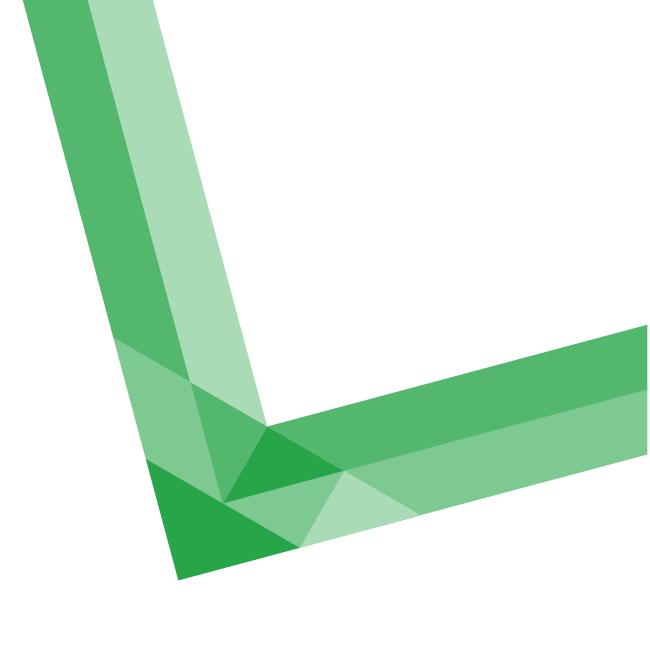


## COMMUNICATING SEAMLESSLY WITH BHASHINI

An overview of India's Language Interface

October 2023



#### **Acknowledgements**

#### **Contributors**

Sohini Dutta, Abhishek Modi, Akshata Jain and Abheya Arora

#### **Technical Reviewers**

Anantha Narayan and Rathish Balakrishnan

#### Disclaimer

This report has been produced by a team from Sattva Consulting as a product for the Sattva Knowledge Institute (SKI).. The authors take full responsibility for the contents and conclusions. Any participation of industry experts and affiliates who were consulted and acknowledged here, does not necessarily imply endorsement of the report's contents or conclusions. To quote this primer, please mention: Sattva Knowledge Institute, *Global Toolkits for FLN Assessments, October 2023*. Use of the report's figures, tables or diagrams, must fully credit the respective copyright owner where indicated. Reproduction must be in original form with no adaptions or derivatives. For use of any images in the report please contact the respective copyright holders directly for permission.

This work is licensed under the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 International License:

Attribution - You may give appropriate credit, provide a link to the license, indicate if any changes were made.

Non-Commercial - You may not use the material for commercial purposes.

Share A Like - If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



To view a copy of this license, visit <a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a>

#### **About Sattva Knowledge Institute**

Sattva Knowledge Institute (SKI), established in 2022, is our official knowledge platform at Sattva. The SKI platform aims to guide investment decisions for impact, shedding light on urgent problems and high potential solutions, so that stakeholders can build greater awareness and a bias towards concerted action. Our focus is on offering solutions over symptoms, carefully curating strong evidence-based research, and engaging decision-makers actively with our insights. Overall, SKI aims to shift intent and action toward greater impact by influencing leaders with knowledge. All of our content proactively leverages the capabilities, experience and proprietary data from across Sattva.

Editing: Anagha Wankhede and Arshiya Gupta | Design: Usha Sondhi Kundu; cognitive.designs@gmail.com

## **CONTENTS**

1	Executive Summary	04
2	Background	06
3	Overview of Bhashini	09
4	Use Cases	22
5	Governance and Implementation	29
6	Role of Stakeholders	33
7	References	35

# **EXECUTIVE SUMMARY**



India is a linguistically diverse country with 121 major languages and 18 written scripts. This diversity creates bottlenecks in last-mile service delivery for a large part of the population. These include difficulty in interpretation and communication, delay in document translations and limited accessibility of social welfare schemes.

Bhashini is an open, interoperable and innovation-augmenting digital public platform which aims to address these challenges. It has three fundamental components: **technology**, **regulations** and **ecosystem**. **Technology** consists of multiple layers and APIs that form the building blocks of the platform. The datasets and models are mainly regulated and governed under the **National Data Sharing and Accessibility Policy (NDSAP)**, and the ecosystem covers a diverse set of stakeholders, ranging from **central ministries to citizens**. The building blocks and development of the platform have encouraged the creation of reference applications like Anuvaad, an open-source document translation and digitisation platform, and marketplace applications like Jugalbandi that powers conversational AI solutions across domains.

Bhashini has the potential to **increase awareness and accessibility** for the poor and vulnerable population across sectors like education, health, financial inclusion and livelihoods. It will enable students to access online classes, digitised print material and various applications to aid their learning process, and also assist workers, especially migrant workers, in navigating job portals, obtaining skills and information. It will also expand the scope of healthtech, and make administrative and logistical tasks more efficient. The diverse language catering will also enable awareness and easy adoption of various schemes and financial initiatives among marginalised sections.

The Ministry of Electronics and Information Technology (MeitY) steers the activities of Bhashini and facilitates its smooth implementation. **Digital India Bhashini Division**, under the Digital India Corporation (created by MeitY), works to develop and maintain a public digital platform that can create and nurture an ecosystem.

