Cambridge Judge Business School

EkStep Foundation

Frugal AI Pathways: AI for Low-Resource Languages Workshop Report

Deployment lessons from the Cambridge workshop, 3 July 2026









FRUGAL AI HUB at UNIVERSITY OF CAMBRIDGE Judge Business School

EkStep

1. Authors and publication information

Serish Venkata Gandikota — Co-founder & Co-Lead, Frugal AI Hub · Honorary Fellow, Centre for India and Global Business, Cambridge Judge Business School

Krishna Raju Jampana — Director - Solutions & Growth, Digital Transformation, EkStep Foundation

Santosh Kevlani — Head, Voice and Language AI, EkStep Foundation

Jaideep Prabhu — Professor of Marketing and Nehru Professor of Indian Business, Cambridge Judge Business School · Co-Director, Centre for India and Global Business

Dr. Arjuna Sathiaseelan — Technology Lead, Frugal AI Hub · Visiting Fellow, Cambridge Judge Business School

Elizabeth Osta — Co-founder & Co-Lead, Frugal AI Hub · Visiting Fellow, Cambridge Judge Business School

Publication month/year:
October 2026

2. Acknowledgements

This report draws on the Frugal AI Pathways: AI for Low-Resource Languages workshop convened by the Frugal AI Hub at Cambridge Judge Business School in partnership with EkStep Foundation on 3 July 2026. The authors thank the practitioners who shared their implementation journeys, the adopter teams who brought live deployment challenges to the workshop, and all participants who contributed their experience, questions and perspectives during the working sessions.

We particularly acknowledge Mitesh Khapra, Dr. Sarbani Banerjee Belur, Dr. Arjuna Sathiaseelan, Kato Steven Mubiru and Santosh Kevlani for sharing practitioner experience and technical insights. Special thanks go to Dr. Sarbani Banerjee Belur and Kato Steven Mubiru for bringing the Soliga and Swahili problem statements into the working sessions, enabling participants to work with live implementation questions rather than hypothetical cases.

We also acknowledge the wider teams at EkStep Foundation and the Frugal AI Hub, the organisations represented in the room, and the workshop participants whose contributions shaped the working observations and open questions in this report.

The workshop and report contribute to the 100 AI Diffusion Pathways initiative, which seeks to capture and share implementation know-how so that future adopters can build on what others have already learned.

3. Suggested citation

Gandikota, S.V., Jampana, K.R., Kevlani, S., Prabhu, J., Sathiaseelan, A. and Osta, E. (2026). Frugal AI Pathways: AI for Low-Resource Languages: Workshop Report. Deployment lessons from the Cambridge workshop, 3 July 2026. Frugal AI Hub at Cambridge Judge Business School and EkStep Foundation, September 2026.

4. Disclaimer / report status

This workshop report synthesises observations, discussions, lived experiences and know how shared during the Frugal AI Pathways: AI for Low-Resource Languages workshop held at Cambridge Judge Business School on 3 July 2026. The views and examples reported reflect contributions made during the workshop and do not necessarily represent the institutional positions of all participating organisations. The working observations and open questions are intended as inputs for further testing, learning and refinement as additional deployment pathways are documented.

Table of Contents

Executive Summary 05

A. About the Workshop 07

A.1 Convenors, venue, and date 07

A.2 Why this workshop was convened 07

A.3 Who was in the room 08

A.4 How the day was structured 09

A.5 From the initial design to the final design 09

A.6 What the workshop produced 10

B. Working Observations 12

B1. Lived experience carries knowledge that case studies often omit 12

B2. Context determines what working voice AI means 12

B3. Technology is not the binding constraint 14

B4. Start with the person, their problem and purpose 15

B5. Community participation shapes the design and governance of AI systems 16

B6. Scaling requires Institutional Anchors 17

B7. Ownership, accountability and operating economics are not after thoughts 19

C. Open Questions 21

C1. What level of data and technical capability is sufficient for the intended use? 21

C2. How should quality and trust be established where conventional language resources are limited? 21

C3. Who defines the purpose of the capability, and whose priorities count? 22

C4. What does community ownership mean in practice? 22

C5. Which institution should anchor the deployment and remain accountable? 23

C6. What operating model can sustain the service after the initial project? 23

C7. What infrastructure and local capability should remain in place? 23

D. Next Steps 25

E. Source and Verification Note 26

Executive Summary

The Frugal AI Pathways: AI for Low-Resource Languages workshop brought together practitioners with experience of building and successfully deploying Voice & Language AI use cases for low-resource languages and adopters beginning similar journeys. Convened by the Frugal AI Hub in partnership with the EkStep Foundation at the University of Cambridge, the workshop explored a simple question: **How can lived deployment experience help the next adopter move further and faster than they would on their own?**

The workshop was originally designed around the idea that practitioners would share their journeys, including deployment lessons, after which participants would work through a predefined set of questions to develop approaches for their own deployments. During the final planning period, however, the organisers recognised that this approach assumed they already knew the most important questions. It also overlooked the fact that the next adopters themselves were still defining their users, purpose, institutional arrangements and long-term operating models. The workshop was therefore redesigned around question discovery. Rather than beginning with predetermined questions, participants first identified the questions each deployment team would need to answer before a credible pathway could emerge.



