

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 adaptations 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 <http://creativecommons.org/licenses/by-nc-sa/4.0/>

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.

AI4Bharat—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.



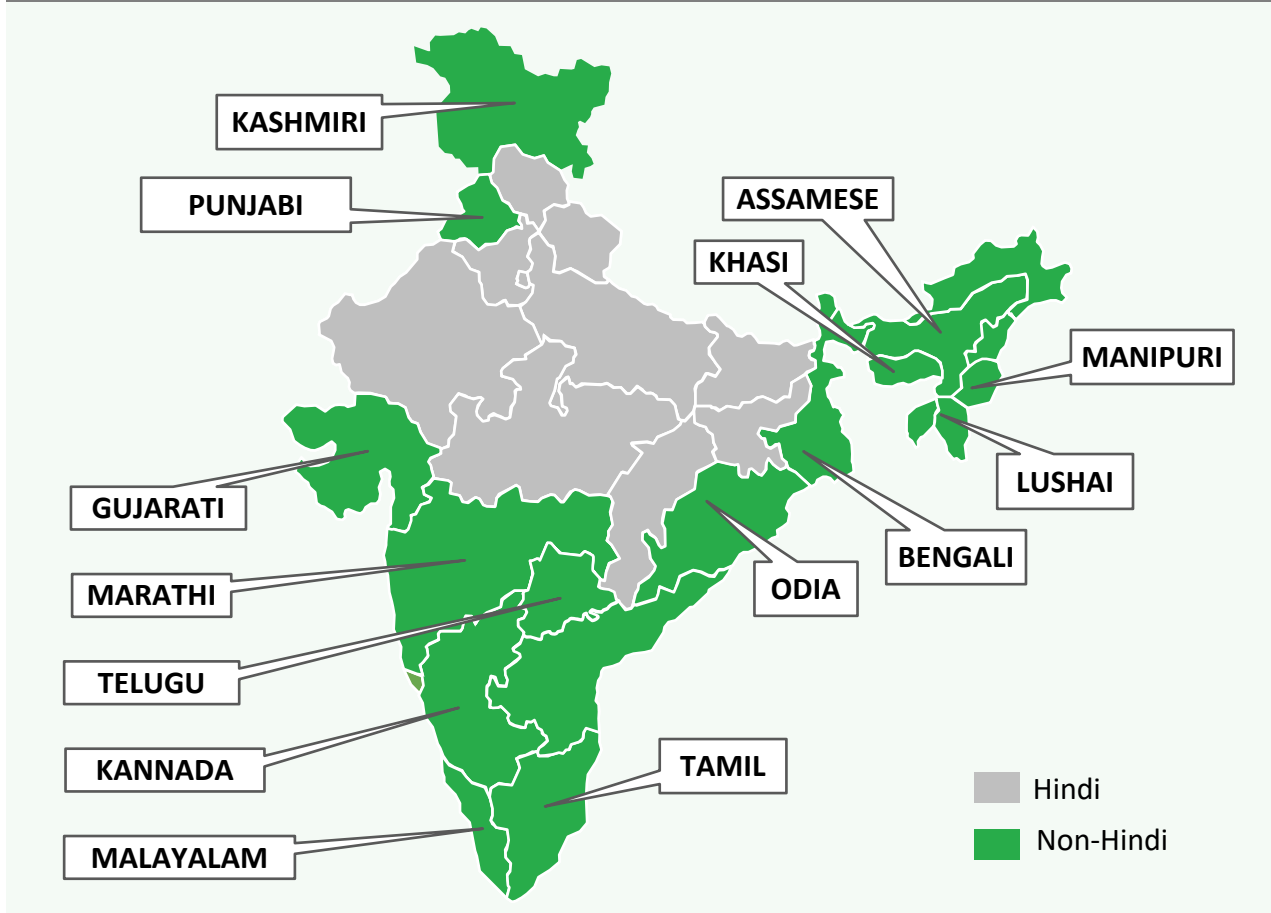
BACKGROUND

मराठी म्हाळी हिन्दी
तेलुगु లిపి മലയാളം
ગુજરાતી
भाषा বাংলা
ಕನ್ನಡ ಸಂಸ್ಕೃತम्
விக்கிப்பீடியா

मराठी म्हाळी हिन्दी
तेलुगु లిపి മലയാളം
गुजराती
भाषা

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

56.4% of the entire population speaks non-Hindi languages as their primary language.



121

Major languages

270

Mother tongues

18

Written Scripts¹

0.914

Linguistic diversity index



- **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.²

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 English³

<33%

of **informal labour** have access to government services.⁴

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**.⁶



914 non-profits

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



>85% cases

are **pending** due to delayed transfer of information from lower (vernacular) to higher (Hindi/English speaking) courts.⁸

Difficulty in interpretation and communication for disabled population

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



20%

Movement Disability



19%

Hearing Disability



19%





Visual Disability

OVERVIEW OF BHASHINI



BHASHINI

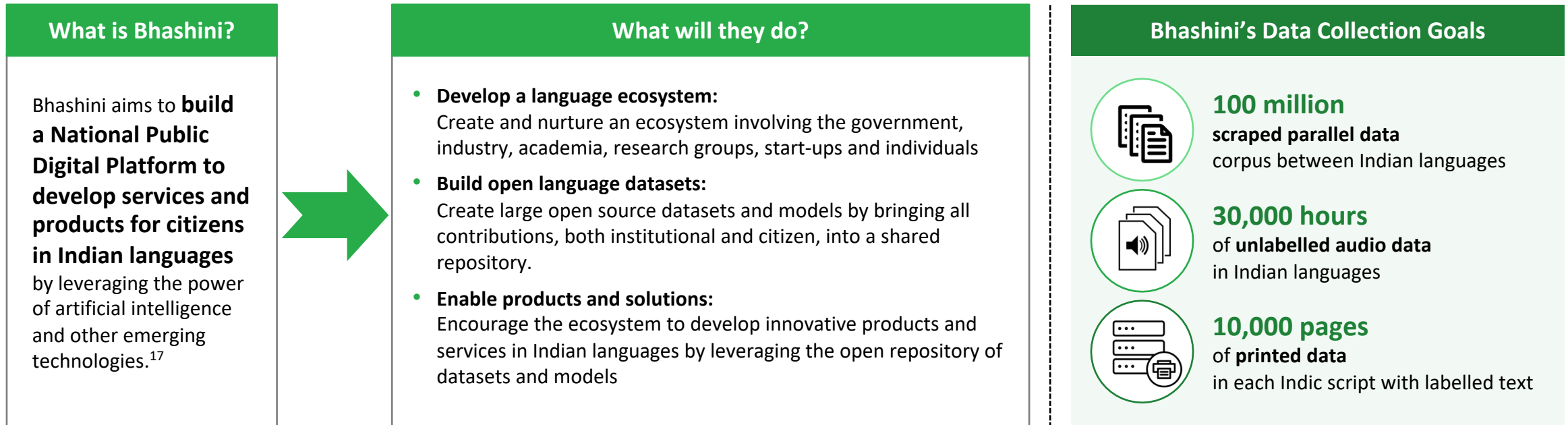
Language-based DPIs and DPGs are being deployed across the world to **optimise service delivery** and enhance the **inclusion of languages in the digital space**.