It also acts as a nodal agency for the Bhashini ecosystem.

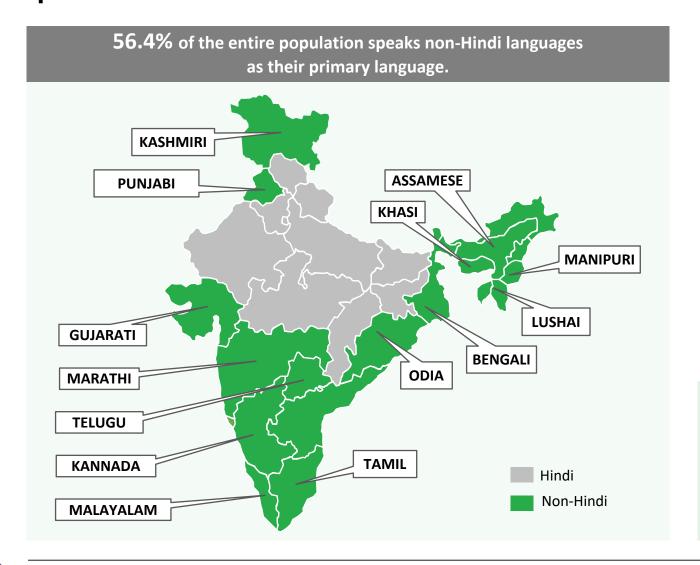
Al4Bharat—a collaborative venture among the stakeholders of Bhashini—creates open datasets and models that enable several language-based digital capabilities. These capabilities and tools can further be used to develop applications and use cases.

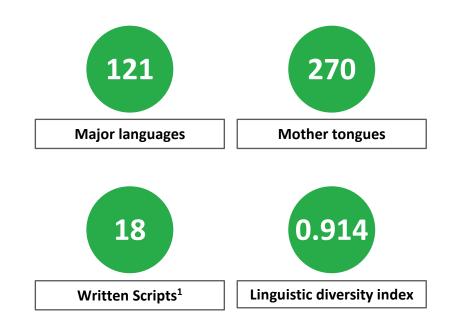
A **multi-stakeholder engagement** with the is needed to support the generation of multiple use cases for Bhashini. Philanthropic organisations, technology players and non-profits ecosystem are the primary drivers of emerging DPGs, including Bhashini. Philanthropy should strengthen infrastructure and encourage innovation in the development of use cases. Technology players should develop and integrate context-specific solutions built using the Bhashini architecture. Further, nonprofits should assess requirements, enable collaboration and assist in deployment of new technologies to achieve social impact.

# **BACKGROUND**



### India is a linguistically diverse country, ranking 4th in the world in terms of number of languages spoken.







- Mother Tongue: The language spoken by an individual's mother and/or in their home
- Language Diversity Index (LDI) is the probability that two people selected from the population at random will have different mother tongues 0 represents no diversity and 1 represents total diversity.<sup>2</sup>

#### Last-mile service delivery to the underserved populations is plagued by several language barriers.

Limited accessibility of social welfare schemes online

**80%** of all online government services are available only in Hindi or English<sup>3</sup>

<33%

of informal labour have access to government services.<sup>4</sup> 14 crore

migrant
population
have difficulty
accessing
services in
vernacular
languages.5

Delay in document translation in various languages

Manual translation speeds are at **250-300** words/hour.<sup>6</sup>



across **124 Aspirational Districts** face high costs
and reduced efficiency
through training and
recruitment of staff
speaking different
languages.<sup>7</sup>



are **pending** due to delayed transfer of information from lower (vernacular) to higher (Hindi/English speaking) courts.<sup>8</sup> Difficulty in interpretation and communication for disabled population

40% growth in app usage per year excludes
2.68 crore disabled population.9



Movement Disability



19% Hearing

**Disability** 



19%
Visual
Disability

# OVERVIEW OF BHASHINI



### Language-based DPIs and DPGs are being deployed across the world to optimise service delivery and enhance the inclusion of languages in the digital space.

Developed in Deployed in stÖry **Story Weaver** Open and free-access, quality storybooks in multiple languages India India Digital Public Goods **Open Foris Collect** Survey setups with an user-friendly interface in a multi-language environment  $(DPGs)^{10}$ Italy Globally LibreOffice Open-source office suite in 120 languages Globally Globally LibreOffice\* **African Storybook** Sub-Saharan Africa Open access to picture storybooks in African languages South Africa

Open Access
Resources

Wîcêhtowin project <sup>11</sup>

- An open access website with resources in over 74 North American indigenous languages
- Over 150 resources and tools like information and links to websites, video/audio repositories, and applications



- An initiative to spur NLP research in African languages
- Creating datasets, models, benchmarks and mentoring across more than 2000 African Languages



3 Digital Initiatives

Zero to Digital<sup>13</sup>

A guide to bring a language online

Welsh Bilingual Tech Toolkit<sup>14</sup>

Toolkit for bilingual software for developers

Inclusion of Inukitut<sup>15</sup>

Masakhane <sup>12</sup>

(an indigenous Canadian language) on Facebook

Motorola's phone<sup>16</sup>

includes indigenous languages from Brazil, Amazon and the US.

