

Enhancing Participatory Development Research in South Asia through LLM Agents System: An Empirically-Grounded Methodological Initiative and Agenda from Field Evidence in Sri Lanka

Xinjie Zhao¹, Hao Wang³, Shyaman Maduranga Sriwarnasinghe¹, Jiacheng Tang⁴,
Shiyun Wang², Sayaka Sugiyama¹, So Morikawa¹

¹ The University of Tokyo, ² University of Copenhagen,

³ China Agricultural University, ⁴ Shandong Normal University

Abstract

The integration of artificial intelligence into development research methodologies offers unprecedented opportunities to address persistent challenges in participatory research, particularly in linguistically diverse regions like South Asia. Drawing on empirical implementation in Sri Lanka's Sinhala-speaking communities, this study presents a methodological framework designed to transform participatory development research in the multilingual context of Sri Lanka's flood-prone Nilwala River Basin. Moving beyond conventional translation and data collection tools, the proposed framework leverages a multi-agent system architecture to redefine how data collection, analysis, and community engagement are conducted in linguistically and culturally complex research settings. This structured, agent-based approach facilitates participatory research that is both scalable and adaptive, ensuring that community perspectives remain central to research outcomes. Field experiences underscore the immense potential of LLM-based systems in addressing long-standing issues in development research across resource-limited regions, delivering both quantitative efficiencies and qualitative improvements in inclusivity. At a broader methodological level, this research advocates for AI-driven participatory research tools that prioritize ethical considerations, cultural sensitivity, and operational efficiency. It highlights strategic pathways for deploying AI systems to reinforce community agency and equitable knowledge generation, offering insights that could inform broader research agendas across the Global South.

1 Introduction

The convergence of artificial intelligence and development research heralds a transformative paradigm shift in participatory methodologies, particularly through the emergence of Large Language Models (LLMs) and their potential to revolutionize community engagement practices (Mohamed et al.,

2024; Skirgård et al., 2023). As these technologies rapidly evolve, their application to development research presents both unprecedented opportunities and complex methodological challenges that demand careful examination (Roberts et al., 2024). This intersection becomes particularly significant in linguistically diverse regions like South Asia, where traditional research approaches have long struggled to bridge communication gaps and cultural divides (Kshetri, 2024; Hassan et al., 2023).

The limitations of conventional participatory research methodologies, heavily dependent on human intermediaries and constrained by resource availability, have historically impeded the scale and effectiveness of development initiatives (Göpferich and Jääskeläinen, 2009). These constraints are particularly evident in regions characterized by complex linguistic landscapes and limited technological infrastructure (Magueresse et al., 2020; Nekoto et al., 2020). However, recent advances in LLM architectures, particularly in few-shot learning and cross-lingual transfer capabilities, offer promising solutions to these longstanding challenges (Raiaan et al., 2024; Wu et al., 2023).

The integration of LLM-based systems into participatory research frameworks raises fundamental questions about the nature of community engagement and knowledge democratization (Hadi et al., 2024; Diab Idris et al., 2024). While these technologies offer powerful tools for bridging linguistic and cultural divides, their deployment must be carefully orchestrated to enhance rather than diminish the participatory nature of development research (Rane et al., 2023; Kovač et al., 2024). This necessitates a nuanced approach that balances technological capabilities with ethical considerations and community agency (Sabarirajan et al., 2024; Ray, 2023).

In this paper, we introduce and test a novel framework (Fig.1) for leveraging LLM-based multi-agent systems in participatory development research, drawing from empirical evidence in Sri

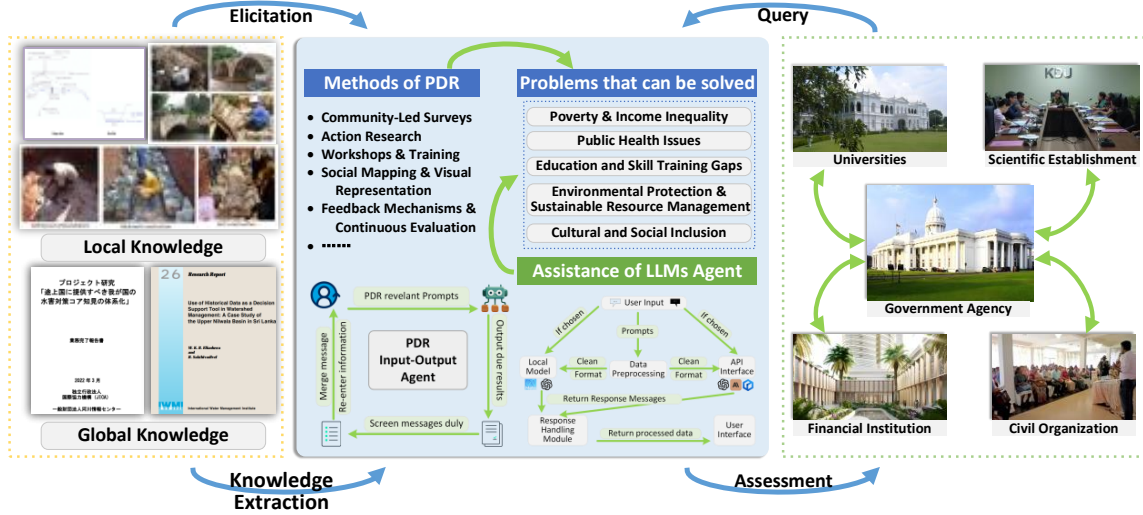


Figure 1: Proposed LLM4Participatory Research Framework

Lanka’s Sinhala-speaking communities (Hashmi et al., 2024; Urwin et al., 2023). Our approach moves beyond simple technological integration to address fundamental questions of community empowerment and knowledge production in Global South contexts (Pfeffer et al., 2013). The urgency of this work is underscored by the increasing complexity of development challenges and the growing need for scalable, culturally sensitive research methodologies (van Rensburg and van der Westhuizen, 2024; Awad et al., 2016). Through critical analysis of both opportunities and challenges, we demonstrate how thoughtfully deployed NLP technologies can enhance human capabilities in development research, potentially leading to more inclusive and impactful outcomes (Ferdaus et al., 2024). Our framework provides a structured approach for implementing LLM-based multi-agent systems while maintaining core principles of participatory research, offering insights for researchers, practitioners, and policymakers working at the intersection of technology and development. We argue that these technologies, when thoughtfully deployed, can enhance rather than replace human capabilities in development research, potentially leading to more inclusive, efficient, and impactful research outcomes.

2 Why South Asia Needs This Now

South Asia stands at a nexus where rapid digitalization meets deeply ingrained linguistic and cultural heterogeneity, presenting formidable challenges but also unparalleled opportunities for participatory research (Rahman, 2024). Growing smart-

phone penetration, expanding internet infrastructures, and the proliferation of digital platforms have catalyzed a democratization of information (Deichmann et al., 2016). Rural communities, previously marginalized due to limited access to communication channels, now experience annual digital literacy growth rates surpassing traditional benchmarks (Kass-Hanna et al., 2022). Despite these advances, the region’s linguistic complexity—home to over 650 languages—remains an enduring obstacle to effective data collection, community engagement, and knowledge co-creation (Hutson et al., 2024). The pervasive phenomenon of code-mixing, where speakers fluidly alternate between languages and dialects, further complicates meaning extraction and translation (Rodríguez Tembrás, 2024). Traditional research paradigms and even earlier-generation NLP tools struggle to handle these intricacies, leading to communication bottlenecks, inflated research costs, and a marginalization of essential local voices (Daramola et al., 2024; Björk Brämberg and Dahlberg, 2013).