| | | | Developed in | Deployed in | |
|---|---|---|---|--------------|--------------------|
| 1 | Digital Public Goods (DPGs) ¹⁰ |  Story Weaver | Open and free-access, quality storybooks in multiple languages | India | India |
| | |  Open Foris Collect | Survey setups with an user-friendly interface in a multi-language environment | Italy | Globally |
| | |  LibreOffice | Open-source office suite in 120 languages | Globally | Globally |
| | |  African Storybook | Open access to picture storybooks in African languages | South Africa | Sub-Saharan Africa |

| | | | | | | | |
|---|-----------------------|--|---|--|-------------------------------|--|--|
| 2 | Open Access Resources | Wicêhtowin project¹¹ | <ul style="list-style-type: none"> 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 |  | Masakhane¹² | <ul style="list-style-type: none"> 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 Digital¹³ A guide to bring a language online | Welsh Bilingual Tech Toolkit¹⁴ Toolkit for bilingual software for developers | Inclusion of Inuktitut¹⁵ (an indigenous Canadian language) on Facebook | Motorola's phone¹⁶ 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.

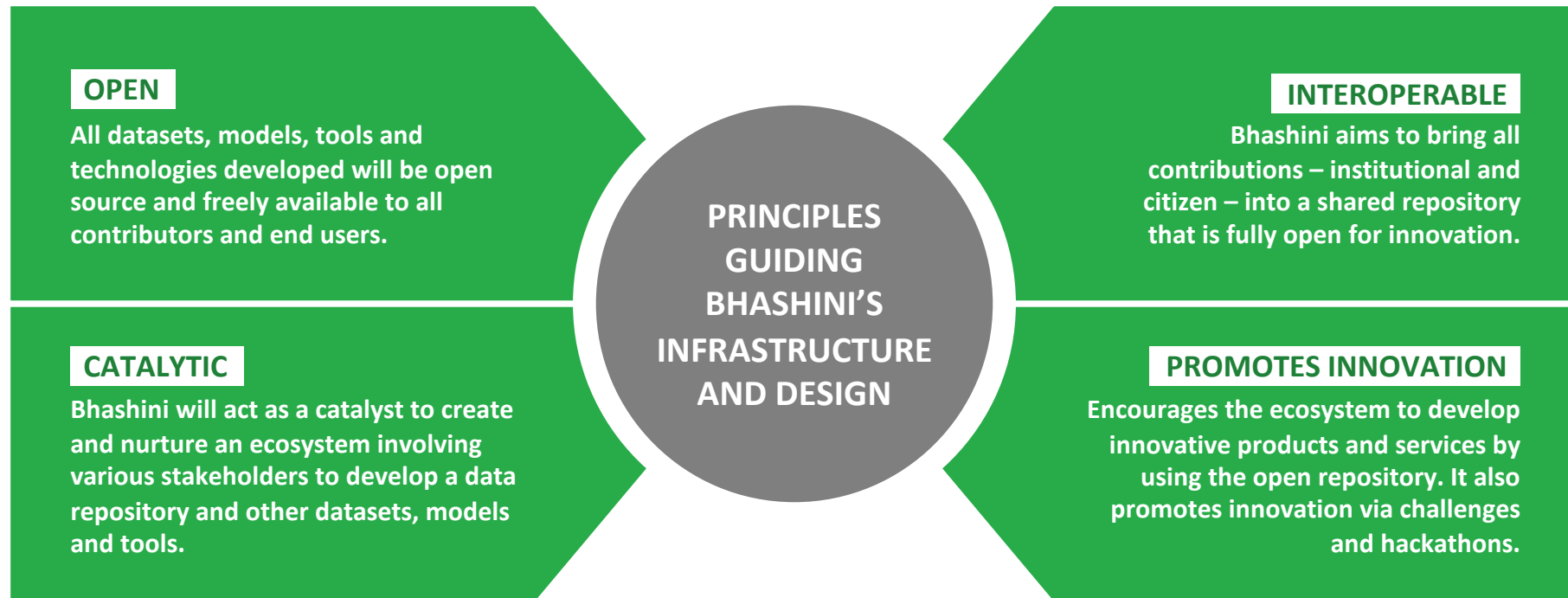


- Bhashini is managed by the Ministry of Electronics & Information Technology (MeitY), Government of India, with extensive support from Technology Development for Indian Languages (TDIL).
- Bhashini takes inspiration from other mature and emerging digital public infrastructure such as Unified Payments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

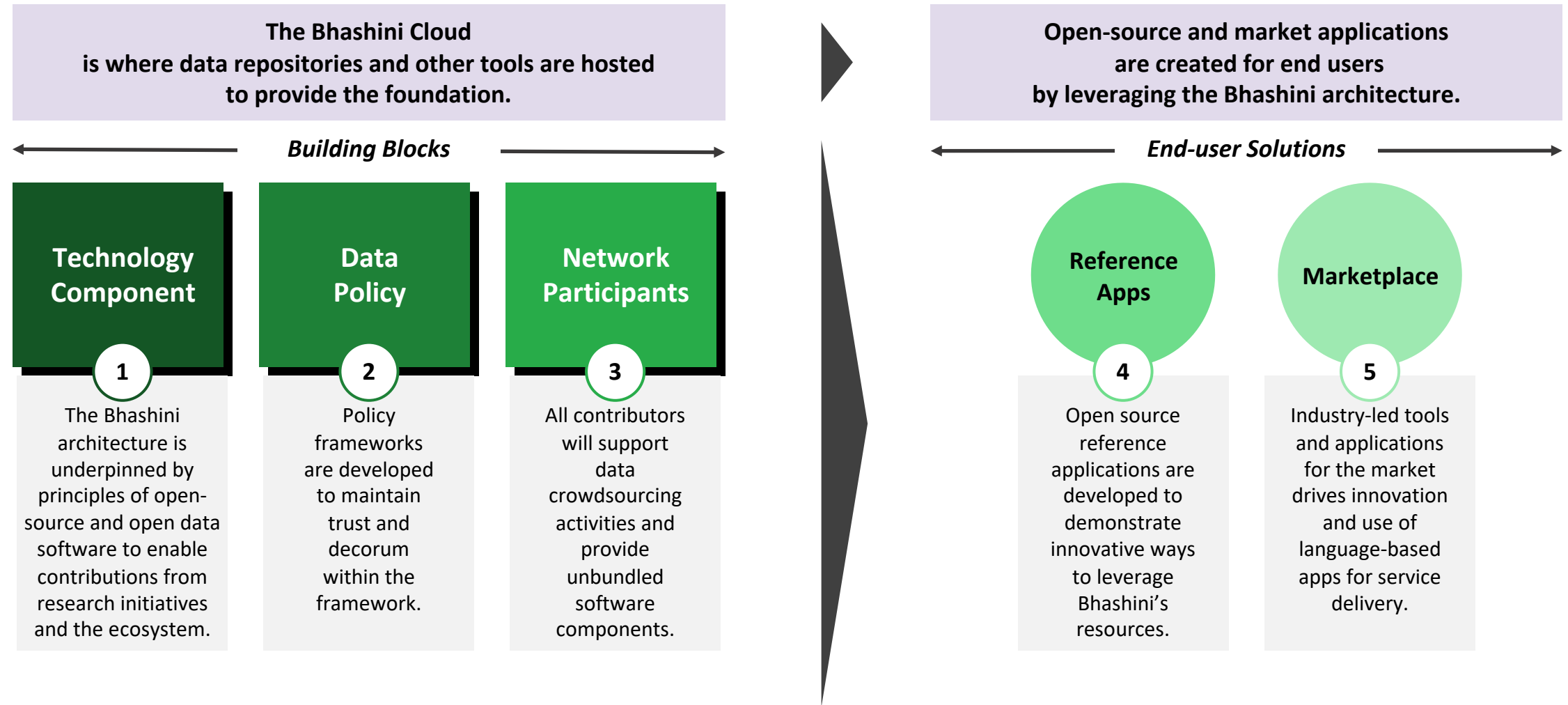
Bhashini enables access to the internet in Indian languages and facilitates last-mile service delivery.