### Bhashini is a public Digital Platform which develops services and products in Indian languages by leveraging artificial intelligence.

#### What is Bhashini?

Bhashini aims to build a National Public Digital Platform to develop services and products for citizens in Indian languages by leveraging the power of artificial intelligence and other emerging technologies. 17



#### What will they do?

- Develop a language ecosystem:
   Create and nurture an ecosystem involving the government, industry, academia, research groups, start-ups and individuals
- Build open language datasets:
   Create large open source datasets and models by bringing all contributions, both institutional and citizen, into a shared repository.
- Enable products and solutions:
   Encourage the ecosystem to develop innovative products and services in Indian languages by leveraging the open repository of datasets and models

#### **Bhashini's Data Collection Goals**



#### 100 million

scraped parallel data corpus between Indian languages



#### **30,000** hours

of **unlabelled audio data** in Indian languages



#### **10,000** pages

of **printed data** in each Indic script with labelled text



- Bhashini is managed by the Ministry of Electronics & Information Technology (MeitY), Government of India, with extensive support from Technology Development for Indian Languages (TDIL).
- Bhashini takes inspiration from other mature and emerging digital public infrastructure such as Unified Payments Interface (UPI), Ayushman Bharat Digital Mission (ABDM) and National Digital Education Architecture (NDEAR).

#### Bhashini enables access to the internet in Indian languages and facilitates last-mile service delivery.

#### **Key Challenges**

#### LIMITED ACCESSIBILITY

of social welfare schemes online

#### **DELAY IN DOCUMENT TRANSLATION**

in various languages

#### **DIFFICULTY IN INTERPRETATION** AND COMMUNICATION for disabled population

#### How will Bhashini solve it?

- Enables availability of existing web content in Indian languages to widen reach and reception
- Curates user-centric assistive technologies (such as chatbots) in local languages to facilitate web access and increase usage
- Increases efficiency and reduces costs of human resources through machine translation optimised in Indian languages
- Fast-tracks translation and transliteration services, ensuring timely delivery of services, such as judicial services
- Facilitates inclusivity by enabling access to essential services for the disabled population
- Enables technological innovations due to the open and interoperable nature of the architecture

### The Bhashini architecture is guided by four principles, which enable the orchestration of the entire ecosystem.<sup>17</sup>

#### **OPEN**

All datasets, models, tools and technologies developed will be open source and freely available to all contributors and end users.

#### CATALYTIC

Bhashini will act as a catalyst to create and nurture an ecosystem involving various stakeholders to develop a data repository and other datasets, models and tools. PRINCIPLES
GUIDING
BHASHINI'S
INFRASTRUCTURE
AND DESIGN

#### **INTEROPERABLE**

Bhashini aims to bring all contributions – institutional and citizen – into a shared repository that is fully open for innovation.

#### **PROMOTES INNOVATION**

Encourages the ecosystem to develop innovative products and services by using the open repository. It also promotes innovation via challenges and hackathons.

### Three fundamental building blocks – technology, regulations and ecosystem – build the Bhashini infrastructure.

The Bhashini Cloud is where data repositories and other tools are hosted to provide the foundation.

**Building Blocks** 

### **Technology Component**

1

The Bhashini architecture is underpinned by principles of opensource and open data software to enable contributions from research initiatives and the ecosystem.

Data Policy

Policy
frameworks
are developed
to maintain
trust and
decorum
within the
framework.

Network Participants

All contributors
will support
data
crowdsourcing
activities and
provide
unbundled
software

components.

Open-source and market applications are created for end users by leveraging the Bhashini architecture.

**End-user Solutions** 

Reference Apps

4

Open source reference applications are developed to demonstrate innovative ways to leverage Bhashini's resources.

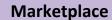
Marketplace

5

Industry-led tools and applications for the market drives innovation and use of language-based apps for service delivery.

#### The Bhashini architecture takes a multi-layered approach to build diverse language solutions for diverse problems.

**Technology Component** 





**Reference Apps** 

Ecosystem Layer:

Market tools and applications

Reference Apps Layer:

Open-source applications to demonstrate innovations

Crowdsourcing portals

**Dashboard** 

Catalogs



**Benchmarking Tools** 





**Open APIs** 



Foundation Layer:

Applications and utilities enabling orchestration among ecosystem partners

**Building Blocks** 



**DATA REPOSITORY** 

Unified Language Contribution API (ULCA): Contribute, Label, Verify, Evaluate



**Training Datasets** 



**Benchmarking Datasets** 

**MODEL REPOSITORY** 

Data and Model Repository:

Open-source data and models driving language technologies in Indian languages



NLTM Cloud: Hosts the foundation layer



A Building Block is a reusable package of business or technological functionality. The functionality can be reused in multiple use cases and solutions with marginal effort, thereby cutting down on design and development time. In substance, a building block may be comprised of data, applications, or a set of interfaces. 18

### Bhashini's building blocks include a large-scale data and model repository built from various sources, which drive language technologies in Indian languages

**Technology Component** 

1 )

#### **Unified Language Contribution API (ULCA)**

ULCA standardises data and model contributions for benchmarking to remove data silos, duplication, and ensure dataset discoverability and quality.

