

Team DISTIL

(Dialectal Speech Transcription in Indian Languages)



Sathvik Udupa



SPIRE LAB

Saurabh Kumar



SPIRE LAB

Srinivasa Raghavan



Jiatong Shi (Mentor)



Carnegie Mellon University
Language Technologies Institute

Megh Makwana



Soumi Maiti



Carnegie Mellon University
Language Technologies Institute

Manjunath K E



Motivation

In countries, like India, with low literacy and high rates of digital adoption, digital interfaces will need to be

**voice driven & conversational
in local languages**

India has **337 million**^{1,2} **low literate, vernacular speaking internet users**, the largest population globally.

Non-English users are on the rise with **9 out of 10³ new internet users** likely to be **Indian language users**.

Voice-search is a need, not a convenience with Google recording a **270%⁴ y-o-y growth in voice search** in India, with **Hindi voice search alone growing by 400%⁵** over the same period.



Bankdidi

আপনাকে কিভাবে সাহায্য করতে পারি?

How can I help you?



Kamaladevi

এই মাসের ঋণের EMI আমাকে কক্ষন দিতে হবে?

When do I need to pay my loan EMI for this month?



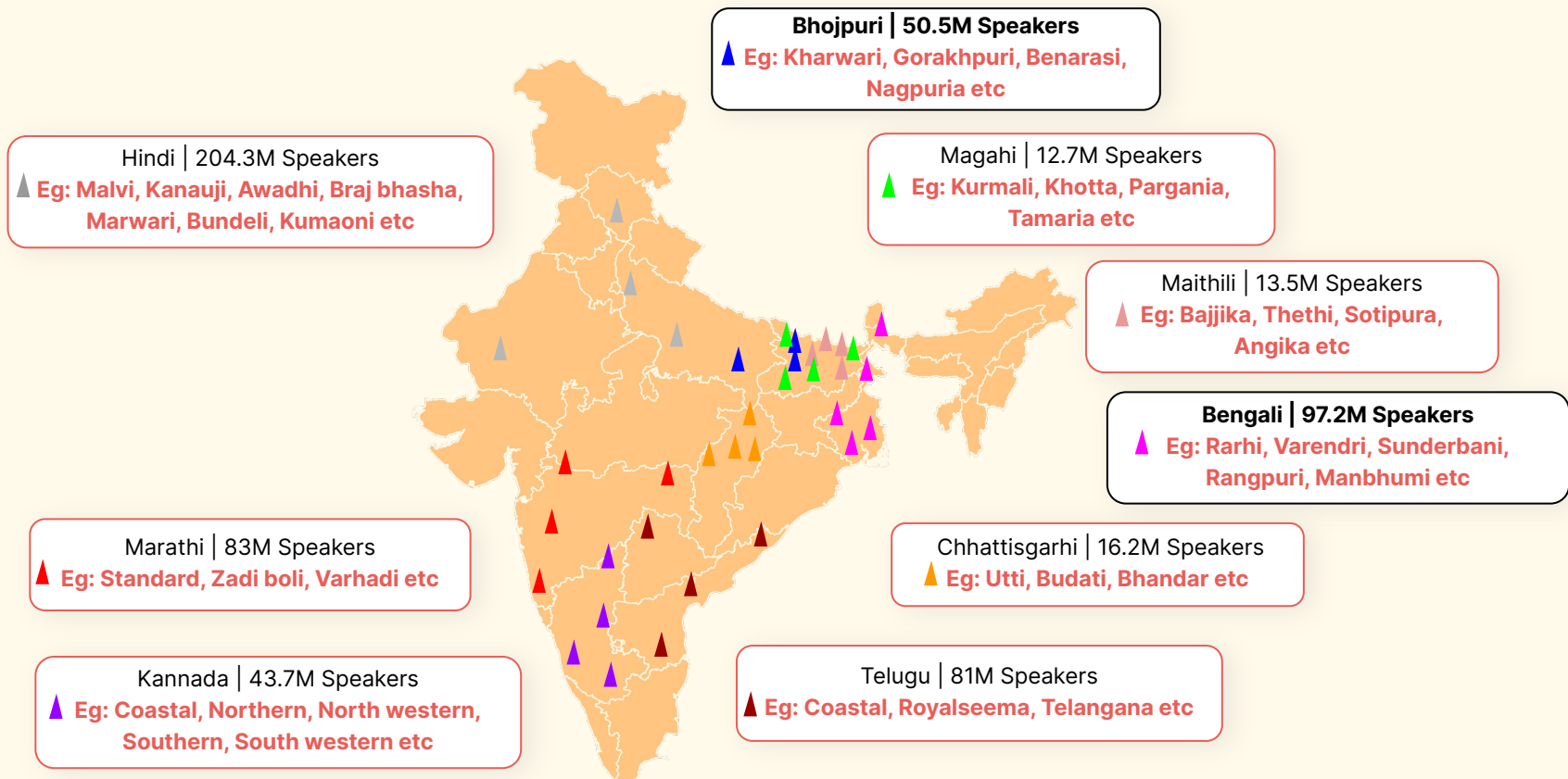
Source:

1. National Sample Survey Office - Ministry of Statistics and Programme Implementation (2015), Social Consumption: Education;

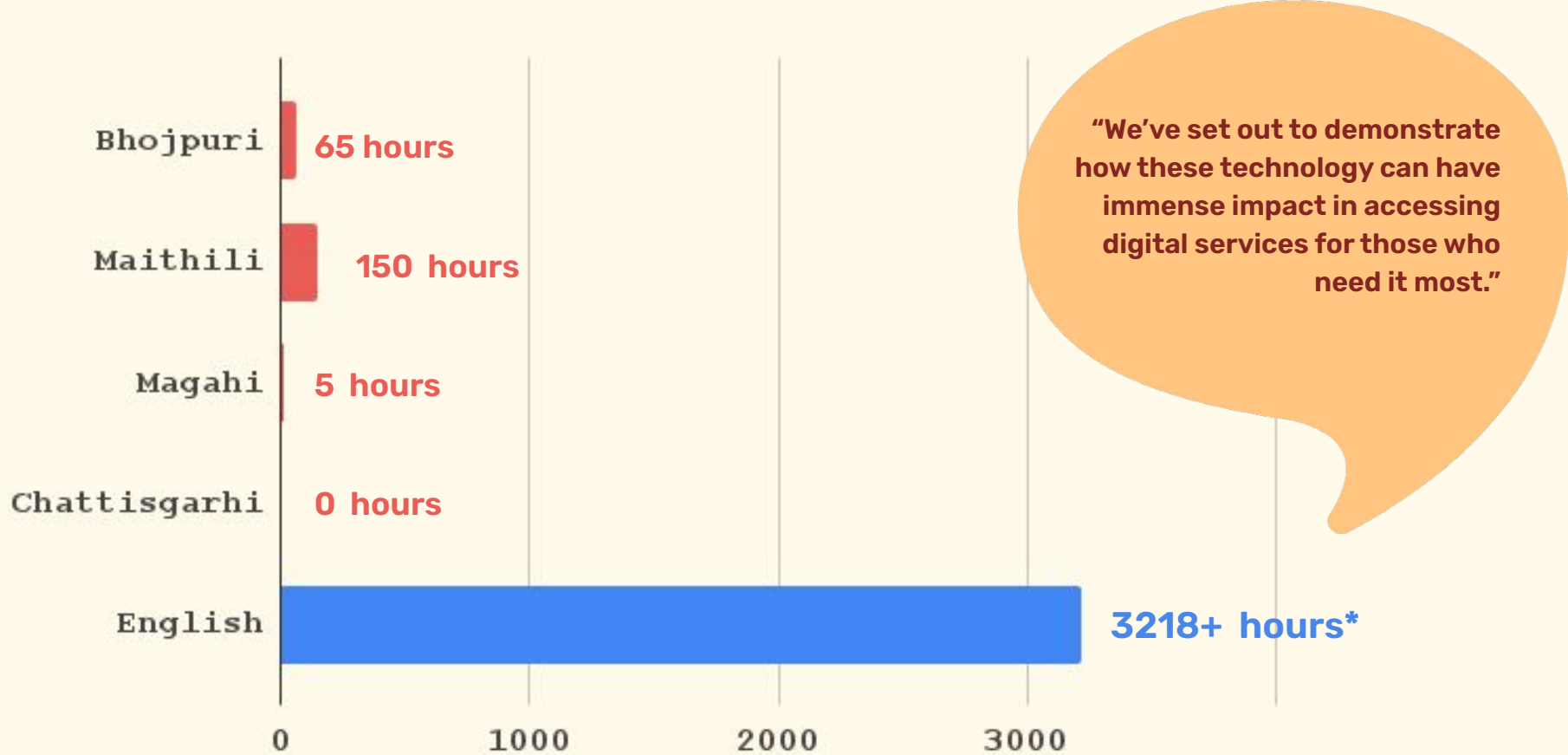
2. Kantar IMRB (2018), ICUBE: Digital adoption and usage trends;