| Key Challenges | How will Bhashini solve it? |
|--|--|
| LIMITED ACCESSIBILITY of social welfare schemes online | <ul style="list-style-type: none"> 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 |
| DELAY IN DOCUMENT TRANSLATION in various languages | <ul style="list-style-type: none"> 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 |
| DIFFICULTY IN INTERPRETATION AND COMMUNICATION for disabled population | <ul style="list-style-type: none"> 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.¹⁷

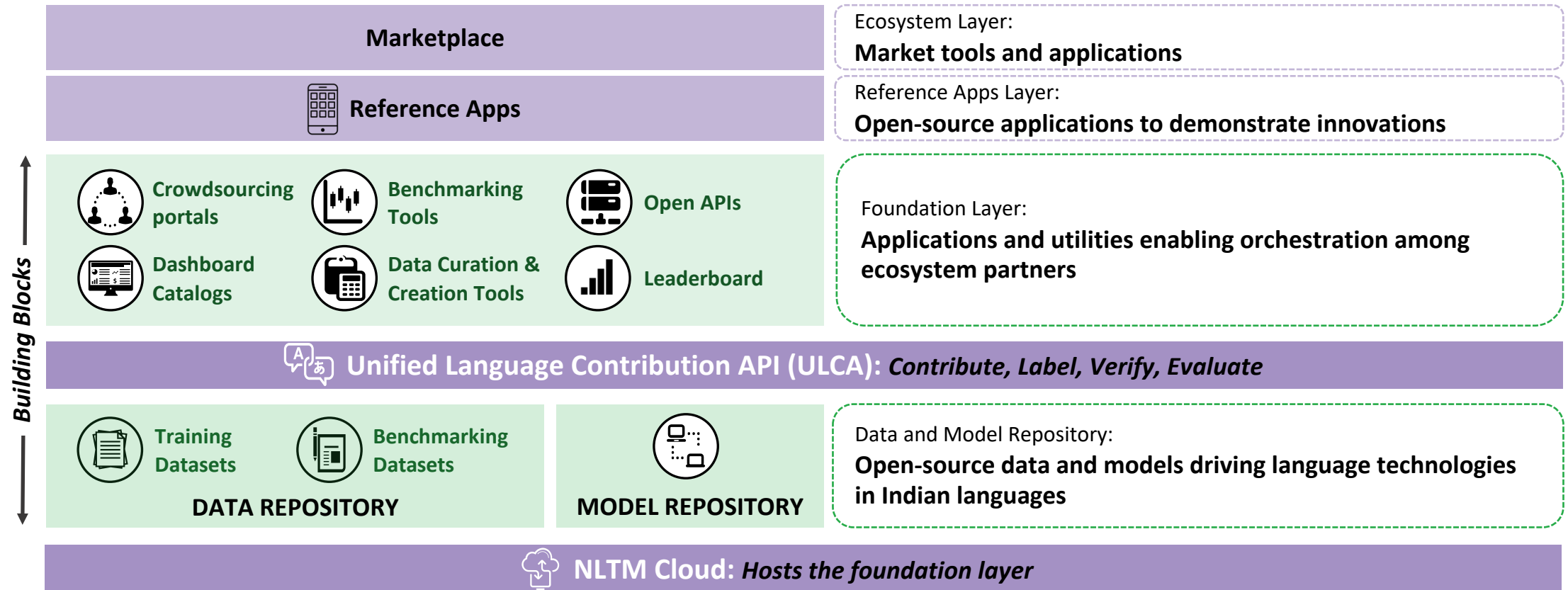


Three fundamental building blocks – technology, regulations and ecosystem – build the Bhashini infrastructure.



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

Technology Component 1



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.

Bhashadaan ¹⁹



Type audios
and validate
transcriptions



Record
sentences and
validate audios



Translate texts
and validate
translations



Label images
and validate
descriptions

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.



