



Future of the Commons: A Conversation on Artificial Intelligence, Indian Languages and Archives

Conference Report

05-07 July, 2024 | Pune

Centre for Internet and Society, Bengaluru
and
Maharashtra Knowledge Corporation Ltd, Pune

Report prepared by Soni Wadhwa, Puthiya Purayil Sneha, Garima Agrawal & Nishant Shankar

Review and Editing by Anubha Sinha, Isha Suri, Pavan Santosh, Subodh Kulkarni & Tanveer
Hasan

Copy-editing and Design by Chris

Table of Contents

Schedule and Proceedings

Day 1

Welcome and Introduction

Conversation on the Theme of Future of the Commons

Panel Discussion 1: Framing AI and Indian Languages in Empirical and Real Possibilities of Achievable Outcomes

Panel Discussion 2: Digital Inclusion for Indian Languages: Navigating the Roadblocks

Visit to the Centre for Development in Advanced Computing, Pune

Day 2

Talk: On LLMs and Openness

Workshop 1: Indian Languages and AI

Roundtable and Conclusion

Keynote Address: Digitalisation and Archives in India

Panel Discussion 3: Digitisation and Archiving in India

Day 3

Workshop 2: Indian Languages and Archives

Share out from Workshop 2

Panel Discussion 4: Thinking of Next steps: Integration of AI, Languages and Digitization in Work Practices

Conclusion

Appendix 1: Concept Note

Appendix 2: List of Resources

Appendix 3: List of Attendees

Future of the Commons: A Conversation on Artificial Intelligence, Indian Languages and Archives

Day 1

Welcome and Introduction

In his opening remarks, Mr. Pranab Sharma, Senior Programme Coordinator, Maharashtra Knowledge Corporation Limited (MKCL) spoke about the organisation, their expectations from the event, general trends in the growth of artificial intelligence across sectors in India, concerns and opportunities in furthering better educational systems in the country.

The second set of opening remarks by Tanveer Hasan, Executive Director, Centre for Internet and Society (CIS), clarified the objectives behind the programme: to see how the discourse of AI speaks to Indian languages, and to bring people from different domains together so they can develop and apply the insights gathered from the conference to their own projects. Another objective was to foreground the idea that discourse around technology need not be a solution-based one but an inclusive one that examines the idea of technology critically. The questions that the conference participants were urged to deliberate on were: why is there a need to think about the digital commons? What is the Indian momentum to AI? How can the Indian imagination of AI add to the mainstream of global/western ideas of AI?

Open Conversation: An Introduction to the Theme of the Future of the Commons

Panellists: Anubha Sinha, Isha Suri, Puthiya Purayil Sneha & Soni Wadhwa

The panellists shared the thought behind putting together this event, especially the focus on AI in Indian languages. The aim was to foster dialogue to shape the development of technology in Indian languages, which is supposed to be an inclusive conversation among diverse stakeholders, but hasn't necessarily been so. As recent mainstream tech and policy discourse has highlighted, India has been planning to extend support towards developments in AI with supply of computing power, high quality datasets and supportive legal frameworks. However, how much of it is going to be channelled towards Indian languages is for all stakeholders to think about.

One of the things that are of critical interest to those closely observing the debates around AI and Large Language Models (LLMs) is how biases, infrastructure gaps and limited access to such technologies can impact decisions concerning society. While similar questions of impact of new technologies on the human condition keep arising with every turn in technology, the AI turn still requires a renewed sense of alertness around human-machine interactions materialising around us.

In discussing the role of digital commons in building AI technologies, the conversation focused on how a lot of AI systems are based on open source technologies and data but many prominent stakeholders including corporations are not transparent about the processes involved in utilising these resources. It is important therefore for advocates of open source/open knowledge movements to engage with this erasure of the role of open source data and technology in the development of AI.

Another nuance that needs to be examined is the role India can play in the development of AI. Is India going to be a site of localization or is it only going to be a market to sell to, or is it going to be a site to hire from for its skills and capacities? One perspective to emerge from the discussions on Indian momentum in AI is that there is a lot of hype around India becoming the next hub of AI, but this requires critical reflection. A large part of the innovation is controlled by the corporations in the West and not by any players in India. A related question is on what the Indian imagination of AI would be? One would be its availability in multiple Indian languages. However, that requires a lot of infrastructure and advanced computing capacities. We also ought to look at the critical question of uneven access to technology in India. It should address questions such as: where is AI needed? Why? How can it be made available? What resources would it need? A sensitivity to changing contexts and vocabularies in Indian languages would also need to be factored into evaluating the uniqueness of Indian contribution to AI development. If AI literacy skills could be disseminated in Indian languages, that would be an added advantage.