Emerging LLMs and advanced NLP architectures, however, offer a pathway to transcend these limitations. State-of-the-art models, when fine-tuned and adapted through few-shot and transfer learning approaches, can now handle morphologically complex languages and capture semantic subtleties even under severe training data constraints (Tomec and Gričar, 2024; Parovic, 2024). These technological capabilities enable more equitable, scalable, and culturally sensitive research methods that respect local communication patterns and linguistic realities. Crucially, these tools do not

merely solve technical challenges; they reshape the participatory research paradigm. By facilitating real-time, multilingual engagement and generating culturally resonant research activities, LLM-based systems empower communities to more actively co-produce knowledge (Kar et al., 2024), while substantially cutting resource overheads. Beyond operational efficiency, this signifies a fundamental shift toward recognizing community agency, acknowledging indigenous knowledge systems, and enhancing the overall authenticity and credibility of development research (Brown, 2024; Dutta et al., 2024). This enhanced research environment supports more sustainable interventions. Researchers can allocate fewer resources to language mediation and more to iterative engagement cycles, iterative validation, and capacity building. The outcome is a more inclusive, trusting, and impactful participatory ecosystem, where community voices shape the research agenda, and shared knowledge guides more relevant and equitable development strategies (Ullah et al., 2024; Matras et al., 2023).

3 Proposed LLM4Participatory Research Framework

Our LLM4Participatory Research Framework is designed as a modular, multi-agent ecosystem that integrates LLMs with multimodal AI capabilities to address the multifaceted challenges of participatory research in linguistically diverse settings. Each agent type is specialized for a distinct function, yet all collaborate seamlessly through well-defined interaction protocols, shared data ontologies, and unified prompt engineering guidelines. Agent interactions are orchestrated using standardized APIs, message passing interfaces, and carefully designed prompt templates, ensuring that the cognitive pipeline remains flexible, transparent, and adaptable.

3.1 Core Components

Participatory Research Design and Analytics Agents (PRDAA): These agents guide the research methodology from inception to continuous refinement. They employ LLM-driven prompt engineering to generate and adapt research instruments (e.g., surveys, interview guides) informed by pre-curated cultural knowledge repositories. During pre-field pilot studies, PRDAAs utilize reinforcement learning-based feedback loops to iteratively refine instruments. For instance, the prompts used

may define constraints such as “Generate three culturally relevant survey questions about water usage practices, each incorporating code-mixed Sinhala-English phrases.” The agents then analyze responses and community feedback, calibrating lexical and thematic complexity to ensure that instruments are both linguistically accessible and culturally resonant (Rane et al., 2024; Agathos et al., 2024).

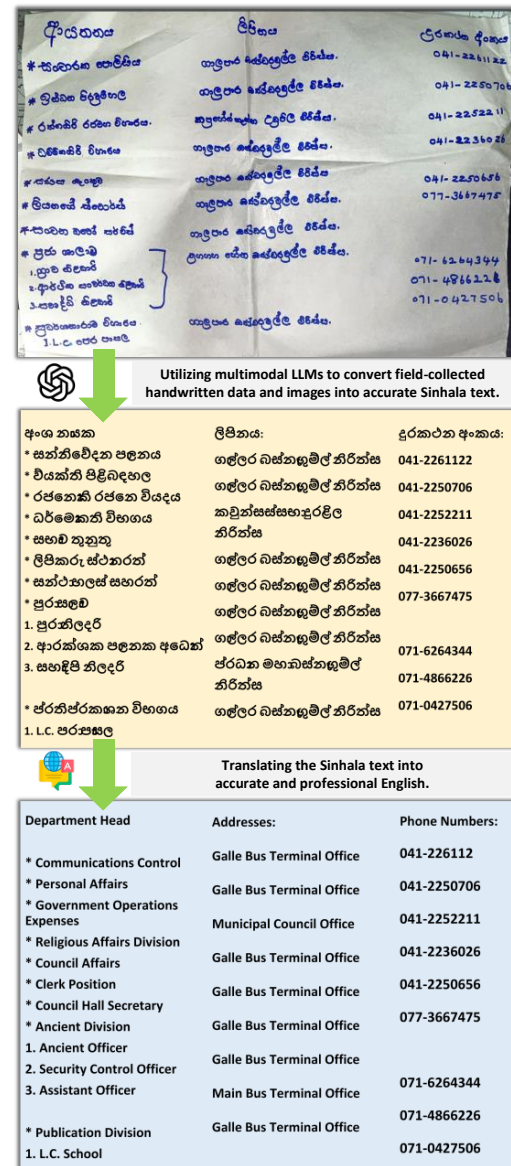


Figure 2: LLM-agent-empowered real-time summary and translation during a participatory workshop.

Socio-Semantic Mediation Agents (SSMA): SSMA's specialize in real-time, code-mixed translation, interpretation, and semantic alignment. They combine transformer-based multilingual LLMs with domain adaptation layers and specialized tok-

enization schemes to handle code-mixing. The underlying algorithms utilize attention-based context retrieval and fine-grained subword embeddings for Indo-Aryan and Dravidian language families. This enables them to preserve semantic nuance across languages, dialects, and honorific forms (Mohamed et al., 2024; Sitaram et al., 2020). By continuously updating a cultural knowledge graph, SSMA ensure fidelity to local ontologies, social hierarchies, and linguistic registers. For instance, when encountering an unexpected code-mixed utterance, the SSMA applies a disambiguation sub-module that uses few-shot prompt examples to infer the correct semantic interpretation before generating a coherent translation or summary (Dowlagar and Mamidi, 2023; Ye, 2024).

Ethnographic Intelligence Agents (EIA): EIAs integrate LLM-based natural language understanding with multimodal feature extraction to capture the richness of ethnographic data. Beyond handling textual inputs, EIAs incorporate audio and visual signals—such as speaker intonation and gesture cues—through multimodal transformers. By aligning textual embeddings with non-verbal cues and contextual metadata, these agents can infer deeper cultural subtexts. Algorithmically, EIAs utilize contrastive learning methods to align representations of linguistic and non-linguistic signals, ensuring that the ethnographic narrative remains coherent and contextually faithful (Yang, 2024; Sadia et al., 2024; Lee et al., 2024).

Community Engagement Orchestration Agents (CEOA): CEOAs manage the ethical and relational dimensions of the research. These agents are configured with ethical protocols, informed consent modules, and data sovereignty guidelines. Their internal logic includes rule-based inference systems that ensure compliance with community-established protocols. For example, CEOAs generate prompts to clarify participant consent forms in code-mixed language or to guide researchers through culturally sensitive topics. They also track and document interactions in a transparent ledger, providing stakeholders with an audit trail of engagement activities (Ninan et al., 2024; Chow and Li, 2024; Guo et al., 2023).

