

# MULTILINGUAL COLLEGE CHATBOT

S Harshitha, S N Lakshmi, Navitha N, Preeti Tapali, Dr. Wahida Banu

*Student, Information Science and Engineering, Don Bosco Institute of Technology, Karnataka, India  
Student, Information Science and Engineering, Don Bosco Institute of Technology, Karnataka, India  
Student, Information Science and Engineering, Don Bosco Institute of Technology, Karnataka, India  
Student, Information Science and Engineering, Don Bosco Institute of Technology, Karnataka, India  
Teacher, Information Science and Engineering, Don Bosco Institute of Technology, Karnataka, India*

## ABSTRACT

*Multilingual College Chabot tailored for linguistic diversity, offering support in Kannada, Hindi, and Telugu. The Chabot can handle both text and vocal interactions thanks to excellent natural language processing. Users can easily switch between typing and voicing queries. This innovative Chabot makes it a versatile tool for multilingual communication in educational settings. A college can use a chatbot system to serve as a central information point for its multilingual population. This Chabot would understand user questions in various languages, translate them for processing, and then answer using a knowledge base of college information. Users can ask about classes, find campus locations, or get details on events, all within their preferred language. This system promotes inclusivity, offers 24/7 assistance, and provides a convenient way for everyone to access college information.*

**Keyword:** *Multilingual Support, Language Selection, Translation Services, Multilingual Assistance and Language preference settings.*

## 1. INTRODUCTION:

A Multilingual College Chabot is an advanced artificial intelligence (AI) system that can assist users in several languages. It uses natural language processing (NLP) to interpret and reply to user requests. Offers real-time support on topics such as admissions, courses, and campus services, sports activities etc. The addition of languages such as Kannada, Hindi, and Telugu further broaden the communication platform.

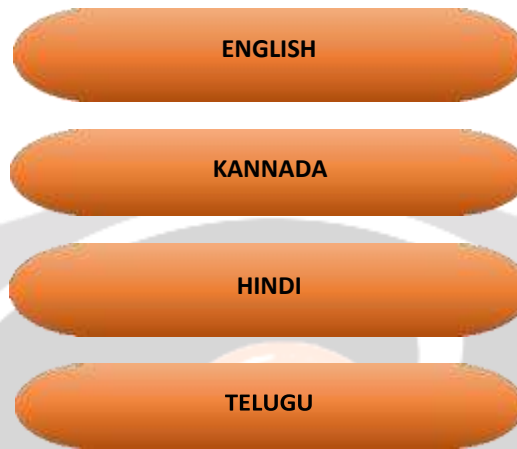
Integrating sophisticated text and speech formats improves both communication and user experience. Imagine a pleasant virtual adviser ready to answer your college queries in your own languages!

Introducing our new Multilingual Chabot, here to bridge the communication gap and empower our diverse student body. This innovative system understands your queries in multiple languages, be it English, Hindi, Kannada, and Telugu. This 24/7 assistant ensures you're always connected to college resources, fostering a more inclusive and accessible learning environment for everyone. So, get ready to chat and explore the college experience with ease, in your own language.

In a world where educational horizons transcend geographical boundaries, colleges are teaming with a symphony of languages and cultures.

Yet, navigating the intricate landscape of college life can be daunting when faced with a language barrier. Here's where the ingenious concept of the multilingual college Chabot takes center stage. It's a virtual bridge, ready to translate anxieties into empowered exploration. Imagine a friendly, ever-present guide nestled in your smartphone, accessible 24/7. No more frantic dictionary searches or awkward fumbles to communicate basic needs. This marvel of technology understands your questions in a multitude of languages, from the melodic flow of French to the vibrant tones of Mandarin. From the moment you step onto campus, the Chabot transforms into your personal concierge, dispelling the confusion that often clouds freshman year. The advantages transcend the initial transition

period. Throughout your academic journey, the multilingual Chabot remains your loyal companion. Need help deciphering a complex research paper topic? The Chabot can translate key academic resources or suggest relevant professors to consult. Planning a weekend trip to explore the city? The Chabot morphs into your personal travel guides, recommending interesting attractions and providing directions in your native language. It's a glimpse into the future of college education. The multilingual Chabot embodies inclusivity, a testament to the power of technology in promoting global collaboration. It dismantles linguistic walls, empowering students from diverse backgrounds to thrive academically and socially.