In India, we need a better understanding of what we mean by AI. How many Indian languages have the resources for development of AI? What are the ways in which questions of climate justice can be brought to bear upon the development of technology? What are the ways in which AI can democratise the process of bringing collective ownership to data and technology? What are the ways in which AI dilutes that process of collective ownership? The Indian imagination of AI therefore should not be a consumerist one, in which Indians are only users of AI. A lot needs to be done to improve and expand the public discourse on AI and digital commons in India.

Frame AI and Indian languages in Empirical and Real Possibilities of Achievable Outcomes

Panellists: Arjun Ghosh, Jameela Sahiba, Rohit Saluja, Plaban Kumar Bhowmick

Facilitator: Aakansha Natani

The ideas that emerged in this panel were a mix of philosophical, pedagogic, historical, technical or workflow process-related, and focussed on solutions, resources, and challenges to the deployment and use of AI in Indian language contexts.

The philosophical idea put forth was that before examining AI for Indian languages, one must ask if language is just words and text, or whether it's about objects and events. Humans learn languages in specific contexts. The disappointments or expectations with AI in Indian languages ought to pay serious attention to the idea of contexts. The pedagogic question was that of language as a tool of knowledge transfer and its usage in teaching and learning. The historical dimension alerted the participants to the historical conditions of the invention of the printing press in the West, and its project of documenting and digitising centuries of heritage and the commons. India, on the other hand, has a history of colonisation in which innovation came via English; it is the pioneers in Indian languages that took initiatives to make technology available to these languages. In addition to the inequality resulting from colonisation, there is also the fact that in India, only the digital-born content of the last 20 years has been documented and digitised. These historical realities should inform one's expectations and strategies towards development of AI technologies.

The technical or workflow-related processes question offered various provocations. One, should projects be designed intensively, working on one language at a time, with a deep dive approach? Or should one get expansive and work on many languages simultaneously? Two, what kind of processes should be designed for accommodating audio-visual content into datasets for the languages selected? Three, are there any benchmarks for how much data is required for Indic language computing projects compared to English; and how do project directors decide this? A strong suggestion was that projects and datasets in Indian languages should speak to one another.

The discussion of parameters for solutions to existing problems and resources involved exchanging information about specific tools included style transfer among fonts in Indian languages, feedback mechanisms which could be embedded in pilots and other ways for projects to invite suggestions. The challenges put forth by the panellists were regarding the lack of funding/resources and limited imagination of ways to incorporate users' feedback into the projects. These insights offered meaningful provocations and ways of thinking about the development of AI for Indian languages

Digital Inclusion for Indian Languages: Navigating the Roadblocks

Panellists: Ashwini Lele, Chandrakant Dhutadmal, Kiran Kumar & Rohit Kumar

Facilitator: Radhika Mamidi

The theme of challenges was expanded upon by the next panel that addressed roadblocks in the development of AI in Indian languages. These included problems such as lack of people, especially volunteers to work on projects; challenges in identifying use cases for Indian languages and difficulties due to language sets not understanding each other. A successful case study/example was of an Indian language project that focused on community building rather than technology products.

One of the major problems discussed when it comes to the development of AI is that there is a lot of data generated but there is no way to access it. For instance, in the banking sector, call centres and chatbots collect a lot of language and voice data that are supposed to be used for quality control and training purposes. But one doubts if such data ever gets used for these said purposes. Thus, a lack of custody for this data has been a major roadblock in development of technology. The other problem is the lack of adequate information on public funded projects such as AI4Bharat. How is the data being collected? What are the incentives for people who are creating such data?

Having information easily available in local languages to support crucial sectors such as healthcare, security, social protection and so on is the need of the hour. With regard to skill-building for the development of technology, one must sensitise students to the need for digitisation. Thanks to NEP, this attention to Indian languages can be made possible. Similarly, all kinds of volunteers who work hard on processing and cleaning data need to be acknowledged for the role they play in the development of technology. Simultaneously, the question of data protection and privacy also needs to be addressed. While the proliferation of cell phones makes it easier to collect data in forms such as audio, privacy remains a huge concern.

