



Including an ignored language into the AI landscape is a team sport

The Bhili Story



10 Million
Bhili Voices



No AI
footprint



A live government
service in 100 days!

#100AIDiffusionPathways

PLAYER 01:

The champion who seeded other champions

PARIMAL SINGH

Project Director, PoCRA
(Nanaji Deshmukh Krishi
Sanjivani Project)





Parimal Singh, Project Director of PoCRA, decided that Bhili-speaking farmers shouldn't be locked out of the benefits of MahaVISTAAR, (Maharashtra's agri-advisory app) simply because the ecosystem hadn't built for them.

.....

PLAYER 02:

The person who set the terms worth building to

.....

DR. MITTALI SETHI

District Collector,
Nandurbar





Dr. Mittali Sethi, believed that tribal languages belonged inside AI, not on its margins.

She had already made Korku, another underserved language, the first language in India to carry IIT Bombay's breastfeeding tutorials. She learned Korku from a broken sentence at a door in Melghat that got her inside a house she'd expected to be kept out of.

For Bhili, this principled customer, didn't write the code, she set the standard that the code had to meet. She would not settle until it really worked.

.....

PLAYER 03:

The trust spine

.....

KAMLESH MAHALE

Poet and Karya's Regional
Coordinator, Nandurbar





Kamlesh Mahale, poet and Karya's (data collection agency) regional coordinator, is a powerful example of what 'community trust' looks like.

Being local, he did not recruit through any system but through twenty years of knowing whose word could be depended on. His participation sparked community ownership and discussions around what the community's own language was worth.

.....

PLAYER 04:

The language community – The Voices

RAVISHANKAR PADVI
MANISHA WARGHADE
TULSIDAS VALVI
BHIMSINGH VALVI
DINESH VALVI
KSHITIJA DESHMUKH
and other contributors who
championed the process





Manisha Warghade, associate professor in Navapur, had worked on tribal languages for years. Yet when she typed her own mother tongue into a screen, she found that the words weren't there. This absence is where the Bhili AI model actually began.

Agricultural officers and community members like Manisha Warghade, Ravishankar Padvi, and Tulsidas Valvi gathered under three ceiling fans and thought not about crops, but about why their language mattered. Ravishankar Padvi, the Taluka Agricultural Officer offered that 'Dehwali Bhili' had

drifted under the schooling system. Hindi words had slipped in, unnoticed. A language with no script would disappear with the older speakers.

Tulsidas Valvi, Asst. Taluka Agri Officer, offered words like; Pamro (a kind of tree), Phaito (a bamboo platform on which grains are stored), Hikam (a method to store and ripen raw mangoes), each one not just a name, but a story.

The contributors named the risks of building a language AI model before any engineer did. Under deadlines, Marathi and Gujarati would slip in because the model would not be able to tell the difference and a distorted Bhili would emerge. The contributors feared corpus corruption more than the technologists did. This is what it looks like to treat the community and its contributors as builders, not data sources.

.....

PLAYER 05:

The committed application

MahaVISTAAR

Maharashtra's AI
farm-advisory platform -
the landing strip that
made the plane align





Most low-resource language AI projects end as a paper and a benchmark. This one transformed into deployment , because there was a landing strip already waiting for it.

MahaVISTAAR, Maharashtra's AI farm-advisory platform under PoCRA, carries university-vetted crop advice straight into a farmer's phone. As Jagadish Babu, EkStep's COO said, "If

the landing strip is there, the plane aligns itself."

'Bhili' became a use case for live deployment because a farmer in Nandurbar could actually call it. Starting from the hardest dialect in the most remote district, rather than the easiest language with the most speakers made this use case an example of using AI for social inclusion and justice.

.....

