

Low-Resource Language AI Adoption Package

India Version

EkStep Voice AI

Copyright & open access license



© 2026 EkStep Foundation. Some rights reserved.

This document draws on EkStep Foundation's experience in building and deploying Voice AI solutions through Project Astitva in Nandurbar and other initiatives. Individuals and organisations listed under "Acknowledgements and Inputs Invited" provided feedback, source material, or field documentation that this package draws on. Their inclusion in that list reflects their contribution to the underlying work and does not constitute authorship, co-ownership, or licensing of any content under the terms below.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits sharing and adaptation for any purpose, including commercially, provided that appropriate credit is given.

Full license: https://creativecommons.org/licenses/by/4.0/

Suggested attribution and credit: "Low-Resource Language AI Adoption Package for India," EkStep Foundation.

Acknowledgements and inputs invited

This package draws directly on documented field experience, workshop transcripts, and technical guidance from the following individuals and institutions. Further input is sought from a wider ecosystem to strengthen the sections that remain thin, and to validate the sections drawn from a single source or a single district's experience.

The package brings together the experiences and work of following individuals and institutions in no particular order of importance.

Santosh Kevlani - EkStep Foundation
Mitesh M. Khapra - AI4Bharat, IIT Madras
Parimal Singh - IAS 2004,Secretary, Agriculture Department, Maharashtra
Dr Mittali Sethi- IAS (2017), District Collector,
Champion, Project Astitva (Nandurbar, Bhili), Maharashtra
PxD (Precision Development) and team - Steering committee role, MahaPMU workshop
BalaKrishnan A, Vanashree Nair - Reviewer from EkStep & COSS
AI4Bharat, IIT Madras - Technical partner, model-building, benchmarking (Shoonya, IndicTrans2, IndicASR, IndicTTS)
Amitabh Nag - CEO, DIBD (BHASHINI)
Bhashini / Digital India NLTM - National hosting platform, ULCA open data repository
Karya (karya.in) - Data collection platform, Nandurbar reference implementation
State Tribal Research Institutes - AdiVanni-Ministry of Tribal Affairs,Maharashtra (Pune), Gujarat (Vadodara), Madhya Pradesh (Bhopal), Karnataka (Mysuru), Jharkhand (Ranchi), Odisha (Bhubaneswar) etc
MahaVISTAAR / Maharashtra State Agriculture Department, POCRA - Deployment reference
Government of NCT of Delhi and other district/state administration partners - As applicable to specific state adoptions
Gates Foundation - Funder audience and reference point for diligence framing

This is a working document, and we welcome comments, corrections, and additional named contributors from across the ecosystem to help shape the version that follows.

Opening note

This package exists because a language almost went missing from the story of what India's AI infrastructure can do. It stayed on the map because a District Collector, a handful of community elders, a technical partner, and people who had never recorded their own voice before decided it would.

> Dehwali Bhili is a tribal language spoken in parts of states of Maharashtra, Gujarat, Rajasthan and MP in India. It has no standard script of its own. The village life happens in it while Government offices run in Marathi. A woman who only speaks Dehwali Bhili has never filed a grievance in her life, not because she is not capable compared to others, but because the system has simply never been asked to interact with her in her own language. That gap repeats everywhere, language by language, wherever India's digital infrastructure was built for the languages that already had readers, typists, and training data.

What follows is a written account of how that gap closed in one place and how it can help another district, institution, language community to adapt from it. Some of what's here is precise: real thresholds, real costs, a real week-by-week runbook for the first month. Some of it is still being learned, and this document says so plainly wherever that's true.

If you are a District Collector, a Secretary or a program lead ready to take your language forward, this package is designed for you. It tries to give you exactly what you need, at the level of detail you need it, how and where to start, and to be honest with you about where it still falls short.

How to use this guide

What this package is

A working package for any Indian state, district, university, or NGOs with local administrative backing to bring a low-resource or tribal language onto India's language AI infrastructure. It draws on the experience of bringing Bhili language to AI map in Nandurbar, alongside insights from consultations with individuals and institutions working with low-resource languages.

How it's organized

Eight clusters, arranged in the order a decision maker actually moves through the decision, in the sequence of a person's own thinking rather than the sequence a technical pipeline runs:

Cluster	Answers	No.
1. The Case	Why this is needed, who benefits, and what "solved" looks like	05
2. Leadership & Ecosystem	Who the champion is, and who else has to be in the room	09
3. Process and Expected Outcome	How data becomes a deployed model, and what "good enough" means	13
4. Getting Ready	The pre-start checklist, budget, and personal time commitment	20
5. The Champion's Role	What a champion should do, delegate and stay out of.	23
6. Setting Expectations & Milestones	The realistic timeline and where it can slip	27
7. Running & Monitoring	Review cadence, quality control, testing, and the documented failure patterns to watch for	29
8. Closing & Sustaining	Sign-off requirements, what "done" means, and systems for continuity beyond the champion.	33

How to use this package

Start with the sections most relevant to your responsibilities.

A Secretary or senior officer deciding whether to commit, the Summary at the top of each cluster, together with Cluster 1 in full, will give you a clear picture of the commitment in about fifteen minutes.

A newly appointed Nodal Officer should begin with Cluster 5, which provides the week-by-week implementation runbook.

A funder or reviewer assessing the proposal should focus on Clusters 3, 4, and 8. These contain the key figures, benchmarks, and completion criteria needed to evaluate the plan and assess progress.

How cross-references work

A reference like "see Cluster 7.2" points to a specific subsection inside that cluster.

Cluster 1 : The Case

Summary

Hundreds of millions of people in India speak languages that public services, health information, and agricultural advisories do not adequately serve. When essential information is unavailable in a language people understand, their ability to access government services is limited. Voice AI in people's own languages offers a way to bridge this gap. The experience of bringing Bhili into language AI in Nandurbar demonstrates that this work can move from intent to implementation in months.

1.1 — The problem, in plain terms

Government services, health information, agricultural advisories, legal aid, and educational platforms are built overwhelmingly for the dominant scheduled languages. Tribal and indigenous communities are effectively excluded from digital governance as a result. When a farmer cannot query a government scheme in her own language, she either goes without the service or relies on an informal translator who may get it wrong. When a frontline health worker must explain a vaccination protocol in a language the patient barely understands, the stakes are real. When an elder cannot interact with a government portal, she is invisible to the state — an outcome created by infrastructure, not by any choice of hers.

This invisibility compounds in a specific, self-reinforcing way: digital infrastructure gets built where training data already exists, and training data gets built where communities are literate, connected, and economically active. Communities outside those conditions stay outside the infrastructure indefinitely, because there is no commercial case for a private developer to build for them, and the gap persists on its own.



1.2 — Who benefits, and how

The initiative can directly benefit the wider language-speaking community, including the personas represented below.



The farmer who cannot describe her problem to a government advisory line in the language she actually thinks in, and either goes without the service or relies on an intermediary who may get it wrong.