Examples of Global, Open-source Language Models



GPT 3.5
by
OpenAI



Bloom
by
Big Science



PaLM
by
Google



Dolly
by
Databricks

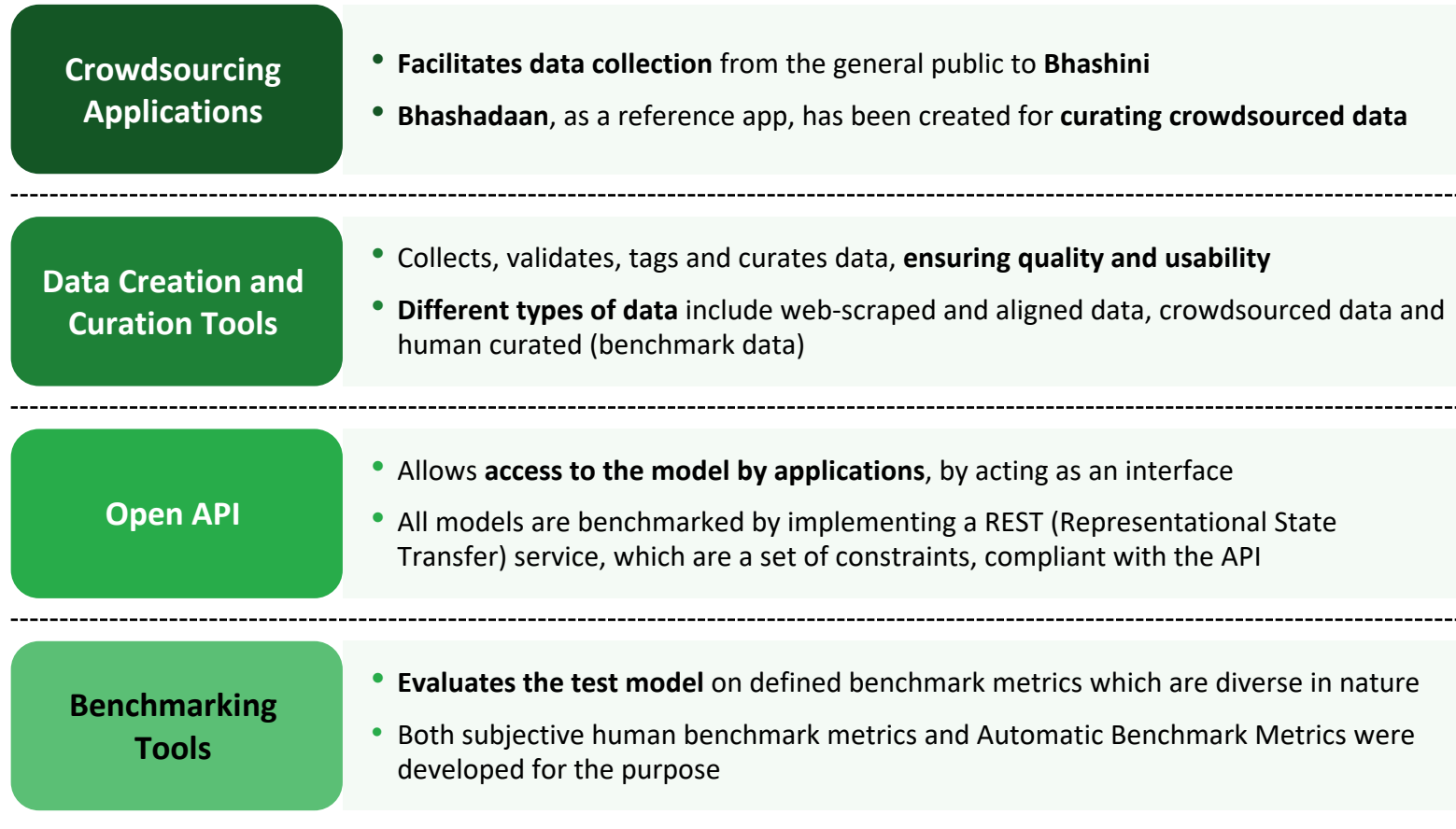


Cerebras
GPT



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

Technology Component 1



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

Data Policy 2

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.

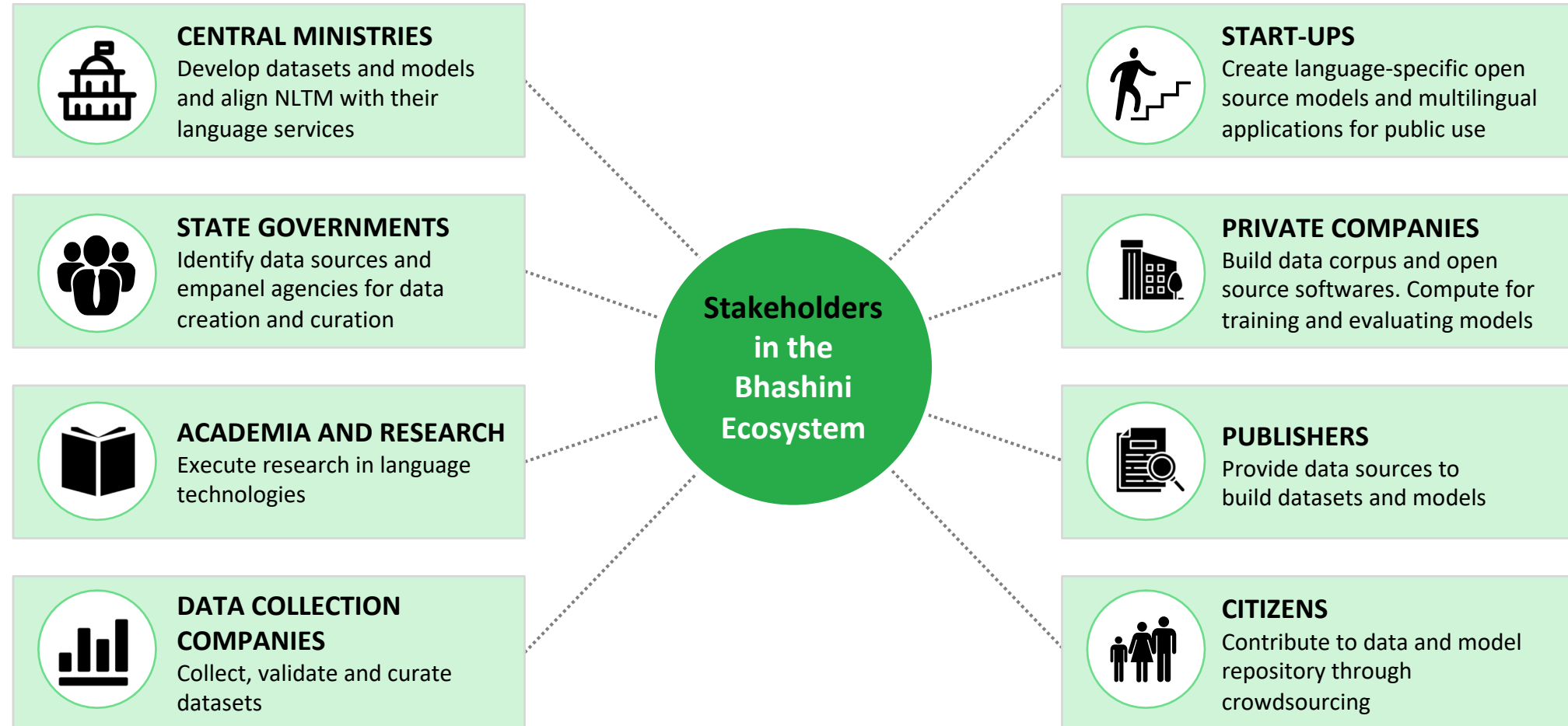
A **linguistic data policy** will also be formulated for NLTM and other such initiatives.