3.2 Integration into Participatory Methods

The integration of our LLM-driven multi-agent framework into participatory research methodologies extends far beyond basic translation or tran-

scription. It is a holistic, context-aware process designed to meaningfully elevate the entire lifecycle of community engagement—from the earliest moments of instrument design to the final phases of data validation and policy recommendation. The guiding principle is that each agent type, while technically distinct, continuously aligns its operational parameters with the evolving socio-cultural and linguistic contours of the communities involved (Fig. 8).

To illustrate this integration, consider the workflow of a community workshop aimed at flood risk assessment in a code-mixed linguistic environment. Initially, the Participatory Research Design and Analytics Agents (PRDAAs) are responsible for selecting and tailoring research instruments—such as surveys or focus group outlines—using prompt-based generation methods that incorporate cultural knowledge repositories and previously annotated corpora. These instruments are not static; rather, they are refined in an iterative manner. For instance, PRDAAs initially produce a series of candidate questions in Sinhala-English code-mixed format, balancing linguistic accessibility with domain specificity. The questions are then tested against synthetic corpora representing likely participant responses. In this simulation step, Socio-Semantic Mediation Agents (SSMA) perform detailed code-mixed translation and semantic alignment checks, ensuring that the initial prompts and questions maintain fidelity to cultural nuances and do not inadvertently skew participant interpretations.

Once the research instruments have passed preliminary tests, they move into the field setting. During live surveys and interviews, PRDAAs dynamically adjust question complexity and phrasing in response to real-time cues from both human researchers and Ethnographic Intelligence Agents (EIAs). If local participants exhibit confusion, fatigue, or hesitation—signaled by vocal intonation changes or subtle body language cues captured and interpreted by EIAs—PRDAAs issue refined prompt directives to SSMA. The SSMA then generate alternative phrasings or linguistic simplifications, ensuring that each question remains culturally resonant and accessible, without sacrificing the analytic integrity of the instrument. This tight feedback loop can occur multiple times within a single interaction, allowing the conversation to flow naturally and responsively, much like a skilled human facilitator adept at shifting linguistic registers

or explanatory strategies.

Workshops and participatory group activities benefit similarly. Community Engagement Orchestration Agents (CEOAs) integrate data from PRDAAs, SSMAAs, and EIAs to propose culturally relevant engagement scripts. For example, if a workshop involves participatory mapping of flood hotspots, CEOAs might recommend starting with a culturally familiar narrative—such as local flood folklore or historical memory—before transitioning to spatial data collection. While participants discuss their lived experiences, EIAs track non-verbal signals indicating trust or discomfort, and SSMAAs ensure that key cultural metaphors and idioms are faithfully preserved in translations and summaries. This coordination embodies a level of anthropologically informed sensitivity: it respects complex social hierarchies, local linguistic honorifics, and the dynamics of multi-generational knowledge transmission, all while operating under strict ethical guidelines that CEOAs enforce and document. The integration protocol also includes a set of formal interaction rules and metadata annotations. Each agent’s output is enriched with contextual tags, which guide subsequent agent operations. These annotations form a semantic layer that human researchers can later review, providing transparency into the decision-making processes of the agents and enabling critical reflection on whether certain prompts, translations, or adjustments influenced participant responses in unintended ways.

3.3 LLM-Agents-Driven Research Workflow

The workflow orchestrated by our multi-agent system unfolds through a series of interlinked phases designed to ensure continuous adaptation, rigorous quality control, and meaningful involvement of local communities. Each phase leverages the strengths of different agent types, while also maintaining pathways for human oversight, ethical review, and methodological triangulation. The goal is a research pipeline that not only collects data efficiently but also enriches the quality, interpretability, and legitimacy of that data in the eyes of both communities and external stakeholders.

Pre-Field Preparation and Instrumentation: Before stepping into the field, the workflow begins with an extensive pre-field instrumentation phase. Here, PRDAAs generate initial drafts of research instruments—surveys, semi-structured interview guides, and community workshop outlines—based

on project goals and available cultural-linguistic corpora. These initial drafts are subjected to synthetic test scenarios: code-mixed test cases are fed into SSMAAs to benchmark translation accuracy and contextual fidelity, while EIAs simulate multimodal inputs (e.g., hypothetical speaker intonations, gesture-based cues) to assess whether the proposed prompts can handle complex ethnographic scenarios. Iterations are performed until a baseline set of instruments meets quality thresholds defined by the research team, including metrics for linguistic clarity, semantic accuracy, and cultural appropriateness.

Adaptive Field Deployment: With baseline instruments in hand, the team moves into the field. Surveys, interviews, and workshops commence, guided by the prepared materials but never locked into them. As participants respond, SSMAAs deploy on-the-fly translation and code-switching adjustments. If a participant uses a regional idiom not encountered in pre-field training data, SSMAAs rely on few-shot prompt adaptation techniques, referencing similar linguistic patterns to generate accurate, context-aware interpretations. Concurrently, EIAs capture non-verbal signals—such as prolonged pauses, changes in vocal pitch, or restless body language—to produce ethnographic annotations. These annotations are fed back into PRDAAs, which may trigger immediate modifications to the research instrument. For instance, if participants appear disengaged, PRDAAs may instruct SSMAAs to simplify the phrasing or incorporate culturally salient metaphors to re-engage the community’s interest.

Ethical Monitoring and Protocol Enforcement: During these field interactions, CEOAs maintain a real-time ethical interaction ledger. This ledger logs every adaptation request, every change in linguistic register, and every potential breach of community protocols. Should a line of questioning veer into sensitive territory—such as local religious traditions or gender-related norms—CEOAs issue alerts prompting the research team to reconsider the approach. If participants request anonymity or display discomfort with certain data-collection practices, CEOAs dynamically adapt the informed consent modules and ensure that new protocols are communicated in accessible, code-mixed language.

Multilingual Thematic Analysis and Iterative Refinement: After field data is collected, it passes through a multilingual thematic analysis pipeline.

PRDAAs and EIAs collaborate to identify recurring narratives, power hierarchies, and cultural themes that emerge from the data. By leveraging transformer-based topic modeling and clustering methods fine-tuned for code-mixed input, the agents reveal patterns that might be missed by single-language or monomodal approaches. This phase also includes a human-in-the-loop feedback cycle, where researchers and local experts evaluate the thematic outputs. Feedback is translated into updated prompt templates and agent-specific instructions. If local stakeholders indicate that a certain theme has been misinterpreted—perhaps a traditional narrative was wrongly associated with risk aversion instead of historical resilience—agents adjust their semantic weighting and cultural context embeddings.