The practitioner journeys presented by AI4Bharat and EkStep, together with the working sessions around the Soliga and Swahili deployments, revealed a consistent pattern. While advances in foundation models, related-language transfer and open infrastructure have significantly lowered technical barriers, the remaining challenges to scale Voice AI use cases at societal scale are increasingly non-technical. Successful deployments depended on understanding the persona and purpose the service was intended to support, engaging communities as active participants in design and governance, establishing legitimate institutional anchors, designing clear ownership and accountability arrangements, and identifying sustainable operating models from the outset. Lived experience and practical know-how — the knowledge accumulated through deploying a real-world use case at scale — emerged as a critical asset that future adopters can reuse and adapt to their context, provided it is deliberately captured and shared.

These observations suggest that AI diffusion involves more than transferring technology. It also requires sharing the lived experiences and practical know-how that practitioners develop over time, enabling future adopters to adapt and apply these insights to their own contexts.

The seven working observations presented in this paper are offered as early observations to be tested for organisations seeking to deploy AI at societal scale and as an initial contribution towards the 100 AI Diffusion Pathways initiative, whose objective is to capture and codify lived experience and practical know-how so that future adopters can build on it rather than begin from scratch.

About Frugal AI Hub

The Frugal AI Hub at Cambridge Judge Business School focuses on how organisations can move from promising AI capabilities to sustained use under real-world constraints. Its work combines research with capacity building, helping adopters define the problem they are trying to solve, understand the institutional and operating conditions around it, learn from practitioners who have faced similar implementation challenges, and adapt those lessons to their own context. Through its collaboration with EkStep on the 100 Pathways initiative, the Hub is exploring how implementation experience can be turned into practical learning that strengthens the capacity of future adopters while generating evidence across different sectors, countries and stages of adoption.

About Ekstep

EkStep Foundation is a philanthropic foundation dedicated to equitably empowering people and unlocking possibilities through technology. Its People + Possibilities approach recognises that while technology creates possibilities, it is people across government, society and markets who shape these possibilities into meaningful impact for end users. EkStep's work spans sectors such as agriculture, livestock, livelihoods and education, and horizontal areas including Voice for AI, Language for AI and Data for AI. EkStep is also one of the founding partners of the 100 Pathways initiative.

About 100 Pathways

100 Pathways is a global initiative to document and share one hundred diffusion pathways by 2030, so that the next deployer of AI in agriculture, health, education, livelihoods or in any sector starts where the last one ended, not where they began. More details about this initiative can be found here: https://100pathways.com/

A. About The Workshop

A1. Convenors, venue, and date

Frugal AI Pathways: AI for Low-Resource Languages was convened **by Frugal AI Hub** in partnership with the **EkStep Foundation on 3 July 2026 at Judge Business School, University of Cambridge.** The workshop brought together organisations working to advance AI for low-resource languages, with a particular focus on translating lived experience into practical guidance for future adopters.



A2. Why this workshop was convened

AI makes it possible for people who have been excluded by digital and literacy barriers to access information and services through voice — a rural farmer can simply speak into a mobile device in their own language and receive relevant, actionable advice.

But this promise is conditional. It holds only where AI systems can work reliably in the languages people use. More than 7,000 languages are documented worldwide, yet digital resources remain heavily concentrated in a relatively small number of high-resource languages. UNDP notes that AI systems are unavailable or underperform for around 1.2 billion native speakers of low-resource languages. As voice becomes an increasingly important interface for accessing information and services, this gap has direct implications for who can benefit from AI.

This is not a gap that will close on its own. If there is no deliberate effort, the digital language divide will not be a temporary lag; it will be baked into the infrastructure of the information age.

The workshop was convened to address this gap by bringing together practitioners who had successfully deployed language AI in low-resource settings with organisations beginning similar work.

Rather than focusing on technical demonstrations, the workshop sought to expose the realities of deploying AI use cases at societal scale: understanding the problem being addressed, making technical and architectural decisions, working with institutions and communities, and sustaining AI capabilities beyond initial deployment. The objective was to



Krishna Jampana

"...critically also what is a real world use case that this language initiative should be attached to because a language effort without a real use case is just like another data set on the wall..."

enable adopters to learn from real deployment experience while developing approaches suited to their own contexts.

A3. Who was in the room

The workshop brought together four groups of participants.

