

Inclusive AI design for justice in low-literacy environments¹

Avanti Durani^{1*,†} and Shivani Sathe^{2,†}

¹ BTGAdvaya, Fort, Mumbai 400 001, Maharashtra, India

² BTGAdvaya, Fort, Mumbai 400 001, Maharashtra, India

Abstract

This paper investigates the potential of inclusive design principles to enhance the adoption of AI-powered legal assistants in low-literacy environments. We focus on developing economies, specifically India, where complex and esoteric language of legal texts make legal information inaccessible to a significant portion of the population, contributing to exploitation and exclusion. Access to legal knowledge can help empower individuals and communities to protect their rights and pursue social justice. With widespread digital penetration, India presents an opportune setting to test AI-based solutions that can help simplify and make legal information more accessible. However, meaningful adoption of these tools requires a participatory process and careful adaptation to local socio-economic conditions, including low literacy rates, English proficiency and spending capacity. This paper outlines how we apply inclusive design principles to develop an AI legal assistant chatbot for informal workers in the construction industry in Maharashtra, India. We focus on strategies to make legal information accessible, comprehensible, and actionable for underserved Indian communities. Through our on-going project, we illustrate how these principles hold the potential to help improve user engagement and promote legal literacy.

Keywords

Access to Justice, GenAI, Legal Literacy, Inclusive Design, Digital Solution, Multilingual, Accessibility, Low literacy, Informal Labour

1. Introduction

This paper explores how generative artificial intelligence (GenAI) can be adapted to bridge legal information gaps in low-literacy populations, specifically in India, through an inclusive design approach. In India, a general lack of legal awareness amongst underprivileged, marginalised and vulnerable communities frequently leads to their exploitation and exclusion from basic services and opportunities. [2, 3] Inadequate knowledge of rights and entitlements often leads to poor decision-making and insecurity. [2, 3] To overcome these challenges, and for larger social justice, there is an urgent need to build practical and inclusive solutions that focus on last-mile delivery of legal information. With new strides in technology, there is an opportunity to harness available tools, including GenAI, to bridge some of these gaps and provide accurate and relevant legal information to all.

While GenAI-based tools hold huge potential to enable access to legal information at scale, their successful adoption in a developing economy context can be challenging. The deployment of AI-powered legal assistants in underserved communities can be hindered by low literacy levels, poor English proficiency, financial scarcity, and other socio-economic barriers. [4, 5, 6, 7] The solution, we believe, is to tailor the development, use and deployment of technology, by applying inclusive design principles, to meet the needs of vulnerable populations.

Before we delve into what these design principles are, it is important to set the context on some of the key barriers to adoption of digital solutions for legal literacy in India today.

¹ AI for Access to Justice Workshop at Jurix 2024, December 11-13, 2024, Brno, Czechia

² Corresponding author.

[†] These authors contributed equally.

avanti@setu.law (A. Durani); shivani@setu.law (S. Sathe)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1.1. Low textual literacy

In India, foundational literacy, i.e., the ability to read and understand age-appropriate text, is a major hurdle in delivering legal information through digital means. Nearly one-third of India's adult population (aged 15 years and above) is illiterate [9], and even among the 'literate', reading skills are often very limited [11]. Over 50% of children aged 10 to 14 years struggle to read basic text. [12] Among those aged 14 to 18 years, about 25% cannot fluently read a Grade II-level text in their regional language, and only slightly over half can read sentences in English. [4] The COVID-19 lockdown exacerbated this learning gap, with children's foundational reading ability regressing to pre-2012 levels. [5]

Foundational reading skills are crucial for performing basic digital tasks like browsing a search engine, setting an alarm, using navigation applications, or finding a video online. [4] A 2023 survey conducted by the ASER Centre confirmed a positive correlation between foundational literacy and digital capability. [4] This makes low reading proficiency one of the most significant challenges in designing technology to deliver complex legal information.

1.2. Limited English proficiency

Delivery of content in English risks alienating vulnerable populations in India, a linguistically diverse country with over 100 different languages and more than 1,000 dialects [13]. English is not the first language for most Indians [7], and English speakers comprise a relatively small, economically privileged segment of the population. A 2019 Lok Foundation–Oxford survey in India found that only 6% of respondents could speak English. More importantly, there were pronounced disparities in English language proficiency across various demographics, including gender, economic status, caste, religion, and geographical location. Men, urban residents, and upper-caste individuals reported much higher proficiency compared to women, rural populations, and socio-economic minority groups. [7]

GenAI systems also face challenges in accommodating linguistic diversity. GenAI, and more specifically large language models (LLM), require high volume and quality of training data to produce accurate and reliable outputs. [15] However, more than half of the content available on the internet is in English. [16] Hindi, the most widely spoken Indian language [13], comprises only 0.1% of the freely available content on the internet [15]. Data in many other Indian languages is simply not available online. [15] The gap in local language training data limits GenAI's ability to comprehend, contextualise and accurately address user requests in languages other than English. [17]