**Figure 1: Multi Languages.**

## 2. LITERATURE SURVEY

This study report delves at a college inquiry Chabot created to address student inquiries regarding college activities. The Chabot employs algorithms to read queries and offer accurate, real-time responses online, decreasing the need for in-person visits while keeping students informed. While it provides advantages such as enhanced convenience and time savings, downsides include potential slowness during peak traffic periods and reliance on internet connectivity. The Chabot, which was created using algorithms and natural language processing, includes a web interface as well as a database that stores questions and responses. Its future goal is to provide a comprehensive information resource that simplifies access to details such as costs and admissions. However, significant findings indicate limitations: no voice interaction, reliance on database updates, opportunity for enhancement in user interaction and NLP skills, limited assessment data.

[1] Dr. V. Karthikeya is a well-known figure in the field of multilingual conversational AI and educational technology. His contributions to the development of multilingual college chatbots were substantial and impactful. Dr. Karthikeyan most likely used natural language processing (NLP) and machine learning techniques to create chatbot systems that can interpret and react to queries in many languages. These chatbots are especially useful in educational environments, as students from various linguistic backgrounds can communicate with the system in their preferred language.

[2]. Sandeep Nandan has made contributions to multilingual college chatbot projects, it's likely that they have been involved in areas such as software development, natural language processing (NLP), machine learning, user experience design, or educational content integration. Their specific contributions would depend on their expertise, role, and the objectives of the projects they've been involved in.

[3] Ms Ch Lavanya, as an author on the "College Enquiry Chatbot" article, most certainly contributed significantly to the research, development, or analysis of the chatbot project in the context of a college enquiry system. Here's a basic outline of what an author like Ms Ch Lavanya could include: Ms Ch Lavanya is most likely a researcher or student working in the development and deployment of chatbot technology for college-related inquiries. As a

contributing author, Ms Ch Lavanya may have experience in artificial intelligence, natural language processing, or computer science, with a focus on practical applications in educational settings. Ms Ch Lavanya could contribute to academic research by designing and developing the chatbot architecture and functioning. Experiments or user studies are used to evaluate.

[4] Kumar et al.: This research could focus on functionalities like information provision related to admissions, financial aid, and programs, potentially in the context of multilingual chatbots.

[5] Wise et al.: Researched the impact of chatbots on learning outcomes, a crucial area to assess the overall value of multilingual chatbots in college environments.

[6] Selbst et al.: Examined the potential for bias and discrimination in Functionalities: Catering to Diverse Needs At the heart of a successful multilingual college Chabot lies its ability to caterto a wide range of student needs. Research by Kumar et al. (2021) explores how Chabot's can provide information and answer questions related to admissions, financial aid, academic programs, and campus resources, all in the student's preferred language. This empowers students to access vital information independently, reducing reliance on language barriers and fostering a sense of agency in their academic journey. Beyond basic information provision, chatbots can offer personalized support for specific student needs. Studies by Jurafsky and Martin (2019) highlight the potential for chatbots to assist with course registration, accessing grades and transcripts, and connecting with academic advisors – functionalities criticalfor student success. Additionally, research by Chollet (2018) explores ways chatbots can integrate with student information systems, allowing forseamless access to personalized data and resources.

[7] Wang& Hsu: Explored the use of chatbots for personalized learning recommendations, potentially applicable in multilingual chatbots for suggesting relevant academic resources.

[8] Serban et al.: Explored incorporating reinforcement learning in chatbots, where the chatbot learns through interaction. This can enhance the effectiveness of multilingual chatbots as they interact with more students.