Iterative Learning and Continuous Improvement: Rather than terminating after a single cycle of data collection and analysis, the workflow encourages continuous learning. New linguistic patterns, emergent cultural idioms, and shifting community priorities feed back into the system. PRDAAs update their instrument-generation models, SSMAAs refine their code-switch adaptation strategies, EIAs improve their multimodal understanding, and CEOAs integrate revised ethical guidelines or local governance structures. Over time, the system becomes more attuned to community-specific realities, and its outputs become increasingly reliable, nuanced, and aligned with local perspectives.

4 Implementation in Field Work and Insights

As is shown in Fig.3 and Appendix.A, to test the feasibility of this novel system, we implemented it in our field research, which focused on enhancing the Early Warning Systems (EWS) for flood management in the Nilwala River Basin, a region prone to recurrent flooding with devastating socio-economic impacts in Sri Lanka. Sri Lanka’s linguistic landscape is emblematic of South Asia’s broader linguistic diversity, characterized by the prevalence of code-mixing and multilingual communication (Mandavilli, 2020). Sinhala, an Indo-Aryan language with agglutinative features and a rich system of honorifics, often intertwines with English and other local dialects in everyday discourse, which poses significant challenges for NLP, as it involves syntactic, lexical, and semantic blending that tradi-

tional language models struggle to interpret accurately. The objective was to employ the proposed system to facilitate participatory development research methods—including surveys, structured and semi-structured interviews, workshops, and other interactive engagements—with stakeholders ranging from national agencies to local communities.

4.1 Practical Experiences and Outcomes

The implementation faced several challenges, particularly in adapting the LLMs to handle Sinhala-specific linguistic features and the pervasive code-mixing in communication. The scarcity of high-quality, annotated Sinhala corpora necessitated innovative approaches, including active learning techniques and data augmentation strategies to enhance the model’s proficiency (Jagosh et al., 2012).

One significant achievement was the development of a hybrid translation approach that combined statistical and neural methods, achieving a 35% improvement in translation accuracy for domain-specific terminology compared to standard multilingual models, which was critical for accurately interpreting participants’ responses during interviews and ensuring that subtle nuances were not lost in translation. During workshops, they assisted in designing interactive activities that resonated with local customs and facilitated real-time feedback collection. In surveys and interviews, the agents helped generate culturally appropriate questions and dynamically adjusted to participants’ inputs, enhancing the depth and authenticity of the data collected. The agents also played a crucial role in the analysis phase. They enabled cross-linguistic comparisons and facilitated the synthesis of complex data into actionable insights (Cemoge et al., 2024). For instance, they helped identify communication bottlenecks between agencies involved in the EWS, revealing that outdated communication methods and bureaucratic procedures were significant barriers to effective disaster management.

4.2 Lessons Learned and Recommendations

Community Involvement is Crucial: Active participation of local stakeholders in the development and refinement of the system was essential. Their input ensured that the agents were culturally attuned and responsive to the community’s needs, enhancing acceptance and effectiveness.

Flexible Adaptation Mechanisms are Necessary: The linguistic diversity and code-mixing practices required the agents to be highly adaptable.



Figure 3: Participatory Field Research with LLM-agent-assisted tools. (Source: Authors’ fieldwork)

Implementing mechanisms for continuous learning and real-time adjustment was critical for handling linguistic variations and unexpected inputs.

Human Oversight Remains Indispensable: While the agents significantly enhanced efficiency and depth, human researchers played a vital role in overseeing the process, interpreting nuanced cultural contexts, and making ethical judgments.

Addressing Technical Challenges: Overcoming the scarcity of linguistic resources demanded innovative technical solutions. Investing in the development of annotated corpora and leveraging transfer learning were effective strategies for enhancing model performance.

4.3 Implementation Considerations for Broader Deployment

The Nilwala River Basin deployment illustrates a scalable and domain-agnostic framework. To adapt it for other South Asian languages and contexts, the modular architecture allows integrating new code-mixing tokenizers, cultural knowledge bases, or domain-specific LLM fine-tunings (Finkel et al., 2022).

Technical Infrastructure: Resource-poor settings demand efficient model architectures. Lightweight LLMs combined with on-device pre-

processing, federated learning, and quantization can mitigate latency and connectivity issues (Qu et al., 2024).

Data Security and Privacy: Incorporating end-to-end encryption and federated learning ensures sensitive community data remains local while still contributing to the global improvement of model quality. CEOAs enforce data usage policies, ensuring that outputs are ethically and legally compliant.

Ethical and Cultural Considerations: The framework’s prompt design explicitly encodes ethical guidelines. CEOAs monitor compliance in real-time, and any deviation triggers a review workflow. Building and maintaining culturally informed knowledge graphs ensures the models reflect community values rather than imposing external biases (Suppadungsuk et al., 2023).

Capacity Building and Institutional Support: Sustained success requires local training programs and policy engagement. By equipping researchers and stakeholders with the skills to interpret, customize, and govern these systems, we foster long-term sustainability and local empowerment. Collaborations with NGOs, government agencies, and academic institutions can institutionalize best practices, streamline resource allocation, and formalize quality assurance standards.

5 Discussion and Future Agenda

The integration of LLM-based multi-agent systems into participatory development research reconfigures the conceptual space at the intersection of technology, community engagement, and anthropological inquiry (Xu et al., 2024). Far from being a mere technical enhancement, this approach prompts us to re-evaluate foundational assumptions about the production, circulation, and legitimation of knowledge in socio-culturally complex contexts. In traditional participatory frameworks, human facilitators, local knowledge brokers, and community spokespersons navigate the intricacies of language, power asymmetries, and cultural semiotics. Our LLM-driven architecture extends this negotiation field, distributing interpretive authority and methodological agency across human and non-human actors. This shift demands that we refine our criteria for epistemic robustness and ethical accountability. By introducing adaptive prompts, multimodal interpretation layers, and code-mixed language models, the research process becomes more dialogic and reflexive, simultaneously more scalable and less deterministic. While existing literature in participatory development and linguistic anthropology has long emphasized the importance of local involvement (Penuel et al., 2020), the emergence of LLM-based agents compels a reconsideration of whose voices are amplified, how biases are mitigated, and under what conditions community knowledge is validated. Methodological rigor thus transcends traditional validation protocols, calling for new evaluative paradigms where model outputs must be continually negotiated, contested, and contextualized by community stakeholders.

These technological trajectories also invoke philosophical questions about the essence of community agency and the nature of equitable development. In harnessing LLMs to broker dialogues between disparate linguistic and cultural systems, we challenge the modernist assumption that technology is a neutral mediator. Instead, AI becomes an evolving participant in a dense socio-technical network—one that can enrich cultural representation, but also requires vigilant governance to prevent the re-inscription of power imbalances. Future research must thus address the deeper normative concerns: how can we ensure that AI-enabled participatory practices bolster rather than diminish local epistemologies and life-worlds? How do we integrate metrics of cultural resilience, trust-

building, and vernacular knowledge sustainability into development assessments (Falcone, 2023)? In charting this future agenda, interdisciplinary collaboration is paramount. Technologists, anthropologists, linguists, and development practitioners must co-design systems that are both contextually resonant and theoretically informed. The promise of these LLM-based frameworks lies not simply in improved data collection or analytical sophistication, but in ushering in a more philosophically coherent paradigm of research—one that values uncertainty, pluralism, and continuous ethical reflection as integral components of knowledge production.