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

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

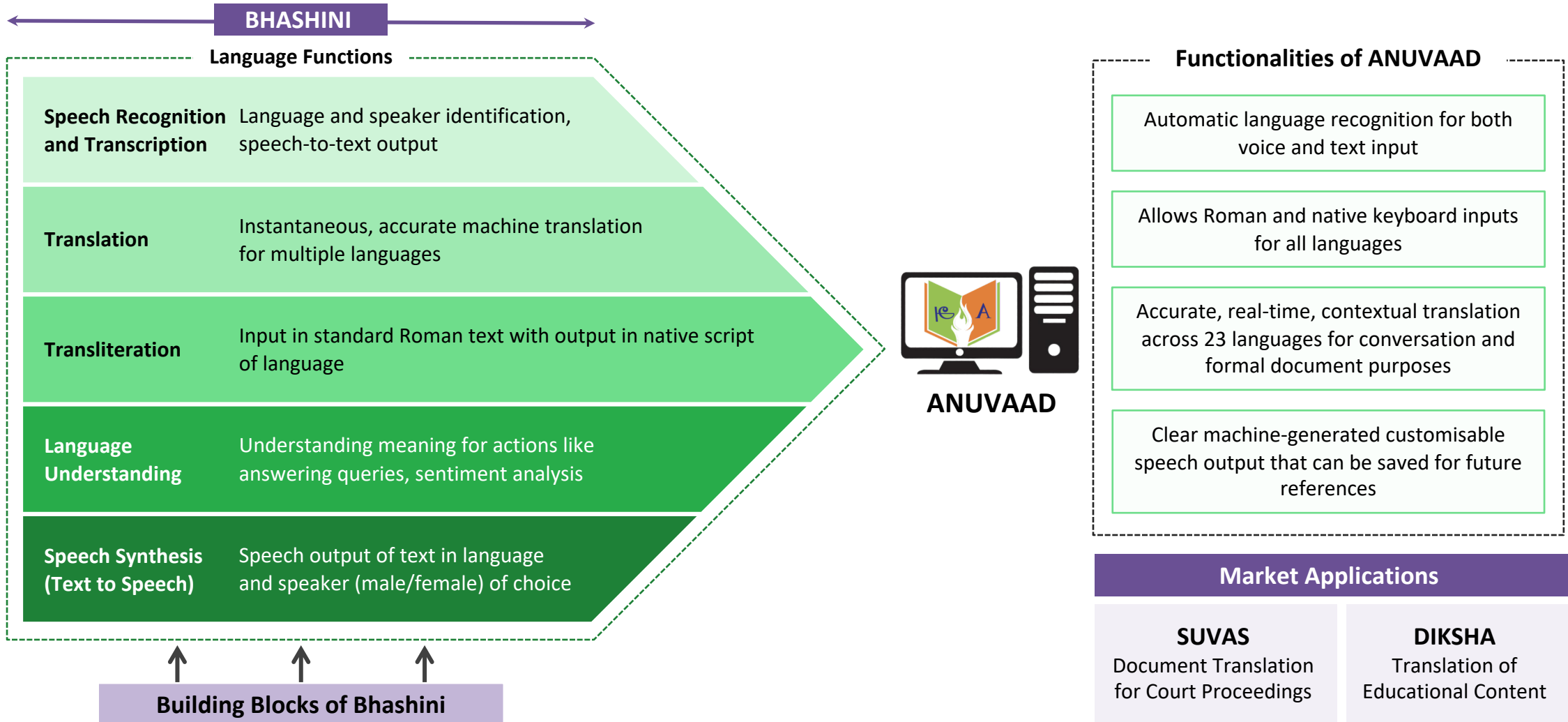
The Bhashini architecture is catalysed and optimised through contributions of **several stakeholders** in the ecosystem.

Network Participants **3**



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

Reference Apps 3



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

Marketplace 5

Jugalbandi, an open platform, combines the power of Large Language Models and the Indian Language Translation models under Bhashini to power conversational AI solutions across domains.