- **Practitioners** who had successfully deployed language & voice AI use cases for low-resource communities, including Mitesh Khapra (AI4Bharat / IIT Madras), presenting on building language AI across Indic languages, Dr. Sarbani Banerjee Belur and Dr. Arjuna Sathiaseelan (IIIT Dharwad / Saving Voices), presenting the Soliga language initiative; Kato Steven Mubiru (Crane AI Labs), presenting on extending Crane AI from Luganda into Swahili and Kenya; and Santosh Kevlani (Head, Voice and Language AI, EkStep Foundation), presenting on Voice AI for the Bhili language.
- **Next adopters** actively developing voice AI use cases for low-resource languages, including teams working on the Soliga language in India and East African language initiatives, who brought live implementation challenges to the workshop.
- **Researchers and technologists** with expertise in language AI, speech technologies and AI deployment.
- **Policy professionals, academics and ecosystem practitioners,** who contributed by challenging assumptions, identifying blind spots, sharing relevant experience and connecting adopters with additional expertise and networks.



A4. How the day was structured

The workshop moved deliberately from listening to doing, across three phases:

- **Practitioner sessions** — 30-minute presentations, each followed by 10 minutes of Q&A, in which practitioners told the full story of how they built and sustained their work: the technical decisions, institutional navigation, community relationships, and human capability involved.

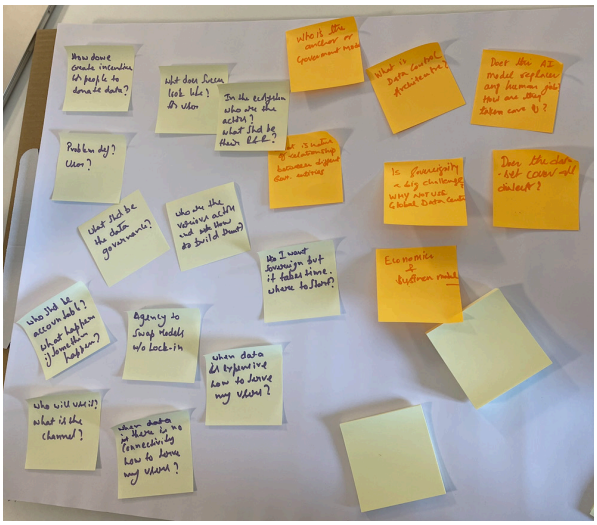
- **Problem statements —** shorter sessions in which next adopters articulated the problem they were trying to solve: who specifically was affected, what success would look like for that person, and why the problem had not been solved already.

- **Working groups —** breakout groups, each anchored by one next adopter, in which the room worked through to come up with a set of questions which need to be addressed to scale their use cases leveraging the insights from the practitioner's sessions.

A5. From the initial design to the final design

The workshop was not originally designed around question discovery. Its initial design assumed that the morning practitioner sessions would be followed by breakout groups working through a predefined set of questions. The intention was to help each next adopter develop an implementation approach by applying the lessons shared by practitioners to a structured framework of questions spanning technical architecture, institutions, ecosystem, workforce and operating model.

During the final planning period, however, the organisers reconsidered this approach. A predefined question set assumed that the organisers already knew the most important implementation questions. It left limited opportunity for participants from technology, government, academia, philanthropy and community organisations to contribute perspectives from their own domains, challenge underlying assumptions or surface implementation questions that had not been anticipated.



There was a second consideration. The adopter teams themselves were still at an early stage of their adoption journeys. Their immediate challenge was not to complete a deployment pathway but to understand what questions needed to be answered before such a pathway could be credibly designed. Asking them to work through a predetermined framework risked narrowing the discussion before the problem had been fully understood.



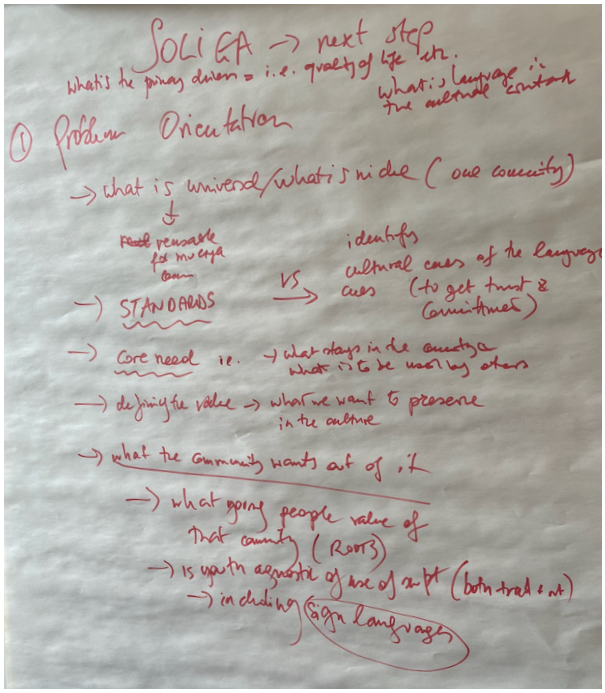
Jaideep Prabhu

"AI inclusion will depend on whether people can access and benefit from AI in their own languages. Frugal AI helps make that possible through practical, locally grounded approaches."