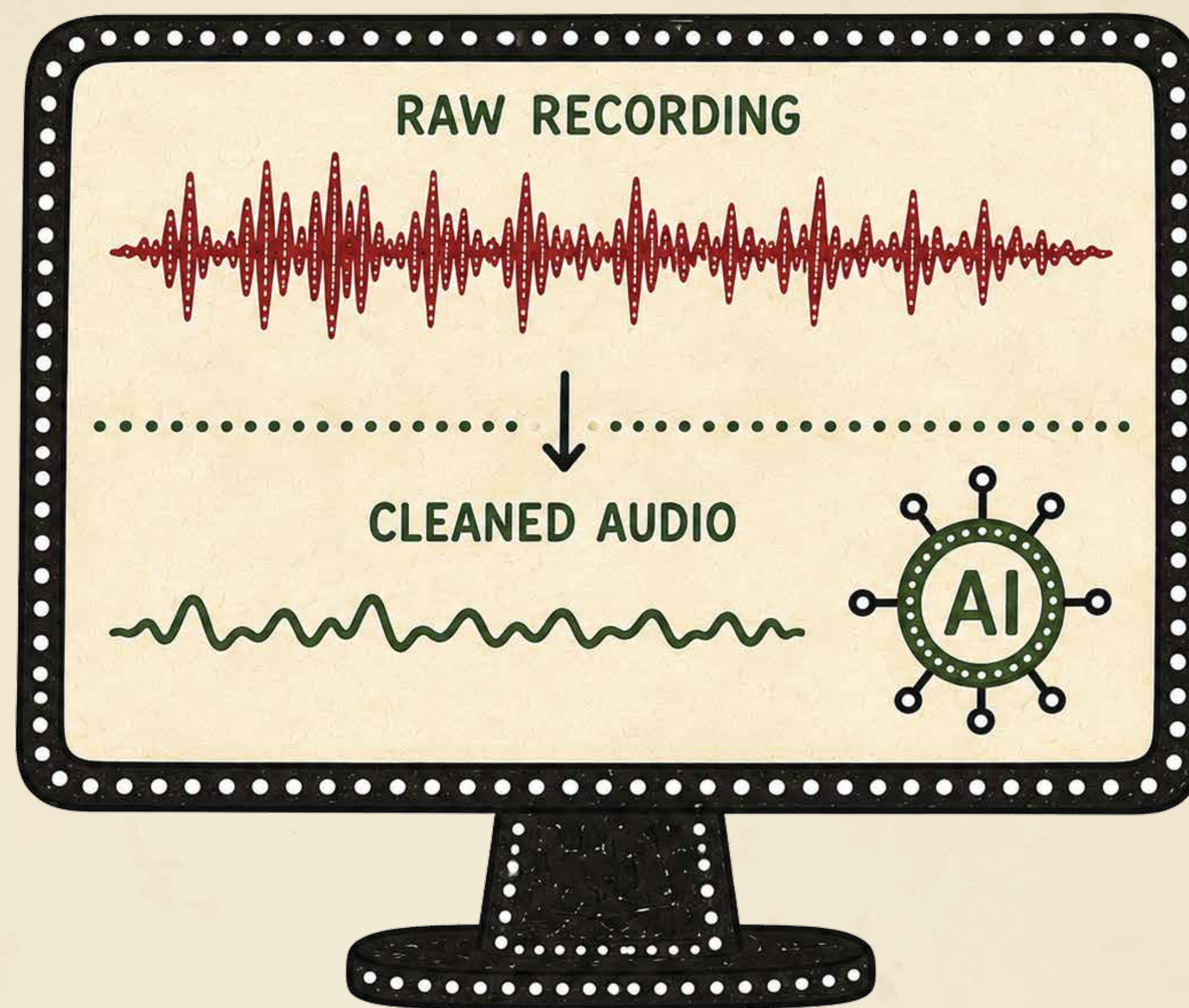
PLAYER 06:

The studio found on day three, & the coder nobody planned for

JAYWANT KAPADE
AND ASHISH

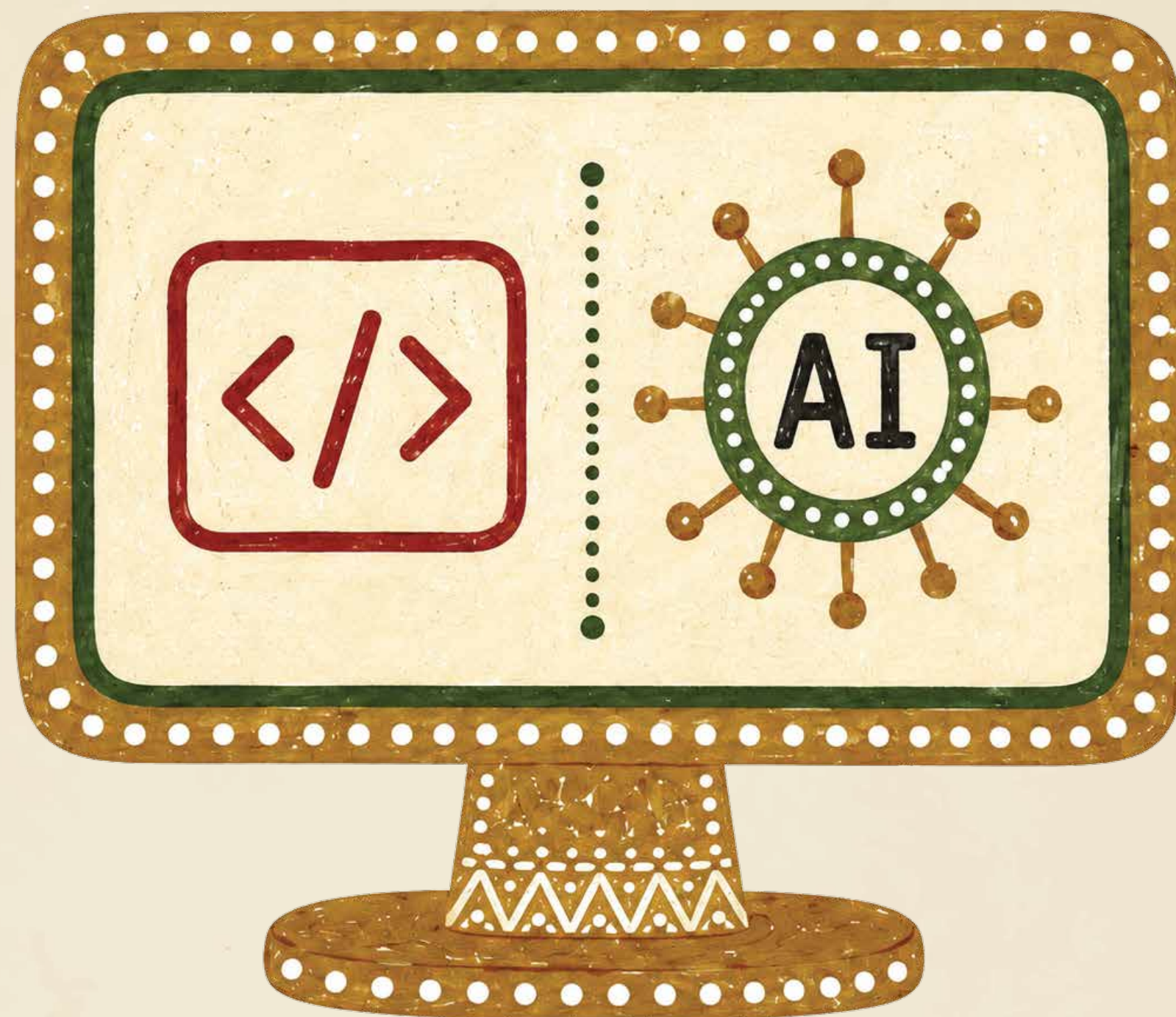
The acoustic keeper and
the volunteer coder





When the project stalled on Day 1, for want of a quiet room, the community pointed the team to the community radio station, Vikas Bharati 90.8 FM, run by Jaywant Kapade. The former Mumbai film-production hand who understood what a clean recording was, cut the same lines three ways and let the teams take their pick.

.....



Then Ashish, a volunteer coder, upset at the pace of the recordings (a 100 lines in two hours) coded an overnight routine to record fifty lines at a stretch with an automatic cut-off between segments.

Both were already in Nandurbar, waiting to be found. This is what it looks like when the design for inclusion activates local expertise and infrastructure, instead of importing it.