3, 4 & 5. Google (2018), Year in Search - India: Insights for brands

Indic languages: Dialectal diversity is very high



* These languages are part of the ongoing voice collection efforts as the part of RESPIN Project



Disparity in availability of data for **English** vs some **low-resource Indian languages**

* well known labeled corpora like Librispeech, Fisher English, Switchboard, TEDLIUM etc

^Maithili: TDIL, ULCA, Magahi: Low-resource corpus , Bhojpuri:ULCA, Low-resource corpus

Dataset - RESPIN*

Language	Dialect	#speakers	#sentences	#utterances	Duration (hours)
Bengali	D1	405	4313	164254	206.40
Bengali	D2	407	4106	197659	271.45
Bengali	D3	419	4300	192024	283.17
Bengali	D4	425	4089	159803	216.14
Bengali	D5	437	5151	157053	236.08
Total		2093	21959	870793	1213.23
Bhojpuri	D1	601	8285	261768	351.25
Bhojpuri	D2	725	7849	325605	417.74
Bhojpuri	D3	705	7782	279246	347.97
Total		2031	23916	866619	1116.95



* **RESPIN**

Speech Recognition in
Agriculture and Finance
for the Poor in India

**IISc Bangalore
(SPIRE Lab)**



भारतीय विज्ञान संस्थान



SPIRE LAB

Navana Tech

NAVANA
TECHNOLOGY FOR THE NEXT BILLION USERS

Funding:

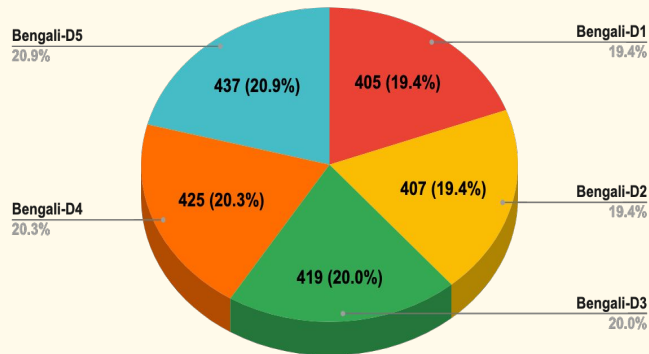
Bill & Melinda

Gates Foundation

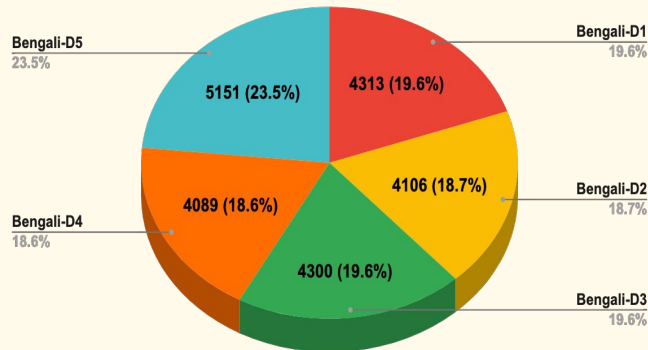
Dataset - RESPIN

Bengali Dialects

No. of speakers

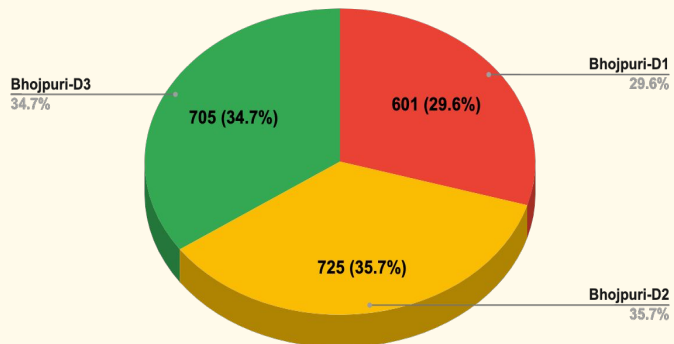


No. of prompts

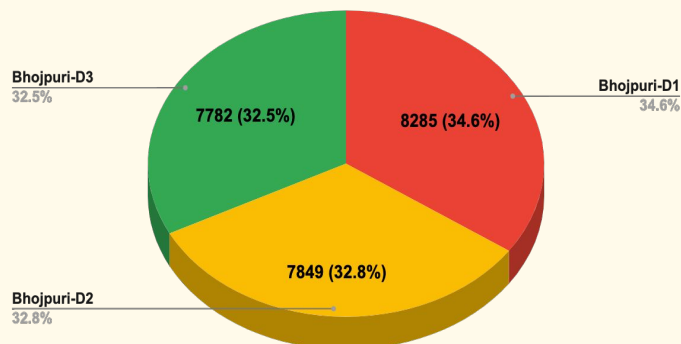


Bhojpuri Dialects

No. of speakers



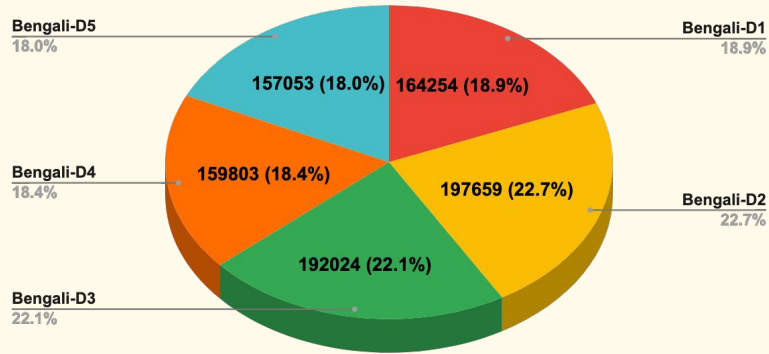
No. of prompts



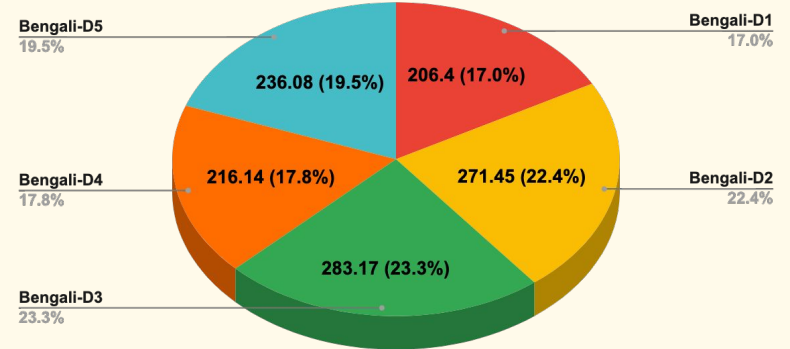
Dataset - RESPIN

Bengali Dialects

No. of utterances