Languages have their own ecosystems. They have questions of class and commerce embedded in them. It will be a tough call to take on which languages to choose for further development in LLMs. Thus, digital inclusion for languages will need to be a balancing act, especially when students and other contributors and stakeholders see English as their priority, and not Indian languages.

The problem of expensive tokenization for low quality results leads to lack of innovation in Indian languages. The problem gets worse with languages that don't have scripts and written material. While the census undercounts languages (that is, it does not count languages with less than 10,000 speakers), documenting languages becomes a great challenge when it comes to creating content in underrepresented/unrecognised languages, which are already overwhelmed by lack of economic opportunities. Political forces also create language wars. Choosing to work on certain languages at the expense of others is going to be a huge factor in the way Indian language LLMs develop. However, the most important factor to consider would

be how we are developing Indian language communities rather than fixate over how to develop technologies in Indian languages. After all, it is the communities that drive technology and not the other way around.

Visit to Centre for Development in Advanced Computing, Pune

The conference participants were generously invited to visit CDAC, Pune, for an overview of some of their work on Indian languages and heritage computing. The team also shared their work on archiving, and building technology products for Indian languages. However, most of these products are not in the public domain and are not easily discoverable. Thus, the question of lack of access continues to be a big one for those working in Indian languages.

A few of these solutions developed by CDAC so far include:

Heritage Computing

- Jatan Virtual Museum Builder: deployed in 10 national museums and state, private museums. Digital collection management system (for managing artefact data).
- Museum Portal, used on Museums of India portal. This application allows for data to be filtered according to different articles, virtual galleries can be created which are curated by museum officers. It also enables 3D renderings of galleries, of sculptures (using photogrammetry) and 2D image formats by adding an annotation layer over the image so researchers won't have to violate the copyright.
- National Cultural Audiovisual Archive: A consortium of 26 organisations contributing heritage data hosted on NIC infrastructure
- Digitalaya Digital library: An ISO certified e-library and archival system SaaS

Language Computing

- Mantra Rajya Sabha: for English to Hindi translation for Rajya Sabha daily proceedings)
- e-Office Solutions: (system used and integrated in software used by govt offices): NLP Components for e-governance
- Transliteration software (Eng to Indian lang, Ind lang to Eng, Ind lang to Ind Lang). [These tools are not publicly available]
- Name Scape - compares names in any language to generate a score (higher score indicates likelihood of names belonging to the same person)
- Named Entity Recognition - compares two databases (name field, address field)
- OCR Software used by bodies such as the High Court of Uttarakhand.

The informal conversations among the participants within and outside the formal sessions enabled a good exchange of ideas and projects currently underway in various capacities in various parts of the country.

Day 2

On LLMs and Openness

Speaker: Sunil Abraham, Director-Policy, Meta

This talk unpacked the discourse around openness, globally and in the Indian context with the growth of movements such as open source and open access, and offered an overview of the approach towards openness adopted by large-scale LLMs such as Meta's Llama.

Workshop 1: Indian Languages and AI

The participants worked in different cohorts on specific themes. They were asked to identify key themes and pointers to the current discourse in AI and Indian languages and to also point towards steps ahead or solutions that can help address the lacunae in the field.

Participants worked in different groups on themes such as:

1. Exploring Policies, Standards, and Stakeholders: Advancing Indian Language LLMs
2. Navigating Global Dependencies, Open Source Potential, and Affordability Challenges in Developing Indian Language LLMs
3. Addressing Data Availability, Licensing, and Sustainability of Commons: Enhancing Transparency and Growth in Indian Language LLM Development
4. Fostering Trust in Indian Language LLMs: Challenges unique to our context
5. Impact of AI technologies - designed for Indian languages - on the efficiency, accessibility, and overall effectiveness of public services delivered through digital platforms.

The challenges identified by the participants included lack of basic technological resources, information asymmetry among languages, morphological complications in Indian languages, and lack of awareness regarding usage of AI in Indian languages. The solutions called for a greater investment in understanding challenges of as many Indian languages as possible, campaigning for public funded projects and tools to be made available in public domain, exchange of data among Indian languages, and prioritising community development over fixation towards technology.

Roundtable and Conclusion

Anubha Sinha delivered closing remarks to the first theme of the conference dedicated to AI and Indian languages. She summarised the takeaways from the sessions and invited suggestions for further development.