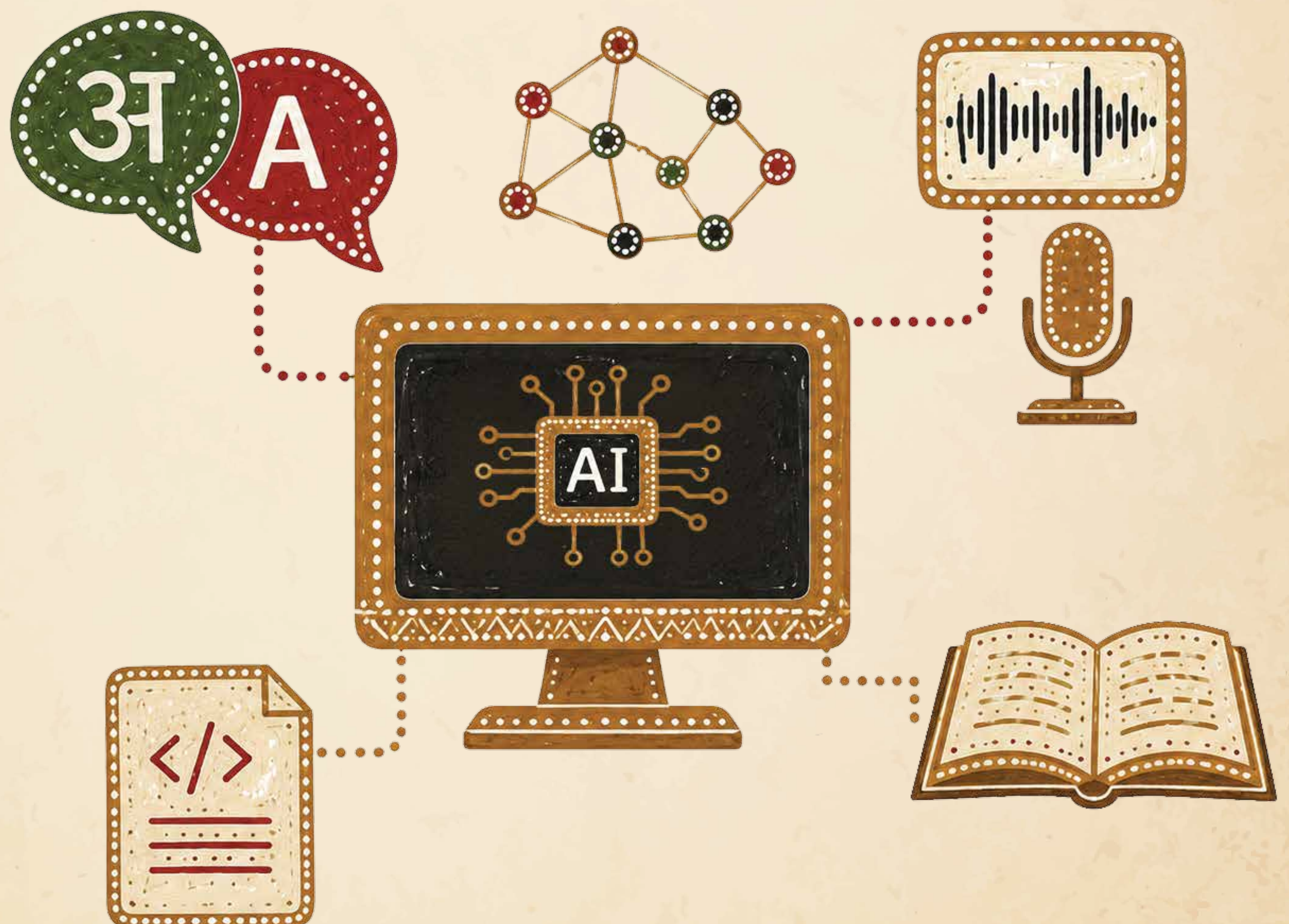
.....

PLAYER 07:

The Builders Who Kept Their Hands Separate

KARYA

Data Company





Karya gathered the data and, by deliberate discipline kept their hands off from grading it. This separation between collecting and grading isn't a technical detail. It's an integrity choice, which made the output trustworthy, not just accurate. This is because the people who gathered the language did not judge whether it had been learned correctly.

PLAYER 08:

The Builders

AI4BHARAT (IIT MADRAS)

AI4Bharat, the
language-AI lab at IIT
Madras built the model.



PLAYER 09:

The Builders

CENTRE FOR OPEN SOCIETAL SYSTEMS

A collaborative initiative between
EkStep Foundation & IIIT, Bangalore.

The Centre for Open
Societal Systems
(COSS) built the
voice
orchestration
layer.



PLAYER 10:

The digital public rails

BHASHINI AND AIKosh

India's AI Powered Language Translation Platform and National AI Library





The Bhili AI model is hosted on Bhashini, India's AI Powered Language Translation Platform.

The model and its data, open by default, lives in AlKosh, the national AI library. And ownership sits with the tribal-welfare department of Maharashtra. Not a company. Not a lab. The community that the language belongs to.

The Bhashini team kept the project's timelines and discipline running.

PLAYER 11:

The Orchestrator

EkStep
Foundation





EkStep through its work in OAN (OpenAgriNet) and involvement in MahaVISTAAR, asked AI4Bharat to create the model, secured Bhashini hosting, built a voice-orchestration layer with COSS (Centre for Open Societal Systems at IIIT Bangalore), that threaded a five-step live call under 1.5 seconds and managed the project till launch.

It did not let the model be judged by a lenient standard, insisting on the same rigour of testing as any other global Voice AI model, carving out a path for other underserved languages.

PLAYER 12:

The keeper of What the machine cannot hold

AMRUT NATHU

PADVI

Bhili Singer, performer of
the songs of Birsa
Munda and the forest.
The one who holds what
the file left out





Amrut Padvi holds the part of Bhili that no file can contain. The Bhili AI model was built to carry crop advice into a phone. The man who sings the songs of Birsa Munda was never in the dataset. Amrut Nathu Padvi doesn't write the wedding songs, the harvest songs, the masked dances that carry Birsa Munda's story, he sings them.

He is the reminder that what diffusion programmes build is not the culture. He is the culture. The machine that learns a language without learning the singer has learned a half-thing.

THE BHILI STORY



Including an ignored language into the AI landscape is a **team sport**



#100AIDiffusionPathways