The frontline health worker (e.g., an ASHA worker) who has to explain a protocol to a patient in a language the patient only partly understands, where a mistranslation carries real consequences.



The elder or low-literacy citizen, often a woman, who has never filed a grievance in her life. She has grievances; every tool through which a citizen would normally raise one is simply a tool she cannot access.



Contributors: the person who speaks into a recorder for the first time. Contributors are paid for their time by the Data Collection Agency directly, and on time because this is livelihood work, not a favour done for the cause. The money this effort spends doesn't leave the district for an outside vendor. It comes back to the people-community who speaks the language for their contribution.

A direct field observation from Nandurbar makes the mechanism concrete: village life happens in Bhili, government camps run in Marathi, and the service quietly fails at exactly that gap — most severely for women, who are frequently the least comfortable in the administrative language and, by the same field account, among the most linguistically precise speakers of their own. Even an Agri extension officer with decades of service experience has to switch to the local language deliberately, because it is the only way to hear the whole of a person's problem.

1.3 — Why AI, specifically

Speech requires no literacy, no device fluency, and no familiarity with administrative language. A citizen who cannot write her own name can still describe her problem, ask her question, and understand a response, provided the

system speaks her language. This makes voice AI in the local language a basic inclusion story rather than a convenience story or a technology showcase. Field evidence from the Bhili work reinforces this directly: people do not type in the language, many struggle to read the dominant script, and a text-based interface would have missed them entirely. Voice was the only channel that actually worked.

> This carries a second dimension beyond access. Agricultural practice, traditional medicine, and oral history in these communities live almost entirely in spoken language, carried by speakers who are often elderly. While each recording session contributes to preserving the language, contributors are often motivated by livelihood opportunities and direct recognition.

1.4 — What "solved" looks like

A concrete, time-bound answer, drawn directly from the lived experiences of other similar initiatives:



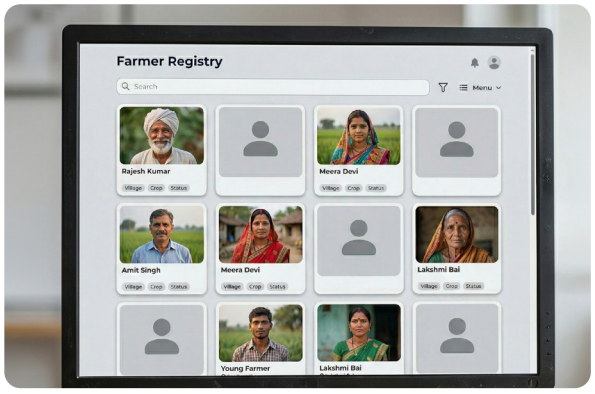
At 3 months:

A first keyword-spotting model is live in a pilot government service. This lightweight model recognises a small, agreed set of 20–30 high-priority spoken words or phrases, such as health service names, government scheme names, crop names, or emergency phrases. Unlike full automatic speech recognition (ASR), which must handle open-ended speech, keyword spotting focuses on a defined vocabulary and therefore requires substantially less training data. Depending on the language, recording conditions, and model approach, an initial prototype may be developed using approximately 200–500 recorded examples per keyword.



At 6 months:

A government department can interact with a live service, such as a crop advisory line, entirely by voice in the target language.The service is functional, though some responses may contain errors.



At 12 months:

The language dataset is publicly available on AIkosh/Bhashini/ULCA for other researchers and developers to build on. The language model is publicly available on Bhashini/Huggingface. The tested real service is now live to a larger public. Another use case is identified for testing. A methodology handbook of dataset collection & model building is published for the next adopter. Evaluation metrics are reported and model quality improvement targets set for the following round.

On quality and evaluation for all three milestones:

This section defines the deliverables at each milestone. Cluster 3 (Process and Expected Outcome) defines the corresponding quality criteria, evaluation metrics (WER, BLEU, and MOS), and benchmark thresholds required for each deliverable to be considered complete.

Metric	Full Name	What It Measures	How It Works	Common Use Case
WER	Word Error Rate	Accuracy of text transcriptions.	Counts substitutions, deletions, and insertions compared to a human reference. Lower the better.	Speech-to-Text (ASR)
BLEU	Bilingual Evaluation Understudy	Similarity between a machine-generated translation and human translations.	Measures overlapping phrases (n-grams). Higher is better.	Machine Translation, Text Generation
MOS	Mean Opinion Score	Human-perceived quality (naturalness, clarity, or fluency).	Humans rate the output on a scale from 1 (bad) to 5 (excellent). Higher is better.	Text-to-Speech (TTS), Audio Quality

Cluster 2 :
Leadership & Ecosystem

Summary

A low-resource language initiative needs more than a sound process. It needs a committed champion who takes personal responsibility for its progress, keeps stakeholders aligned, who becomes the face of the project and provides continuity across the project. Without this leadership, even a well-designed plan risks losing direction or stalling during implementation.

2.1 — Why a champion, and why this person



Why the mission needs a champion at all. A project of this kind spans multiple departments, an outside community's trust, and a multi-year timeline — none of which moves on the strength of a document or a committee alone. It needs one person whose care is genuinely aligned to the outcome, who carries the project's credibility into every room it enters, and who is willing to be identified with its success or failure. That person is what turns a plan into something that actually moves.

Why a district collector or secretary, department of tribal affairs, tends to be the right fit. These posts meet the criteria the role demands, specifically:



Alignment of care -
A genuine stake in the population this serves, held directly rather than delegated



Direct influence -
Day-to-day authority over the departments and machinery the project depends on



Sufficient power to convene and mobilize resources -
The standing to bring cross-department stakeholders and budgets to the table on their own authority



Local community respect -
Legitimacy in the eyes of the language community itself, earned locally rather than supplied by an outside vendor or NGO

The champion may also be someone at a more senior level in the administrative hierarchy. A Chief Secretary, Agriculture Secretary, or equivalent authority can fulfil this role, provided they meet the above four criteria. The role depends on the person's commitment and ability to lead, not their specific designation.

2.2 — Who else needs to come along





The champion cannot carry this alone. Once the champion is in place, the following stakeholders complete the ecosystem needed for the project to actually succeed.

External vendors are marked so it's clear at a glance which roles are contracted in versus which are part of the core governance or community structure.