Parallel text corpus

Monolingual text corpus

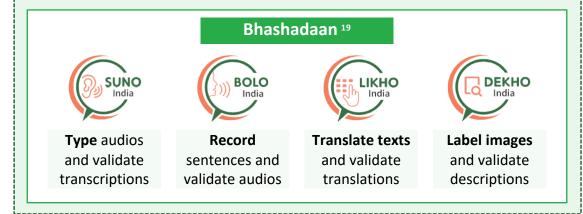
**ASR/TTS corpus** 

**OCR** corpus

**NLU Datasets** 

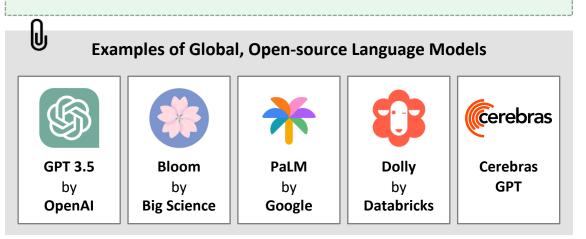
#### **Data Repository**

- The data repository will ingest data from government agencies, the web, previous language projects, NGOs and crowdsourcing.
- Crowdsourcing tools like Bhashadaan will help create the largest data repository powering language technologies.



#### **Model Repository**

- Research groups, start-ups and volunteers working on languages technologies will contribute to model repository.
- Open source models, such as MT, OCR, ASR, TTS, will be created for Indian languages.



### The foundation layer, housed on the Bhashini cloud, consists of applications and utilities which enables orchestration among ecosystem partners.

**Technology Component** 

1

### **Crowdsourcing Applications**

- Facilitates data collection from the general public to Bhashini
- Bhashadaan, as a reference app, has been created for curating crowdsourced data

### Data Creation and Curation Tools

- Collects, validates, tags and curates data, ensuring quality and usability
- **Different types of data** include web-scraped and aligned data, crowdsourced data and human curated (benchmark data)

#### **Open API**

- Allows access to the model by applications, by acting as an interface
- All models are benchmarked by implementing a REST (Representational State Transfer) service, which are a set of constraints, compliant with the API

#### Benchmarking Tools

- Evaluates the test model on defined benchmark metrics which are diverse in nature
- Both subjective human benchmark metrics and Automatic Benchmark Metrics were developed for the purpose

#### Data policies for Bhashini enable access and security of large amounts of collected data for model creation.

**Data Policy** 

#### National Data Sharing and Accessibility Policy (NDSAP)

- Notified by the Ministry of Science and Technology (MoST)
- Allows the open-sharing of non-sensitive and non-personal data created using public funds
- Applies to all data and information created, generated, collected and archived data by government funds
- Aims to promote data sharing and enabling access to go owned data for national planning and development

#### **Audio and Visual Recordings**

- For recordings available in the **public** domain, it is assumed that they can be used, unless explicitly opted out.
- For **private** recordings, explicit consent is required to use the recordings for datasets and models.

A linguistic data policy will also be formulated for NLTM and other such initiatives.

Bhashini will only use original data sources that fall under the following licensing categories:

- Creative Commons or any open licensed data
- Copyright-free or public domain data
- Publicly available government data permitted to be used by government departments and ministries
- **Privately-held data with explicit permission** from copyright holders
- Publicly available licensed data permitted to be used by the copyright owner

### The Bhashini architecture is catalysed and optimised through contributions of several stakeholders in the ecosystem.

**Stakeholders** 

in the Bhashini

**Ecosystem** 

**Network Participants** 

3



#### **CENTRAL MINISTRIES**

Develop datasets and models and align NLTM with their language services



#### **STATE GOVERNMENTS**

Identify data sources and empanel agencies for data creation and curation



#### **ACADEMIA AND RESEARCH**

Execute research in language technologies



### DATA COLLECTION COMPANIES

Collect, validate and curate datasets



#### **START-UPS**

Create language-specific open source models and multilingual applications for public use



#### **PRIVATE COMPANIES**

Build data corpus and open source softwares. Compute for training and evaluating models



#### **PUBLISHERS**

Provide data sources to build datasets and models



#### **CITIZENS**

Contribute to data and model repository through crowdsourcing

### Reference applications, such as Anuvaad, are built using Bhashini's building blocks to enable market players to imagine potential use cases.<sup>20</sup>

**Reference Apps** 

3

#### **BHASHINI Language Functions** Speech Recognition Language and speaker identification, speech-to-text output and Transcription Instantaneous, accurate machine translation **Translation** for multiple languages Input in standard Roman text with output in native script **Transliteration** of language Understanding meaning for actions like Language **Understanding** answering queries, sentiment analysis

Speech output of text in language

**Building Blocks of Bhashini** 

and speaker (male/female) of choice

**Speech Synthesis** 

(Text to Speech)



#### **Functionalities of ANUVAAD**

Automatic language recognition for both voice and text input

Allows Roman and native keyboard inputs for all languages

Accurate, real-time, contextual translation across 23 languages for conversation and formal document purposes

Clear machine-generated customisable speech output that can be saved for future references