The workshop design was therefore revised. Instead of beginning with predefined questions, participants were first asked to generate and prioritise the implementation questions they believed each adopter would need to answer. Only after agreeing on those questions did the groups begin working towards concrete answers. The organisers suggested broad dimensions to stimulate discussion, but deliberately left the questions themselves to emerge from the collective experience of the room.

A6. What the workshop produced

The workshop produced a set of questions grounded in the Soliga and Swahili cases, a comparison with the Bhili experience, and a clearer account of what each case still needed before a pathway could be agreed. One recurring gap concerned the locally accountable institution that would carry the service after the initial technical or research work. Bhili had an institutional home through the district administration and the Department of Agriculture, Maharashtra. Soliga and Swahili had important actors involved, while the continuing service owner remained open.



The day also produced a more precise agenda for the next phase. Each case needs to define its user and purpose, decide what capability is sufficient for that use, identify the appropriate local anchor, agree the roles of technical and community partners, and select a bounded commitment that the responsible actors can approve. At this stage, the workshop output consists of question sets, working observations and conditions that need to be resolved as the adopters move towards deploying their use cases.



These outputs provide an initial contribution to the 100 AI Diffusion Pathways initiative. As further adopters work through similar deployment challenges, the questions and lessons can be tested, refined and added to the diffusion pathways. Recurring decisions and roles across several such pathways may, over time, provide the evidence for a more developed playbook.

B. Working Observations

B1. Lived experience carries knowledge that case studies often omit

The workshop reinforced that practitioners who have deployed AI use cases at societal scale accumulate a kind of implementation know-how that is rarely captured in conventional case studies. Case studies typically describe what happened in one particular deployment: its context, actions, and results. What they often miss are the assumptions that proved wrong, the informal relationships that unlocked progress, the decisions that had to be reversed, and the reasoning behind those changes.

The Bhili deployment illustrated this clearly. Santosh Kevlani, described how the initial effort began before the use case and delivery ecosystem had been fully defined. Early speech data was collected using smartphones in relatively clean recording environments, only to discover later that many farmers would access the service through feature phones transmitting at 8 kHz. The team subsequently returned to the field to collect telephony speech, extend validation, improve vocabulary coverage and reduce latency. As Santosh reflected, "Not solved in one go. Solved by failing well." The value of this experience lies not simply in the final system, but in understanding where the original assumptions broke down, why they changed, and how the team adapted.

The workshop suggests that AI diffusion depends not simply on documenting deployments, but on documenting the lived experience and practical know-how that practitioners accumulate through deployment. This includes the assumptions that proved wrong, the course corrections that followed, the informal relationships that unlocked progress, and the reasoning behind key decisions. Capturing this know-how allows future adopters to benefit from experience without having to rediscover the same lessons for themselves.

B2. Context determines what working voice AI means

The workshop reinforced that **Voice & Language AI should be understood as a horizontal capability, but not a generic one**. Like identity, payments or maps, it is intended to be reused across many applications while allowing each vertical solution to focus on its own domain expertise. However, the

workshop also demonstrated that horizontal capabilities do not evolve in isolation. They evolve by solving real problems emerging from successive deployments, each exposing new linguistic, technical and operational requirements that strengthen the capability for future adopters.

The Bhili deployment illustrates this well. Supporting MahaVISTAAR for Bhili took far more than adding another language to a list. It meant extending the language capability with new speech data, vocabulary, and conversational patterns rooted in how these farmers actually talk. Words for crops, trees, storage practices, local farming methods — these weren't edge cases to handle. They were the vocabulary through which farmers explained their reality. The voice capability got better at understanding how people actually talk about that expertise and it evolved through the demands of the use case rather than independently of it.



The workshop also demonstrated that context determines what "working" Voice AI actually means. Bhili had to operate across feature-phone calls, messaging applications, mobile apps and extension workers. Soliga needed to function offline across dispersed hamlets with limited power and connectivity. Although both deployments sought to enable Voice AI, the implementation requirements - and therefore **the definition of success** - **were fundamentally different.**

This became equally apparent when participants discussed technical performance. AI4Bharat's presentation on large-scale speech datasets reflected the objective of building broadly reusable language capability across many Indic languages.

By contrast, the Bhili deployment focused on delivering a trusted agricultural advisory service for one community and continued improving through field deployment.

The Soliga team demonstrated that a community-valued offline prototype could be built with as little as five hours of speech data, accepting a higher word error rate because the objective was community engagement rather than benchmark performance.



Dr. Arjuna Sathiaseelan

"This entire Soliga project was done with just five hours of data. We might have got thirty-five per cent error rates, which is fine, because the community was very happy with what they heard."

B3. Technology is not the binding constraint

One of the clearest observations from the workshop was that technology is increasingly the smaller part of deploying Voice & Language AI. Open models, related-language transfer, shared infrastructure and public platforms have substantially lowered the effort required to build and adapt language models. The challenge is no longer whether a model can be trained, but whether it can be deployed as a trusted and sustained service enabling specific use cases.