Stakeholder	Type	One-Line Mandate
Steering Committee	Governance	Cross-department authority; approves budget and direction; Reviews quarterly
Project Management Unit (PMU)	Governance	Split into a Management arm (procurement, vendor management, reporting) and a Quality arm (validation standards, sign-off) — kept independent of each other
District/Local Administration	Governance	Convenes, funds, and champions the effort (see Cluster 5)
District-Level Line Departments	Governance	Hold the ground and provide continuity between phases (Agriculture, Education, Health, Tribal Welfare)
Core Application Owner (e.g., State Agriculture Department Eg:POCRA-MahaVISTAAR Application)	Governance	Deploys the outcome via a real, live service; owns fine-tuning and upkeep
Ecosystem Orchestrator (e.g., PxD / EkStep Foundation-equivalent)	Governance	Leads orchestration across the whole ecosystem — the neutral party with no vendor stake
Linguistic Experts	Community	Validate content and resolve dialect & Language related questions
Community Contributors & Local Community Head/Influencing Members	Community	Create, validate, and own the language; legitimize the effort and mobilize participation — the centre of the ecosystem rather than a stakeholder at the periphery
Data Collection Agency	External Vendor	Supplies the collection platform, dashboards, and field task management — see 2.3 for detail
Technical/Model-Building Partner	External Vendor	Builds and benchmarks the underlying ASR/NMT/TTS models — see 2.3 for detail
Model Hosting Platform	External Vendor	Provides open, sovereign hosting on the national platform — see 2.3 for detail

Convening sequence matters: community leadership and the Core Application Owner should be in the room before a Data Collection vendor is contracted. Contracting collection first and looking for a deployment owner later is a documented failure pattern, and it is exactly what Cluster 4's non-negotiable readiness gate exists to prevent.

2.3 — External technical entities: data collection agency, model builders, model hosters

These three roles are always external, contracted in rather than built internally, and each keeps a distinct scope.



Data Collection Agency — Gather speech and text data at the required volume and diversity.

- Examples: Karya, JoshTalks, Shoonya-integrated vendors, BhashaDaan (volunteer channel)
- In scope: Field dataset collection, Transcription, On-platform task management,Proprietary toolkit for data collection. First-pass data validation
- Out of scope: Final validation sign-off, model-building



Model Builders — Train, benchmark, and maintain the language models

- Examples: AI4Bharat, IIT Madras
- In scope: Model training, benchmarking against shared standards, retraining cycles, Provide feedback about any data collection misses / requirements
- Out of scope: Field collection, community convening, deployment ownership



Model Hosters — Provide the infrastructure where the trained model lives and is served.

- Examples: Bhashini/ State Data centers
- In scope: Sovereign, publicly accessible hosting; API/service availability
- Out of scope: Model quality decisions, data collection, annotation

Cluster 3 :
Process and Expected Outcome

Summary

The work moves through five stages after preparation — data collection, model building, hosting, application integration, and sign-off — each owned by a different party. Three progressively larger outputs are possible depending on how far through those stages a project runs, and three standard, named metrics define "good enough" at each stage.

3.1 — The process: from data to deployment

Six sequential stages, each with a distinct owner (full scope and named examples in Cluster 2.3):



STAGE 0. PREPARATION

Before any data collection begins, conduct secondary research to understand the language, population, dialects, geographic coverage, and key language characteristics. Also map existing datasets and audio sources available through radio stations, Krishi Vigyan Kendras, government institutions, and other open-market/public sources to avoid duplication and identify gaps. : Ensure decisions on the script, dialect and sub-dialects are made, responsibilities are assigned, core team including Nodal Officer is setup, the location and scope of data collection is defined, and the institutes and vendors needed for every subsequent stage are identified and aligned. This stage is covered in Cluster 4 (Getting Ready) and Cluster 5 (the champion's specific role in it) in full detail.



STAGE 1. DATA COLLECTION

Speech and text are collected through a structured field-based process, guided by predefined diversity criteria and domain splits outlined in section 3.4. This approach helps preserve linguistic context, natural usage, and culturally relevant expressions within the collected data. Three quality tracks — collection, checking, and validation — run in parallel through digital dashboards in real time, so problems surface early instead of forcing expensive re-collection at the end. The data collection process should ensure proper consent is taken from the contributors before starting the

process and all relevant compliance related to DPDP Act 2027 should be adhered to.

STAGE 2. MODEL BUILDING

The validated dataset moves through transcription into training of the ASR, NMT, and TTS models. (see Cluster 7.2 for the full review chain). The technical partner benchmarks each model against the shared metrics in 3.3 before it's considered ready to move forward. This stage is model-agnostic by design — the underlying language capability is built to be general-purpose and reusable across future use cases, staying open to any deployment rather than being locked to the first one.

STAGE 3. HOSTING

The trained model moves to open, sovereign, publicly accessible infrastructure (Bhashini/State DCs), rather than staying locked inside any single vendor's platform. This is what allows a second department, or a different state entirely, to build on the same underlying language capability later without renegotiating access.

STAGE 4. APPLICATION INTEGRATION

The model is integrated into an existing service that already reaches its intended users, such as an agricultural IVR, a frontline health worker's tool, or a public helpline. Real-world use reveals gaps, such as misrecognized words, missing domain-specific vocabulary, dialect variations, the model does not yet handle. These findings are documented and used to guide the next round of model refinement.

Ownership is shared by design: a common foundation model supports general use cases, while each department deploying the model funds and manages the domain-specific adaptations required for its services.