#### **Market Applications**

#### Translation of Educational Content

DIKSHA

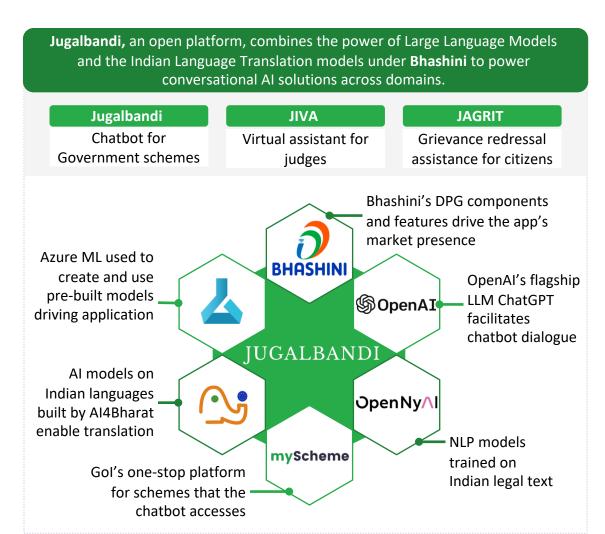
**SUVAS** 

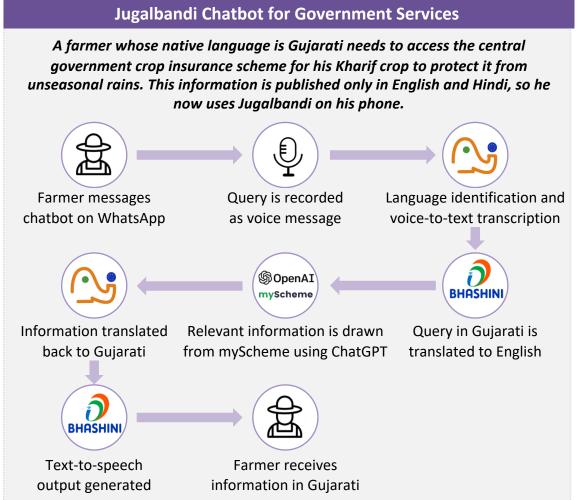
Document Translation for Court Proceedings

### Solution applications, such as Jugalbandi are created by private market players to drive innovation in enhancing service delivery.<sup>21</sup>

Marketplace

5





# **USE CASES**



### Bhashini's capabilities open up use cases across sectors such as education, livelihoods and financial inclusion (1/2).

#### Some capabilities to be built under Bhashini



Automatic Speech Recognition Software

- Processes human conversations into readable text
- An ASR software for Indian English and 11 Indian languages has been developed.



Text-to-speech Synthesis System

- Converts normal language text to speech
- M-Vachak has been made available on Android-based OS and Indus OS



Optical Character Recognition System

- Converts an image of a text to machine-readable text format
- An OCR system has been developed for 13 Indian languages/scripts



Machine Translation

- Automatically converts one language to another without human intervention
- MT engines have been developed to connect English with other Indian languages



**Input tools** 

- Allows entry of text in multiple languages
- Input tools will allow the creation of keyboard interfaces in Indian languages



Named Entity Recognition

Detects and categorises important information in a text



Sentiment Analysis

Analyses the emotion or intent behind a text



Question Sharing

Retrieves answer to a question in a given text in order to respond to search queries



**Summarisation** 

Synthesises large amount of information to their most important parts

### Bhashini's capabilities open up use cases across sectors such as education, livelihoods and financial inclusion (2/2).

Livelihoods		Education		Financial Inclusion		Health	
XIA XIII	Enables real-time translation of web content across skilling platforms and job portals	文门户	Seamless translation of online learning content for children		Helps synthesise information in complex financial documents in multiple Indian languages		Allows patients to resolve health-related queries by typing in their native languages
	Faster translation of <b>job notifications</b> in multiple languages. For example, information available in English can be accessible in Marathi	Q A	Connects teachers with parents speaking different languages to increase parent engagement. For example, a chatbot to answer parents' queries in local languages	QYA	Enables <b>quick redressal of queries</b> in native languages		Real-time translation of multilingual speaker during online health consultation. For example, Tamil-speaking doctor talking to a patient in rural Bihar.
	Multilingual information sharing of schemes and services for migrant workers and farmers	[OCR]	Ability to digitalise and give grades to student responses in OMR sheets		Enable <b>remittance</b> by giving local language voice command	[OCR]	Easy digitalisation of hospital documents and records from district and state hospitals

### Bhashini enables foundational learning and advancement among children through resources available in native languages.

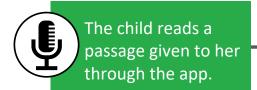


Name: Silme Momin

Age: 8 years Grade: Third

**Location:** A village in Meghalaya **Aim:** Learn English and strengthen FLN

Silme is a Grade 3 student studying in the government school in her village. Her recent **FLN evaluation puts her below the average reading level** for her grade. So, her parents and teachers suggested the FLN enhancement application that is built upon Bhashini's infrastructure.







The app assesses and analyses the words that were mispronounced, missed or half-spoken.



The app gives the result and a targeted set of words for the child to practice.

#### Challenges and opportunities enabled by Bhashini

#### **Existing Challenges**

- Restriction to educational resources made available by schools, often unavailable in native languages
- Lack of accessible learning material causes disinterest among children, thereby causing deficit in learning levels.
- Inability to provide targeted attention to students in large-scale, public school classrooms causes students to fall back and eventually drop out.