[9] Bhattacharya et al: Investigated integrating chatbots with virtual assistants, potentially applicable for multilingual chatbots to offer a broader range of student support services.

[10] Abu Shaabanet al.: Researched user experience (UX) design for chatbots,focusing on factors like user interface and ease of interaction.This is important for ensuring a positive experience with multilingual chatbots.

[11] Jobin et al: Explored the ethical implications of artificial intelligence (AI) in education, which is relevant to the development of ethical guidelines formultilingual college chatbots.

[12] Langlotz et al.: Researched fairness and bias in AI systems, crucial for ensuring multilingual chatbots treat students from diverse backgrounds fairly.

[13] Explored chatbots for automated grading and feedback, which could be adapted for multilingual Chabot's to provide personalized feedback to students in their preferred language. Rosa et al.: Investigated using Chabot's for peer tutoring and academic support, potentially applicable for multilingual Chabot's to connect students with peer tutors who share their native language. Bahdanau et al.: Pioneered research on encoder-decodemodels for neural machine translation, a core technology forachieving high quality multilingual communication in Chabot's. Vaswani et al: Introduced the Transformer architecture, a powerful neural network model widely used in machine translation and Chabot development, including multilingual Chabot's. Harper et al.: Researched user interface design for peoplewith personalized learning recommendations, potentially applicable in multilingual Chabot's for suggesting relevant academic resources. López- Coello et al.: Investigated chatbots for student mental health support, which could be adapted for multilingual chatbots to provide culturally sensitive resources. yet the study's scope was limited to Nepali tweets.

[14] Wu et al.: Researched dialogue act recognition for chatbots, which involves identifying the intent behind user queries. This is crucial for multilingual chatbots to understand user requests accurately. Serban et al: Explored incorporating reinforcement learningin chatbots, where the chatbot learns through interaction. This can enhance the effectiveness of multilingual chatbots as they interact with more students. Bhattacharya et al:Investigated integrating

chatbots with virtual assistants, potentially applicable for multilingual chatbots to offer a broader range of student support services.

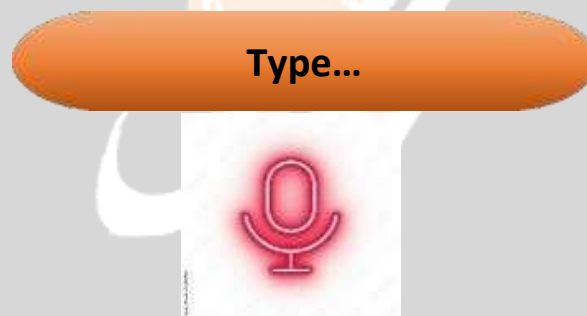
[15] Abu Shaabanet al.: Researched user experience (UX) design for chatbots, focusing on factors like user interface and ease of interaction. This is important for ensuring a positive experience with multilingual chatbots.

### 3. METHODOLOGY OVERVIEW:

The methodology involves developing robust NLU models for each language, integrating machine translation, and localizing responses to cultural nuances. User experience design prioritizes intuitive interfaces and personalized interactions. Continuous improvement via user feedback ensures relevance and effectiveness over time, enabling seamless communication in English, Hindi, Kannada, and Telugu.

**Table 1: Languages.**

<b>Languages</b>
<b>English</b>
<b>Kannada</b>
<b>Hindi</b>
<b>Telugu</b>



**Figure 2: Speech and Text Format.**

**3.1 DATA COLLECTION:** Data collection for the multilingual college chatbot involves gathering diverse datasets in English, Hindi, Kannada, and Telugu to train the natural language understanding (NLU) models effectively. For English, publicly available corpora like the Common Crawl or Wikipedia can be utilized, while for Hindi, Kannada, and Telugu, resources such as government documents, news articles, and literature are valuable sources. Additionally, crowdsourcing platforms and academic institutions may provide labelled datasets for specific domains relevant to college life, such as course information, campus events, and student services. These datasets need to be curated and annotated by native speakers to capture the nuances and variations of each language. Furthermore, including user-generated content, such as chat logs and social media interactions, adds to the training data and improves the chatbot's capacity to grasp colloquial language and slang. Careful consideration of data privacy and ethical guidelines is essential throughout the data collection process to ensure compliance and protect user confidentiality. The multilingual college chatbot, which uses varied datasets in multiple languages, can effectively read user inquiries and deliver contextually relevant responses, promoting effective communication and participation within the college community.