Additionally, processing queries in vernacular languages may also be more expensive. GenAI models process sentences by dissecting them into bite-sized units called 'tokens'. When a GenAI model is not acquainted with a specific language, it tends to produce a higher number of tokens. GenAI companies charge users based on the number of input/output tokens. Therefore, a higher number of tokens makes it costlier to cater to end-users, acting as a potential access barrier for the underprivileged. [18]

1.3. Socio-economic barriers to digital access

While language and literacy make it difficult for people to use technology, unequal access to technology further exacerbates this paradigm. In 2023, there were 821 million active internet users in India. [19] This number has been consistently growing and is expected to reach 900 million by 2025. [20] Smartphone penetration is also increasing rapidly; 95% adoption is expected by 2030. [21] However, in 2023, despite this rise, 45% of the population still lacked internet access. [19] According to a 2022 report, about 40% of mobile subscribers did not use smartphones [6] and most households reported not having a computer. [4, 6] Access to technology in India is influenced by socio-economic factors like caste, class, gender, and place of residence. The most marginalised tend to be those with the least digital access, risking further exclusion from technology-driven welfare initiatives and opportunities. [6]

1.3.1. Gender

A 2023 Report by Sattva Consulting states that India accounts for almost 50% of the global gender digital divide. [10] Indian women are 15% less likely to own a mobile phone, and 33% less likely to use mobile internet services than men. [22] Factors like lower literacy rates amongst women and restrictive gender norms [10] reduce the frequency of women's digital engagement [6] and ultimately limit their ability to use digital services [4].

1.3.2. Geographical location

The digital divide between rural and urban populations in India is also pronounced. In rural areas, internet availability is either intermittent, poor or non-existent. [6] According to a 2022 Report by Oxfam only 31% of rural residents use the internet, as opposed to 67% of the urban population. [6] Another study by the Internet and Mobile Association of India (IAMAI) and Kantar revealed that 21% of rural users access the internet from someone else's mobile phone [19], which makes assured and timely access to technology difficult.

1.3.3. Income level

Affordability plays a key role in determining digital access. There is a positive correlation between internet access and income levels. [6] In 2020, amongst the poorest 20% of households, only 2.7% had access to a computer and 8.9% had internet access. In contrast, among the top 20% of households, these figures rose significantly to 27.6% and 50.5%, respectively. [23] Further, reports suggest that economic hardships or loss of employment often force low income households to forgo internet use. [6]

The effective deployment of AI-based legal literacy tools in India requires adapting to several socio-economic barriers discussed above. To overcome these challenges, we launched an AI and Access to Justice Initiative in India called 'Project Setu'. Project Setu is dedicated to making legal information accessible to everyone, especially marginalised communities. It uses AI-powered tools, designed by legal experts, to bridge gaps in understanding and provide accurate, relevant legal knowledge. The platform is adaptable, curated for different legal areas and tailored to the needs of vulnerable individuals. By working closely with non-governmental and civil society organisations, Project Setu ensures that its solutions address real-world challenges. Project Setu is more than just a tool, it is a step towards enabling access to justice for all.

2. Designing AI-based assistants for legal literacy

Project Setu aims to democratise legal knowledge for underserved populations. To achieve this, it works on two levels—(i) as a community-based platform to promote collaboration between the legal, civil society and technology communities; and (ii) to accelerate the development of issue-specific, AI-based chatbots using inclusive design principles. In this section, we illustrate how Project Setu develops digital solutions through one of our on-going projects on informal labour rights (ILR).

The ILR project is being developed in partnership with a labour helpline non-governmental organisation (NGO) that works at the frontlines with informal workers. For this project, we are building industry-wise information kits in the form of conversational chatbots that can be used by workers at the time of onboarding or on the job. The chatbots can provide workers access to reliable legal information to demand fair treatment. The chatbot can also assist the NGO's helpline attendants, most of whom are not trained lawyers, in their casework.

The informal sector constitutes over 90% of India's overall workforce [41], and contributed nearly 75% to the construction industry's Gross Added Value (GAV) in 2017-18. [42, 43] Informal workers often belong to socially and economically underprivileged communities [41], including many migrants who move from rural areas to cities seeking better livelihood opportunities [44]. Given their limited education, skills and training, they frequently find themselves in demanding jobs with low wages, poor working conditions, and no access to social protection [44]. The sector is

characterised by the absence of contractual arrangements, formal guarantees and other welfare benefits generally offered to employees. [41, 42] For example, in the construction sector, 10.7 million workers in India did not receive minimum employment protection and benefits, according to a 2022 Oxfam publication. [43]

In the following subsections we exemplify the key steps in our project workflow and highlight our deliberate design choices at each stage. Specifically, we adopt a participatory approach and emphasise content localisation during the ideation stage. During the development stage, we incorporate accessibility features that can enhance usability of the tool for our target audience. Finally, we leverage a networked and offline support model at the deployment stage to improve implementation and uptake of the tool.

2.1. User-centricity during the project initiation and ideation phase

One of the main ways to ensure adoption of digital solutions in low-literacy environments is through user-centric approaches. Several studies in the fields of education, migration, agriculture, healthcare, and legaltech show positive evidence of this. [24, 25, 26, 27] The studies suggest that user-centricity can be achieved in two ways—participatory methods [11] and content localisation [28, 11]. Participatory methods include co-designing solutions in a way that involves stakeholders and beneficiaries of the solution in the development process. Localisation means creating meaningful content in familiar formats [11] to enhance access to information [28].

