply minimum standards for ethical, non-extractive practice and share findings back with the community in accessible formats to strengthen downward accountability. Where it is used to assess an existing document produced by others — as in a research or evidence synthesis setting — this step covers transparent recording of the assessment and, where possible, publishing it alongside the source; the 'share back' obligation falls on the original intervention, not the reviewer.

Part 1: context, design and funding (Process)

Category (principles)	Question	Score (0–3/N/A)	Evidence and justification
1. Context and Risk (P2, P5)	1a. Was local/Indigenous knowledge documented as a primary driver of the risk assessment?		
1. Context and Risk (P2, P5)	1b. Did vulnerability baseline assessments clearly inform design choices?		
1. Context and Risk (P2, P5)	1c. Were structural inequalities and marginalised groups mapped prior to implementation?		
Category 1 average			
2. Design and Decision Making (P1, P2)	2a. Did the community have access to, and influence over, decision making processes? (A specific, corroborated instance of a community declining or redirecting a proposal scores as highly as a formally named approval protocol.)		
2. Design and Decision Making (P1, P2)	2b. Did local stakeholders, especially marginalised groups, have approval or veto power?		
2. Design and Decision Making (P1, P2)	2c. Were local communities represented in core design teams (membership logs, roles)?		
2. Design and Decision Making (P1, P2)	2d. Was inclusive public participation documented (town halls, consultations)?		
Category 2 average			
3. Funding Attributes (P3)	3a. Did funds reach local entities directly?		
3. Funding Attributes (P3)	3b. Was funding aligned with community-defined priorities?		
3. Funding Attributes (P3)	3c. If the project was supported by external funding, is there evidence some funding was allocated as overhead or flexible funding to local organisations? (Score Not Applicable if the project has no external/international funding component at all.)		
Category 3 average			
Process average (mean of Category 1–3 averages)			

Part 2: Actions, capacity and learning (Impact)

Category (principles)	Question	Score (0–3/N/A)	Evidence and justification
4. Actions and Outcomes (P2, P6)	4a. Were outcomes achieved through leadership by local actors rather than external consultants?		
4. Actions and Outcomes (P2, P6)	4b. How robust is the evidence that outcomes were equitably distributed across socioeconomic groups?		
4. Actions and Outcomes (P2, P6)	4c. Is there evidence of mechanisms to enable changes during implementation, based on evolving community priorities? (A single well-evidenced instance of a plan changing in response to feedback is sufficient.)		
4. Actions and Outcomes (P2, P6)	4d. Is there evidence of roles, responsibilities and formal decision-making power for marginalised groups within the project?		
4. Actions and Outcomes (P2, P6)	4e. How clearly does the data demonstrate that final outcomes align with community priorities (including priorities that evolved during the project)?		
Category 4 average			
5. Capacity and Sustainability (P4)	5a. Is there evidence that capacity for analysis, planning and maintenance has been transferred to local organisations?		
5. Capacity and Sustainability (P4)	5b. Is there evidence of continued operation and functionality of activities beyond the original funding or implementation period?		
Category 5 average			
6. Learning and Transparency (P6, P7, P8)	6a. Is there evidence of existence and/or implementation of iterative feedback loops resulting in verified changes to plans or budgets? (Either the existence of a mechanism, or evidence it was used once, is sufficient.)		
6. Learning and Transparency (P6, P7, P8)	6b. Is there evidence of multi-directional learning, up and down the delivery chain?		
6. Learning and Transparency (P6, P7, P8)	6c. How robust is the documentation of the community's perception of results and finance/budget utilisation?		
Category 6 average			
Impact average (mean of Category 4–6 averages)			
Overall classification (from the interpretation and action guide)			

Scoring scale

Score	Level	Evidence threshold
0	No Evidence	No verifiable documentation; claims unsupported by records or traceable sources.
1	Limited Evidence	Some indication of LLA principles, but partial or descriptive; without triangulation.
2	Substantive Evidence	Clear, documented, verifiable evidence, including in-depth anecdotal testimony.
3	Robust/Triangulated Evidence	Multiple sources — records, observations, local testimonies — demonstrating consistent practice and clear outcomes.
N/A	Not Applicable	The question's underlying mechanism cannot structurally apply to this source — not merely unmentioned, but genuinely inapplicable.