The Bhili deployment illustrates this shift. The team built on AI4Bharat's language models, leveraged Marathi as a related language, used Bhashini for public hosting and integrated the capability into the existing MahaVISTAAR platform. Santosh Kevlani noted that adapting and fine-tuning the model took roughly 24–36 hours. The overwhelming majority of the effort, however, lay elsewhere: collecting representative speech data, validating the system with farmers, integrating telephony, expanding vocabulary, coordinating government departments and implementation partners, and continuously improving the service based on field use.



Mitesh Khapra
Professor at Indian Institute of Technology, Madras
Lead Bodhan.AI

"Once foundational language resources exist, models can often be adapted relatively quickly...The challenge is building the foundations which everyone ignores."

Mitesh Khapra described the same phenomenon from the perspective of language foundations. Once foundational language resources exist, models can often be adapted relatively quickly. Building those foundations—representative datasets, quality assurance processes, community participation and evaluation benchmarks—takes considerably longer. As he observed, **"The challenge is building the foundations which everyone ignores."**

The workshop suggests that the binding constraints to scaling AI for societal use cases now lie primarily outside the model itself. They lie in the quality of data, the legitimacy of institutions, the trust of communities, the accountability of service providers, and the operating models needed to sustain services over time. These are precisely the dimensions where implementation experience and practitioner judgement become indispensable.

B4. Start with the person, their problem and purpose

The workshop consistently demonstrated that societal scale AI deployments begin with a clearly defined person facing a clearly defined problem. Technology follows from that understanding; it does not precede it. The first design decision is not the model, the platform or the interface, but identifying **who** the service is for, **what** problem they are trying to solve, and **why** that problem matters.

The Bhili deployment exemplified this approach. It began with specific users—a farmer seeking trusted agricultural advice, a low-literacy worker using a keypad phone, and an extension worker more comfortable communicating in Bhili than in the official language. These personas immediately shaped the service itself: the trusted source of information, the preferred interaction channel, the language, the device, and the overall service design all followed from understanding the person and the task they needed to accomplish.

By contrast, the Soliga and Swahili discussions demonstrated what happens when the person and purpose remain open.



Dr. Sarbani Banerjee Belur
Asst. Professor, IIIT Dharwad, Co-Founder, Saving Voices

"Global AI speech models are designed around assumptions of heavy compute and stable high-speed internet, which are non-existent in forest hamlets."

The Soliga case began from a different position. IIIT Dharwad researchers had already been working on Soliga speech technology for several years, with published work on speech data collection and text-to-speech models by 2023. The current phase brings that earlier technical work together with Saving Voices, Dr. Sarbani Banerjee Belur, Dr. Arjuna Sathiaseelan and the wider IIIT Dharwad team. By the time of the workshop, the case therefore had technical experience and established community engagement behind it, while the purpose of the next stage remained open.



The Soliga initiative surfaced multiple legitimate aspirations: preserving songs and oral traditions, documenting traditional farming practices, supporting intergenerational language transfer, and potentially extending to traditional medicine and biodiversity knowledge. The discussions also raised sign language and showed that elders, diaspora families and younger members of the community could have different priorities. Each represented a different use case requiring different data, permissions, interfaces, institutional arrangements and long-term owners.

Similarly, the Swahili discussions repeatedly returned to the question: who is the primary user? Large commercial farmers, smallholders, extension workers and service providers each face different problems, use different channels and depend on different sources of trust. Until that question was answered, participants found it difficult to assess technology choices, institutional arrangements or service design.

The workshop suggests that the first step in designing societal AI is to define the persona, the problem and the purpose. User, task, interaction channel and trusted source are not independent design decisions—they emerge together. Once these are clearly understood, decisions about models, platforms and interfaces become considerably easier to evaluate.

B5. Community participation shapes the design and governance of AI systems

The workshop demonstrated that communities contribute far more than speech recordings or validation data. They influence how an AI capability is designed, deployed and governed. Their participation shapes interaction models, establishes trust, determines acceptable uses of community knowledge and helps define the governance arrangements needed for long-term sustainability.

The Bhili deployment illustrated this through continuous engagement with local officials and community members, who helped resolve questions around dialect, script, recruitment and trust. Santosh Kevlani also argued for a community governance mechanism capable of overseeing data collection,

consent, and deployment. He also emphasised that the community governance council should take decisions on which material needs restriction vs material that can be openly shared.

The Soliga discussions showed that community participation can also reshape the design of the technology itself. Community members wanted to hear their own recorded voices broadcast back into their hamlets and wanted to interact directly with the Raspberry Pi devices. These requests did not emerge from technical requirements; they emerged from how the community wanted to experience and engage with the technology.



Kato Steven Mubiru
Co-Founder and CEO, Crane AI Labs

"Our journey from Luganda to Swahili has reinforced a simple lesson: building language AI is only the beginning. The real work is making it useful, trusted and sustainable in the lives of the people it is meant to serve.."