2.1.1. Identification of the legal knowledge gap through fieldwork

We began the ILR project by identifying a local partner organisation. Our selection criteria included the organisation's level of integration within a target community and their role in supporting community needs. Next, we conducted online discussions with their leadership team to establish alignment on the scope of work. After these initial discussions, we met with the field staff at an industrial suburb in Mumbai, Sakinaka, to conduct an on-ground analysis. We conducted informal focused group discussions (FGD) and key informant interviews (KII) with the stakeholders and potential beneficiaries of the tool. We also studied case summaries prepared by the NGO's helpline attendants. These sessions helped us gain insight into three critical design considerations—(i) the main legal issue(s); (ii) the target users of the tool; and (iii) an in-depth understanding of their background details and specific barriers to access.

The field staff identified construction workers as one of their most challenging beneficiary segments, which also contributed to the bulk of their casework. From the beneficiary interviews, it was apparent that most of them are illiterate and migrants from other states. They seek work on a daily basis by standing at designated points along street corners. In terms of the process for securing a job, a contractor, or third-party intermediary, arrives at the street corner, verbally discusses a wage rate with the worker, and takes them to the work site. However, at the end of a day's work, the workers often receive less wages than the amount promised to them. Moreover, they have no awareness of any welfare programs or benefits intended for their sector. Based on this fieldwork, and other research findings, the NGO and our team collectively decided to create a prototype for construction workers in the state of Maharashtra, India. In India, labour regulation is governed by laws framed by both the Central Government and individual state governments, resulting in variations and nuances across the country. For our pilot project, to ensure relevance and precision, we decided to limit the geographical scope of the prototype to the state of Maharashtra.

This localised prototype exemplifies Project Setu's model of building issue-specific AI tools tailored to a distinct set of users, rather than attempting to create a broad-based solution for a wider, heterogeneous audience. This focused approach allows us to leverage the deep expertise of legal professionals and NGO partners within a particular context, ensuring the tool's legal accuracy and practical applicability. This approach also enables us to iteratively refine and improve the tool's capabilities while maintaining user-centricity.

2.1.2. Curation of the legal database

The next step in the process was to start putting together the various components of our technology stack for the chatbot. Project Setu's technology stack comprises six layers—(i) storage; (ii) privacy filter; (iii) knowledge database; (iv) LLM; (v) language and audio API; and (vi) user interface. In this subsection we focus on the curation of the knowledge database.

Since Project Setu uses a retrieval augmented generation (RAG) method for generating natural language responses and reducing hallucinations [45], it was important to have an expertly curated knowledge base that interacts with our LLM. Moreover, the prompt that we used to instruct the AI requires it to double-check and cross reference each answer before responding, an additional guardrail against inaccuracy.

Research shows that content that is tailored and contextualised to a user's needs reinforces its relevance and encourages engagement with the tool. [28] For the ILR project, the insights from the fieldwork helped determine the knowledge database of the chatbot. We approached legal experts with the relevant domain knowledge to study our findings and develop a strategy for legal document curation. The knowledge database was created by sourcing legal documents from trusted online sources that are used by lawyers in the ordinary course of their work. We began by sourcing legislative acts for the informal labour sector that apply to all states in India, as well as those that are specific to the state of Maharashtra. This totalled to about 32 legislations. Currently, we are in the process of adding rules framed under these legislations and ad-hoc welfare schemes, both at the pan-India and Maharashtra-state level, to the database.

These initial steps of the ILR project described above are important for ensuring that the intervention is relevant [24, 28] and responsive [29] to user needs. The participatory process [28] is important for scalability of the tool as well. It can enhance the perceived value [27] of the tool for various participant groups. It can help unearth the challenges and barriers faced by vulnerable groups to develop mitigation strategies, [29] and highlight enabling factors [24] that can broaden the reach of the solution. In the next subsection, we discuss accessibility features which can support in scaling the solution.

The initial fieldwork is a critical step in developing a strategy around the scope and design of the tool. It helps ensure that the solution is collaboratively designed with a locally trusted partner organisation, and is relevant, contextually appropriate, and directly addresses the needs of their beneficiaries. In fact, the data collection at this stage can also serve as a baseline and a metric to evaluate the performance of the tool post deployment.

2.2. Accessibility features during the development phase

A key consideration during development of a digital solution is its accessibility. Studies show that features that reflect a natural extension of a user's behaviour [30] can help increase the potential for digital inclusion and empowerment. [24] Given our findings from the fieldwork and the target user base, we considered three technological features to enable greater uptake—(i) a simplified interface for users to engage with the content; [24, 31, 11] (ii) allowing users to help generate the content in an interactive or prompted manner; [11] and (iii) most crucially, audio, multilingual and affordable options for content dissemination. [31, 11, 29, 30]

2.2.1. Simplified and interactive interface

In India, WhatsApp is ubiquitous with smartphone usage and is often the preferred communication medium among groups like our target users. [38] Considering the rapidly growing rate of internet and smartphone users in India [21], Project Setu uses WhatsApp as the chatbot interface. In practical terms, a user does not have to download any application to access the tool, but simply stores the WhatsApp contact number on their phone and can query the chatbot when required, at their convenience.