#### Opportunities opened by Bhashini

- Enables access to existing learning content and applications in native languages by translation of web content
- Targeted and customised learning material for students to generate interest in learning
- Self-paced learning through accessible learning material develops FLN abilities and leads to the overall improvement of country's FLN status.

#### Bhashini enhances business opportunities for small entrepreneurs through multilingual conversations.



Name: Anthuvan Sethupathi **Profession:** Artisan (handlooms)

**Location:** Tamil Nadu Languages Known: Tamil

Aim: Sell his products across India

Anthuvan participated in an exhibition in Delhi, and had a stall displaying all his products. Several individuals came to the stall to enquire about them. But the customers didn't know Tamil and Anthuvan didn't know Hindi or English.







App Z recognises the Hindi words spoken by the buyer and converts and gives a voice output in Tamil.





The process repeats and gives the answer to the customer in Hindi.

#### Challenges and opportunities enabled by Bhashini

#### **Existing Challenges**

- Sellers face difficulties in communicating with buyers who do not speak their native languages
- Reduced business opportunities for small entrepreneurs unable to widen reach for their products.

#### Opportunities opened by Bhashini

- Provides a language translation software, enabling communication between buyers and sellers speaking different languages.
- Enables access to e-commerce opportunities for small entrepreneurs through accessible and comprehensible web services.

#### Bhashini enables access to teleconsultation for those with no knowledge of English.



Name: Sunita Tiwari Profession: Labourer

**Status:** 22 weeks pregnant **Location:** Jamshedpur

Sunita has been facing severe discomfort. Maternal mortality rates and neonatal mortality rates are high in her area, so she does not want to be lax. Her local clinic is overburdened and understaffed. She has been unable to get an appointment for two weeks.



The application connects her with a doctor in Mumbai and forwards her query in English.



The doctor provides consultation and prescribes medicine.



It reaches
Sunita in
her
regional
language.



Follow-ups are scheduled and Sunita receives notification in their regional language

#### Challenges and opportunities enabled by Bhashini

#### **Existing Challenges**

- Patients are unable to access specialists and doctors who speak different languages, thereby limiting access to treatment
- Teleconsultation is inaccessible due to its functions being available solely in English

#### **Opportunities opened by Bhashini**

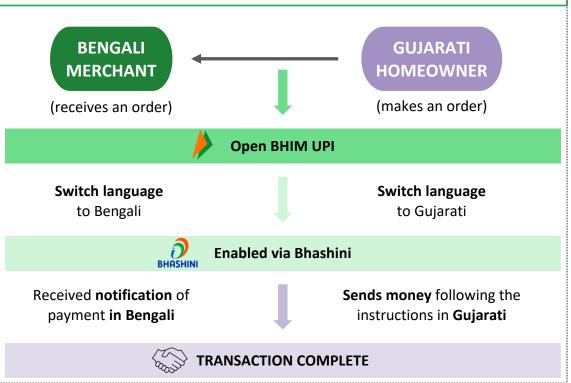
- Teleconsultation facilities are available in native languages, thereby widening access to healthcare services
- Provides faster real-time translation services to facilitate conversations between patients and doctors speaking different languages

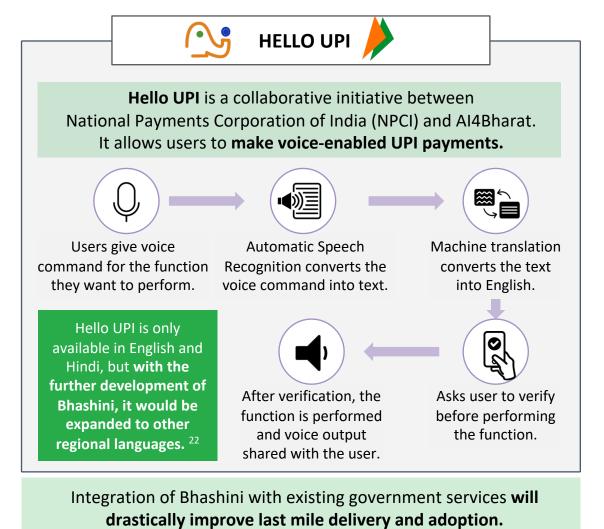
### Bhashini enables financial inclusion by making payment services, such as BHIM UPI, available in multiple Indian languages.



**Bhashini** has been integrated with BHIM UPI, making the service available in 20 local Indian languages.

A Gujarati homeowner has placed an order with a Bengali merchant. Both of them only know their regional languages. So, they use the language feature of BHIM-UPI to make the payment.