The discussions also highlighted that different groups within the community - elders, younger speakers and diaspora families - held different expectations about what the technology should preserve and how it should be used.

The workshop suggests that community participation should be viewed as a design and governance function rather than simply a source of data. It determines how AI systems are experienced, what knowledge should remain under community control, and the governance arrangements needed to sustain trust over time.

B6. Scaling requires Institutional Anchors

The workshop highlighted an important distinction between the organisation that **orchestrates** a deployment and the institution that ultimately **owns and sustains** the service. Orchestration is about bringing together capabilities, coordinating actors and enabling implementation.

It does not replace the locally legitimate institution that remains accountable for the service once temporary projects, partners or funding have come to an end.

The Bhili deployment illustrates this distinction. It depended on a network of organisations including the District Administration, State Agriculture Department, AI4Bharat, Bhashini, Karya, community contributors, linguistic experts and EkStep Foundation.

Each played a distinct role within the deployment ecosystem. EkStep acted as the orchestrator—bringing together the ecosystem, connecting language

AI capabilities to the MahaVISTAAR farmer advisory service, coordinating implementation across partners, and ensuring the programme progressed from language development to deployment.

The District Administration provided the mandate, mobilisation and funding required to initiate the effort, while the State Agriculture Department and MahaVISTAAR provided the institutional home through which the service was deployed, operated and sustained.



Santosh Kevlani

"It cannot be EkStep. EkStep is just an orchestrator and supporter. It has to be someone like an institute, a government or an NGO who cares for that language and says, 'I care for the language.' Other stakeholders will keep coming and going, but I will be there."

The Soliga and Swahili discussions reinforced the same principle. Soliga has community demand, a local champion, a long-standing NGO and a committed research team, yet the institution that will ultimately own and sustain the service remains undefined. Soliga discussions were also about creating the environment and ecosystem for self-sufficiency and self-organisation. Swahili has technical capability and potential deployment pathways, but participants repeatedly returned to the question of which locally accountable institution would validate information, assume responsibility for the service and sustain it over time.

The workshop suggests that societal scale AI use case deployments require two complementary forms of leadership. An orchestrator assembles capabilities, convenes the right actors and helps translate ideas into implementation. An institutional anchor provides the legitimacy, decision-making authority, resources and continuity needed to own and sustain the service over time. While the public agricultural system fulfilled this role in the Bhili deployment, the workshop concluded that the appropriate institutional anchor will vary across contexts.

A government department, local authority, cooperative, NGO, university or another locally legitimate service owner can all fulfill this role, provided they remain accountable for the service long after the initial deployment effort has ended.



Elizabeth Osta

"The starting point isn't the technology, it's the purpose. Whether we are preserving cultural knowledge or delivering critical advice determines the entire pathway to success."

B7. Ownership, accountability and operating economics are not after thoughts

The workshop demonstrated that scaling AI use cases extends beyond developing a technically capable language model. Once AI becomes part of a public or community service, questions of ownership, accountability and long-term operating economics become integral to the design of that service rather than considerations to be addressed after implementation. These decisions determine whether an AI capability becomes a trusted, sustainable service or remains a technical demonstration/proof of concept.

The **MahaVISTAAR** farmer advisory service illustrates this principle. Farmers receive advice which originates from authoritative government sources, validated by the State Agriculture Department and extension workers. This operating model allows the underlying Bhili language capability to evolve while ensuring that accountability for the advice remains with the institution responsible for delivering the public service.



Santosh Kevlani

"Who is going to take responsibility if an advisory to a farmer goes wrong? If it says spray fifty grams and the answer is five hundred grams, and the crop is lost, who is accountable for that?"

The workshop also highlighted that the operating economics of low-resource language AI do not naturally align with commercial incentives. As speaker populations become smaller, field collection becomes more difficult and formal language resources become scarcer, the unit cost of creating representative language resources increases. None of the deployments presented identified a direct-to-user commercial model capable of sustaining the full cost of service delivery. Long-term viability therefore depends on identifying, from the outset, who funds the service, who operates it, and who remains accountable for it over time.

The workshop suggests that **ownership, accountability and operating economics should not be an after thought.**

Questions such as

- Who owns the service?
- Who is accountable when something goes wrong?
- Who funds ongoing operations?
- Who continues after the initial deployment?

are not implementation details - they are design decisions that determine whether AI can be deployed responsibly and sustained at societal scale.



C. Open Questions

Several questions became sharper during the day without reaching a common answer. Some arose during the practitioner presentations and technical discussions, while others emerged directly from the Soliga and Swahili working groups. Taken together, they form a research and capacity-building agenda for the next phase.

C1. What level of data and technical capability is sufficient for the intended use?

