

# Enhancing Accessibility to Government Services in India through Natural Language Processing (NLP): Bridging Language Barriers and Supporting Literacy

Abhijay Parija

**Abstract:** *Natural Language Processing (NLP), a subset of artificial intelligence, has advanced from rule-based models to deep learning algorithms that can understand and respond to natural human language. By leveraging this technology, government portals can automate chatbots, and make services more accessible to those who face language barriers and literacy challenges. This paper examines the challenges faced by internal migrants and illiterate individuals in India while attempting to access essential services such as healthcare, education and financial aid. Two NLP-based solutions are presented: an on-site terminal to assist users in government offices and a remote, web-based service accessible through personal devices such as phones and laptops. Both of the solutions listed aim to translate spoken language and simplify the submission of documents. Despite some initial costs and maintenance, the long-term benefits of enhanced accessibility and reduced operational burdens justify these costs. Ultimately, adopting NLP in government services can contribute to a more inclusive and equitable society, and improve the quality of life for millions.*

**Keywords:** Natural Language Processing (NLP), Artificial Intelligence (AI), Deep Learning, Chatbots, Language Barriers

## 1. Introduction

### NLP as an emerging subset of AI

Natural language processing (NLP) is a component of artificial intelligence that enables computers to understand and communicate with human language. Natural language processing is used to analyse text and speech data effectively. With the help of machine learning NLPs can understand and respond to different dialects, grammatical irregularities, etc (Amazon Web Services). Given their diverse abilities NLPs may be key in helping us automate more capable chatbots; classify and extract text; process, analyse, and archive large documents.

Early iterations of NLP programs used rule-based models that worked on pre-programmed rules and did not involve any AI presence. Therefore, these early NLPs could only respond to a limited number of prompts. The second generation of NLPs were statistical NLPs which involved analysing the statistical likelihood for the respective text and voice data and assigning a probable meaning to those elements. Finally, a new up-and-coming form of NLPs is Deep Learning NLPs that train and test raw and unprocessed data (Holdsworth). Being the successor of statistical NLP, deep learning NLPs involve the use of neural network models that attempt to mimic how the human brain understands and responds to spoken or written language, through the use of math and algorithms (IBM, "What is a Neural Network?").

An application of NLP is preserving endangered languages by creating accessible digital archives for these "rare languages" that have a lesser representation relative to the larger population. The Europarl parallel corpus is a collection that assists in developing translation systems for storing uncommon languages (Khurana et al.). By combining these datasets with statistical or deep learning techniques, translating these languages can be made more efficient and straightforward. Consequently, this can be beneficial for

individuals who speak local languages that are not generally spoken by a significant portion of the population in a particular nation. Google Translate is an excellent instance of a successful NLP system that utilizes extensive databases and employs deep learning techniques. Through this Google Translate has been able to translate up to 133 languages. This demonstrates that if such technology is employed wisely it may generate major benefits in terms of assisting individuals to communicate across language barriers.

In addition to its effectiveness in data translation, NLP is capable of processing vast quantities of unstructured data, analyzing and making predictions based on the data that is provided to. Typically, dealing with unstructured data transcripts like this would have required carefully examining each input by an individual. This would render the process tedious and inefficient. NLPs have already been utilized in these scenarios and have yielded favourable outcomes. An example of this is the healthcare industry utilizing NLPs to analyze medical records, research articles, and clinical notes to assist doctors in enhancing the treatment of patients. Furthermore, the analysis of information on patient data has greatly facilitated medical professionals in swiftly recognising illnesses. (Khatri) Governments have also utilized NLPs to carry out extensive data analysis. (Khatri) For instance, the government of India back in 2021, had made use of NLP techniques in order to develop an e-governance platform that aimed at improving smart city initiatives and managing COVID-19 responses. Platforms such as this made use of NLPs to convert public voice inputs into text, categorize feedback and aid governments in automatically notifying relevant government departments of actions regarding this (Leelavathy and Nithya).

### Lack of access to government services for the public

One of the main roles of a welfare state is to ensure that all citizens can access and use the services necessary for them to thrive. The Indian government offers its residents a range of

important services including, but not limited to, healthcare, financial assistance, pensions, transportation, and education. Various state-specific initiatives foster entrepreneurship, scholarships, and pensions. Residents of India are eligible to use the programs offered by the Central government, as well as specific programs offered by their State's government without hindrance. However, residents of the country who do not speak the predominant language of the region face difficulties in accessing these vital programs.

According to a report by the Ministry of Statistics and Program Implementation, 28.9% of the Indian population migrated from 2020 to 2021 (Indian Government). Internal migrants can often face linguistic difficulties as they might not be able to speak the native language of that region. Speaking the native language might be essential to avail of the vital services listed above. Such a demographic of people can face difficulties in taking advantage of state-specific schemes as well.