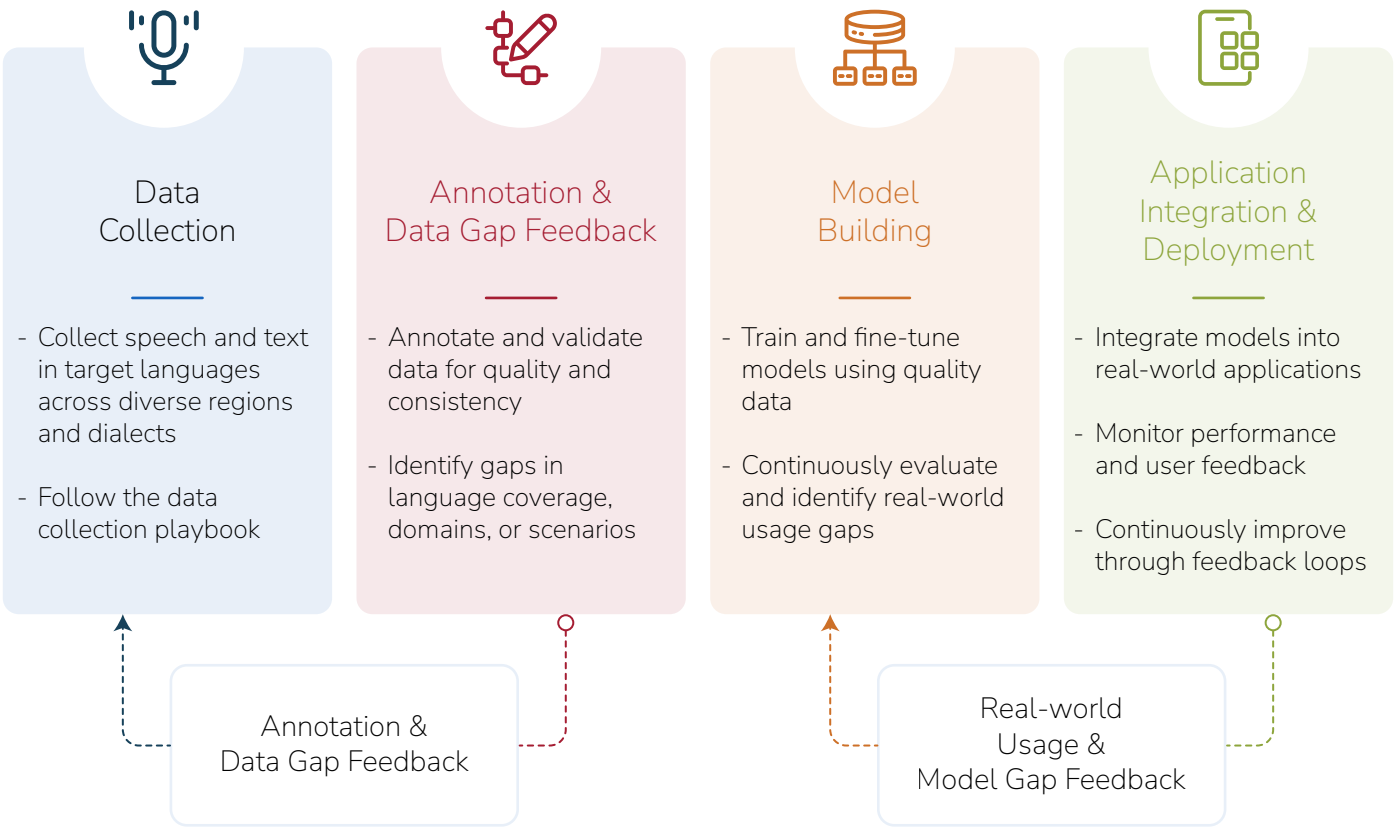
STAGE 5. TESTING AND FIELD VALIDATION

A dedicated testing phase brings together field testers, language experts, and engineering teams. Field testers are first briefed on the testing protocol and do not need to understand the underlying model or technical architecture. Using simple, everyday questions, they assess whether the system understands their speech, uses natural and locally appropriate language, and responds clearly. Language and engineering teams separately evaluate linguistic and technical performance and analyse issues identified during field testing. Findings may trigger data, model, or application refinements, followed by another testing cycle before final acceptance.

STAGE 6. SIGN-OFF

Once the model is live, usable, and has passed use-case acceptance testing (Cluster 7.5), the project formally signs off against the gates in Cluster 8.1. Sign-off includes releasing both the dataset and the model to the open commons (Bhashini/ULCA, under CC BY 4.0 or equivalent) — a formal step for both, rather than an informal afterthought for either. Full sign-off mechanics, including the two-cycle completion threshold, are covered in Cluster 8; this stage exists here to keep the process sequence complete.

The process runs as a loop with adjacent feedback, in a circuit rather than a straight line:



Continuous Improvement Loop

Feedback from each stage drives data quality, model performance, and real-world impact.

Model Building feeds back directly to Data Collection when transcription surfaces gaps in the underlying data. Application Integration feeds back directly to Model Building when real usage surfaces gaps in the model. Issues surfaced in deployment are expected to route through Model Building first, keeping each stage's feedback local to its immediate neighbor rather than skipping ahead.

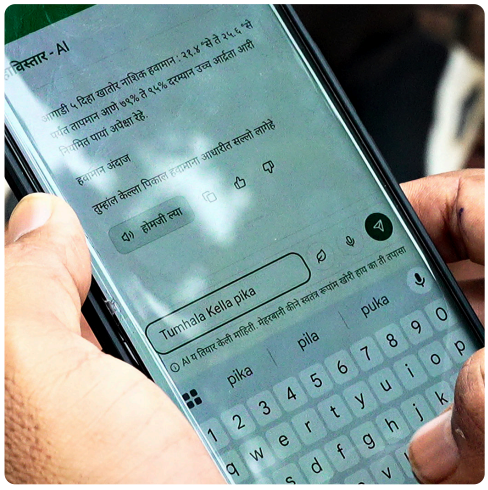
3.2 — The three output tiers

Tier	What Exists at the End	Typically Corresponds To
Dataset Collection	A validated, usable speech/text corpus for the language, ready for someone to train on (requirement mentioned in section 3.4)	End of Stage 1 (Data Collection)
Model Building	The above, plus trained ASR/NMT/TTS models with reported evaluation scores	End of Stage 2 (Model Building)
Live Deployment Integration	The above, integrated into a real government service that citizens actually use, with a feedback loop running	End of Stage 4 (Application Integration)
Signed Off and Open	The above, formally signed off against Cluster 8's gates, with the dataset and model released to the open commons	End of Stage 5 (Sign-off)

The project tier should be clearly defined at the beginning. Cluster 1.4 explains what each tier is expected to achieve at 3, 6, and 12 months. It is important to distinguish between completing a dataset and building a working, deployable model. Treating these as the same milestone can create unrealistic expectations for funders and government partners.

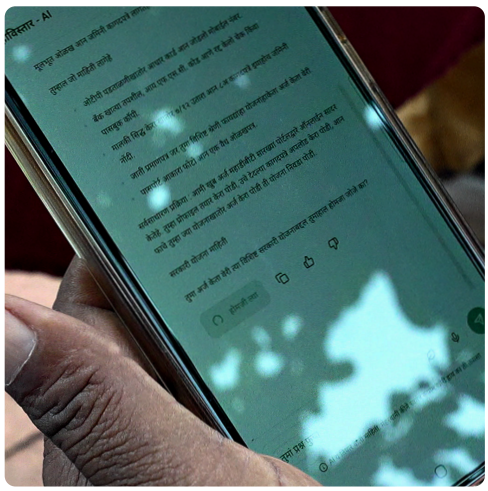
3.3 — Benchmarks that define "Good Enough"

Three standard metrics, one per core AI capability, with real numeric thresholds:



ASR (Automatic Speech Recognition) →Word Error Rate (WER) — the percentage of words the AI transcribes incorrectly.

- Below 20% = generally usable for government service applications
- Below 10% = production-grade
- Early tribal language models typically achieve 30–50% WER — useful for keyword spotting and simple commands, with open-ended transcription still a stage away



NMT (Translation) →Bilingual Evaluation Understudy (BLEU) Score — how closely the AI's translation matches human reference translations, on a 0–100 scale.

- Above 20 = usable translation for domain-specific content
- Above 35 = approaches human-level fluency for constrained domains
- First-generation models for low-resource Indian languages typically score 10–25



TTS (Voice Output) →Mean Opinion Score (MOS) — human listeners rate the naturalness of synthesized speech, 1–5 scale.