A conversation with the chatbot is activated by the user by typing a simple key word, such as 'hi', in the chat. The chatbot responds with an introductory message to the user, explaining the legal domain that it can assist with and requesting the user to select their preferred language of communication. Once the user selects a language, the chatbot follows up with a short tutorial message for the user in the chosen language, in text and audio. The user can then proceed to type their query or record a voice message. There is no limit to the number of questions that a user can ask the chatbot. The AI has been instructed to respond in the following format: (i) title of the specific law, legal principle, or statute relevant to the query; (ii) a summary of the relevant legal information based on the law and the user's context; (iii) a detailed explanation of how the law applies to the user's situation and any procedural remedies available to them; and (iv) references to specific legal provisions where necessary. An example of a response provided by the chatbot is below.

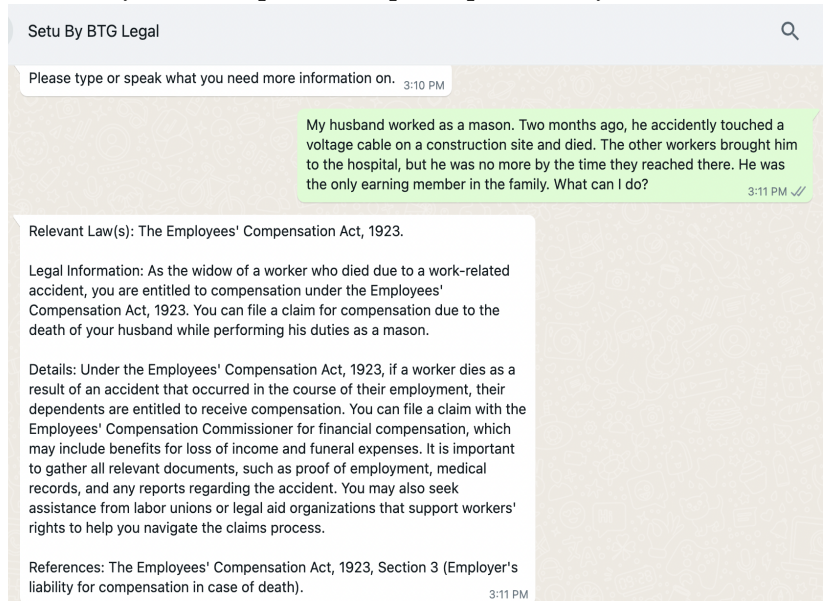


Figure 1: Screenshot of response provided by the chatbot built in Project Setu's ILR project to a legal query posed by a user.

The AI is instructed to cache up to three preceding user messages to help it build context for generating a response. It is also instructed to send out certain error messages if the queries are either incomplete, ambiguous or out of the scope of the chatbot. This helps ensure that the chatbot does not attempt to answer queries which are unclear or beyond its purview, reducing the possibility of erroneous responses.

The conversation flow repeats in the same manner either until a user mentions that they are satisfied with the answer or there is a gap of a few minutes from their last query. At this point, the chatbot sends a message to the user requesting feedback and then closes the conversation with a message that links the user to the NGO for any further support.

2.2.2. Multilingual, audio-enabled and affordable

As discussed in Section 1.2., Indians belonging to weaker socio-economic strata are likely to exhibit low levels of English proficiency. In fact, research suggests that many Indian internet users consider vernacular digital content to be more reliable than English content. [34] Primary challenges in adoption of online services cited by Indian language internet users include limited availability of content in vernacular languages [34, 35] and difficulty in navigating English keyboards to input queries in regional languages. [34] The ILR chatbot offers multilingual functionality to users, allowing them to ask questions and receive legal information in 22 languages via text and 14 languages via audio. According to the last census in India (2011), 97% of the population speaks at least one of these 22 languages [13], making our chatbot accessible to a wide variety of Indian users. A user can select their preferred language at the beginning of the chat, and also switch to another

language mid-conversation by typing in the same key word that prompted the chat flow to begin (e.g. 'hi').

Furthermore, the lack of reading proficiency has led Indians to opt for non-textual modes of obtaining information. [11] Indians' preference for voice-based communication is seen in the growing popularity of voice commands for completing basic tasks on a smartphone [19, 32]. About 82% of smartphone users opt for voice-activated technology [32] and about 64% users opt for voice assistants to get daily news and updates [33]. The audio feature of the chatbot allows the user to ask their query through the voice note feature on WhatsApp, and receive responses in the form of voice notes as well, in their selected language. This helps break access barriers for users with no or low reading capabilities.