Jugalbandi

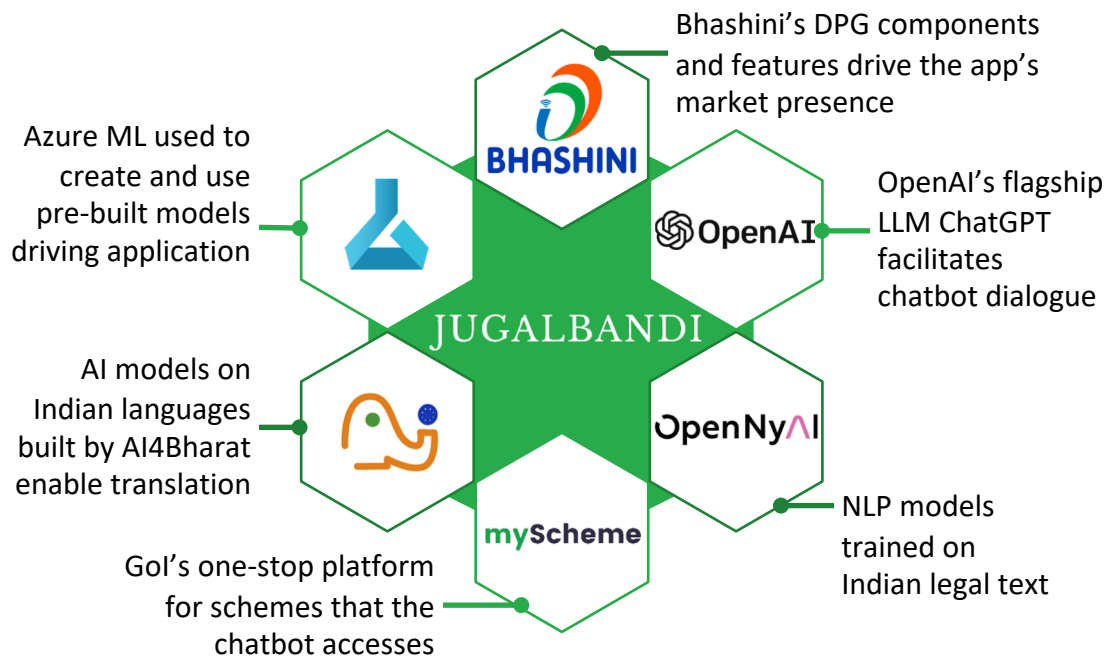
Chatbot for Government schemes

JIVA

Virtual assistant for judges

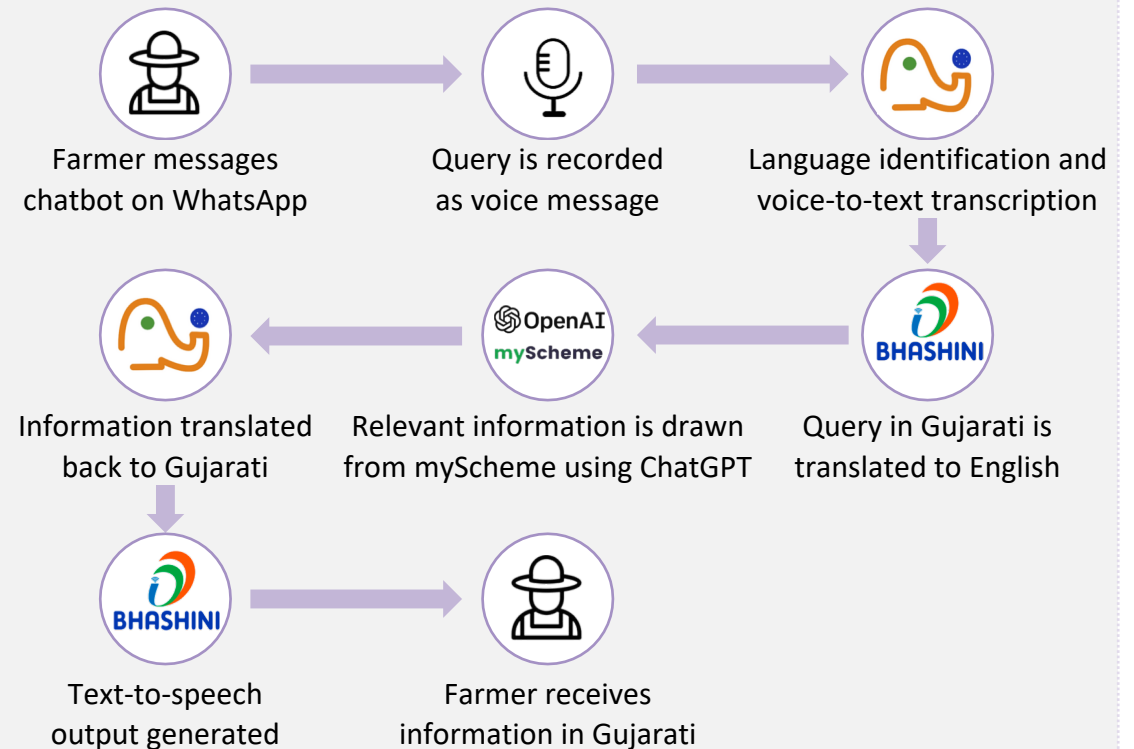
JAGRIT

Grievance redressal assistance for citizens



Jugalbandi Chatbot for Government Services

A farmer whose native language is Gujarati needs to access the central government crop insurance scheme for his Kharif crop to protect it from unseasonal rains. This information is published only in English and Hindi, so he now uses Jugalbandi on his phone.

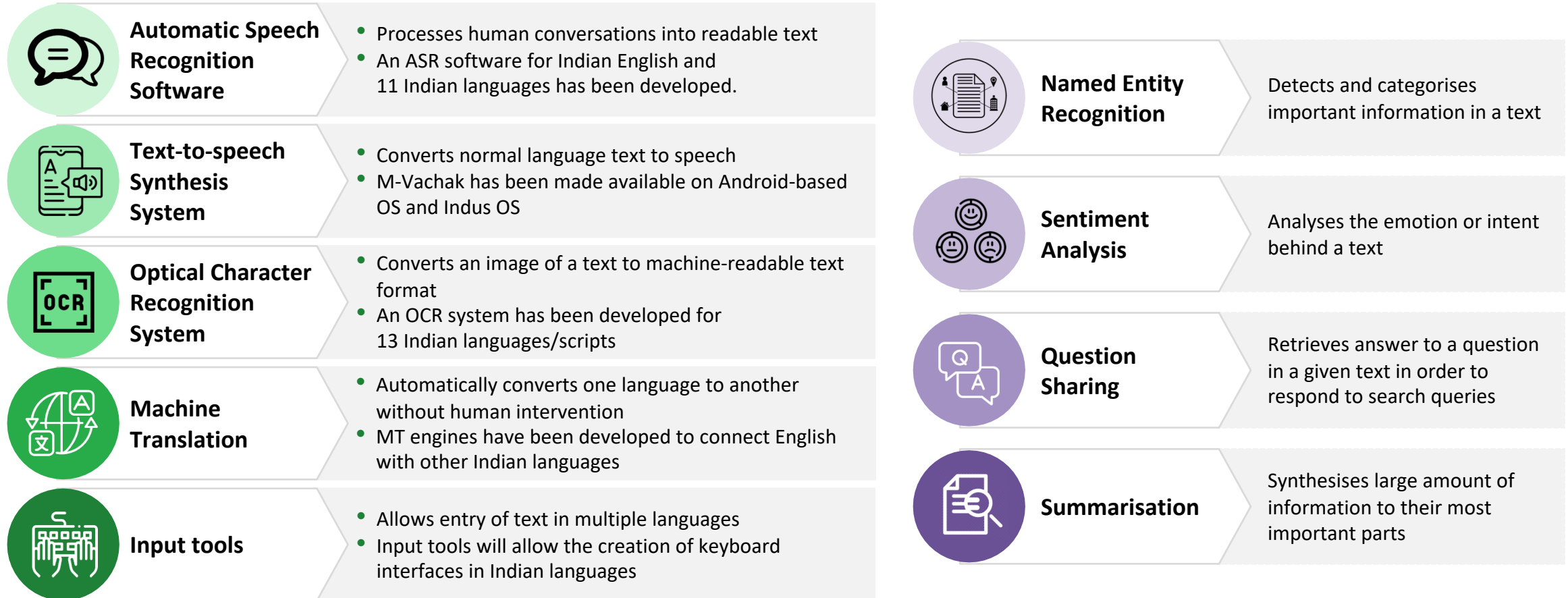


USE CASES















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



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

| Livelihoods | | Education | | Financial Inclusion | | Health | |
|---|--|---|---|---|---|---|--|
|  | Enables real-time translation of web content across skilling platforms and job portals |  | 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 job notifications in multiple languages. For example, information available in English can be accessible in Marathi |  | Connects teachers with parents speaking different languages to increase parent engagement . For example, a chatbot to answer parents' queries in local languages |  | Enables quick redressal of queries 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 |  | Ability to digitalise and give grades to student responses in OMR sheets |  | Enable remittance by giving local language voice command |  | Easy digitalisation of hospital documents and records from district and state hospitals |

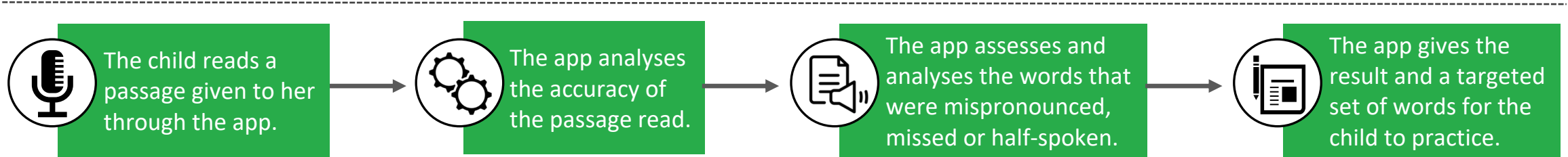
Bhashini enables foundational learning and advancement among children through resources available in native languages.

PROFILE



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.



Challenges and opportunities enabled by Bhashini

| Existing Challenges | Opportunities opened by Bhashini |
|--|--|
| <ul style="list-style-type: none">• 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. | <ul style="list-style-type: none">• 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. |

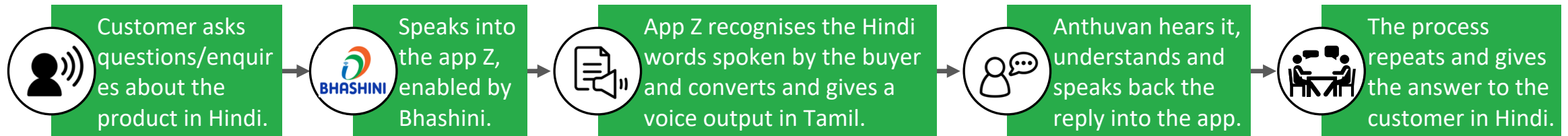
Bhashini enhances business opportunities for small entrepreneurs through multilingual conversations.

PROFILE



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.**



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.

PROFILE



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.