6 Conclusion

The introduction of LLM-based multi-agent architectures into participatory research settings in South Asia signals a profound transformation, offering new avenues for bridging linguistic divides and socio-cultural complexities without reducing communities to passive data sources. Rather than replacing traditional methods, these technologies complement and extend established participatory principles: human facilitators remain indispensable ethical and interpretive anchors, while LLM-based agents broaden the scope, adaptability, and depth of engagements. The real significance of this paradigm lies in how it reconfigures the relational field of development research. By treating language models as interlocutors that adapt to local idioms, cultural protocols, and conceptual frames, the process moves closer to what humanistic inquiry has always sought: a genuine dialogic co-production of meaning. This approach transcends conventional efficiency metrics, orienting research toward a deeper, ethically engaged form of knowledge-making.

The path forward necessitates sustained reflection and critical praxis. Cross-sectoral alliances and supportive institutional frameworks are required to ensure that technology-enhanced participatory models do not inadvertently replicate existing inequalities or impose external epistemologies. Ultimately, the convergence of advanced NLP, anthropological rigor, and participatory ethos challenges the prevailing boundaries of development research. It opens the door to an epistemically plural and ethically attentive mode of inquiry, one that holds promise for more inclusive, contextually authentic, and transformative engagements with communities worldwide.

References

- Leonidas Agathos, Andreas Avgoustis, Xristiana Krylesi, Aikaterini Makridou, Ilias Tzanis, Despoina Mouratidis, Katia Lida Kermanidis, and Andreas Kanavos. 2024. [Bridging linguistic gaps: Developing a greek text simplification dataset](#). *Information*, 15(8):500.
- Germine H Awad, Erika A Patall, Kadie R Rackley, and Erin D Reilly. 2016. [Recommendations for culturally sensitive research methods](#). *Journal of Educational and Psychological Consultation*, 26(3):283–303.
- Elisabeth Björk Brämberg and Karin Dahlberg. 2013. [Interpreters in cross-cultural interviews: a three-way coconstruction of data](#). *Qualitative health research*, 23(2):241–247.
- Nik Bear Brown. 2024. [Enhancing trust in llms: Algorithms for comparing and interpreting llms](#).
- Tapumih Cemoge, Alexei Ivanov, and Isaiah Moreno. 2024. [A two-stage fine-tuning approach for democratizing multilingual large language models in automated literature review tasks](#).
- James CL Chow and Kay Li. 2024. [Ethical considerations in human-centered ai: Advancing oncology chatbots through large language models](#). *JMIR Bioinformatics and Biotechnology*, 5:e64406.
- Gideon Oluseyi Daramola, Adetomi Adewumi, Boma Sonimiteim Jacks, and Olakunle Abayomi Ajala. 2024. [Navigating complexities: a review of communication barriers in multinational energy projects](#). *International Journal of Applied Research in Social Sciences*, 6(4):685–697.
- Uwe Deichmann, Aparajita Goyal, and Deepak Mishra. 2016. [Will digital technologies transform agriculture in developing countries?](#) *Agricultural Economics*, 47(S1):21–33.
- Mohamed Diab Idris, Xiaohua Feng, and Vladimir Dyo. 2024. [Revolutionizing higher education: Unleashing the potential of large language models for strategic transformation](#). *IEEE Access*, 12:67738–67757.
- Suman Dowlagar and Radhika Mamidi. 2023. [A code-mixed task-oriented dialog dataset for medical domain](#). *Computer Speech & Language*, 78:101449.
- Manaswita Dutta, Tina MD Mello, Yesi Cheng, Niladri Sekhar Dash, Ranita Nandi, Aparna Dutt, and Arpita Bose. 2024. [Universal and language-specific connected speech characteristics of bilingual speakers with alzheimer’s disease: Insights from case studies of structurally distinct languages](#). *Journal of Speech, Language, and Hearing Research*, 67(4):1143–1164.
- Pasquale Marcello Falcone. 2023. [Sustainable energy policies in developing countries: a review of challenges and opportunities](#). *Energies*, 16(18):6682.
- Md Meftahul Ferdaus, Mahdi Abdelguerfi, Elias Ioup, Kendall N. Niles, Ken Pathak, and Steven Sloan. 2024. [Towards trustworthy ai: A review of ethical and robust large language models](#).
- Raphael Finkel, Daniel Kaufman, and Ahmed Shamim. 2022. [Analyzing code-mixing in linguistic corpora using kratylos](#). *J. Comput. Cult. Herit.*, 15(1).
- Susanne Göpferich and Riitta Jääskeläinen. 2009. [Process research into the development of translation competence: Where are we, and where do we need to go?](#) *Across languages and cultures*, 10(2):169–191.
- Zishan Guo, Renren Jin, Chuang Liu, Yufei Huang, Dan Shi, Supryadi, Linhao Yu, Yan Liu, Jiaxuan Li, Bojian Xiong, and Deyi Xiong. 2023. [Evaluating large language models: A comprehensive survey](#).
- Muhammad Usman Hadi, Qasem Al Tashi, Abbas Shah, Rizwan Qureshi, Amgad Muneer, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, et al. 2024. [Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects](#). *Authorea Preprints*.
- Ehtesham Hashmi, Sule Yildirim Yayilgan, Ibrahim A. Hameed, Muhammad Mudassar Yamin, Mohib Ullah, and Mohamed Abomhara. 2024. [Enhancing multilingual hate speech detection: From language-specific insights to cross-linguistic integration](#). *IEEE Access*, 12:121507–121537.
- Mohammad Hassan, Prakash Rai, and Sabin Maharjan. 2023. [Empowering south asian agricultural communities: A comprehensive approach to iot-driven agriculture through awareness, training, and collaboration](#). *Quarterly Journal of Emerging Technologies and Innovations*, 8(3):18–32.
- James Hutson, Pace Ellsworth, and Matt Ellsworth. 2024. [Preserving linguistic diversity in the digital age: a scalable model for cultural heritage continuity](#). *Journal of Contemporary Language Research*, 3(1).
- Justin Jagosh, Ann C Macaulay, Pierre Pluye, JON Salsberg, Paula L Bush, JIM Henderson, Erin Sirett, Geoff Wong, Margaret Cargo, Carol P Herbert, et al. 2012. [Uncovering the benefits of participatory research: implications of a realist review for health research and practice](#). *The Milbank Quarterly*, 90(2):311–346.
- Indrajit Kar, Zonunfeli Ralte, Maheshakumara Shivakumara, Rana Roy, and Arti Kumari. 2024. [Agents are all you need: Elevating trading dynamics with advanced generative ai-driven conversational llm agents and tools](#). In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pages 1–6. IEEE.
- Josephine Kass-Hanna, Angela C Lyons, and Fan Liu. 2022. [Building financial resilience through financial and digital literacy in south asia and sub-saharan africa](#). *Emerging Markets Review*, 51:100846.

