Beyond the cloud: Unlocking edge AI for LMICs – qualitative study

Term of Reference

1. Introduction

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Find more at gsma.com.

The GSMA Mobile for Development (M4D) foundation operates at the intersection of the mobile ecosystem and the development sector. Our aim is to stimulate digital innovation and deliver both sustainable business and large-scale socio-economic impact. Our research and insights platform, in-market expertise and community of partners push forward digital innovations and implementations that empower underserved populations. Find out more at gsma.com/solutionsand-impact/connectivity-for-good/mobile-for-development/.

The Central Insights Unit (CIU) sits at the core of GSMA Mobile for Development (M4D) and produces in-depth research on the role and impact of mobile and digital technologies in advancing sustainable and inclusive development. The CIU engages with public and private sector practitioners to generate unique insights and analysis on emerging innovations in technology for development. Through our insights, we support international donors to build expertise and capacity as they seek to implement digitisation initiatives in low- and middle-income countries through partnerships within the digital ecosystem

2. Background, scope and objectives

Background

The global AI ecosystem today is largely built on centralised, cloud-based infrastructure. Most LMICs face structural barriers to using that infrastructure at scale including costs, latency, foreign exchange access, data sovereignty and unstable power. Edge AI – running inference on or near where the data is generated – has become technically more viable due to rapidly improving smartphone chips, on-device microcontrollers, and model-efficiency techniques. In Africa in particular, mobile is the dominant digital channel, and smartphones projections are projected to keep rising, widening the base for on-device capability.

Despite the growing emergence of edge AI, there is still limited evidence of the potential and viability of edge AI in resource-constrained environments, where economic, technical and market conditions differ significantly from high-income contexts. Existing studies, including recent GSMA research, have examined the demands of model training, dominant compute pathways, and large-scale cloud infrastructure, but have not assessed the market readiness or operational feasibility of edge AI in these settings. At the same time, an increasing number of innovators are developing edge AI solutions to enable scalable AI applications. However, significant gaps remain in understanding the technology needs, economic feasibility, operational models, priority use cases, and policy frameworks required to scale edge AI sustainably in LMICs.

Scope

The GSMA is undertaking a mixed-methods study to address evidence gaps on the feasibility, economics, and enabling conditions for edge AI in LMICs. The research aims to quantify market readiness and unit economics, document real-world deployments, and generate actionable guidance for industry, investors, and policymakers. Findings will feed into a public GSMA report.

This ToR covers the qualitative workstream. The GSMA is seeking a supplier to generate grounded evidence on how edge AI is being designed, deployed, and sustained in LMICs, with a focus on real-world feasibility, adoption drivers, and the enabling conditions for scale. The work should surface practical trade-offs and operating models, and translate them into actionable recommendations for industry, policymakers, and funders. The qualitative analysis will complement the quantitative workstream by contextualising numerical estimates with real-world evidence.

For the purposes of this study, edge AI refers to the deployment of AI models on or near the device where data is generated. This includes smartphones (across capability tiers), embedded and IoT devices (e.g. sensors, single-board computers, microcontrollers), as well as optimisation techniques (e.g. TinyML, quantisation, model compression) that enable AI inference on constrained devices. The study will consider this full spectrum but place particular emphasis on smartphones and mobile devices, given their central role in LMIC digital ecosystems.

The primary audience for this project includes industry, i.e. startups, MNOs, device manufacturers, infrastructure providers, and sources of capital, i.e. VCs, DFIs, corporate investors. The secondary audience includes governments and policy makers, donor organisations and academia and training providers.

Objectives

The qualitative stream of the study aims to:

- Identify and analyse emerging edge AI deployments, focusing on how on-device or near-device inference addresses user needs in low-resource environments.
- Document trade-offs and benefits when shifting workloads from cloud to edge (latency, unreliable connectivity, test-time compute demands of reasoning models, cost efficiency, privacy).
- Assess the potential of distributed learning methods (e.g., federated learning) to leverage edge compute for training in ways that support local relevance and data sovereignty.
- Consider the social and developmental impacts of edge AI deployments, including their potential to accelerate AI adoption, support local innovation, and promote equitable technology access.
- Map policy and regulatory enablers/risks (data protection, security, pro-innovation regulation) and the skills/hardware–software integration gaps that constrain scale.
- Identify actionable recommendations to support sustainable and inclusive edge AI deployments in LMICs.

Note: While a full ecosystem analysis is out of scope, the supplier should flag key dependencies on global technology providers, local MNO strategies, and government choices that shape edge AI ecosystems, and highlight issues for potential follow-up research.