The participants suggested further themes for discussion, including :

- Hallucination by AI and its consequences
- Why use deep learning models in AI for welfare when standard regression suffices?
- Possibilities of data centralisation
- AI development in the context of literacy rate in the country
- Career paths in Indian languages technology development and how that affects the development of its economics

Keynote Address

Speaker: P Sainath, People's Archive of Rural India

The keynote address by P Sainath, journalist, author and founder-editor of the People's Archives of Rural India (PARI), introduced the second theme of the conference, which was dedicated to archives in India and Indian languages, the growth of digitisation and now the advent of emerging technologies such as AI.

In his talk, Sainath unpacked the making of PARI, its mission and identity. He pointed out that the archive is available in 15 Indian languages so that the stories reach as many linguistic communities as possible. The same is accomplished by human translators rather than machine translation applications. Elaborating on the nuances of translation, he pointed out several unintentionally humorous translations of certain components when left to machines. The translator community within PARI is very strong; they debate and discuss specific words and their connotations and sometimes change certain words later in order to do justice to the story. The process also keeps the community honest, as they participate in the creation of the archive.

His talk delved into the nuances of access and visibility of Indian languages, especially given large-scale digitalisation, but also the implications of these changes on low-resource languages. He also discussed some of the socio-political factors determining these advancements, the role of data as an entity and its evolution, the growth of misinformation and what this means for the growth of new technologies such as AI. Coming to archives, he also discussed the romanticisation of archives and libraries seen over the years, and how that has impacted access to these memory institutions. He ended with an appeal for better awareness and engagement with open access movements, and the need to make libraries, archives and other such knowledge repositories, especially those in Indian languages, open and accessible to the general public.

Digitisation and Archiving in India

Panellists: Smita Khator, Ishita Shah, Ranjani Prasad & Dharmendra Saha

Facilitator: Soni Wadhwa

The challenges highlighted by the panellists in the processes of archiving touched upon various aspects of language. For example, politics of religion hurts a language because a language does not equal a religion. Similarly, working with multiple languages requires a lot of creativity to come up with words for concepts that do not exist in certain languages. For instance, the idea of a salt-pan is conceivable in Tamil but not in Punjabi. Translators are thus forced to think on their toes and figure out ways to work around it. Another issue is the loss of language as a result of a loss of a tradition. When a tradition dies, a whole set of words associated with the culture or practice dies, as it happened with millet farming. In such circumstances, it becomes difficult to create data. What then is left to archive?

The challenges put forth by technology or digitisation are not any fewer. Technology poses its own challenge. For instance, PDFs are not the last stop of digitisation, as these too will become obsolete in the face of other types of documentation. Another challenge is to locate solutions that do not create huge dependence on technology or require huge funds. For archives such as PARI, using machine translation is likely to become a way forward in order to be able to reach people in their languages but to what extent such tools are useful remains to be seen. Archives such as Rekhta struggle with receiving substantial CSR funds because language is not seen as pressing an issue as livelihood-related concerns, for instance.

Not every archive can have tangible consequences. Therefore, the utility of an archive cannot be judged on the basis of what kind of afterlife it provides to the digital avatars of its collections. Archives bring an element of care for things, which contributes to an overall ethos of care for knowledge. They are also a form of working towards social justice, as they emerge when the need to be proactive about preservation is felt at its most intense. For instance, when a village is about to submerge because of an infrastructure project, there is a need to document every space within the village.

Day 3

Workshop 2: Archives and Indian Languages

The workshops were preceded by a demonstration by Prof. Radhika Mamidi from IIIT Hyderabad regarding their technology Himangy and Disco-MT which is a speech to speech machine translation software.

The participants worked in different cohorts on specific themes. They were asked to identify key themes and pointers to the current discourse in AI and Indian languages and to also point towards solutions that can help address the lacunae in the field.

The themes worked upon by different cohorts included:

- Open-Source Solutions to Archival Challenges: What is Working, and What is Not?
- Archives, Infrastructure and Digitisation
- Creating Indian Language Archives: Visibility, Financial Sustainability and Resources
- AI and Big-Tech: What can they do for Digitisation and Multilingual Archives
- Archives, Equity and Community

Overall, the groups identified problems they have been facing in their projects as developers and facilitators and suggested certain possibilities that could empower them in overcoming their challenges. While the problems identified specific restrictions in terms of pressure on resources and labour, lack of ways to disseminate and attract engagement with the archives, or lack of democratisation and interaction among the different archives, a wish-list that would help in overcoming these challenges included greater creativity with financial and technological resources.