The audio and multilingual capabilities of our tool are supported via the Bhashini API, developed by the Ministry of Electronics and Information Security, under Digital India's National Language Translation Mission. [39] The Bhashini API is currently available at no cost for third-party users, enabling Project Setu to optimise on cost while building digital solutions. Affordability is a key indicator of dissemination of digital solutions [11] with our target audience. Project Setu's solutions are provided at no-cost to the end user, further enhancing accessibility and ensuring continued usage.

2.3. On-going assistance during the deployment stage

For semi-literate and illiterate users, continuous assistance is important to establish meaningful and sustained engagement with any intervention. [31, 29] A collaborative and networked model can be instrumental for such engagement. [24, 28] It allows various stakeholders in an ecosystem to leverage each other's resources and capacities to plug any gaps in the uptake of a solution. [31] Similarly, digital solutions deployed with offline or multi-channel support options, can help ensure continued participation of users. [29] In this section, we focus on our networked approach for testing, training and periodic review of the chatbot, and providing offline support channels for users.

2.3.1. Testing, training, and periodic review

The participatory model adopted at the beginning of the ILR project doubles as the deployment network for the chatbot. Each stakeholders' capability is leveraged to help disseminate the tool in a systematic manner. The NGO acts as a pathway to its beneficiaries, playing a vital role in fostering user trust and facilitating seamless adoption and scalability of Project Setu's chatbot.

The first stage of deployment involves training and testing, and this is done iteratively with the help of the NGO. For the ILR project, we are in the process of conducting an orientation session with the NGO staff to test our prototype's functionality. We will then incorporate their feedback in terms of any immediate improvements. Thereafter, we will seek the support of the NGO to facilitate training sessions with the end-users on how to use the tool. The physical presence of the NGO staff in these sessions is important because it can help signal confidence for the beneficiaries, who in the absence of a discerning ability, often rely on trusted partners to guide them.

The testing phase will be repeated several times, each time increasing the number of users and real-world queries to help improve the chatbot's components. Such an incremental approach will allow for reinforcing feedback and valuable insights from user behaviours and patterns [11] to foster more sustained integration of the tool. The user queries and responses generated during testing will be shared with subject matter legal experts for review and auditing. Their feedback on accuracy and completeness of the responses generated by the AI will also be incorporated to improve the chatbot. This lawyer-driven AI-training will help improve not just the accuracy of the tool, but also help incorporate relevant and nuanced legal context in a supervised manner. This step is particularly significant given the subtleties of Indian laws, state-wise variations, and the complex interconnections between the different types of legal documents. It will also help further localise the tool per the requirements of the users. Post deployment, we see a role for legal experts to help update

the knowledge database for any changes in the law and audit the chatbot responses for legal accuracy on a periodic basis to ensure continued reliability.

2.3.2. Resources for offline support

As mentioned in Section 2.2.1., we incorporate forward linkages at the end of chatbot conversation flow to connect users to the NGO for any offline support. For example, in the case of the ILR project, once the user confirms that they require no further assistance from the chatbot, they receive the contact details of our partner NGO for any additional support. This feature helps establish a robust support ecosystem throughout the lifecycle of the tool. It also helps monitor the tool's utility and identify improvement and growth opportunities over time. Crucially, it connects the user with human touchpoints, fostering a sense of trust and comfort and allowing users to navigate legal crises more effectively.

3. Conclusion

We believe that the value of an inclusive, community-oriented approach to AI design will be vital for access to justice. Inclusive design principles offer a path towards overcoming challenges in accessibility of legal information, such as low literacy levels, limited English proficiency, and socio-economic barriers to digital access. Through our research and early evidence from the on-going ILR project, we are in the process of assessing how participatory methods, content localisation, accessibility enhancing features, and a multimodal delivery and monitoring mechanism can lead to an improvement in legal literacy for underserved communities. We conclude by considering potential limitations of our ILR project as well as research directions that can help us expand this work.

3.1. Limitations

An important limitation of this study is that the pilot is currently on-going and, therefore, we do not yet have any implementation results or any quantitative or qualitative metrics to evaluate the project. In this section, we highlight some of the potential limitations we anticipate encountering with respect to our digital solution. First, we discuss limitations on uptake due to user characteristics or behaviour patterns and second, we consider technical or design limitations of the digital solution, in its current form.

3.1.1. Limitations on account of user characteristics

Studies indicate that despite careful design considerations the operation of such tools may yet require more intermediate, and not merely foundational, digital competencies and skills. [31] One way to mitigate this hurdle may be through supplementary upskilling programmes on digital literacy for the target users.

Second, the barriers to access may also be exacerbated by poor digital connectivity and infrastructure that often correlate with our target groups. [6, 31] Therefore, creating low data consuming tools seem likelier to work in such environments.

One additional aspect to be mindful of when using digital technologies is the community's perceptions and associations with that technology itself. While WhatsApp is one of the most ubiquitous messaging applications in India [38], its perceived role in the spread of misinformation cannot be ignored [40]. The deployment of any tool must be followed up with a well-structured socialisation process as well.

3.1.2. Technical and design limitations

As discussed in Section 1.2., the GenAI or LLM component used by our technology stack is predominantly trained on English data. [36] This imbalance in the underlying training data may lead

to creeping in of certain biases or incomprehension on the part of the AI, which reinforces the need for a human-supervised training approach, in lieu of an indigenous LLMs.