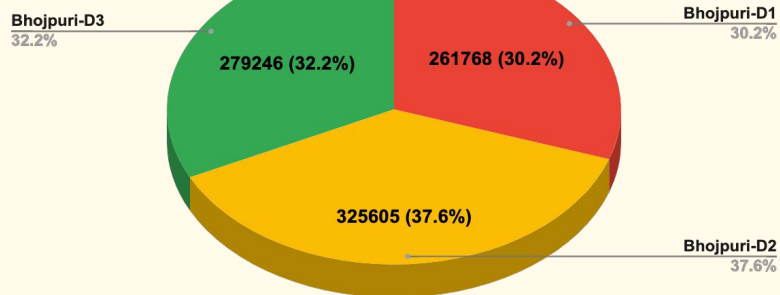


Duration in hours

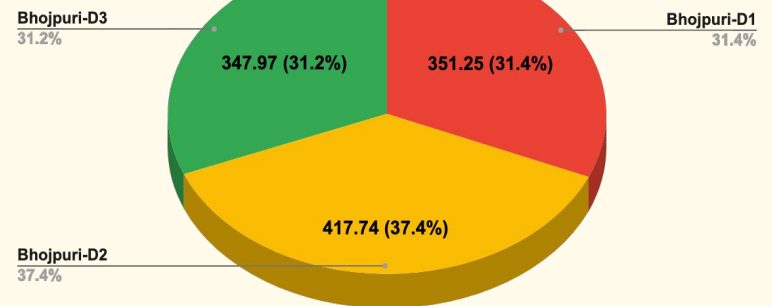


Bhojpuri Dialects

No. of utterances



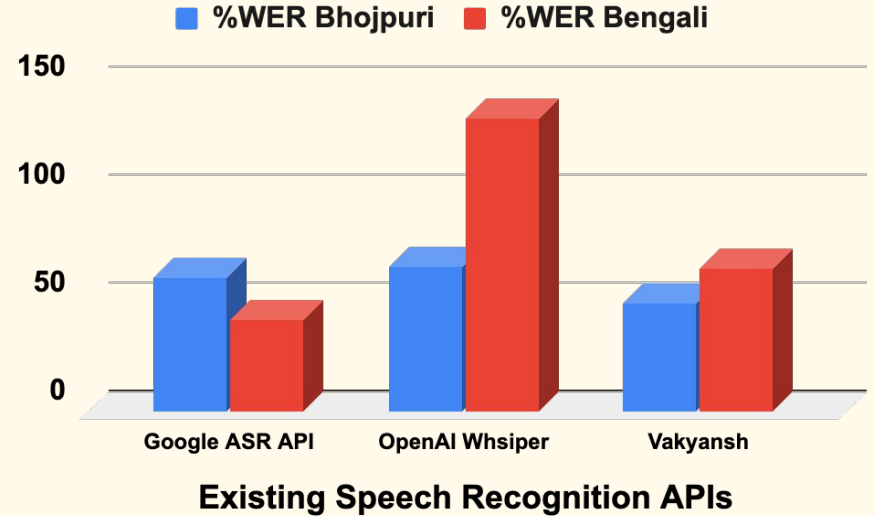
Duration in hours



Existing model performance

Bhojpuri (With Hindi Models)	WER (%)	CER (%)
Google Cloud ASR API ¹	61.42	37.54
Whisper - large (v1) ²	66.54	36.15
Vakyansh ³	49.6	37.54

Bengali	WER (%)	CER (%)
Google Cloud ASR API ¹	41.67	16.37
Whisper - large (v1) ²	135.25	130.06
Vakyansh ³	65.93	23.72