- 3.5 or above = generally acceptable for public service use
- 4.0–4.5 = studio-quality (typical for well-resourced scheduled languages)
- Community TTS models for tribal languages typically achieve 2.5–3.5 at first iteration

A system that performs well on these metrics in controlled conditions may still underperform in real-world dialect or noise conditions, so field testing stays a standing supplement. Testing methodology and use-case acceptance criteria are in Cluster 7.5.

3.4 — Dataset collection requirements

Below table gives the estimate of dataset collection required with different quality outputs. A realistic starting point lies somewhere between “Must Have” and “Good to have”

Dataset targets
What you need to collect per language

	Must Have	Good to Have	Great to Have
Speech data	100 hrs	500 hrs	1,000 hrs+
Studio speech	10 hrs	40 hrs	80 hrs+
Translation Pairs	100K	500K	1M+
Dictionary Entries	20K	50K	100K
Quality (tentative)*	Low (60–70%)	Average (75–85%)	Good (>90%)
Cost (tentative) INR	40L–50L	140–180L	275–350L
Timeline (tentative)	6–8 months	10–15 months	18–24 months

* Quality – % of users rating the overall call experience as pass (binary test). In Testing scale (1–5 with 5 being best). Must Have (3–3.5). Good to have (4–4.5). Good (4.7+).

Data requirement Source: AI4Bharat DMU framework — the destination for a fully production-ready system.

Data collection should also follow a certain break up of who to collect from and in which proportion. The below table gives the details of the breakup.

Age Mix	
6–18 years	20%
18–45 years	40%
45–60 years	30%
60+ years	10%

Speech Style	
Read	10%
Spontaneous*	45%
Conversational**	45%

Domain Split	
Generic	70%
Others	30%

* Spontaneous are more impromptu speaking on a given topic or subject. E.g. Sometimes speakers are shown a picture and asked to speak about it.

** Conversational speech is two way communication with a specific objective. E.g. Mimicking calling to a government agriculture officer and requesting a visit to the farm to monitor the crop damage due to floods.

It is further recommended that 50% of audio speech is captured through telephony channels to represent the real world scenario of voice model/bot being used largely on telephony lines.

The above breakup should also be equally representative of both the genders.

For a tribal language with zero prior digital resources (E.g. the Bhili starting condition), the practical minimum for a first functional model is about 300 hours of Audio data along with 200K translation pairs.

Cluster 4 :
Getting Ready

Summary

Before committing, a real institutional owner and a real deployment use case need to exist, and seven concrete items need to be true or in motion. Personnel needs are well defined; budget and personal time commitment are the two areas still being sharpened as real experience accumulates.

4.1 — Pre-start checklist

- Institutional anchor committed for 2–3 years (department, university, or established NGO — a role that outlasts one official's tenure)
- A committed deployment application identified before any data collection starts — the one non-negotiable gate
- Local/state administrative authority able to convene cross-department stakeholders (champion identified per Cluster 2.1)
- Script/orthography decision made, or an owner assigned to make it
- Technical partner, data collection agency, and model hoster identified (see Cluster 2.3 for scope of each)
- Funding secured for the first stage of data collection (Cluster 6.1's Language survey through Pilot collection steps)
- Nodal officer designated by name and post (post survives personnel transfer)



A "No" on any item is not a disqualifier but a task to be completed to avoid any gap in expectations and rework of any efforts.

Foundation-building for one language is a one-time exercise spanning roughly 6–9 months of active effort — a useful expectation to set with all parties before work begins.

4.2 — Budget categories to plan for

The reference figures below apply to a language with a speaker base in the 10–20 lakh range, without a mature production-grade dataset — no native script infrastructure, moderate-to-minimal existing media, and no open-source datasets of production quality.

Cost Component	Reference Estimate	Basis
NMT - Translation (~200,000 sentences, ~10 words each)	~₹40 lakh	₹18 – ₹25/sentence pair
ASR — Read Speech (100 hrs)	~₹2 lakh	₹1500 – ₹2,500/hour
ASR — Conversational/Spontaneous Telephony & In Person + Voice Cmmands (200 hrs)	~₹4 lakh	₹1500 – ₹2,500/hour
Existing Media Labelling (200 hrs) — To be clarified (Social Media Data–YouTube Annotation)	(not included below)	₹1,000/hour
Audio Transcription + Review (300 hrs)	~₹20 lakh	₹110 – ₹150/min transcription + review
Total ASR + Transcription/Review (300 hrs)	~26 lakhs	-
TTS (20 studio hours)	~₹4 lakh	₹15,000 – ₹20,000/ studio hour
Subtotal — NMT + ASR + TTS	~₹70 lakh	-
Miscellaneous (20% contingency)	~₹14 lakh	20% of subtotal
Overall Project Cost	₹84 lakh - ₹1 Cr	-

This figure rises as speaker count falls, and falls as speaker count rises. As the speaker base drops below the 10–20 lakh range, per-language cost tends to rise — fewer available contributors and voice artists mean higher search and coordination effort per hour of usable data, and existing media (a source of "free" reference material at higher speaker counts) tends to thin out further. Conversely, a language with a speaker base above this range, more existing media, and open-source datasets already available can cost meaningfully less to reach the same output tier, since less new foundational work is required. Treat ₹1.8 crore as a mid-point reference to adjust from, rather than a fixed number to quote for every language.

Beyond the dataset-building cost above, the budget should also be planned for super reviewer compensation (distinct from the collection cost above) and post-deployment maintenance (see Cluster 8) — this table covers foundation-building only. There would be separate and additional costs to upgrade and fine tune the models for new use cases. Contributor payment should be direct and timely, framed as livelihood rather than honorarium. Any government officers contributing outside office hours may need separate compensation and need to be budgeted for explicitly as part of the contributor cost.

4.3 — Personal bandwidth this actually demands

The know-how for this kind of effort already exists. The document covers the checklist (4.1), the champion's role (Cluster 5), and the ecosystem it depends on (Cluster 2). The champion leading this effort inherits an already-built process; the personal time cost is mainly about steering it, rather than inventing it from scratch.

Assuming a full-time Nodal Officer is already in place, the personal bandwidth required from the champion stays within 5–6 hours a week for most of the project's duration. The exception is the first 3–4 weeks, where setting up the team, convening meetings with different stakeholders, and aligning with the community on core decisions can take 10–12 hours a week. As the project progresses past that initial window, the champion's role narrows to overseeing direction, making executive decisions, and resolving roadblocks, staying above day-to-day work rather than running it.

Cluster 5 :
The Champion's Role

Summary

In the first two weeks, three things fall to the champion personally and cannot be handed off — naming the Nodal Officer, personally convening the community, setting up and issuing the formal order that makes the review mechanism real. Everything after that routes through the team that's been stood up, with the champion's role shifting to a fortnightly check-in rather than daily execution.