Scoring reflects the strength of evidence, not a value judgement of the intervention. Each score should carry a brief justification referencing specific sources. Process categories (1–3) and Impact categories (4–6) are averaged separately before synthesis, and Not Applicable questions are excluded from their category average entirely, rather than being counted as a zero.

Use Not Applicable when a question's underlying mechanism genuinely cannot apply — for example, a funding-transparency question where a project is entirely self- or locally-funded, or an outcome question where no implementation phase has yet occurred — not merely because the source does not mention it. Score 0 only when the mechanism could have applied and the source had a genuine opportunity to document it, but does not. A simple test is to ask 'If this had happened, would the source likely have said so?' If yes, and it says nothing, score 0. If the answer is that the source had no occasion to address this at all, the correct code is N/A.

If two or more of the three categories feeding a domain (Process or Impact) are entirely N/A for a given source, do not compute an average for that domain from the single remaining category. Report that domain as 'Not Classifiable' instead. If this is true of both domains simultaneously, the source lacks a basis for scoring altogether; recommend it for exclusion from the working corpus rather than forcing a classification (see interpretation and action guide).

Interpretation and action guide

High = domain average ≥ 1.5 ; Low = domain average < 1.5 .

Scenario	Interpretation	Recommended next step
High Process/ High Impact	Robust LLA — the 'gold standard': strong local ownership with verifiable, equitable outcomes.	Scale and influence: use as a case study to shape policy and funding models elsewhere.
High Process/ Low Impact	Strong Process LLA — strong local leadership; outcomes not yet materialised (often a time lag).	Patience and monitoring: maintain patient funding (P3) to allow results to emerge.
Low process/ High impact	Top-Down Impact — results achieved through external drivers; weak local legitimacy and high post-donor risk.	Transfer ownership: pivot to investing in local capabilities (P4).
Low Process/ Low Impact	No Evidence of LLA — neither approach nor outcomes align with LLA principles as documented.	Fundamental redesign: restart the context and risk assessment using local/Indigenous knowledge as the primary driver.
Not Classifiable (one domain)	More than two of three categories are N/A on one side — insufficient basis to trust a number there, though the other side may still be genuine.	Report the classifiable side only; do not force a number on the Not Classifiable side. Common for research studying a phenomenon rather than an implemented project.
Excluded from corpus	More than two of three categories are N/A on BOTH sides simultaneously — no basis for scoring at all.	Recommend removal from the working corpus; list separately in any summary; exclude from the denominator of any classification-percentage calculation.