- Grgur Kovač, Rémy Portelas, Peter Ford Dominey, and Pierre-Yves Oudeyer. 2024. [The socialai school: a framework leveraging developmental psychology toward artificial socio-cultural agents](#). *Frontiers in Neurobotics*, 18:1396359.
- Nir Kshetri. 2024. [Linguistic challenges in generative artificial intelligence: Implications for low-resource languages in the developing world](#).
- Sangmin Lee, Minzhi Li, Bolin Lai, Wenqi Jia, Fiona Ryan, Xu Cao, Ozgur Kara, Bikram Boote, Weiyan Shi, Diyi Yang, and James M. Rehg. 2024. [Towards social ai: A survey on understanding social interactions](#).
- Alexandre Magueresse, Vincent Carles, and Evan Heetderks. 2020. [Low-resource languages: A review of past work and future challenges](#).
- Sujay Rao Mandavilli. 2020. [Towards a comprehensive compendium of factors impacting language dynamics in post-globalized scenarios: Presenting principles, paradigms and frameworks for use in the emerging science of language dynamics](#). *ELK Asia Pacific Journal of Social Sciences*, 6(3).
- Yaron Matras, Rebecca Tipton, and Leonie Gaiser. 2023. [Agency and multilingualism in public health care: how practitioners draw on local experiences and encounters](#). In *Understanding the Dynamics of Language and Multilingualism in Professional Contexts*, pages 46–60. Edward Elgar Publishing.
- Yasir Abdelgadir Mohamed, Akbar Khanan, Mohamed Bashir, Abdul Hakim H. M. Mohamed, Mousab A. E. Adiel, and Muawia A. Elsadig. 2024. [The impact of artificial intelligence on language translation: A review](#). *IEEE Access*, 12:25553–25579.
- Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohunge, Solomon Oluwale Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, Sackey Freshia, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa, Mofe Adeyemi, Masabata Mokgesi-Seling, Lawrence Okegbemi, Laura Jane Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroko Orife, Ignatius Ezeani, Idris Abdulkabir Dangana, Herman Kamper, Hady Elshahar, Goodness Duru, Ghollah Kioko, Espoir Murhabazi, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Emezue, Bonaventure Dossou, Blessing Sibanda, Blessing Ito Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. [Participatory research for low-resourced machine translation: A case study in african languages](#).
- Johan Ninan, Stewart Clegg, Ashwin Mahalingam, and Shankar Sankaran. 2024. [Governance through trust: Community engagement in an australian city rebuilding precinct](#). *Project Management Journal*, 55(1):16–30.
- Marinela Parovic. 2024. [Improving Parameter-Efficient Cross-Lingual Transfer for Low-Resource Languages](#). Ph.D. thesis.
- William R Penue, Robbin Riedy, Michael S Barber, Donald J Peurach, Whitney A LeBouef, and Tiffany Clark. 2020. [Principles of collaborative education research with stakeholders: Toward requirements for a new research and development infrastructure](#). *Review of Educational Research*, 90(5):627–674.
- Karin Pfeffer, Isa Baud, Eric Denis, Dianne Scott, and John Sydenstricker-Neto. 2013. [Participatory spatial knowledge management tools: empowerment and up-scaling or exclusion?](#) *Information, Communication & Society*, 16(2):258–285.
- Guanqiao Qu, Qiyuan Chen, Wei Wei, Zheng Lin, Xi-anhao Chen, and Kaibin Huang. 2024. [Mobile edge intelligence for large language models: A contemporary survey](#).
- Qazi Arka Rahman. 2024. [Reconceptualizing south asia: Bangladeshi literature and the politics of representation](#).
- Mohaimenul Azam Khan Raiaan, Md. Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. 2024. [A review on large language models: Architectures, applications, taxonomies, open issues and challenges](#). *IEEE Access*, 12:26839–26874.
- Nitin Liladhar Rane, Mallikarjuna Paramesha, Saurabh P Choudhary, and Jayesh Rane. 2024. [Artificial intelligence, machine learning, and deep learning for advanced business strategies: a review](#). *Partners Universal International Innovation Journal*, 2(3):147–171.
- Nitin Liladhar Rane, Abhijeet Tawde, Saurabh P Choudhary, and Jayesh Rane. 2023. [Contribution and performance of chatgpt and other large language models \(llm\) for scientific and research advancements: a double-edged sword](#). *International Research Journal of Modernization in Engineering Technology and Science*, 5(10):875–899.
- Partha Pratim Ray. 2023. [Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope](#). *Internet of Things and Cyber-Physical Systems*, 3:121–154.
- John Roberts, Max Baker, and Jane Andrew. 2024. [Artificial intelligence and qualitative research: The promise and perils of large language model \(llm\)‘assistance’](#). *Critical Perspectives on Accounting*, 99:102722.
- Vanesa Rodríguez Tembrás. 2024. [Code-Switching in Bilingual Medical Consultations \(Galician-Spanish\)](#). Ph.D. thesis.

- A Sabarirajan, Latha Thamma Reddi, Sandeep Rangineni, R Regin, S Suman Rajest, and P Paramasivan. 2024. [Leveraging mis technologies for preserving india's cultural heritage on digitization, accessibility, and sustainability](#). In *Data-Driven Intelligent Business Sustainability*, pages 122–135. IGI Global.
- Bareera Sadia, Farah Adeeba, Sana Shams, and Kashif Javed. 2024. [Meeting the challenge: A benchmark corpus for automated urdu meeting summarization](#). *Information Processing & Management*, 61(4):103734.
- Sunayana Sitaram, Khyathi Raghavi Chandu, Sai Krishna Rallabandi, and Alan W Black. 2020. [A survey of code-switched speech and language processing](#).
- Hedvig Skirgård, Hannah J Haynie, Damián E Blasi, Harald Hammarström, Jeremy Collins, Jay J Latarche, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, et al. 2023. [Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss](#). *Science Advances*, 9(16):eadg6175.
- S Suppadungsuk, C Thongprayoon, J Miao, S Tangpanithandee, I Craici, W Cheungpasitporn, et al. 2023. [Ethical implications of chatbot utilization in nephrology](#). *Journal of Personalized Medicine*, 13(9).
- Tina Tomec and Sergej Gričar. 2024. [Risk language barriers in a globalized world: insights from female managers from slovenia](#). *Strategic Management-International Journal of Strategic Management and Decision Support Systems in Strategic Management*, 29(2).
- Ehsan Ullah, Anil Parwani, Mirza Mansoor Baig, and Rajendra Singh. 2024. [Challenges and barriers of using large language models \(llm\) such as chatgpt for diagnostic medicine with a focus on digital pathology—a recent scoping review](#). *Diagnostic pathology*, 19(1):43.
- Eliza Urwin, Aisalkyn Botoeva, Rosario Arias, Oscar Vargas, and Pamina Firchow. 2023. [Flipping the power dynamics in measurement and evaluation: International aid and the potential for a grounded accountability model](#). *Negotiation Journal*, 39(4):401–426.
- Zander Janse van Rensburg and Sonja van der Westhuizen. 2024. [Ethical ai integration in academia: Developing a literacy-driven framework for llms in south african higher education](#). In *AI Approaches to Literacy in Higher Education*, pages 23–48. IGI Global.
- Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. 2023. [Autogen: Enabling next-gen llm applications via multi-agent conversation](#).
- Haowen Xu, Jinghui Yuan, Anye Zhou, Guanhao Xu, Wan Li, Xuegang Ban, and Xinyue Ye. 2024. [Genai-powered multi-agent paradigm for smart urban mobility: Opportunities and challenges for integrating large language models \(llms\) and retrieval-augmented generation \(rag\) with intelligent transportation systems](#).
- Qian Yang. 2024. *Scottish-Chinese students' language use in Chinese complementary school classroom: a translanguaging perspective*. Ph.D. thesis, University of Glasgow.
- Zixi Ye. 2024. Language barriers in intercultural communication and their translation strategies. In *International Conference on Finance and Economics*, volume 6.