Thinking of Next steps: Integration of AI, Languages and Digitization in Work Practices

Panellists: Vidula Tokekar, Sreechand Tavva, Pavitra Jayaraman, Plaban Kumar Bhowmick

Facilitator: Vishnu Vardhan

The panellists drew upon experiences from their respective fields of work, to discuss the impact of digitisation and now the advent of AI tools, as well as the evolving role of Indian languages. On the question of how the panellists use AI for translation, it was pointed out that using English as an intermediary language remains immensely problematic. Machine translation addresses semantics but not the gaps in how humans experience things and express themselves. Some tools such as Anuvaad, Narakeet, Chitrlekha, Glyphic, Polly, and

Jugalbandi are used with the help of human supervision. The panellists also elaborated on challenges specific to their own contexts, and how the use of translations may be impacted by the same, such as in the case of gendered aspects of mental health, tools used in primary education or ways to build data and narratives around natural resources. Another important point of discussion included the use of bridge languages, which usually are more widely spoken and may have better resources; but it could also lead to loss of meaning in translation. The conversation also focussed on the role of the community in language preservation, either by evolving new ways of using a language or contributing to knowledge repositories and open technological resources.

The panellists shared that the content from their work is all on Creative Commons licences and invited participants to engage with the same. The possibility of developing glossaries in Indian languages, especially for technical professions was discussed, and ways to test these glossaries by opening them up for the public to contribute to them. The need for more work on bridge languages, and developing context for machine translations through the use of speech-based tools was also discussed. A number of tools, platforms and resources were discussed and it was suggested that CIS could compile a list of these and share it with the community.

Conclusion

The final session of the conference had closing remarks by Soni Wadhwa and a vote of thanks by Tanveer Hasan.

In her remarks, Soni Wadhwa pointed out that thematically speaking, the gathered community looked at archives from the points of view of democratisation of archiving processes. They also touched upon this democratisation by looking at archival projects beyond state records, as these were from various domains. They realised that archives of all kinds - community, personal, institutional - have political responsibilities in the way they approach a variety of causes such as: promoting a language, preservation of knowledges of ethnic communities within biospheres, voicing stories of marginalised groups, collecting oral histories of certain communities and so on.

The closing remarks discussed the theme of archives when contrasted with technology conversations even though digitality was an underlying concern for both. While technology and AI related discussions were about problems, solutions, demand for products, resources, and everything that is missing to design and develop better technology products; archives conversations were about a strong sense of purpose, an ethos of care, an approach of empathy, a mission to work with communities, and were mindful of privileges, agency, and opportunities that those working with archives come with.

This contrast was not intended at making a value judgement by making one thematic concern emerge as a better kind of conversation to have but to point to some of the modalities of how these conversations take place, and understanding how technology and archiving questions can intersect and speak to each other. Arguably, archiving and archivists have a lens of purpose and resource constraints that can work as an ideal testing ground for technology

solutions to be tested. That archival projects exist across domains is an added advantage that can help in piloting technology in various contexts, and with datasets and corpora.

The energy of the archives theme of the conference indicated that a lot more needs to be done. Suggestions shared during the discussion on how to address some of the lacunae in this sector also offered insights on ways forward for the community.

In his vote of thanks, Tanveer Hasan reiterated the reasons behind the organisation of the conference. He pointed out that putting this conference together was a labour of love, a lot of stress and could have also been a risk. The CIS team really wanted it to be creative, at the level of building a community of practitioners and thinkers, and also wanted to address a propensity towards tech solutionism often seen in the sector: solutions need not only come from peripheral vision, and a broader field of view is needed for this to become a reality. He ended his vote of thanks by inviting questions and collaboration while working towards ensuring that institutional and individual energies come together in the process.

Appendix - 1|Concept Note

Future of the Commons: A Conversation on Artificial Intelligence, Indian Languages and Archives

Organised by: Centre for Internet and Society, Bengaluru and Maharashtra Knowledge Corporation Limited (MKCL), Pune

Dates: Friday, 5 July 2024 to Sunday, 7 July 2024| 9am-5:30pm

Venue: MKCL, ICC trade tower, Senapati Bapat Road, Pune

Overview

This three day event, scheduled for July 5-7, 2024, aims to bring together a diverse group of stakeholders to explore the evolving relationship between artificial intelligence (AI), Indian languages and digital commons.