5.1 — What's the Champion's To-do check list



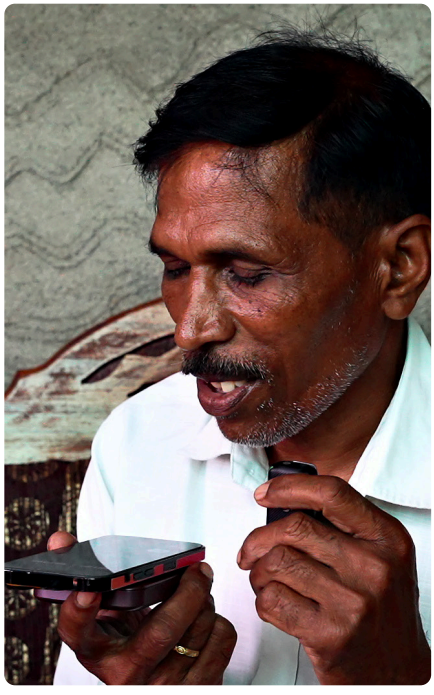
- Confirm the district's target language. This decision anchors everything downstream and needs to come directly from the champion, ahead of any team inference.
- Nominate the Nodal Officer and stand up their team. This is the single most time-sensitive action in the entire project — everything else routes through this appointment. It should be post-based rather than person-based, so the project survives the eventual transfer of either the champion or the Nodal Officer.
- Personally convene the meetings with community leaders/elders, language experts, and tribal welfare associations. The champion's institutional weight is what makes this land — a Nodal Officer sending an invitation reads very differently from the Collector or equivalent authority personally showing up to say this effort is not being done for the community, it is only possible with the community. Budget 3–4 such meetings; keep the first one personal, and give this step the time it needs rather than compressing it under timeline pressure.
- Issue the formal departmental order establishing the review mechanism. A written order survives a transfer or a change in priorities; a verbal agreement rarely does.

-

5.2 — What the Champion convenes vs. What goes to the nodal officer

The Champion Does Personally	Goes to the Nodal Officer's Team
Nominate the Nodal Officer	Day-to-day coordination of the effort, working with line departments to Identify pool of data collectors
Chair the first round of community-expert meetings	Going one level deeper into data collection requirements (types, quantity)
Issue the departmental order for the review mechanism	Gathering existing community resources (text, audio/video, digital content)
Attend the fortnightly review	Running the weekly review with the project team and community
Co-sign the formal scope document	Drafting the scope document itself, jointly with the orchestrating partner
Co-sign MoUs with technical and data partners	Day-to-day vendor coordination once contracted

5.3 — What the Champion enables without owning



- Script or dialect decisions: The call belongs to the Linguistic Expert, along with the community members, informed directly by the community meetings the champion convened. The champion's role is to create the room where that decision gets made, leaving the decision itself to the people equipped to make it.
- Annotation, model-building, and benchmarking belong to the technical partner. Of the six stages in Cluster 3.1, the champion's office holds direct responsibility for Stage 1 (Data Collection) and its on-site quality checks. Stages 2 through 5 (Model Building, Hosting, Application Integration, Sign-off) belong to the technical partner and model hoster. Extending authority into those stages is as much a failure mode as under-investing in the one the champion does own.

- The champion's authority is convening, funding, and continuity — enabling decision-making rather than making it directly. Every role above that requires domain expertise (script, dialect, data quality thresholds, model architecture) sits with someone else by design.

5.4 — The full week-by-week sequence

This is the detailed runbook underneath 5.1–5.3, in sequence.

Steps 1–4: Get Started (Week 1 onward)



1. Identify — Nominate the Nodal Officer; stand up a small coordinating team under them
2. Understand — The team goes one level deeper into what data collection actually requires for this language before moving further.
3. Convene experts — The champion personally meets with tribal language enthusiasts, language experts, tribal welfare associations, authors, researchers, and respected community members. Extract: script preference, which dialect district- region to start with, community-flagged challenges, existing resources, names of resourceful individuals, preservation preferences beyond language itself, and a directory of important people.
4. Collect what exists — The team gathers community-held resources; the orchestrating partner pulls relevant open-source datasets in parallel.



In Parallel — Identify Contributor Roles

Role	What They Do
Audio Contributor	Speakers who record the raw data
Voice Artists	Speakers who record professional-quality voice in studio settings (Identifying any community led radio station if exists and leveraging that)
Annotator/Transcriber/Translator	Converts the raw recording into transcribed and/or translated text
Linguistic Expert	Decides script and dialect
Validator	Checks transcriptions and translations for accuracy and fluency; joins once collection begins, in consultation with language experts and community
Super Reviewer	Audits Validator output; accountable for the dataset's language purity

Requirements/scope document — Jointly written (the champion's office + orchestrating partner), covering dialects, script, regions, contributor distribution, existing inventory, and the review/audit process. The orchestrating partner adds technical specification before converting this into an RFP-equivalent document for partner selection.

Work order — Formal work orders should be issued to every team and institution engaged on the project, in place of verbal agreement. Once a work order is issued and data collection agency reports back, it should be reviewed with the steering committee before the next stage proceeds.

Review mechanism — A dedicated review mechanism, involving the wider stakeholder group, should be issued as a formal departmental order in week one. Full cadence detail — who reviews what, and how often — is set out in Cluster 7.1.

Immediate kickoff checklist (First two weeks) — Executed by the PMU / Nodal Officer's team, unless marked otherwise

- [PMU Team] Set up a single coordination channel, such as a WhatsApp group, with all nodal officers and stakeholders.
- [PMU Team] Share the data collection checklist with nodal officers in writing.
- [PMU Team] Hold a one-to-one walkthrough with each nodal officer to cover the checklist items.
- [PMU Team] Lock in the review cadence starting day one, ahead of any missed deadline.
- [Champion] Confirm the district's target language today, independent of anything else being resolved. (Already listed as non-delegable in 5.1 — repeated here for sequence.)
- [PMU Team] Begin identifying contributors, a linguistic expert, and a reviewer today as well.
- [PMU Team] Bring in validators once data collection formally starts — that point is when they become relevant.

5.4a — Best Practice: Build for continuity

It is highly recommended that it runs on its institutional mandate rather than on any one person's energy. That means writing every item in 5.1–5.4 into the departmental order itself instead of leaving it as informal practice. Done this way, the project is designed to outlast a transfer: the next person to hold the post can pick it up directly from the order itself, keeping momentum intact without needing to reconstruct the approach from memory or handover conversations.

Cluster 6 :
Setting Expectations & Milestones

Summary

Overall data collection timeline may run into 9 months under normal conditions based on the experience in Nandurbar (Bhili). Model fine-tuning runs in parallel with every collection cycle (about a week each time); application integration is a one-time effort of a couple of weeks to a month, done once after the first model-building cycle use case.

6.1 — Data collection timeline, milestones, and common delays

Below timelines are based on AI4Bharat's published methodology broken into steps. Key milestones from the Nandurbar experience are integrated with each step here.