Second, we have deliberately excluded judicial decisions (i.e., case law) from our database at this stage, due to several challenges related to legal accuracy, accessibility and affordability. Our interviews with legaltech experts in India highlighted that Indian judgments are fairly lengthy and unstructured. Each document contains several components including facts, decisions, references, ancillary observations, dissenting views, citations, etc. Observations made in judgements are typically fact-specific and may not apply universally. Moreover, the frequency with which new judgements are passed, or repealed, by numerous judicial and quasi-judicial bodies across the country can be challenging to track, validate and update. Without systematic and regular annotation of these documents, there is a risk of irregularities in the AI's responses. From an affordability perspective, the collation of judicial decisions is a resource intensive exercise that could drive up overall costs of the tool. Given these constraints and due to the lack of structured, annotated datasets, we have refrained from using judgements in our database at this stage. Once efforts within the Indian legaltech ecosystem towards organising and annotating judgments advance, we may consider incorporating case law into our knowledge database.

Currently, the chatbot uses a RAG method to generate responses to a user query, as discussed in Section 2.1.2. It lacks a dynamic response function to identify gaps in the user query and request specific, additional pieces of information to fill in those gaps and build more context for a targeted response. This may impact its ability to assess applicability of specific laws to a certain degree, especially in a complex domain like informal labour which is largely unregulated and varied. For example, many labour legislations are applicable only to certain types of establishments, which meet specific criteria such as scope of operations, number of workers, etc. The chatbot would have to elicit these factual details from the user to make an accurate assessment of whether the legislation would apply to them. In the future, we hope to build in dynamic response capability to improve the AI's contextual understanding and responses.

3.2. Scope for further research

Experimentation in legal tech can foster opportunities for innovation and new possibilities in overcoming barriers to legal literacy and many of the limitations we have discussed in Section 3.1. Over time, digital solutions may enhance not only access to legal information but also supply opportunities for upskilling, creating demand for newer, more sophisticated tools as user needs evolve. Further, as excluded populations are enabled to participate in legal systems via digital solutions, it can foster changes to people's lives and provide a gateway for more specialised social justice opportunities. [31] Such experiments can also generate learnings on the psychological impacts of AI [37] and strategies to mitigate digital exclusion from poorly managed digitisation [29]. This research will be vital for ensuring that technology-based interventions remain equitable and inclusive.

Our projects can serve as valuable knowledge resources for technology improvements and policy outcomes in the future. [11] For example, the RAG method can be systematically tested and studied for accuracy in the legal domain. The documentation of our project process can help set out guardrails and protocols for responsible digital and AI practices. The aggregate, anonymised data collected from users can help inform larger legal and socio-economic policy reforms as well. The review of responses can highlight real-life trends at the individual and community level to help design appropriate policy interventions.

Lastly, Project Setu will also benefit from a baseline study that can estimate and help identify populations that are excluded from the legal system due to lack of legal literacy, in order to develop more targeted and effective interventions. A deeper understanding of gaps in legal access will help in designing future iterations of Project Setu and similar legal literacy initiatives.

Acknowledgements

We would like to express our gratitude to the team at BTGAdvaya for their financial support for Project Setu and Abhishikta Mallick for her feedback and editorial suggestions, which helped enhance the depth and rigour of this paper. We would also like to extend our appreciation towards India LabourLine, our NGO partner for the ILR project.