Since ChatGPT's launch, multiple initiatives have been taken to develop generative AI systems in Indian languages, for a variety of uses. The process of training generative AI models requires huge quantities of data. Openly available training data - or digital commons - are critical to developing legitimate, well-performing and versatile AI models.

Archives and museums, as rich repositories of cultural heritage material, have been actively digitising their collections. Collectively, the contribution of public memory and cultural heritage institutions and mass-digitisation programmes undertaken by state and civil society organisations, has been crucial to the development of a digital commons in Indian languages.

Set against this backdrop, the event will delve into ongoing development of 'localised' or 'indic' generative AI models, state of digital commons, and role of public institutions such as archives and digitisation movements; and explore the interdependencies and tensions between them.

Proposed Themes and Topics

- **Ongoing Efforts and Innovations:** A look at the current state of research and development of 'localised' generative AI systems.
 - ❖ Use-cases of GenAI products in Indian languages
 - ❖ Challenges with availability of training datasets
 - ❖ Dependency on Big-tech for generative AI models
 - ❖ Open source development of generative AI models
- **Development and 'use' of digital commons:** How should we digitise, preserve, and use knowledge and cultural heritage for training AI models? What are the practices and challenges in archives, digitisation movements, especially in Indian languages?
 - ❖ Role of open knowledge movements and Galleries, Libraries, Archives and Museums (GLAM), born digital and community archiving practices
 - ❖ Tensions with ownership, access, and privacy of underlying data
- **Future of generative AI in Indian languages:** How to design generative AI systems which do not just 'take' from the open movement and public institutions, but also contribute.

- ❖ Incentivising expansion, foregrounding diversity and safeguarding of digital commons
- ❖ Role of law and policy, ethics and algorithmic justice
- ❖ Ensuring that openness movements and communities such as Wikimedia projects are safeguarded against unfair extraction

Format and Participants

The event will feature a mix of panel discussions, roundtable discussions, coworking/brainstorming sessions, demonstrations and interactive activities over three days. Participants will include:

- **Policymakers and CSOs:** Legal and Policy Analysts working on language-related policy reform and regulations. Civil society organisations and advocates working on language conservation, revitalisation and reform.
- **Academia and Researchers:** Researchers working on a diverse set of areas including (but not limited to) AI and ML, open knowledge movements, data governance, privacy and regulation, ethics of emerging technologies, computational social sciences etc. Linguists, social sciences and humanities subject experts working on language conservation and revitalisation.
- **Technologists, Archivists and Creators:** Practitioners with a focus on AI/ML, Indic language computing, open source technologies etc. Archivists, curators and digitisation experts. Creative practitioners using emerging technologies to develop Indian language content.
- **GLAM and A2K Activists:** Cultural and memory institutions, GLAM professionals working with Indian language collections.
- **Wikimedians:** Content creators, editors and specific language user groups.

Objectives

- **Understanding Perceptions and Applications:** Assess how AI is currently perceived with a focus on localised and Indic initiatives. What is the future of developing and deploying GenAI for Indian languages?
- **Exploring Ongoing Efforts:** Highlight ongoing efforts in AI that pertain to Indian languages? What is the role of open source in developing AI for Indic languages? What is the India open source AI ecosystem like?
- **Addressing Data Challenges:** Discuss limited availability of high-quality, diverse and representative datasets for Indian languages, current efforts in building robust, localised datasets, and challenges to the same.
- **Facilitating Multidisciplinary Collaboration:** Foster dialogue and collaboration among researchers, technologists, linguists, legal experts, policymakers, and content creators to shape the future of AI in Indian languages.
- **Understanding downstream uses of AI applications:** Unpacking the impact of AI across diverse sectors that work with Indian languages, such as the algorithmic curation of cultural experiences in the GLAM sector or the mediation of creative labour through AI.

Appendix - 2| List of Resources

The following list is a compilation of projects, tools and software mentioned and discussed at the conference.