Step	Duration	Key Milestone	Common Delays
Language survey	1–2 months	Speakers, dialects, and script status mapped	Script disputes
Standards creation	1–2 months	Orthography policy and transcription guidelines agreed	Institutional sign-off delays
Pilot collection	1 month	Data collection tested at small scale for quality; contributor pool registered	Community trust-building takes longer than planned
Full collection	6–9 months	Foundational dataset collected; first keyword-spotting model live within weeks of collection starting. Transcription and validation happens in parallel. Two to three cycles of collection and model testing	Validator availability; contributor dropout; quality failures requiring re-collection

Total: 9+ months for data collection under normal conditions.

These timelines cover data collection only.

It is recommended that data collection happens in 2-3 tranches after initial testing rather than a single tranche. After every tranche, the ASR/TTS model needs to be

updated, followed by a round of testing. This will help reveal any surprises and unaccounted gaps in data collection.

The process of Model building and application deployment run in parallel with collection, on rolling batches of collected data, per the feedback loop in Cluster 3.1, rather than starting only once collection finishes. A first keyword-spotting model can go live within weeks of collection starting, well before the full 6–9 month collection window closes.

Two of these parallel activities have known durations of their own:

- Model fine-tuning — After every round of data collection, the model builder (e.g., AI4Bharat) fine-tunes the model against the newly collected data. This typically takes about a week, and repeats with every subsequent data collection cycle as a recurring cost.
- Application integration — unlike fine-tuning, this is a one-time activity, done once at the end of the first model-building cycle. The model hoster (e.g., Bhashini) typically needs a couple of weeks to complete the integration, run it through testing, and set up staging. Once the system passes testing, the full stack is deployed to production. Subsequent model updates from fine-tuning cycles follow the same established integration and deployment pipeline.

6.2 — Quality Bar

The quality bar at each step ties to the benchmark thresholds in Cluster 3.3 (WER, BLEU, MOS) and the use-case acceptance approach in Cluster 7.5. Ask the technical partner to report these three metrics throughout Cluster 3.1's Model Building and Sign-off stages, tracking them continuously rather than checking only at the end. Along with these metrics, There should be parallel First testing step (Cluster 7.5) as a genuine supplement rather than a formality.

Cluster 7 :
Running & Monitoring

Summary

A fixed review cadence keeps the project honest week to week. Progress and quality are tracked separately, because a project can be on schedule and still be collecting the wrong thing. Seven documented failure patterns account for most of what actually derails an effort like this.
Recognizing them early is cheaper than recovering from them.

Following is the recommendation on Review cadence and the validation mechanism.

7.1 — Standing review cadence

Issued as a formal departmental order by the champion in week one (Cluster 5.4):

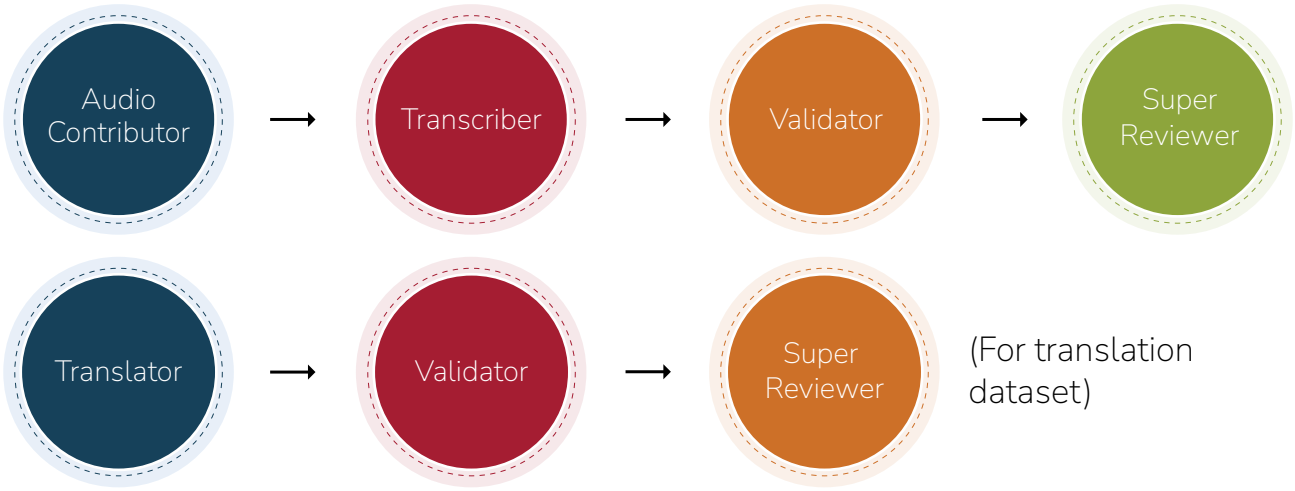
- Weekly — Nodal Officer + project team + wider community
- Fortnightly — The champion — progress, issue resolution, accountability
- Monthly — Line Secretary — progress/outcomes/priorities across all district language projects

7.2 — Review & validation mechanism

Quality is checked at two independent layers, contractually separated so no single party grades its own work:

- Layer 1 — The collection agency's own first-pass validation, built into its scope of work.
- Layer 2 — An independent Super Reviewer (E.g. university linguist, State Tribal Research Institute, or credentialed expert), appointed by the government on a contract separate from the collection agency's.

Within this, data moves through a defined chain, with escalation criteria at each handoff:



Role definitions are set out in full in Cluster 5.4. What's specific to this chain is the escalation logic between handoffs: a Validator's rejection routes back to the Transcriber/Translator with specific correction guidance rather than a bare pass/fail flag, and the Super Reviewer's audit feeds systemic error patterns back into the Transcriber/Translator's ongoing calibration, addressing the pattern rather than just the individual item.

One central program office sets the transcription guidelines, metadata schema, and quality bar, applied uniformly across every field team, holding the same standard regardless of team. This is a hard rule: quality control always runs against a standardised, language-specific rubric agreed in advance, since a collection agency grading its own work by its own informal standard is exactly the gap this two-layer structure exists to close.

7.3 — Progress metrics vs. Quality metrics

These are tracked separately on purpose — a project can be fully on schedule and still be building the wrong dataset, and the two failure modes need different fixes.

	Progress Metrics	Quality Metrics
Answers	Are we on pace?	Is what we're producing actually good?
Tracked by	Nodal Officer, weekly	Quality PMU / Super Reviewer, at every collection and review gate
Examples	Hours/sentences collected against target; contributors onboarded; milestones hit vs. planned dates	Rejection rate at each review layer; diversity gate compliance (age, gender, dialect, channel); WER/BLEU/MOS once a model exists
A good number here means	The pipeline is moving	The pipeline is producing something usable

A project can look healthy on progress metrics alone while quietly failing on quality — reaching a volume target with data that's skewed toward one demographic, for instance. Both volume and quality metrics should be reviewed alongside for every review.