The cases involved different ambitions and therefore different requirements. AI4Bharat described the data needed to build broadly useful and benchmarked speech capability, while Bhili focused on an agricultural advisory service and Soliga demonstrated a narrower community-defined use with a much smaller dataset. The discussions also raised questions about latency, connectivity, devices and whether a service requires speech alone or a combination of speech, text and visual interfaces.

Related-language transfer adds another variable. Bhili benefited from Marathi and an existing AI4Bharat base, while AI4Bharat cautioned that target-language data remains necessary. Future pathways need to record how the intended task, accepted error, user consequence, delivery channel and related-language resources affect the amount of data and technical work required.

C2. How should quality and trust be established where conventional language resources are limited?

Validation becomes more difficult where a language has no agreed script, dictionary, teaching profession or widely recognised linguistic authority. **The workshop considered the possible roles of elders, subject experts, community validators and speakers of related languages**, while questions remained about how disagreements should be resolved and how performance should be assessed consistently.

The same issue extends beyond language accuracy. In an agricultural advisory service, users also need to know where an answer comes from and which institution stands behind it.

Future work therefore needs to examine language validation, service accuracy, source provenance and user trust together, particularly where an incorrect answer could have consequences for livelihoods or access to services.

C3. Who defines the purpose of the capability, and whose priorities count?

The Soliga discussion surfaced several legitimate aspirations, including preserving songs and oral traditions, documenting farming practices, supporting intergenerational language use, and exploring applications connected to traditional medicine and biodiversity knowledge. Participants also raised sign language and different expectations among elders, younger community members and diaspora families. Each purpose could lead to a different use case, with different data, interfaces, permissions and institutional arrangements.

The Swahili group faced a related question around the primary user. Smallholder farmers, larger commercial farmers, extension workers and service providers have different needs, channels and sources of trust.

Future pathways need a way to identify whose problem is being addressed, what outcome is being sought and how competing priorities are resolved before technical and service choices are fixed.

C4. What does community ownership mean in practice?

Participants used community ownership to describe several different arrangements: informed consent, payment for contribution, local storage, access rights, participation in governance, control of hardware and infrastructure, and rights over future commercial or public use. The Soliga discussion also raised questions about what should be returned to the community, including recordings, tools and access to the technology itself.



Ownership also affects which knowledge can travel. Open resources may support wider reuse, while cultural, medicinal, ecological or ritual knowledge may require restricted access.

Future work needs to distinguish the permissions attached to software, speech recordings and community knowledge, identify who has authority to grant those permissions, and establish how they may change over time.

C5. Which institution should anchor the deployment and remain accountable?

Both next-adopter discussions returned to the institution that would remain once the initial research, technical support or grant funding had ended. Such an institution needs sufficient local legitimacy, authority and continuity to make decisions, hold resources and accept responsibility for the service.

The appropriate form may differ by case. Soliga may require a community organisation, public body, university, NGO or another institution connected to a clearly defined use. In the Swahili case, participants discussed who could play this role among Crane AI Labs, Hello Tractor, agricultural institutions, extension services and national or county government.

Future pathways need to establish who owns the service, who validates the information it provides, who carries responsibility when something goes wrong, and what happens if a technical partner leaves.

C6. What operating model can sustain the service after the initial project?

The breakout groups identified continuing costs that remain after an initial model or demonstration has been developed, including data collection, hosting, telephony, maintenance, human validation and user support. The Swahili discussion also asked who pays whom across the value chain, how high connectivity and telephony costs affect delivery, and how a new service should work alongside existing publicly funded agricultural systems.

Future pathways need to identify the buyer or budget owner, the continuing operator and the payment flows while the service is being designed. They also need to test which combination of public funding, philanthropy, commercial revenue, cooperative arrangements or existing service budgets can sustain the capability once temporary funding ends.

C7. What infrastructure and local capability should remain in place?

The cases raised different questions about where technology should run and who should be able to operate it. Soliga explored offline use on low-cost hardware in communities with limited connectivity. The Swahili discussion considered local and international cloud infrastructure, data residency, privacy, feature-phone access, messaging applications and other delivery channels.



Serish Gandikota

"The hardest part of AI is increasingly not building the capability, but making it scale and endure. That is why implementation know-how matters as much as the technology itself."

Future work needs to examine which technical components can be shared across settings and which skills, infrastructure and operating responsibilities need to remain locally available.

D. Next Steps

The Cambridge workshop demonstrated that organisations learn faster from practitioners who have successfully navigated similar challenges. The next phase is to operationalise this learning by bringing together a community of global organisations actively deploying Voice & Language AI use cases and systematically connecting them with practitioners who have already solved comparable problems.



So the immediate priority is a **discovery-first** approach. A series of virtual **Adopter Discovery Sessions** will identify organisations working on AI for low resource languages, understand their use cases, stage of adoption and challenges, and create a growing community of adopters to inform future in-person convenings.

Rather than bringing organisations together around a broad theme, future convenings will curate adopters facing similar challenges alongside practitioners with relevant lived experience. The objective is to help adopters learn from practitioners who have successfully navigated similar challenges, adapt those learnings to their own context, commit to clear next steps, and contribute their own experience back to the community as they progress.