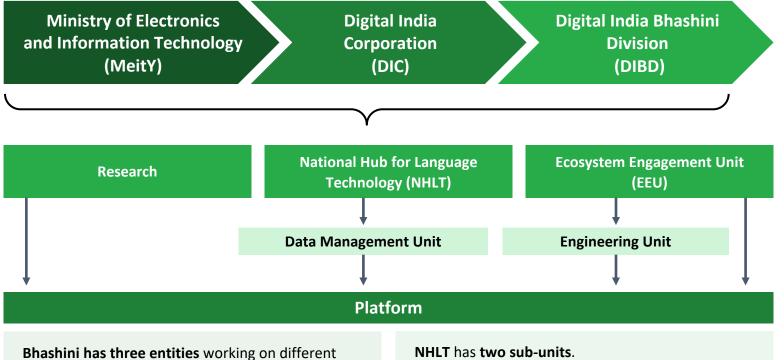




# GOVERNANCE AND IMPLEMENTATION



### The Ministry of Electronics and Information Technology steers the activities of Bhashini, and facilitates its smooth implementation.



**Bhashini has three entities** working on different objectives.

- Research works towards the development of language technology.
- **NHLT** is the foundational layer that develops tools and processes for training and datasets.
- EEU coordinates inter-ministerial and international collaborations.
- The Engineering Unit works towards operationalising and maintaining the foundation of the Bhashini platform through software development agencies and an in-house team.
- **DMU** creates an efficient pipeline to manage data collection and curation processes, and also assists in the creation and validation of each dataset.

- MeitY has two committees: Apex and Executive, which work towards setting the goals, and monitoring the functioning of Bhashini respectively.
- DIC has been set up by MeitY to assist the Digital India Programme.<sup>23</sup>
- DIBD, under DIC, acts as the nodal agency to develop and maintain a public digital platform that can create and nurture an ecosystem.<sup>24</sup>

#### **Key People**



MR. AMITABH NAG CEO, Digital India Bhasini Division (DIBD)

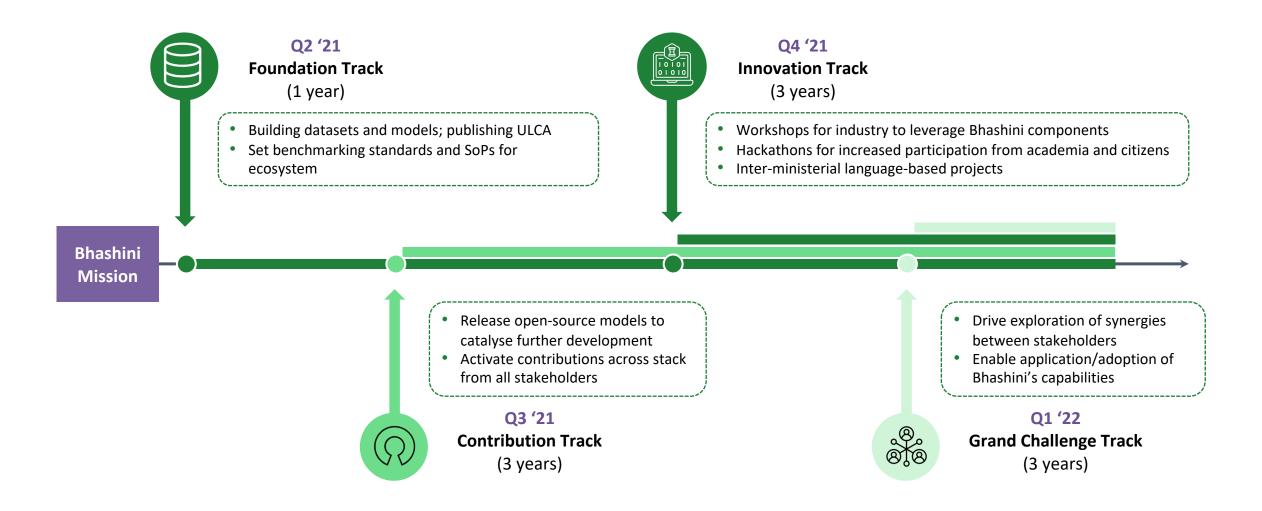


MR. ABHISHEK SINGH MD & CEO, Digital India Corporation (DIC)



**DR. SWARAN LATA**Advisor, NLTM, MeitY

### Bhashini is a three-year mission which has been segregated across different tracks comprising a set of activities



#### Open datasets and models are built through collaborative ventures among stakeholders of the Bhashini ecosystem, Al4Bharat is an example of such a collaboration.



#### **AI4BHARAT**

**Applications** 

AI-based tools leveraging the ML models

Al4Bharat, or Artificial Intelligence for Bharat <sup>25</sup>, focuses on building open source datasets, models and applications in Indic languages by enabling an innovation ecosystem.

#### **Datasets**

Public language corpora across 22 scheduled languages

**14 datasets** – both for training and evaluation purposes – have been created so far

#### Models

Trained on language datasets

**10 models** across various linguistic capabilities have been trained on the existing datasets



#### **Translation**

**Optimises** web content

#### Text-to-Speech

Optimises assistive technologies

#### **Transliteration**

*Increases variety of* text input methods

#### Language **Understanding**

Enhances user-application through diversity

Generates natural language

Functions enabled by resources

**Language Generation** 

texts for Indian users



**Interdisciplinary** network of research experts and technology providers



• Anuvaad: Document-translation application

• Chitralekha: Video transcription application

• **Shoonya:** Efficiency-improving interface

#### NANDAN NILEKANI

Co-Founder of Infosys; **Primary Sponsor for** AI4Bharat



Supports research activities through unrestricted grants



Supports mentorship and software engineering



Provides access to technology and researchers



**Speech Recognition** 

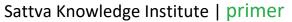
Enables access for

*lower literacy users* 

Sign Language

Enables access for

disabled population



# ROLE OF STAKEHOLDERS

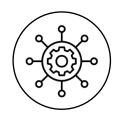


#### A multi-stakeholder engagement is required among philanthropic organisations, technology service providers and non-profits to strengthen the Bhashini ecosystem and foster innovation.