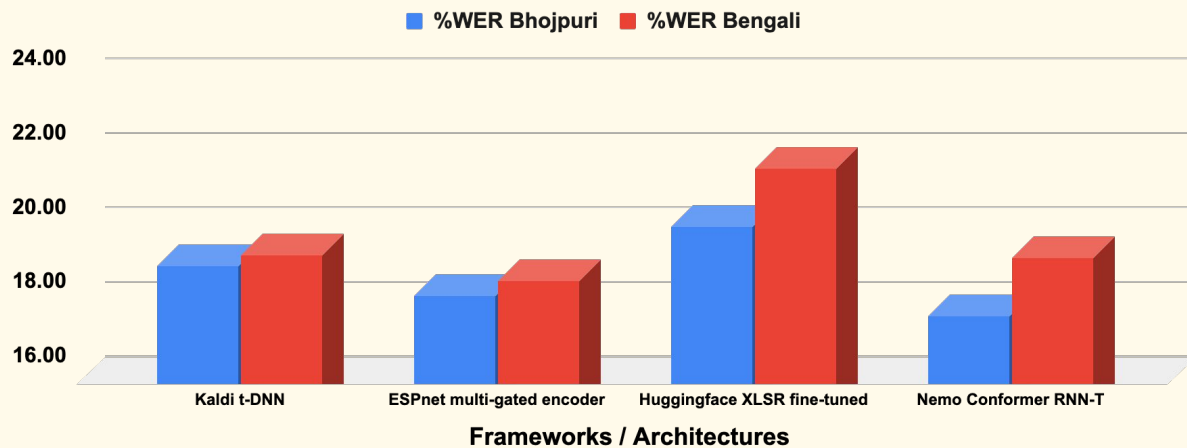


1 <https://cloud.google.com/speech-to-text>

2 <https://github.com/openai/whisper>

3 <https://github.com/Open-Speech-EkStep/vakyansh-models>

Results on test set



Framework / Architecture	%WER Bengali	%WER Bhojpuri
Kaldi TDNN	19.27	19.00
ESPnet multi-gated encoder	18.60	18.20
Huggingface XLSR fine tuned	21.63	20.06
Nemo conformer RNN-T	19.21	17.65

Dialect specific models - Bengali

Kaldi: Dialect-specific AMs with common LM

	% W E R	% C E R	Dialect level CER				
			D1	D2	D3	D4	D5
D1	21.83	7.61	7.0	4.7	7.8	9.9	8.8
D2	22.74	8.19	8.7	4.0	7.8	11.7	8.7
D3	22.47	8.11	8.6	3.8	6.8	12.0	9.3
D4	23.49	8.64	9.2	6.4	9.4	8.1	10.4
D5	21.96	7.9	8.8	4.2	8.0	11.1	7.3

Kaldi: Dialect-specific both AMs and LMs

	% W E R	% C E R	Dialect level CER				
			D1	D2	D3	D4	D5
D1	28.18	9.91	7.9	6.2	10.5	13.2	12.2
D2	28.72	10.32	11.4	4.3	9.5	14.3	12.2
D3	28.2	10.33	11.7	4.8	7.6	14.7	12.8
D4	37.41	14.38	16.4	12.8	17.6	8.6	17.1
D5	30.32	11.01	12.7	6.5	12.0	15.4	8.2

ESPnet: Dialect-specific AMs without LM

	% W E R	% C E R	Dialect level CER				
			D1	D2	D3	D4	D5
D1	38.33	14.40	12.3	9.4	15.2	17.6	18.7
D2	40.11	15.54	17.9	8.4	13.9	18.8	19.5
D3	44.30	18.90	22.8	10.7	14.6	23.8	23.0
D4	50.37	20.35	23.3	18.4	22.5	13.4	25.5
D5	40.81	15.21	17.2	10.8	15.9	19.7	12.1