References

- [1] M. Hagan, Good AI Legal Help, Bad AI Legal Help: Establishing quality standards for responses to people's legal problem stories, JURIX 2023: 36th International Conference on Legal Knowledge and Information Systems, AI and Access to Justice Workshop, 2023. URL: <https://ssrn.com/abstract=4696936>.
- [2] A. Kumar, Status of Legal Literacy in India: A Report on National Legal Literacy, Mission Notion Press, 2013. URL: https://www.researchgate.net/publication/256054889_STATUS_OF_LEGAL_LITERACY_IN_INDIA_A_REPORT_ON_NATIONAL_LEGAL_LITERACY_MISSION.
- [3] Press Trust of India, Absence of legal awareness root cause of rights' deprivation, 2019. URL: https://www.business-standard.com/article/pti-stories/absence-of-legal-awareness-root-cause-of-rights-deprivation-119081800664_1.html.
- [4] ASER, Annual Status of Education Report (Rural) 2022, 2023. URL: <https://asercentre.org/wp-content/uploads/2022/12/aserreport2022-1.pdf>.
- [5] ASER, Annual Status of Education Report (Rural) 2023: Beyond Basics, 2024. URL: <https://asercentre.org/wp-content/uploads/2022/12/ASER-2023-Report-1.pdf>.
- [6] A. Mahendru, M. Dutta, and P.R. Mishra, Inequality Report 2022: Digital Divide, Oxfam India, 2022. URL: https://d1ns4ht6ytuzzo.cloudfront.net/oxfamdata/oxfamdatapublic/2022-12/Digital%20Divide%20India%20Inequality%20Report%202022_PRINT%20with%20cropmarks.pdf?3l.73PGQrpQfYrnwWeoXV3BFjhETfA_p.
- [7] Rukmini. S, In India, who speaks in English, and where?, 2019. URL: <https://www.livemint.com/news/india/in-india-who-speaks-in-english-and-where-1557814101428.html>.
- [8] A. Kapoor, A. Jhalani, N. Vinayak and S. Zutshi, State of Foundational Literacy and Numeracy in India, Institute for Competitiveness, 2021. URL: https://www.competitiveness.in/wp-content/uploads/2021/12/Report_on_state_of_foundational_learning_and_numeracy_web_version.pdf.
- [9] World Bank Group, Gender Data Portal, Literacy rate (%), UNESCO Institute for Statistics (UIS), Stat Bulk Data Download Service: <https://apiportal.uis.unesco.org/bdds>, 2024. URL: <https://genderdata.worldbank.org/en/indicator/se-adt?gender=total>.
- [10] Dr. S. Gaur, H. Kumari, S. Mathur, and S. Rout, Reimagining Possibilities: Reducing the Gender Digital Divide in India, Sattva Knowledge Institute, 2023. URL: <https://www.sattva.co.in/ski/reducing-gender-digital-divide-india/>.
- [11] S.K. Agarwal, A. Kumar, A.A., Nanavati, A. and N. Rajput, User-generated content creation and dissemination in rural areas, Information Technologies and International Development, Vol. 6, No. 2, pp. 21-37, 2003. URL: <https://itidjournal.org/index.php/itid/article/download/518/518-1389-2-PB.pdf>.
- [12] EduAnalytics, India learning poverty brief, 2019. URL: <https://thedocs.worldbank.org/en/doc/386361571223575213-0090022019/original/SASSACININDLPBRIEF.pdf>.
- [13] Government of India, Census of India 2011, Paper 1 of 2018: Language, URL: https://censusindia.gov.in/nada/index.php/catalog/42458/download/46089/C-16_25062018.pdf.

- [14] U. Krishna, Bolstering Internationalised Domain Name Uptake in India, ICRIER, 2021. URL: [https://icrier.org/pdf/Bolstering Internationalised Domain Name Uptake in %20India .pdf](https://icrier.org/pdf/Bolstering_Internationalised_Domain_Name_Uptake_in_%20India_.pdf).
- [15] The Economist, India has a unique opportunity to lead in AI, 2024. URL: <https://www.economist.com/asia/2024/10/03/india-has-a-unique-opportunity-to-lead-in-ai>.
- [16] M. Barot, S. Patel and T. Patel, Dominance of English over the local languages on Internet, JETIR, Volume 8, Issue 10, ISSN-2349-5162, 2021. URL: <https://www.jetir.org/papers/JETIR2110480.pdf>.
- [17] M. Khan, Large language models are leaving non-English speakers behind, 2024. URL: <https://wired.me/technology/english-dominance-in-llms/>.
- [18] J. Tidy, ChatGPT bias: 3 ways non-English speakers are being left behind, 2023. URL: <https://medium.com/@joetidy/chatgpt-bias-3-ways-non-english-speakers-are-being-left-behind-799b4898eee6>.
- [19] Internet and Mobile Association of India (IAMAI) and Kantar, Internet in India 2023, 2024. URL: <https://www.iamai.in/sites/default/files/research/INTERNET%20IN%20INDIA%202023.pdf>.
- [20] Internet and Mobile Association of India (IAMAI) and Kantar, Internet in India 2022, 2023. URL: [https://www.iamai.in/sites/default/files/research/Internet%20in%20India%202022_Print%20v ersion.pdf](https://www.iamai.in/sites/default/files/research/Internet%20in%20India%202022_Print%20version.pdf).
- [21] GMSA, The Mobile Economy Asia Pacific 2024, 2024. URL: <https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2024/07/240724-Mobile-Economy-Asia-Pacific-2024-FINAL.pdf>.
- [22] M. Nikore, India's Gendered Digital Divide: How the Absence of Digital Access is Leaving Women Behind, Observer Research Foundation, 2021. URL: <https://www.orfonline.org/expert-speak/indias-gendered-digital-divide#:~:text=Indian%20women%20are%2015%20percent.41%20percent%20of%20adult%20men>.
- [23] P. Kundu, Indian education can't go online – only 8% of homes with young members have computer with net link, 2020. URL: <https://scroll.in/article/960939/indian-education-cant-go-online-only-8-of-homes-with-school-children-have-computer-with-net-link>.
- [24] U. Kenny, Á. Regan, Co-designing a smartphone app for and with farmers: Empathising with end-users' values and needs, Journal of Rural Studies, Volume 82, Pages 148-160, ISSN 0743-0167, 2021. URL <https://doi.org/10.1016/j.jrurstud.2020.12.009>.
- [25] R. B. Jones, P. Stallard, S. S. Agha, S. Rice, A. Werner-Seidler, K. Stasiak, J. Kahn, S. A. Simpson, M. Alvarez-Jimenez., F. Rice, R. Evans and S. Merry, Practitioner review: Co-design of digital mental health technologies with children and young people, Journal of Child Psychology and Psychiatry / J Child Psychol Psychiatr, 61: 928-940, 2020. URL <https://doi.org/10.1111/jcpp.13258>.
- [26] G. Concilio, G. Costa, M. Karimi, M. Vitaller del Olmo, and O. Kehagia, Co-Designing with Migrants' Easier Access to Public Services: A Technological Perspective Social Sciences 11, no. 2: 54, 2022. URL: <https://doi.org/10.3390/socsci11020054>.
- [27] M. Đurić, T. Martinec, M. Porobija and M. Štorga, Designing For Legal Practitioners: Lessons Learned from Legal Tech Development and Implementation. Proceedings of the Design Society 3 (2023): 1377–86, 2023. URL: <https://doi.org/10.1017/pds.2023.138>.
- [28] J. Suresh K. and Dr. D. Shobana, A Study on Challenges and Solutions for Implementing Digital Literacy Programmes in India's Rural Areas, 2024. URL: https://www.researchgate.net/publication/382671754_A_Study_on_Challenges_and_Solutions_for_Implementing_Digital_Literacy_Programmes_in_India's_Rural_Areas.
- [29] Hernandez, K., Flynn, J., He, J. and Alsahi, H, Towards digital inclusion in rural transformation, Rome, FAO, 2024. URL: <https://doi.org/10.4060/cc9816en>.