A Authors' Field Works Assisted by LLM agents system

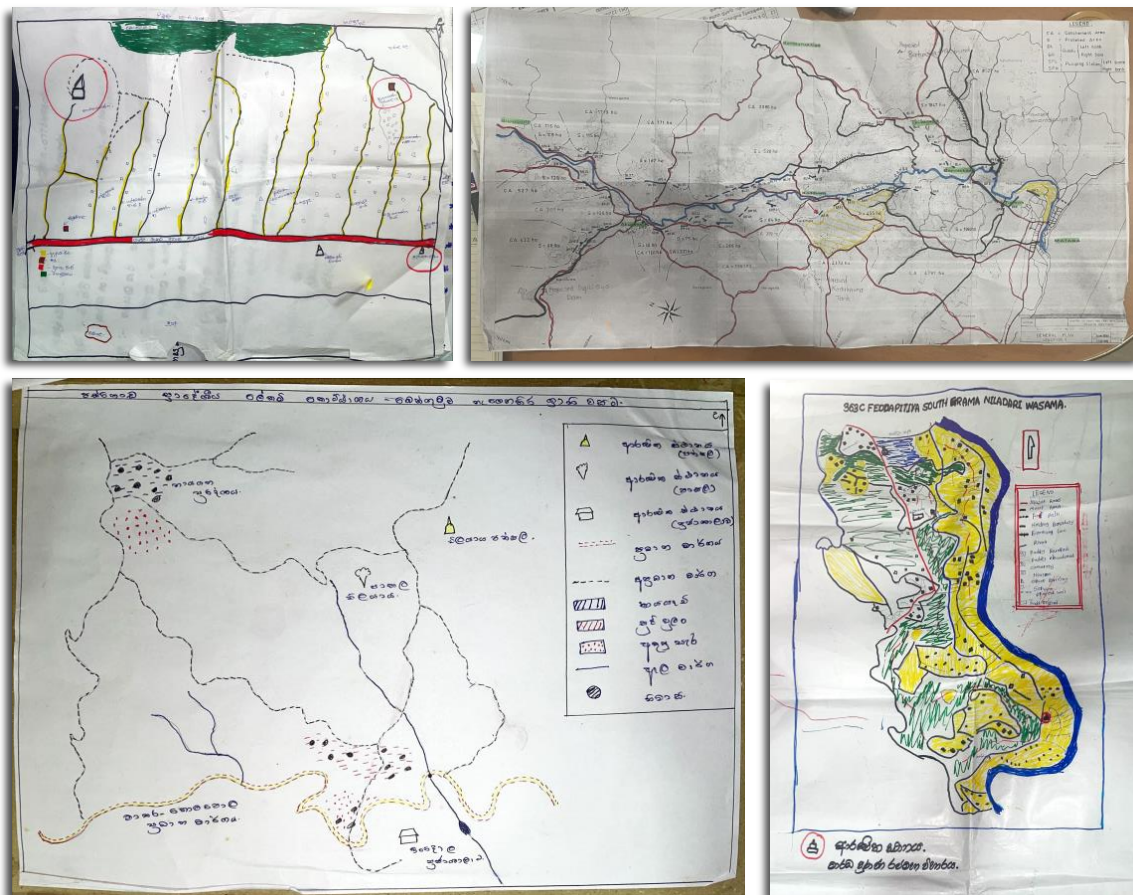


Figure 4: Collected participatory workshop results.

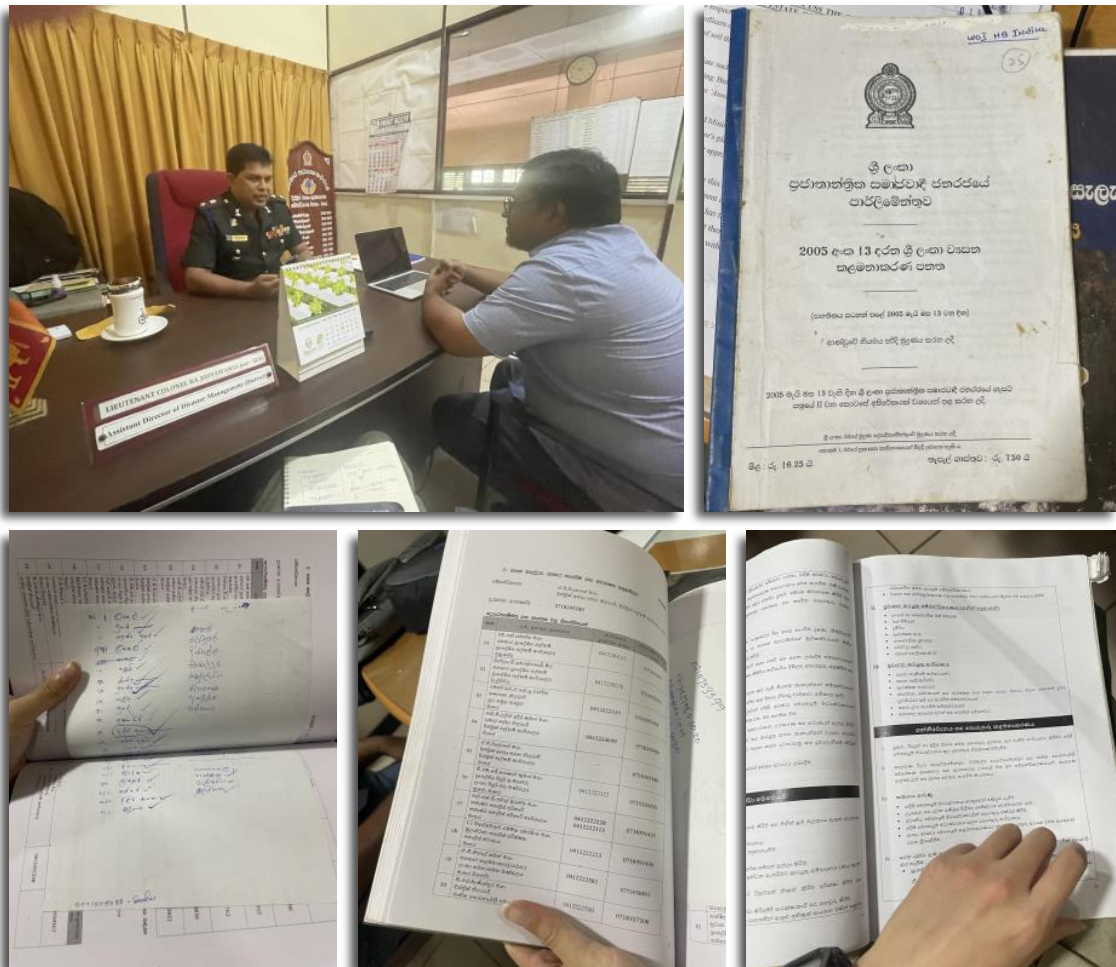


Figure 5: Participatory interview with local DMC (Disaster Management Centre)