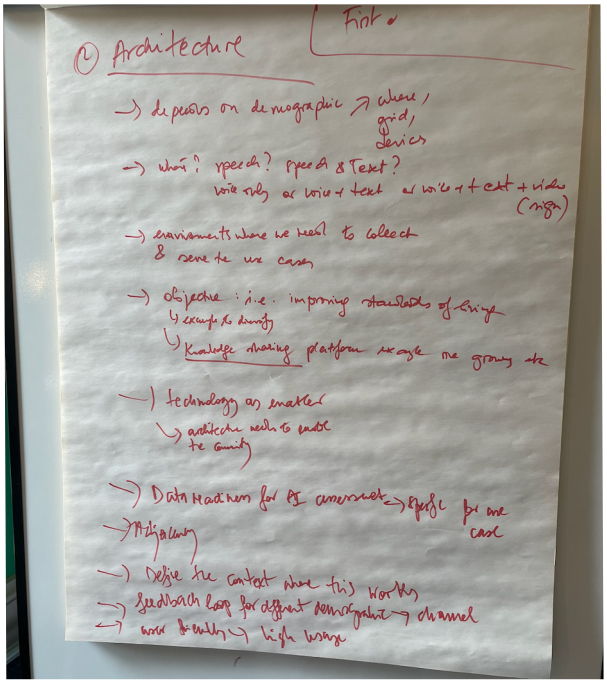
Over time, this creates a self-reinforcing diffusion system in which each adopter benefits from the lived experience and practical know-how of previous adopters while strengthening the knowledge available for those that follow. The result is a growing library of diffusion pathways capturing the know-how of scaling Voice and Language AI use cases that enables every new adopter to move further and faster than the last.

Explore the pathways:
https://cube.100pathways.com/explore

E. Source And Verification Note

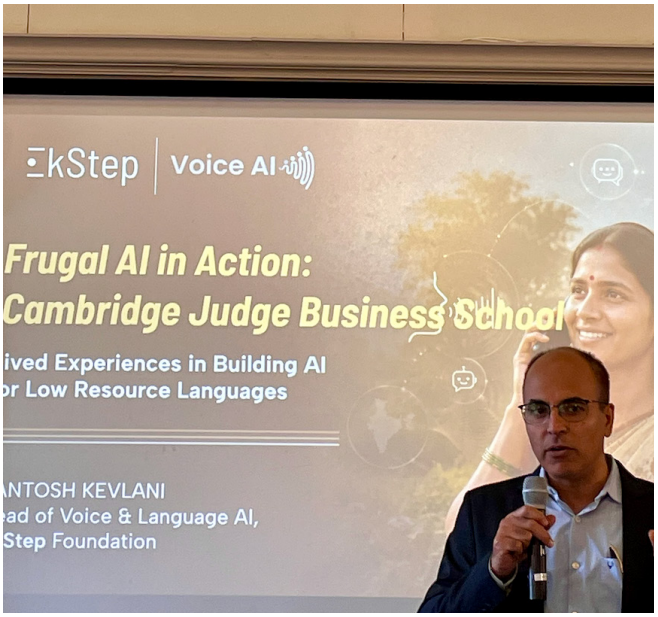
This report is a synthesis of the presentations, working-group discussions and deployment experiences shared during the Frugal AI Pathways: AI for Low-Resource Languages workshop held at Cambridge Judge Business School on 3 July 2026.

Statements relating to the Bhili, Soliga, Swahili and AI4Bharat experiences are based primarily on practitioner presentations and discussions during the workshop. Quantitative examples — including data volumes, model adaptation times, error rates and deployment characteristics — are reported as presented by the respective practitioners and should be understood as context-specific observations rather than independently benchmarked comparisons.



External contextual information used in the report should be attributed to the relevant public source. The global language count draws on Ethnologue, while discussion of the digital under-representation and AI performance of low-resource languages draws on UNDP.

The working observations and open questions are intended as propositions for further testing as additional deployment pathways are documented, rather than as universal conclusions.


kStep | Voice AI
Frugal AI in Action:
Cambridge Judge Business School
ived Experiences in Building AI
or Low Resource Languages
NTOSH KEVLANI
ead of Voice & Language AI,
Step Foundation




UNIVERSITY OF CAMBRIDGE
Judge Business School






CAMBRIDGE JUDGE BUSINESS SCHOOL
Frugal AI Hub
An Overvie
PREPARED FOR
Workshop on Frugal AI Pathways –
AI for Low-Resource Languages
3rd July 2026

Frugal AI Hub at Cambridge Judge Business School:

https://frugalai.org/

EkStep Foundation :

https://ekstep.org/

100 AI Diffusion Pathways :

https://100pathways.com/

Publication date :

October 2026

FRUGAL AI HUB at UNIVERSITY OF CAMBRIDGE Judge Business School

EkStep