Sr No	Project	Description	Link
1	AI4Bharat	Research lab at IIT Madras developing open-source datasets, tools, models and applications for Indian languages	https://ai4bharat.iitm.ac.in/
2	Dasubhashitam	App for Telugu audiobooks	https://www.dasubhashitam.com/title-type/telugu-audio-books
3	People's Archive of India, PARI	Archive of stories from rural India	https://ruralindiaonline.org/en/
4	Rekhta Foundation	Archive for Urdu language and literature	https://www.rekhta.org/
5	Keystone Foundation	Archive about the Nilgiris' Biosphere	https://keystone-foundation.org/
6	Curating for Culture	Consultants in archive building	https://sites.google.com/curatingforculture.com/curatingforculture
7	Himangy	Indian language to Indian language machine translation	https://ssmt.iiit.ac.in/translate
8	India Water Portal	Open data project about water resources	https://www.indiawaterportal.org/
9	Pratham	Educational resources for learners	https://www.pratham.org/
10	White Swan Foundation	Archive of articles on mental health in various Indian languages	https://www.whiteswanfoundation.org/
11	Translation Panacea	Corporation providing translation services	http://www.translationpanacea.in/

12	Chitralkha	Open source video subtitling software for Indian languages	https://chitralkha.ai4bharat.org/
13	Anuvaad	Open source document translation software	https://www.anuvaad.org/
14	Narakeet	Online text to speech video maker	https://www.narakeet.com/
15	Glyphic	Document automation tools	https://www.glyphicsoftware.com/
16	Polly	Software for hybrid teams	https://www.polly.ai/
17	Jugalbandi	free and open platform that powers conversational AI solutions in any domain.	https://www.jugalbandi.ai/

Appendix - 3| List of Attendees

No.	Name	Affiliation
1	Aakansha Natani	International Institute of Information Technology, Hyderabad
2	Radhika Mamidi	International Institute of Information Technology, Hyderabad
3	Neelima Motwani	CHM College, Ulhasnagar & IIT-M AI4BHARAT
4	Kiran Kumar	Shreyam, Bengaluru
5	Anivar Aravind	Public interest technologist, Bengaluru
6	Karan Vanni	Language advocate, AI4BHARAT
7	Chandrakant Dhutadmal	Centre for Development in Advanced Computing, Pune
8	Vidula Tokekar	Translation Panacea, Pune
9	Pavitra Jayaraman	The PARC
10	Arjun Ghosh	Centre for Humanities and Social Sciences, IIT Delhi
11	Jameela Sahiba	The Dialogue, New Delhi
12	Rohit Kumar	The Quantum Hub, New Delhi
13	Rohit Saluja	Indian Institute of Technology, Mandi
14	Sunil Abraham	Meta
15	Sharika Parmar	Flame University, Pune
16	Ranjani Prasad	Keystone Foundation Archives, Nilgiris
17	Shivakumar Jolad	Flame University, Pune
18	Ishita Shah	Independent Curator, Designer and Archivist, Bengaluru
19	Dharmendra Saha	Rekhta Foundation, New Delhi
20	Smita Khator	People's Archive of Rural India, Kolkata
21	Plaban Kumar Bhowmick	Indian Institute of Technology, Kharagpur
22	Shraddha Chorgi	Pratham, Pune
23	Bodhisattwa Mandal	Bangla Wikimedian
24	Chinmayee Mishra	Odia Wikimedian
25	Ojas Kadu	Archives at National Centre for Biological Sciences, Bengaluru
26	Sreechand Tavva	Arghyam, Bengaluru
27	Ashwini Lele	Marathi Wikimedian
28	Surya Kumar Pandey	Indira Gandhi Rashtriya Manav Sangrahalaya, Bhopal
29	Selvasivagurunathan M.	Tamil Wikimedian
30	Vishnu Vardhan	Board Member, Centre for Internet and Society, Bengaluru
31	P. Sainath	People's Archive of Rural India

32	Prasad Joshi	Archives at Pune Nagar Vachan Mandir Library
33	Sushrut Kulkarni	Translator and Content Writer
34	Pranab Sharma	Maharashtra Knowledge Corporation Limited (MKCL),Pune
35	Soumya Ranjan Pani	Maharashtra Knowledge Corporation Limited (MKCL),Pune
36	Soni Wadhwa	SRM University, Andhra Pradesh
37	Puthiya Purayil Sneha	Centre for Internet and Society, Bengaluru
38	Anubha Sinha	Centre for Internet and Society, Bengaluru
39	Isha Suri	Centre for Internet and Society, Bengaluru
40	Tanveer Hasan	Centre for Internet and Society, Bengaluru
41	Pavan Santhosh	Centre for Internet and Society, Bengaluru
42	Subodh Kulkarni	Centre for Internet and Society, Bengaluru
43	Nishant Shankar	Centre for Internet and Society, Bengaluru
44	Garima Agrawal	Centre for Internet and Society, Bengaluru