References

- Akampurira, P (2024) Examining 'Locally Led' Climate Action and Local Leadership: A Case Study of Local Representative Committees (WCCPCs) and their Multistakeholder Collaborations in Kenya's Decentralised Climate Governance. Wageningen University, Wageningen, The Netherlands. <https://edepot.wur.nl/672311>
- Beauchamp, E, Sainsbury, NC, Greene, S and Chaigneau, T (2021) Aligning resilience and wellbeing outcomes for locally-led adaptation in Tanzania, *Sustainability*, 13(16), 8976. doi:10.3390/su13168976.
- Bedelian, C, Mulwa, J, Sumari, BK, Rogers, P, Gravesen, ML and Funder, M (2024) Locally-led climate change adaptation works: Here are eight ways to support it. Danish Institute for International Studies, Copenhagen. <https://research.diis.dk/en/publications/locally-led-climate-change-adaptation-works-here-are-eight-ways-t/>
- Bennett, NJ, Kadfak, A and Dearden, P (2016) Community-based scenario planning: a process for vulnerability analysis and adaptation planning to social–ecological change in coastal communities, *Environment, Development and Sustainability*, 18(6), pp.1771–1799. doi:10.1007/s10668-015-9707-1.
- Bryan, E and Behrman, JA (2013) Community-based adaptation to climate change: A theoretical framework, overview of key issues and discussion of gender differentiated priorities and participation. CGIAR Systemwide Program on Collective Action and Property Rights (CAPRI), (109), p. 36. doi:10.2499/capriwp109.
- Buggy, L and McNamara, KE (2016) The need to reinterpret “community” for climate change adaptation: a case study of Pele Island, Vanuatu. *Climate and Development*, 8(3), pp.270–280. Doi:10.1080/17565529.2015.1041445.
- CARE International (2010) Community-Based Adaptation Framework. CARE International, Geneva.
- Carodenuto, S, Schwarz, B, Nelson, A, Bome, G and Andre, G (2022) Practice-based knowledge for REDD+ in Vanuatu. *Society & Natural Resources*, 35(2), pp.220–241. Doi:10.1080/08941920.2021.2011996.
- Clarke, T, McNamara, KE, Clissold, R and Nunn, PD (2019) Community-based adaptation to climate change: lessons from Tanna Island, Vanuatu. *Island Studies Journal*, 14(1), pp.59–80. Doi:10.24043/isj.80.
- Clissold, R and McNamara, KE (2019) Exploring local perspectives on the performance of a community-based adaptation project on Aniwa, Vanuatu, *Climate and Development*, 12(5), pp.457–468. doi:10.1080/17565529.2019.1640656.
- Fischer, F (2012) Participatory Governance: From Theory To Practice, in Levi-Faur, D (ed.) *The Oxford Handbook of Governance*. Oxford Academic, pp.457–471.
- Ford, JD, Sherman, M, Berrang-Ford, L, Llanos, A, Carcamo, C, Harper, S, Lwasa, S, Namanya, D, Marcello, T, Maillet, M, et al. (2018) Preparing for the health impacts of climate change in Indigenous communities: The role of community-based adaptation, *Global Environmental Change*, 49, pp.129–139. doi:10.1016/j.gloenvcha.2018.02.006.
- Fraser, N (2006) Reframing justice in a globalizing world, in: *Nationalism and global solidarities*. Routledge, pp.168–186. doi:10.4324/9780203085981.
- Gero, A, Méheux, K and Dominey-Howes, D (2011) Integrating community based disaster risk reduction and climate change adaptation: examples from the Pacific, *Natural Hazards and Earth System Sciences*, 11(1), pp.101–113. doi:10.5194/nhess-11-101-2011.
- Global Center on Adaptation (2023) 2023 Stories of Resilience: Lessons from Local Adaptation Practice. Rotterdam. <https://gca.org/reports/2023-stories-of-resilience-lessons-from-local-adaptation-practice/>
- Lukes, S (1974) *Power: a radical view*. Macmillan, London.
- Mannke, F (2011) Key themes of local adaptation to climate change: Results from mapping community-based initiatives in Africa. In: WL Filho (Ed.), *Experiences of climate change adaptation in Africa*. Springer, Springer Berlin, Heidelberg, pp. 17–32. doi: 10.1007/978-3-642-22315-0
- McNamara, KE (2013) Taking stock of community-based climate-change adaptation projects in the Pacific, *Asia Pacific Viewpoint*, 54(3), pp.398–405. doi:10.1111/apv.12033.