Research questions

It will seek to answer the following research questions:

Deployment and economics	- What use cases and design choices are succeeding and under what conditions (device class/tier, connectivity, power, data constraints) do they succeed or fail? - What do early deployments tell us about cost drives (capex/opex, maintenance, support) and design choices that improve viability? - What commercial and partnership models (e.g. device bundling, data-plan incentives, managed edge services) are emerging, and how sustainable do practitioners see them?
Adoption and impact	- How do users, startups, and operators perceive the value of edge AI compared to cloud-only models (e.g., cost, latency, reliability, privacy)? - What evidence or perceptions exist on inclusion impacts (e.g., affordability for women and rural users, handset financing, local-language UX)? -
Risks and enabling conditions	- What risks (privacy, security, vendor lock-in) are practitioners most concerned about, and how are they mitigated in practice?
Inclusion, safety and trust	- How do gender and other inclusion factors shape adoption/outcomes, and which practical interventions (handset financing, targeting, local-language UX) can help close gaps? - What policy/regulatory gaps or reforms do stakeholders see as critical for scaling (e.g. data protection, spectrum, procurement)? - What skills/training gaps are most often cited as barriers to sustainable deployment?

Geography

This study will focus primarily on Africa, given the high need for alternative compute approaches in the region and comparatively lower connectivity environments. The supplier should also look for relevant insights and case studies from across LMIC regions (e.g., South/Southeast Asia, Latin America). The qualitative component will develop case studies and country spotlights to document operational realities across a diverse set of LMIC contexts.

3. Anticipated approach and delivery timeline

The below is a proposed approach but we welcome further ideas and proposed methodology from the supplier.

Phase	Activities and deliverables	Timeline (indicative)
	Key activities	

Phase 1: Inception	- Kick off call to discuss methodology, scope, timelines and methodological approach - Review of literature on edge AI in LMICs (deployment patterns, device classes, skills and policy considerations)* - Stakeholder mapping and outreach plan for key-informant interviews - Case study selection framework (criteria to pick ~5 diverse edge AI use cases)	Mid-October
	Deliverables - Inception report/deck with literature review highlights, stakeholder map, and case study selection framework	
Phase 2: Data collection, analysis and interim findings	**Key activities** - 20-25 semi-structured interviews across stakeholder types - Draft at least 5 case studies of edge AI deployments (mix of sectors/device types/deployment models) - Build 3-5 country spotlights to contextualise the quantitative findings	November-January
	Deliverables - Interview tracker and detailed summary of interview notes - Interim report (word document) with emerging insights and gaps; draft case studies	
Phase 3: Synthesis, validation and final report	**Key activities** - Consolidating insights, analysis and report drafting	Early February
	Deliverables - Final report (word document) with insights, case studies and recommendations	

*Note: GSMA has initial internal material for the literature review, the supplier will be expected to review and incorporate, proposing improvements where needed.

4. Supplier requirements

Key requirements

The GSMA is searching for an individual consultant to deliver the qualitative workstream. Ideally they will have:

Essential:

- Strong track record in qualitative research in LMIC contexts (interviews, case studies, thematic synthesis).
- Subject-matter familiarity with AI/edge computing/IoT or related digital development fields.

- Experience embedding gender and inclusion lenses in analysis.
- Awareness of policy and ecosystem dynamics in LMICs.
- Strong stakeholder networks (local innovators, MNOs, policymakers, tech providers).

Desirable:

- Experience working with MNOs, device makers, or big tech; understanding of commercial models (bundling, data plans, edge services).
- Track record with donors/DFIs and policy audiences.
- Ability to produce outputs accessible to non-technical audiences (e.g., policymakers, funders)

Proposals should include a technical and financial proposal:

Technical proposal

1. A short (1 page) statement of suitability, highlighting recent relevant experience.
2. A short (2-4 page) discussion of the proposed approach including: the analytical frameworks to be used, identified data sources, and initial proposals on case studies.
3. Any proposed changes to the ToR.
4. Details of relevant firm project experience.
5. Gantt chart outlining major project stages and timelines
6. CVs, and location of team members.

Financial proposal

1. Level of effort (person-day) by activity.
2. Fee rates (per day in GBP).
3. Total project cost (GBP), without VAT[1].
4. The Respondent's Total Price is inclusive of all costs, insurances, fees, costs, expenses, liabilities, obligations, risks, and all financial requirements for the performance of Services and provision of Deliverables.
5. Any charge not stated in this Proposal, which extends above to the Total Price, is not permitted.

Due to GSMA compliance requirements, exact project budgets cannot be provided at this stage. You are, however, able to provide a few implementation/budget options that can help us assess value for money and we can align our project scope to the relevant budget after a consultant has been selected.

Proposal assessment and selection process

The proposal will be scored on the following set of criteria:

Criteria	**Importance**	**Weighting**
Cost	Proposal's value for money	20%
Quality	Quality of the research approach outlined in the proposal, including degree to which it addresses the outlined research questions and proposal elements	35%

Bidder's capacity to manage the project on time and on budget	Selection of experienced high-quality research partner(s) and ability to manage the project on time and on budget	30%
Relevant experience	Bidder's experience in successfully conducting similar projects	15%

- Proposals are to be submitted no later than **17:30 BST, Wednesday 24th September 2025** for this work to Daisy Macaskie (dmacaskie@gsma.com) and Eugénie Humeau (ehumeau@gsma.com).
- Clarification questions can be sent to Daisy Macaskie (dmacaskie@gsma.com) and Eugénie Humeau (ehumeau@gsma.com).
- Shortlisted consultants may be contacted for an interview **w/c 29th September**.