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

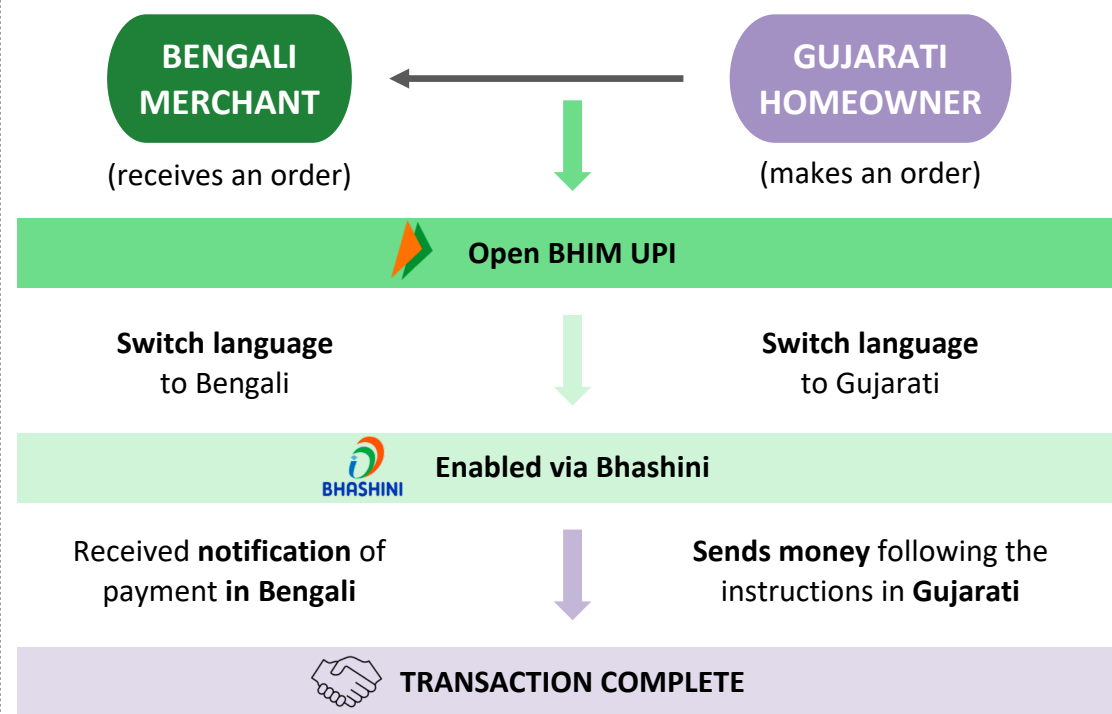
- Teleconsultation facilities are available in native languages, thereby widening access to healthcare services
- Provides faster real-time translation services to facilitate conversations between patients and doctors speaking different languages

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.



HELLO UPI



Hello UPI is a collaborative initiative between National Payments Corporation of India (NPCI) and AI4Bharat. It allows users to **make voice-enabled UPI payments**.



Users give voice command for the function they want to perform.

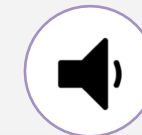


Automatic Speech Recognition converts the voice command into text.



Machine translation converts the text into English.

Hello UPI is only available in English and Hindi, but **with the further development of Bhashini, it would be expanded to other regional languages.** ²²



After verification, the function is performed and voice output shared with the user.



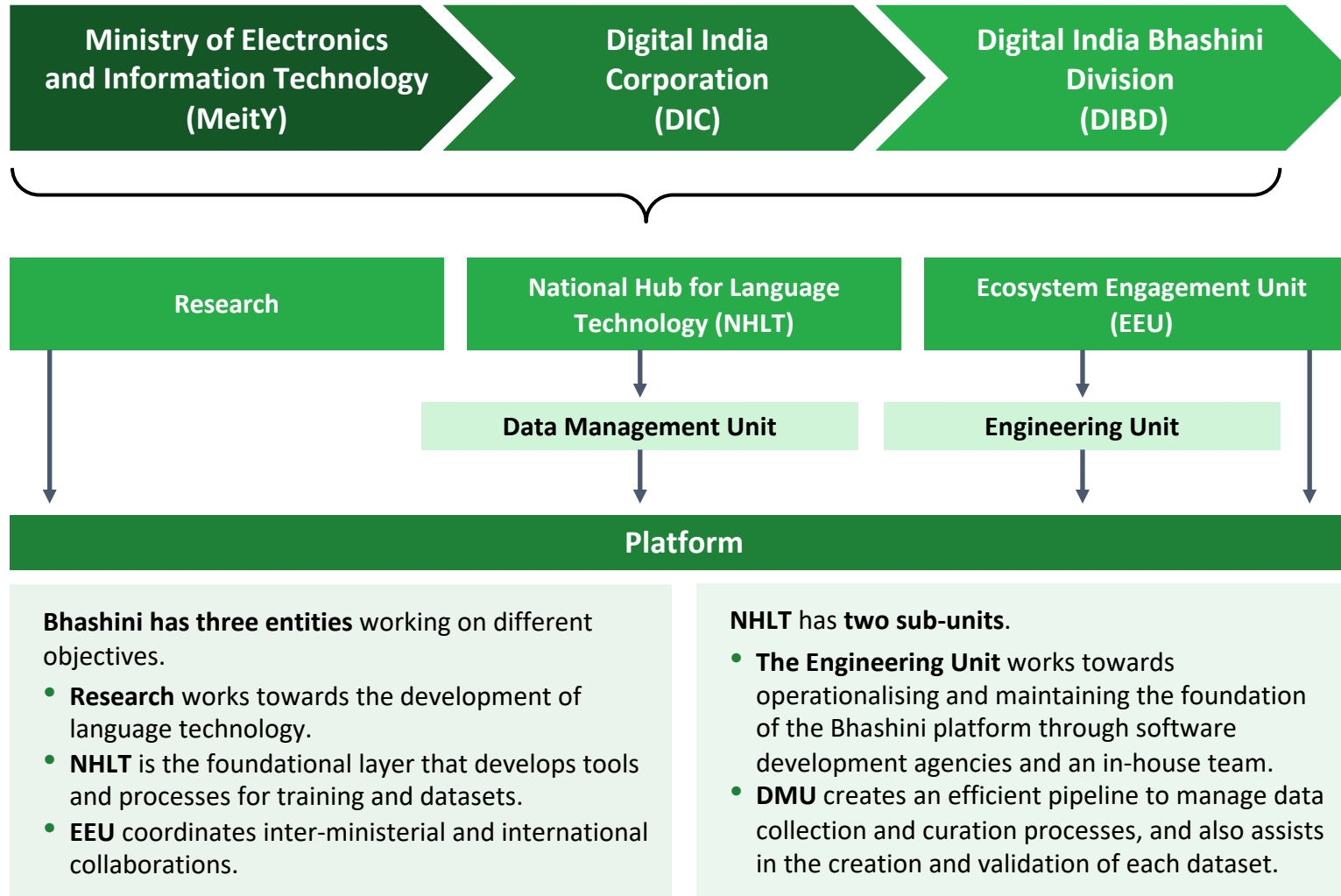
Asks user to verify before performing the function.