- McNamara, KE, Clissold, R, Westoby, R, Piggott-McKellar, AE, Kumar, R, Clarke, T, Namoumou, F, Areki, F, Joseph, E, Warrick, O, et al. (2020) An assessment of community-based adaptation initiatives in the Pacific Islands, *Nature Climate Change*, 10(7), pp.628–639. doi:10.1038/s41558-020-0813-1.
- Mills-Novoa, M (2023) What happens after climate change adaptation projects end: A community-based approach to ex-post assessment of adaptation projects. *Global Environmental Change*, 80, 102655. Doi: 10.1016/j.gloenvcha.2023.102655
- Omukuti, J (2020) Country ownership of adaptation: Stakeholder influence or government control? *Geoforum*, 113, pp.26–38. doi:10.1016/j.geoforum.2020.04.019.
- Rahman, MF, Falzon, D, Robinson, S, Kuhl, L, Westoby, R, Omukuti, J, Schipper, ELF, McNamara, KE, Resurrección, BP, Mfitumukiza, D, et al. (2023) Locally led adaptation: Promise, pitfalls, and possibilities, *Ambio*, 52(10), pp.1543–1557. doi:10.1007/s13280-023-01884-7.
- Rajabali, F, Landry, M-E, Karaho, E, Musiimenta, V and Chooma, B (2024) Intersecting realities: Four approaches to making locally-led adaptation more inclusive. CDKN.
- Rashid, AKMM, Khan, MR and Khan, MR (2013) Community Based Adaptation: Theory and Practice. In Shaw, R, Mallick, F and Islam, A (eds) *Climate Change Adaptation Actions in Bangladesh*. Springer Japan, Tokyo, pp.341–362. doi:10.1007/978-4-431-54249-0_18.
- Reich, Y (2025) A Framework for analyzing and supporting communities on their path to sustainability, *Sustainability*, 17, 7262. doi:10.3390/su17167262
- Roche, C, Cox, J, Rokotuibau, M, Tawake, P and Smith, Y (2020) The Characteristics of Locally Led Development in the Pacific, *Politics and Governance*, 8(4), pp.136–146. doi:10.17645/pag.v8i4.3551.
- Tye, S and Suarez, I (2021) Locally Led Climate Adaptation: What Is Needed to Accelerate Action and Support? World Resources Institute. doi:10.46830/wriwp.20.00039.
- Tye, S, Mukasa, CH, Kratzer, B, Businge, G, Wamalwa, RN, Atieno, R, O'Connor, R, Trivedi, A, Cogger, T and Elwell, N (2023) Strengthening Gender Equity in Locally Led Adaptation Processes in Africa. World Resources Institute. doi:10.46830/wriwp.21.00166.
- Westoby, R, McNamara, KE, Kumar, R and Nunn, PD (2019) From community-based to locally led adaptation: Evidence from Vanuatu, *Ambio*, 49(9), pp.1466–1473. doi:10.1007/s13280-019-01294-8.
- Westoby, R, Clissold, R, McNamara, KE, Ahmed, I, Resurrección, BP, Fernando, N and Huq, S (2021) Locally led adaptation: drivers for appropriate grassroots initiatives, *Local Environment*, 26(2), pp.313–319. doi:10.1080/13549839.2021.1884669.
- Yee, M, Piggott-McKellar, A, McMichael, C and McNamara, KE (2025) Framing locally led adaptation in a planned relocation in Fiji, *Geoforum*, 159, p.104196. doi:10.1016/j.geoforum.2024.104196.

Related reading

- GA-LLA programme page: Generating ambition for locally led adaptation. www.iied.org/generating-ambition-for-locally-led-adaptation
- Mitchell, P (2025) Scaling locally led adaptation: closing the rhetoric-reality gap. IIED, London. www.iied.org/22686iied
- Mitchell, P, Thazin Aung, M and Setaro, L. (2025) Lessons for designing locally led adaptation programmes. IIED, London. www.iied.org/22663iied
- Treichel, P, Joshi, R, Taylor, E, Ridout, M, Mitchell, P and Acuda, A (2025) Locally implemented or locally led? Tracking finance for community climate action. IIED, London. www.iied.org/22674iied
- Pichon, F, Joshi, R, Thazin Aung, M and Setaro, L (2025) Practical guidance for aligning operations with the LLA principles. IIED, London. www.iied.org/22667iied



Knowledge
Products

Research Report

October 2026

**Climate change;
Policy and planning**

Keywords:

Locally led adaptation, resilience, climate adaptation, evidence of impact

Locally led adaptation (LLA) is now widely endorsed in principle, yet the evidence base to show whether it works in practice has not been mapped systematically. This report presents the LLA Evidence and Impact Framework — a structured analytical lens organised around eight LLA principles, two dimensions and six evidence categories — applied, following extensive validation against independent human review, to a verified corpus of 106 academic and grey literature documents published between 2005 and 2026. Using a mixed manual and AI-assisted method, the study identifies where robust LLA evidence concentrates in the literature, where it is absent, and how the framework itself performs. This matters because endorsement of the eight LLA principles is not the same as evidence that they are being followed: without a shared, structured way to weigh the evidence, it is difficult to tell an intervention that is genuinely shaped and governed by local actors from one that merely consults them, or to distinguish a strong claim of success from a well-documented one.

IIED is an international policy and research organisation working with partners globally to build a fairer, more sustainable world. Together, we challenge the destructive economic models, unjust power dynamics, entrenched mindsets and protectionist laws that perpetuate poverty, suppress rights and hinder progress towards a thriving world. We explore solutions to complex economic, social and environmental crises, using research, action and influencing to tackle the root causes of climate change, nature loss and inequality.



International Institute for Environment and Development
44 Southampton Buildings, London WC2A 1AP, UK

Tel: +44 (0)20 3463 7399
www.iied.org

Funded by:



Ministry of Foreign Affairs of the
Netherlands