Additionally, illiteracy is also a prevalent issue in India. Indian adults have a 76% literacy rate, according to World Bank Statistics 2024 (WorldBank). Even though this number is significantly higher than it was in previous years, 24% of adults in the population are illiterate. These days, because of their illiteracy, this group of people find it difficult to access necessary government services like the ones mentioned above. Inadequate reading and writing skills in a particular language may make it difficult to understand and complete government forms.

Functional literacy is also important to look at. While 76% of the population might be literate, having functional literacy and practical use of literacy is different. It is likely that out of the population that is considered "literate" there could be a significant number of people who cannot read or write well enough to function independently in their community. This means a lot of them cannot make use of the essential government services that are stated above.

### **What can NLPs do?**

For all the issues stated above NLPs can provide a viable solution, as they can detect and translate spoken languages. By adding native languages to repositories that support translation, instructions and forms for government services can be easily translated through NLP.

Since NLPs can understand and respond to human speech, chatbots powered by NLP can perfectly imitate humans providing help and support. What this could mean for a government portal is that the chatbots can be operational 24/7 to resolve requests concurrently. They can do this without requiring a staff member unless the issue or the complaint is escalated, which reduces the need for people to operate the helpline. NLPs can translate information from any of the languages and provide a human-like conversation (Woord). By using conversational AI techniques, chatbots with NLP can respond with appropriate actions and helpful comments with contextual answers. This has already been quite widespread as conversational agents such as Amazon's Alexa and Apple's Siri are already doing this well and assisting users in real time. Employing a multilingual version of such

technologies in a government portal can be quite helpful for the citizens of India.

Another benefit of NLPs is that they can identify if the user has provided Personal identifiable information (PII). It can quickly locate such personal information in documents and identify the individuals. So when governments are handling large volumes of PII, NLPs can create versions with the PII obscured for enhanced cyber security. Once a PII is identified, NLPs can automate this redaction process. The process they follow is either replacing or removing PII from documents, ensuring that the sensitive information is either not shared or analysed. They do this simply by detecting and masking the sensitive information with "X" or "\*" (Howal et al.).

When handling large data NLPs can take into account large volumes of digital text and create summaries and synopses for indexing. So when the government is handling large datasets it can provide summaries. NLP does this by making use of extractive and abstractive summarisation techniques. Extractive summarisation involves picking key sentences from the text while abstractive summarisation involves generating new sentences that capture the essence of the original text. This can make everything very resource-efficient. Furthermore, if there is a large set of applicants appealing to the government or writing to the government, summarisation by NLP makes it easier for the government to understand the appeals and respond to the applicants.

### **Our Proposal**

Based on the problems we identified, we propose a potential NLP-based solution. This proposal will differ based on the On-site or Remote requirements. "On-site" refers to translation needs at a point of service - municipal office, public health centre, police station etc. "Remote" refers to a solution that would be available to any citizen using a device with internet access.

### **On-site**

Our onsite proposal plans to address the issue of people trying to submit personal documents to government offices to avail themselves of essential services that might range from rations, healthcare, education, water, etc. People who are illiterate or do not know the local language cannot read or write in the local language of the area, they cannot fill out important government forms and know which documents to specifically provide to the government. Therefore our proposed solution is an on-site terminal that can aid people seeking specific services, submitting documents and receiving printouts, and getting step-by-step instructions on how to avail of a service. The terminal will be powered by NLPs that can aid and detect the language of a person through a speaker and then help them in the language the individual is talking in. The terminal will also have the ability to translate scanned documents and speaker's requests into the local language or national language of the area if further help is required by the individual. To aid illiterate individuals the terminal will also be equipped with a speaker that can speak information to the users.

This can be a device that can be located in every government centre for people to use. The average costs for this device are as follows: \*

Equipment	Costs
Core i3	8000
intel arc a780	19279
RAM - 16gb	3,199
SSD - 1tb	8095
Coolers	7000
UPS	4000
Motherboard - GigaByte	990
Mic	3000
Speaker	3500
Monitor	8000
Scanner + Printer	15000
<b>Tentative Total</b>	<b>80063</b>

Note that these costs are just for hardware, running and hosting an NLP chatbot with voice activation can incur additional costs for the government. Also, these are retail prices found online. Prices would be significantly lower for a large contract.

## 2. Remote

The remote version of this NLP is there to assist the user remotely with their queries. This aims to reduce costs and time for the citizens and the government as it aims to answer their concerns through an online portal.

The government can also use an online version of this NLP device. This will be a web-based service that can be accessed from the government portal itself. This service will be accessible remotely through computers and mobile devices. The website will host general information about queries related to accessing government services such as “how to apply for a metro card” or “what documents do I require to register a ration card”. The web service will also have the option to provide voice commands and can provide help through a chatbot (The voice command can be any language and the chatbot will reply in that language itself). For privacy concerns, the web service won’t be able to accept personally identifiable information and so can’t process documents like the onsite counterpart of the service.

The benefit of this online counterpart is that it can reduce potential costs for individuals as it won’t require their time and effort to go into government centres. The cost for the government would be hiring someone who needs to program this NLP and then also paying maintenance costs for this model.