Integration of Bhashini with existing government services **will drastically improve last mile delivery and adoption.**

GOVERNANCE AND IMPLEMENTATION



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



- **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.²³
- **DIBD**, under DIC, acts as the nodal agency to develop and maintain a public digital platform that can create and nurture an ecosystem.²⁴

Key People



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

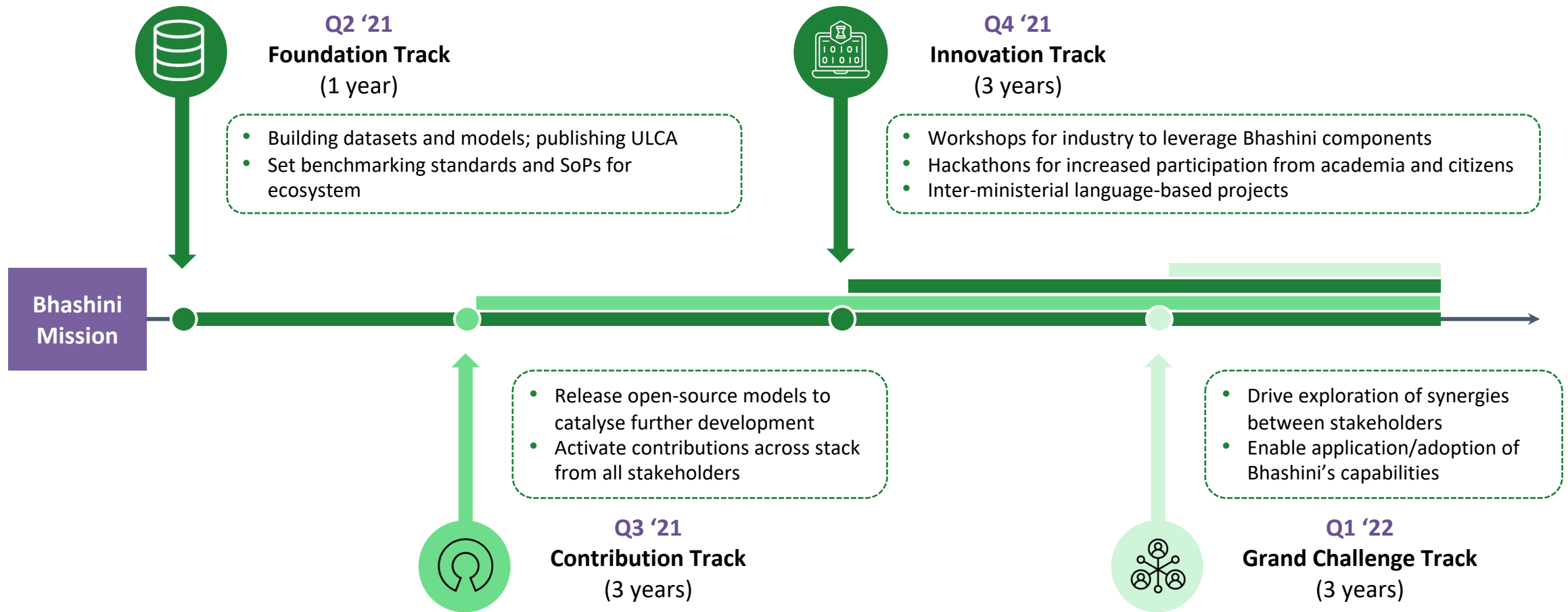


MR. ABHISHEK SINGH
MD & CEO, Digital India Corporation (DIC)



DR. SWARAN LATA
Advisor, NLTM, MeitY

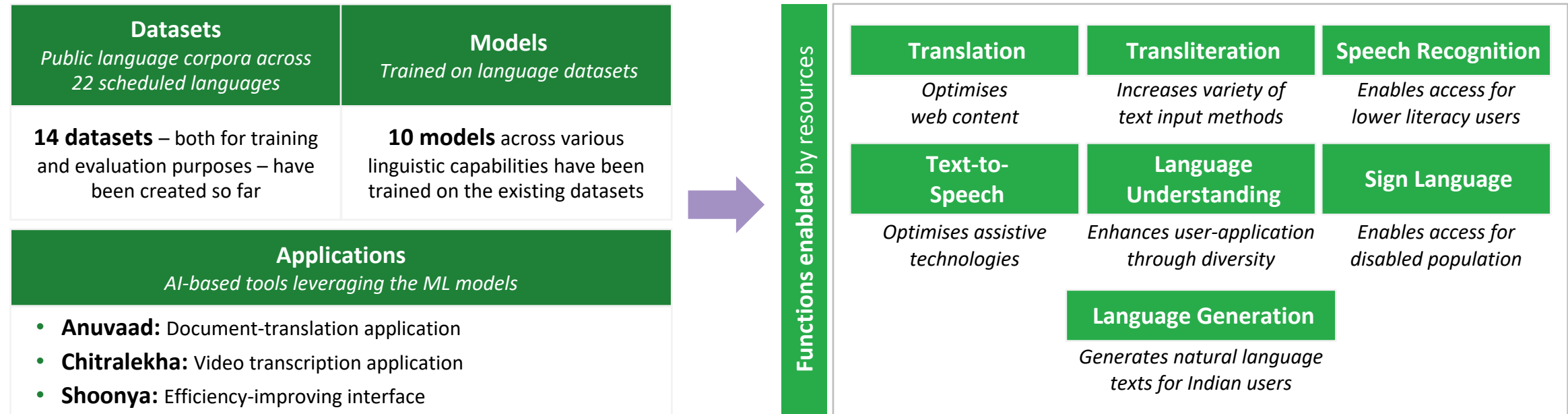
Bhashini is a three-year mission which has been segregated across different tracks comprising a set of activities



Open datasets and models are built through collaborative ventures among stakeholders of the Bhashini ecosystem, AI4Bharat is an example of such a collaboration.



AI4Bharat, or Artificial Intelligence for Bharat ²⁵, focuses on building open source datasets, models and applications in Indic languages by enabling an innovation ecosystem.



ROLE OF STAKEHOLDERS

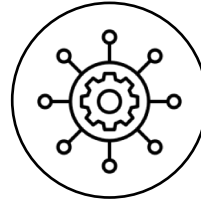


A multi-stakeholder engagement is required among philanthropic organisations, technology service providers and non-profits to strengthen the Bhashini ecosystem and foster innovation.



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



1. [Language Atlas](#), Census of India, 2011.
2. [Ethnologue \(2017 20th edition\)](#), viewed on 25 September 2023.
3. SKI research.
4. Ministry of Labour and Employment, [Annual Periodic Labour Force Survey, 2018-19](#).
5. Government of India, [Economic Survey 2016-17](#).
6. IIT- Madras, [AI4Bharat](#), 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, [DPG Registry](#), 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.
14. 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, [Digital India Corporation \(DIC\)](#), viewed on 22 September 2023.
24. Digital India Corporation, [Digital India Bhashini Division](#), viewed on 22 September 2023.
25. Ibid, 5.