7.4 — Mistakes to avoid: The seven documented failure patterns

These are drawn directly from documented project experience:



1. The community disengages midway. Contributors lose enthusiasm after the initial launch. This can be secured with real-time progress visibility, regular recognition, and — most effectively — showing contributors the actual product of their work as early as possible. Hearing a real result in their own language is the strongest motivator available.



2. The community contributor selection and training isn't properly planned. The community contributor selection criteria may usually overindex on the convenience of finding and training contributors. Given there is usually a tech dependence for data collection: old veterans can easily get missed in the process. They bring in the purest form of spoken language.

   After selection, the coordinators and validators aren't properly informed about the purpose of the data collection exercise and get deployed across different dialects without proper training - creating errors and frustration.



3. Data is collected but never deployed. Models get trained but no government system integrates them, because the deployment platform was never engaged as a partner from the start Or it wasn't ready to be tested when models got built. This is exactly what Cluster 4's readiness checklist exists to prevent — deployment gets identified before data collection begins, ahead of any later discovery.



4. The language expert bottleneck. Linguists and bilingual validators are scarce, and losing even one briefly can stall the project. Build a distributed validator pool of at least 5–10 trained people, with documentation thorough enough that any validator can pick up where another left off.

5. Political or administrative change. See Cluster 5.4a — this is the individual-champion fragility risk that best-practice notes exist to address.



Vendor or platform lock-in. If data, models, or workflows sit exclusively with one commercial platform, the project becomes dependent on that vendor's continued interest and pricing. Insist on open data standards and ULCA-compatible exports from day one — Bhashini's ULCA repository

6. exists specifically to prevent this.

7. Low community adoption of the deployed service. A technically functional system needs real users to count as a success. This usually traces back to poor UX for low-literacy users, a deployment that misses actual community need, or insufficient community involvement in how it was designed. Involve end-users in testing before formal launch, and choose the deployment use case with community input at the center of that decision. Often the missing piece for a low adoption is low awareness - a structured program from the application owner to create awareness among the targeted beneficiaries.

A related, narrower lesson worth carrying alongside these seven: quality checks run only at the end of a pipeline force expensive re-collection when problems surface. Multi-stage review embedded throughout the process, rather than batched at the end, is what makes catching a problem affordable instead of catastrophic.

7.5 —Testing and Use Case acceptance

The three benchmark metrics in Cluster 3.3 (WER, BLEU, MOS) confirm whether a model is technically sound; they leave open whether the deployed service actually works for the purpose it was built for. A fourth layer sits above them: use-case acceptance testing.

The documented approach is direct and simple — run a defined batch of real test interactions (for example, roughly 100 test calls to a live IVR service) and have the deploying department itself judge whether the output was acceptable for the department's officials and for the community. This acceptance judgment is best made by an independent third party, standing apart from the vendor whose model is being evaluated. This is also the stage for formal sign-off (see Cluster 8). A model proceeds to full deployment once it passes the use-case acceptance test, with benchmark scores serving as supporting evidence rather than the sole basis for the decision.

The following template gives good references of testing frameworks to be used by the testers & auditors as well as to capture the testing feedback at every stage of the exercise.

Testing Chat and Voice

Cluster 8 :
Closing & Sustaining

Summary

Sign-off is tied to the three core process outputs — dataset collection, model building, and application integration — rather than to setup or contracting milestones. "Done" means the model is live and usable by the public, with at least two feedback cycles completed. Improvement can continue indefinitely after that point without changing the project's completion status. Surviving the champion's departure depends on the mandate being written into institutions rather than carried in one person's memory.

8.1 — Sign-off gates

Sign-off tracks whether the actual work got done, separately from whether the paperwork to start it is in place. Readiness and contracting items — governance setup, deployment anchor confirmation, community trust-building, the reuse audit — are preconditions covered by Cluster 4's checklist; they sit outside the sign-off gates here. Whether a vendor is contracted, a Tribal Research Institute is engaged, or a technical partner is confirmed is a Cluster 4 readiness question, separate from a Cluster 8 completion question. The three gates below are about output, and the third one repeats every cycle:

Gate	What Must Be True	Who Signs Off
1. Dataset Collection Complete	The phase-wise dataset (per Cluster 3.4's scale targets and Cluster 5's domain split) is collected and validated through the full chain in Cluster 7.2 (Audio Contributor → Annotator/Transcriber/Translator → Validator → Super Reviewer)	Super Reviewer
2. Model Building Complete	ASR/NMT/TTS models are trained and meet the benchmark thresholds in Cluster 3.3 (WER, BLEU, MOS)	Technical partner, audited by the Quality PMU
3. Application Integration & Feedback Complete	The current model version is deployed, live, and usable by the public in the target service; a use-case acceptance test (Cluster 7.5) has passed; usage gaps are captured and fed back into the next round	Independent use-case acceptance test (Cluster 7.5), signed off by the Core Application Owner

Gate 3 is not a one-time event — it repeats every cycle as the loop in Cluster 3.1 runs (Model Building ↔ Data Collection, Deployment ↔ Model Building). The project's completion status depends on how many times Gate 3 has been passed — see 8.2.

8.2 — Definition of "Done"



Done means: the latest version of the model is deployed, live, and usable by the public in the target application, with at least two complete feedback cycles finished — two full passes through Gate 3, each incorporating what the previous cycle's usage gaps surfaced. At that point, the foundation-building effort is complete.

The language's AI journey continues well past this point. Further rounds of improvement can and should continue indefinitely — Cluster 3's process loop is continuous by design, and Cluster 7's ongoing quality assurance exists precisely because a model that was good enough at launch will still need maintenance. Once the public-facing service is live on its latest version and two feedback cycles are behind it, the foundation-building phase is complete, distinct from the language's ongoing relationship with AI. Everything after that point is Cluster 7/8's ongoing operating rhythm, a continuation of maintenance rather than of the original build.

8.3 — What has to exist for this to survive the Champion leaving

Three things, all already established elsewhere in this package and drawn together here:



The mandate lives in a departmental order, independent of the champion's personal memory — every non-delegable action in Cluster 5.1–5.4 should already be written into that order, per Cluster 5's own best-practice note (5.4a).



The dataset and model are openly released, held in public commons rather than inside any single vendor's or individual's control (open licensing under CC BY 4.0 or equivalent) — so the work outlives the specific contracts and people who built it.



A replication handbook is published — this package, once complete, is that instrument. A successor can pick up the project directly from the departmental order and this handbook.

A field perspective worth carrying alongside these three mechanisms: a champion's own role is closer to a midwife's than an owner's — facilitating what needs to be facilitated, then stepping back and becoming redundant to the process, rather than remaining personally indispensable to it. A project still fully dependent on its champion's day-to-day involvement at the point they leave has fallen short of "done" in the sense defined in 8.2, whatever its gate status — sustainability is itself a condition for completion, even though it stands apart from the three named gates above.

Notes

Notes


EkStep
Voice AI