## 3. Discussion

The benefits of the model as stated above are the reduced time and resources required to access government services that will come from the introduction of the devices. Furthermore, it can positively change millions' lives as it makes essential services more accessible to them.

Some potential issues with the on-site model are that it will require some setup costs as well as maintenance over time, which can incur some costs and be expensive for the government. Furthermore, government employees will require some additional training to use the hardware device as otherwise the device can’t be used to its full potential.

For the software solution, we would require to program in the different information about the government and their policies which will inquire costs as professionals need to be hired to do this. Additionally, the software will require maintenance as debugging will be required every week to avoid bugs to prevent inconveniences for the user.

## 4. Conclusion

Implementing NLP into the ways people can access government services can enhance accessibility and efficiency in delivering public services. By making use of such technologies we can support the majority of the population of this diverse country in accessing government services by bridging the linguistic barrier and powering the joint effort to make government services as accessible as possible despite an individual’s native language or literacy levels.

Our proposed on-site and remote NLP solutions will cater to the requirements of our diverse populations. People from any part of the country and any demographic can access government services based on the solutions we provide. The on-site solution with its hardware-based approach, addresses the needs of those requiring important in-person assistance that might require exchanging and submitting documents. In contrast, the remote solution offers a cost-effective method of getting answers to general queries through an online portal.

While setting up devices there might be some initial setup costs and some maintenance costs, but provide long-term benefits that justify these costs. Improved access to essential government services will improve the quality of life of the citizens and will ensure that government services are fully utilised. Additionally, by automating routine inquiries and facilitating communication, NLP technology can help reduce the workload of government employees and make the workspace more efficient.

Ultimately, the adoption of NLP technologies in government services represents a more inclusive, efficient and responsive government. By empowering the citizens of the country with better access to information services, we can work towards a more equitable society for all.

## References

- [1] Amazon Web Services. “What Is Natural Language Processing? - NLP - AWS.” *Amazon Web Services, Inc.*, 2024, [aws.amazon.com/what-is/nlp/](https://aws.amazon.com/what-is/nlp/). Accessed 20 June 2024.
- [2] Howal, Sadanand, et al. “Object Detection for Autonomous Vehicle Using TensorFlow.” *Advances in Intelligent Systems and Computing*, 19 Oct. 2019, pp. 86–93, [https://doi.org/10.1007/978-3-030-30465-2\\_11](https://doi.org/10.1007/978-3-030-30465-2_11). Accessed 30 Aug. 2021.
- [3] IBM. “What Are Neural Networks?” *Www.ibm.com*, IBM, [www.ibm.com/topics/neural-networks](https://www.ibm.com/topics/neural-networks). Accessed 14 June 2024.
- [4] ---. “What Is Natural Language Processing?” *IBM*, 6 June 2024, [www.ibm.com/topics/natural-language-processing](https://www.ibm.com/topics/natural-language-processing). Accessed 14 June 2024.
- [5] Khatri, Manish Raj. “Integration of Natural Language Processing, Self-Service Platforms, Predictive

- Maintenance, and Prescriptive Analytics for Cost Reduction, Personalization, and Real-Time Insights Customer Service and Operational Efficiency.” *International Journal of Information and Cybersecurity*, vol. 7, no. 9, 2 Sept. 2023, pp. 1–30, [publications.dlpress.org/index.php/ijic/article/view/36/32](https://publications.dlpress.org/index.php/ijic/article/view/36/32). Accessed 23 May 2024.
- [6] Khurana, Diksha, et al. “Natural Language Processing: State of the Art, Current Trends and Challenges.” *Multimedia Tools and Applications*, vol. 82, no. 3, 14 July 2022, pp. 3713–3744, [link.springer.com/article/10.1007/s11042-022-13428-4](https://link.springer.com/article/10.1007/s11042-022-13428-4), <https://doi.org/10.1007/s11042-022-13428-4>. Accessed 17 May 2024.
- [7] Leelavathy, S., and M. Nithya. “Public Opinion Mining Using Natural Language Processing Technique for Improvisation towards Smart City.” *International Journal of Speech Technology*, vol. 24, no. 3, 11 Nov. 2020, pp. 561–569, <https://doi.org/10.1007/s10772-020-09766-z>.
- [8] Government of India. *Ministry of Statistics and Programme Implementation*. 2020.
- [9] Woord, Peter. “Computer Assisted Reading in German as a Foreign Language, Developing and Testing an NLP-Based Application.” *Calico Journal*, vol. 28, no. 3, 2011, [www.jstor.org/stable/calicojournal.28.3.662?seq=1](https://www.jstor.org/stable/calicojournal.28.3.662?seq=1). Accessed 19 Aug. 2024.
- [10] WorldBank. “Literacy Rate, Adult Total (% of People Ages 15 and Above) - India | Data.” *Data.worldbank.org*, 24 Apr. 2024, [data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=IN](https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=IN). Accessed 17 July 2024.