Figure 6: Flood sites in Nilwala River Basin Area

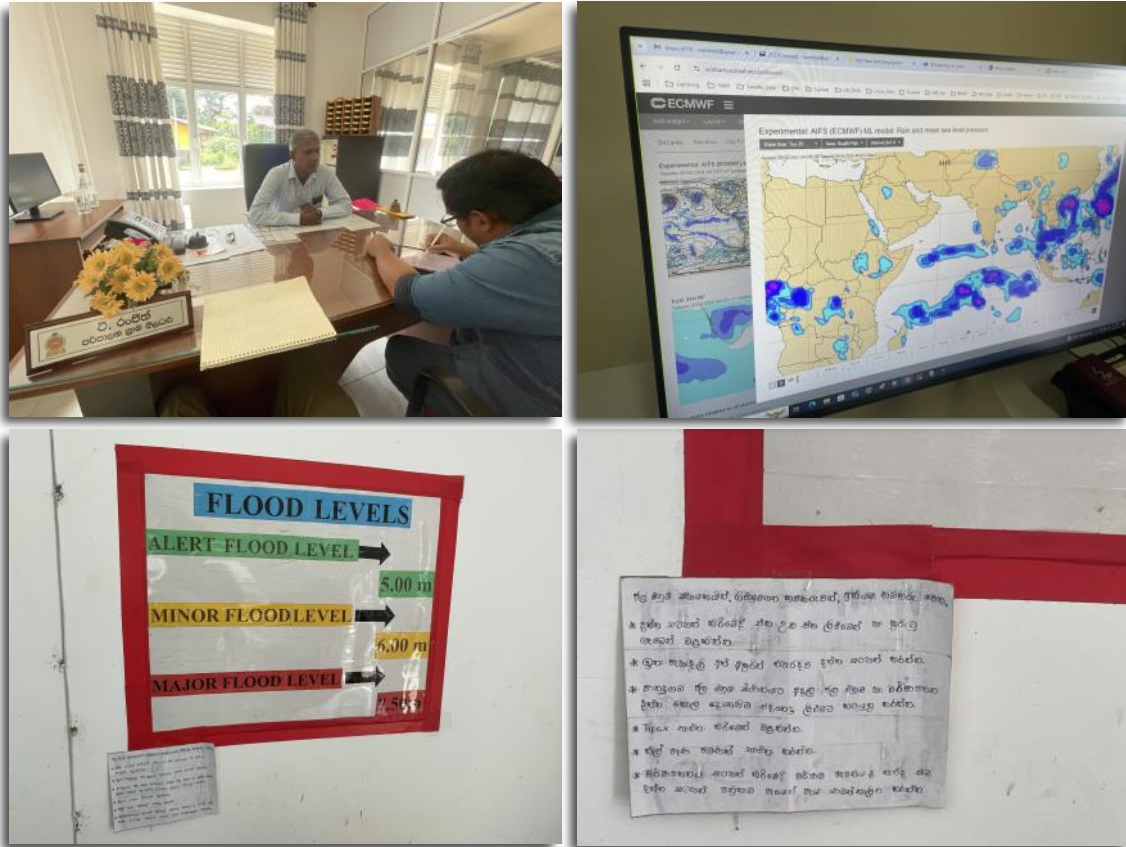


Figure 7: Participatory interview with local government office and Hydrology Department

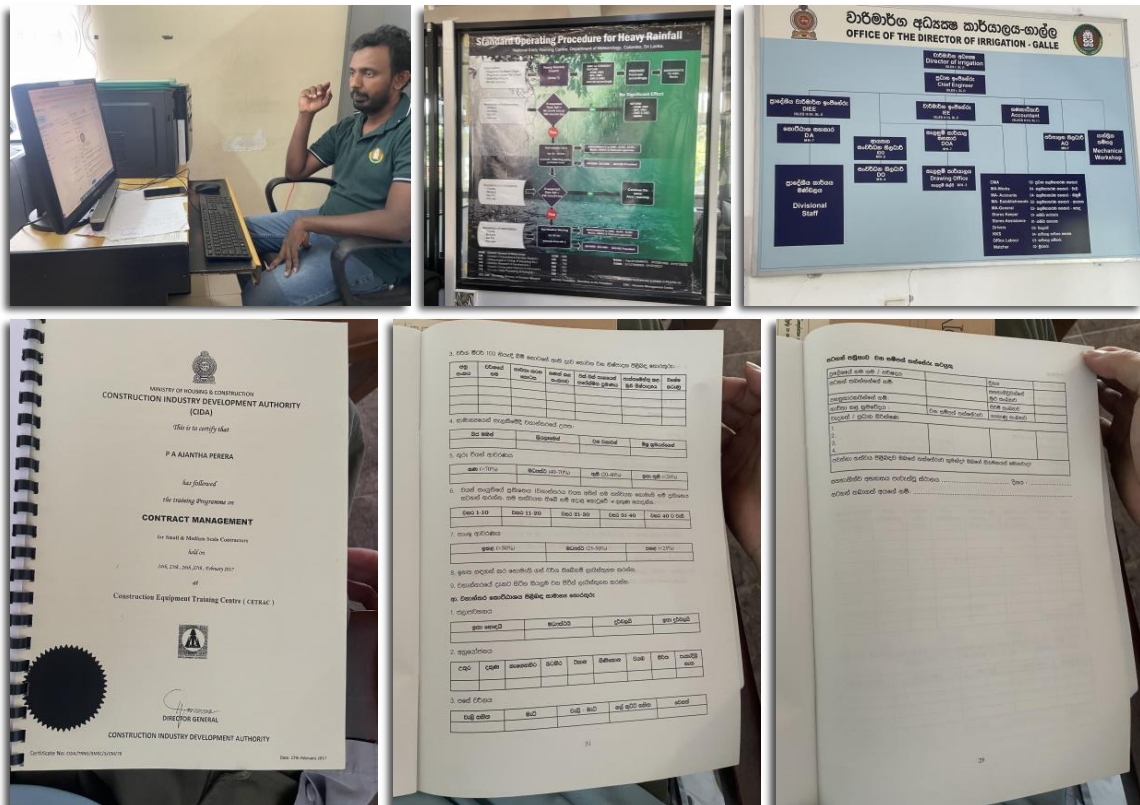


Figure 8: Participatory interview with local Irrigation Department