#### **PHILANTHROPIC ORGANISATIONS**

- Support and organise innovation competitions, such as hackathons, to encourage new, inclusive products in the Bhashini ecosystem
- Provide technical support to the government in building privacy frameworks and safeguards for Bhashini
- Strengthen ICT infrastructure and know-how among non-profits which will enable them in onboarding users on Bhashini-integrated applications



#### **TECHNOLOGY SERVICE PROVIDERS**

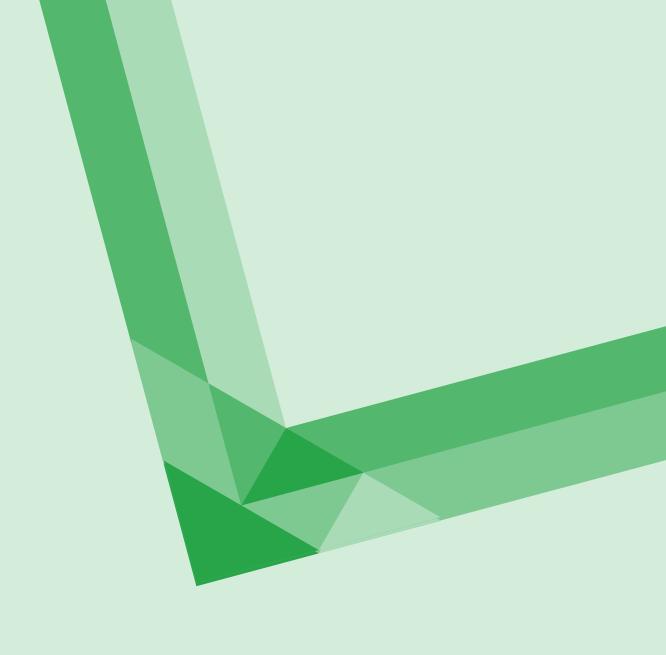
- Build context-specific applications by leveraging the Bhashini architecture to widen reach, and solve for last-mile delivery issues
- Assist in managing and curating language datasets, ensuring data quality, security and compliance with privacy regulations
- Assist organisations and developers in **integrating components** of the Bhashini infrastructure into existing systems and applications, ensuring seamless adoption



#### **NON PROFITS**

- Assess specific requirements and preferences of end-users and beneficiaries to help align tech products built in the Bhashini ecosystem
- Engage with multiple stakeholders on best practices of using DPGs in the social sector, to influence collaboration among users and developers
- Assist in the deployment of new technologies by coordinating pilot programmes with small user groups.

# REFERENCES



- Language Atlas, Census of India, 2011.
- 2. Ethnologue (2017 20th edition), viewed on 25 September 2023.
- SKI research.
- 4. Ministry of Labour and Employment, <u>Annual Periodic Labour Force Survey</u>, <u>2018-19</u>.
- 5. Government of India, Economic Survey 2016-17.
- 6. IIT- Madras, Al4Bharat, viewed on 22 September 2023.
- 7. Piramal Foundation, Aspirational Districts Collaborative, viewed on 25 September 2023.
- 8. Department of Justice, National Judicial Data Grid, viewed on 25 September 2023.
- 9. The Centre for Internet and Society, Accessibility, viewed on 25 September 2023.
- 10. DPG Alliance, <u>DPG Registry</u>, viewed on 22 September 2023.
- 11. Morin Chris 2019, Research team developing Indigenous languages app, viewed on 25 September 2023.
- 12. Masakhane, viewed on 22 September 2023.
- 13. Translation Commons, Global Language Digitization Initiative, viewed on 25 September 2023.
- Welsh Government 2021, Bilingual technology toolkit for good user experience.
- 15. UNESCO, Digital Empowerment Driving the International Decade of Indigenous Languages, viewed on 25 September 2023.
- 16. Ministry of Electronics and Information Technology, Bhashini, viewed on 22 September 2023.
- 17. Ministry of Electronics and Information Technology, National Language Translation Mission Whitepaper, viewed on 22 September 2023.
- 18. Mukherjee and Maruwada 2021, Fast-Tracking Development: A Building Blocks Approach for Digital Public Goods, Center of Global Development.
- 19. Ministry of Electronics and Information Technology, Bhashini, Bhashadaan, viewed on 25 September 2023.
- 20. Anuvaad, viewed on 25 September 2023.
- 21. Jugalbandi, viewed on 22 September 2023.
- 22. Ray, A 2023, Hello, UPI: Use voice commands to send money, pay bills, The Economic Times, viewed on 22 September 2023.
- 23. Ministry of Electronics and Information Technology, <u>Digital India Corporation (DIC)</u>, viewed on 22 September 2023.
- 24. Digital India Corporation, Digital India Bhashini Division, viewed on 22 September 2023.
- 25. Ibid, 5.