- [30] P. Gutch, Ok Google: How is voice making technology more accessible in India? 2020. URL: <https://www.thinkwithgoogle.com/intl/en-apac/future-of-marketing/emerging-technology/ok-google-how-is-voice-making-technology-more-accessible-in-india/>.
- [31] R. Zelezny-Green, S. Vosloo and G. Conole, A Landscape Review: Digital Inclusion for Low-skilled and Low-literate People, United Nations Educational, Scientific and Cultural Organisation (UNESCO), 2018. URL: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000261791&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_0db84645-e500-4c57-85b5-e60e1fe60af2%3F%3D261791eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000261791/PDF/261791eng.pdf#%5B%7B%22num%22%3A20%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C0%2C680%2Cnull%5D.
- [32] Isobar India, The Voice Playbook, P 3, 2021. URL: https://www.mmaglobal.com/files/documents/mma_voice_playbook_0.pdf.
- [33] A. Biswas, Voice Technology in India: Now and Future, Wat Consult, 2019. URL: <https://www.watconsult.com/watinsights/voice/>.
- [34] A. Singh, A. Khaitan, R. Suresh, A. S. Pilla, Indian Languages – Defining India’s Internet, KPMG India and Google, 2017. URL: <https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf>.
- [35] KPMG, #IndiaTrends2018: Trends sharing Digital India, 2018. URL: <https://assets.kpmg.com/content/dam/kpmg/in/pdf/2018/05/IndiaTrends2018-Trends-shaping-Digital-India-Internet.pdf>.
- [36] M. Khan, Large language models are leaving non-English speakers behind, Wired, 2024. URL: <https://wired.me/technology/english-dominance-in-llms/>.
- [37] S. Tully, C. Longoni and G. Appel, Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity, MSI Working Paper Series, 2024. URL: https://thearf-org-unified-admin.s3.amazonaws.com/MSI_Report_24-132.pdf.
- [38] Statista, Most used messenger by brand in India as of September 2024, 2024. URL: <https://www.statista.com/forecasts/1348378/most-used-messenger-by-brand-in-india>.
- [39] Government of India, Bhashini, Ministry of Electronics and Information Security. URL: <https://bhashini.gov.in/ulca>.
- [40] T. McLaughlin, How WhatsApp Fuels Fake News and Violence in India, Wired, 2018. URL: <https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india/>.
- [41] Government of India, Ministry of Labour and Employment, Labour Bureau, Report on Employment in the Informal Sector and Conditions of Informal Employment: Volume IV, 2013-14. URL: https://labour.gov.in/sites/default/files/report_vol_4_final.pdf.
- [42] S V R. Murthy, Measuring Informal Economy in India Indian Experience, International Monetary Fund, 2019. URL: <https://www.imf.org/-/media/Files/Conferences/2019/7th-statistics-forum/session-ii-murthy.ashx>.
- [43] Oxfam India, Overview of Informal Sector, 2022. URL: https://d1ns4ht6ytuzzo.cloudfront.net/oxfamdata/oxfamdatapublic/2022-04/ISW%20Fact%20Sheets%202022_0.pdf?HAYi8EGUzESpZptOAWes007.C_nC3R4J.
- [44] C. Thomas, D. Varma, N. Jayaram, R. Mehrotra, R. R. Subramanian and S. Sugathan, Unlocking the Urban: Reimagining Migrant Lives in Cities Post COVID-19, Aajeevika Bureau, 2020. URL: <https://aajeevika.org/wp-content/uploads/2023/10/Unlocking-the-Urban.pdf>.
- [45] P. Béchar, O. M. Ayala, Reducing hallucination in structured outputs via Retrieval-Augmented Generation, Cornell University, 2024. URL: <https://arxiv.org/abs/2404.08189>.