Dialect specific models - Bhojpuri

Kaldi: Dialect-specific AMs with common LM

Dial	W E R	C E R	Dialect level CER		
			D1	D2	D3
D1	20.28	8.01	7.39	8.28	8.32
D2	19.28	7.33	7.94	6.68	7.33
D3	20.14	7.73	8.39	8.43	6.58

Kaldi: Dialect-specific both AMs and LMs

Dial	W E R	C E R	Dialect level CER		
			D1	D2	D3
D1	24.9	9.7	7.8	10.9	10.6
D2	24.7	9.5	10.8	6.9	10.5
D3	24.3	9.3	10.2	11.2	7.0

ESPnet: Dialect-specific AMs without LM

Dial	W E R	C E R	Dialect level CER		
			D1	D2	D3
D1	31.22	11.95	9.5	12.6	13.6
D2	30.89	11.78	12.8	8.0	13.9
D3	30.02	11.27	12.4	12.8	9.1



Kharagpur

Lakshmi, 36

Meet Lakshmi. She runs a street food stall and has studied till grade 10. She speaks Bengali fluently. She has her own bank account which her husband or daughter help to manage. She uses her smartphone for simple tasks and prefers WhatsApp to chat with her family and friends.

She has to travel around 45 mins to reach the bank for any service/query. She wishes to be able to clarify concerns and fetch details in her fingertips.



Low to Intermediate level of digital confidence

Uses smartphone in a limited capacity for low risk tasks such as messaging, watching videos and making calls.



Early stages of financial literacy

Is part of the banking ecosystem but not actively managing it. Knows basic concepts of banking and would like to have the independence to manage finances on their own.



90% Indians

10% users speak English¹ and the rest are distributed across 22 languages with 100+ dialects.



300+ Million

Emergent users like Lakshmi have access to smartphone.

Lakshmi (Text): Hi

Bot (Image): Namaste, I am bankdidi! I can help you with questions regarding Funds transfer, loan application, loan payments.

Bot (Voicenote): Namaste, I am bankdidi! I can help you with questions regarding Funds transfer, loan application, loan payments. Tap on the mic button and send me a audio message with your questions.

Lakshmi (Voicenote): When do I need to pay my loan EMI for this month?


Bot (Voicenote): I've shown below the details of your loan EMI including EMI Amount and due date for payment. If you have any more doubts, send me a voice note and I'll help you out!

Bot (Text): Payment due in 7 days!

Loan name: Education loan

Loan number: EL-14530295

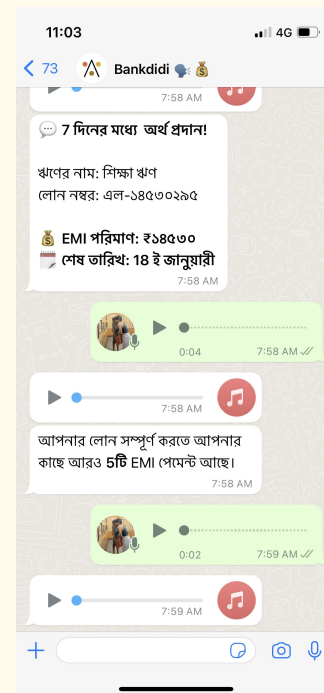
 EMI Amount: ₹2890

 Due date: 18th January

Lakshmi (Voicenote): How many more EMI payments are left to finish my loan?

Bot (Voicenote): You're paying off your loan on time! I've shown below the details of how many payments are pending for you to finish off your loan.

Bot (Text): You have 5 more EMI Payments to complete your loan.



In this demo, Lakshmi enquires about her upcoming EMI and the status of her loan by sending voice messages to the bot.

Try out the bot by sending voice notes

For 🌾 Agriculture and 💰 Banking use cases in 2 languages



Scan for Bhojpuri



Scan for Bengali

Try out Streaming ASR in Bengali and Bhojpuri



Visit our Website

<https://sites.google.com/view/slt-team>

- To know more about our experiments, more demo videos, resources for custom implementations of models and demos