**3.2 DATA PREPROCESSING:** In the methodology for developing a multilingual college chatbot, data processing plays a crucial role in preparing and leveraging the collected datasets effectively. Initially, the gathered data in English, Hindi, Kannada, and Telugu undergoes pre-processing steps such as tokenization, stemming, and lemmatization to standardize the text and remove noise. Subsequently, the processed data is used to train language-specific natural language understanding (NLU) models tailored to each language's linguistic nuances and structures.

These NLU models employ techniques like deep learning and transfer learning to extract relevant features and patterns from the text, enabling accurate interpretation of user queries across languages. Additionally, machine translation systems are integrated into the data processing pipeline to facilitate seamless communication between languages, allowing the chatbot to translate messages between English, Hindi, Kannada, and Telugu in real-time. Furthermore, data augmentation techniques such as back-translation and paraphrasing are applied to augment the training data and improve the robustness and generalization capabilities of the NLU models. Continuous monitoring and evaluation of the data processing pipeline ensure that the chatbot remains effective and up to date in understanding and responding to user inputs across multiple languages, fostering inclusivity and accessibility within the college community.

**3.3 DATASET EXPLORATION:** Data exploration refers to the process of gathering and analyzing information from various sources to understand student needs and shape the chatbot's development. It involves delving into college websites, FAQs, and student support interactions to identify common questions, topics, keywords, and even sentiment. This analysis helps create a comprehensive and diverse dataset in multiple languages.

**3.3.1. Data Sources: College Websites and Resources:** Explore college websites to identify frequently asked questions (FAQs) sections, student handbook information, and departmental resources. Look for content related to academics, admissions, financial aid, student life, and campus facilities. **College Chat Logs:** Analyze transcripts of past interactions with student support services (e.g., email, phone) to understand common student inquiries and concerns. [Type... Consider anonymizing and pre-processing chat logs before analysis. **Surveys and Focus Groups:** Conduct surveys or focus groups with students from diverse backgrounds to understand their preferred communication methods, language needs, and topics they would like chatbot assistance with.

**3.3.2. Data Collection and Preprocessing: Scrape or Extract Text Data:** Use web scraping tools (with proper permissions) to extract relevant text from college websites and resources. **Clean and Pre-process Text:** Remove irrelevant information like HTML tags, punctuation, and stop words (common words like "the", "a"). Perform language detection to identify the language of each data source. **Translation (if needed):** Translate non-English content to your target languages using Machine Translation (MT) services or professional translators. Ensure high-quality translations for better chatbot performance.

**3.3.3. Data Analysis: Topic Modelling:** Apply topic modelling algorithms to identify common themes and topics within the data. This helps categorize chatbot responses and training data.

**Keyword Analysis:** Identify frequently occurring keywords and phrases that students use when seeking information. This helps tailor chatbot responses and trigger relevant knowledge bases.

**Sentiment Analysis (optional):** Analyze the sentiment (positive, negative, neutral) of student queries in chat logs or surveys. This can help the chatbot respond with empathy and adapt its tone depending on the situation.

### 3.3.4. Building the Multilingual Dataset

**Create a Multimodal Dataset:** Combine text data with additional resources like images, videos, or audio recordings (if available) to enhance chatbot responses.

Ensure the dataset reflects the diversity of student languages and learning styles.

**Annotate Data:** Depending on your chosen chatbot development approach, you might need to annotate the data with labels or intents (specific user goals). This helps the chatbot understand the meaning behind user queries.

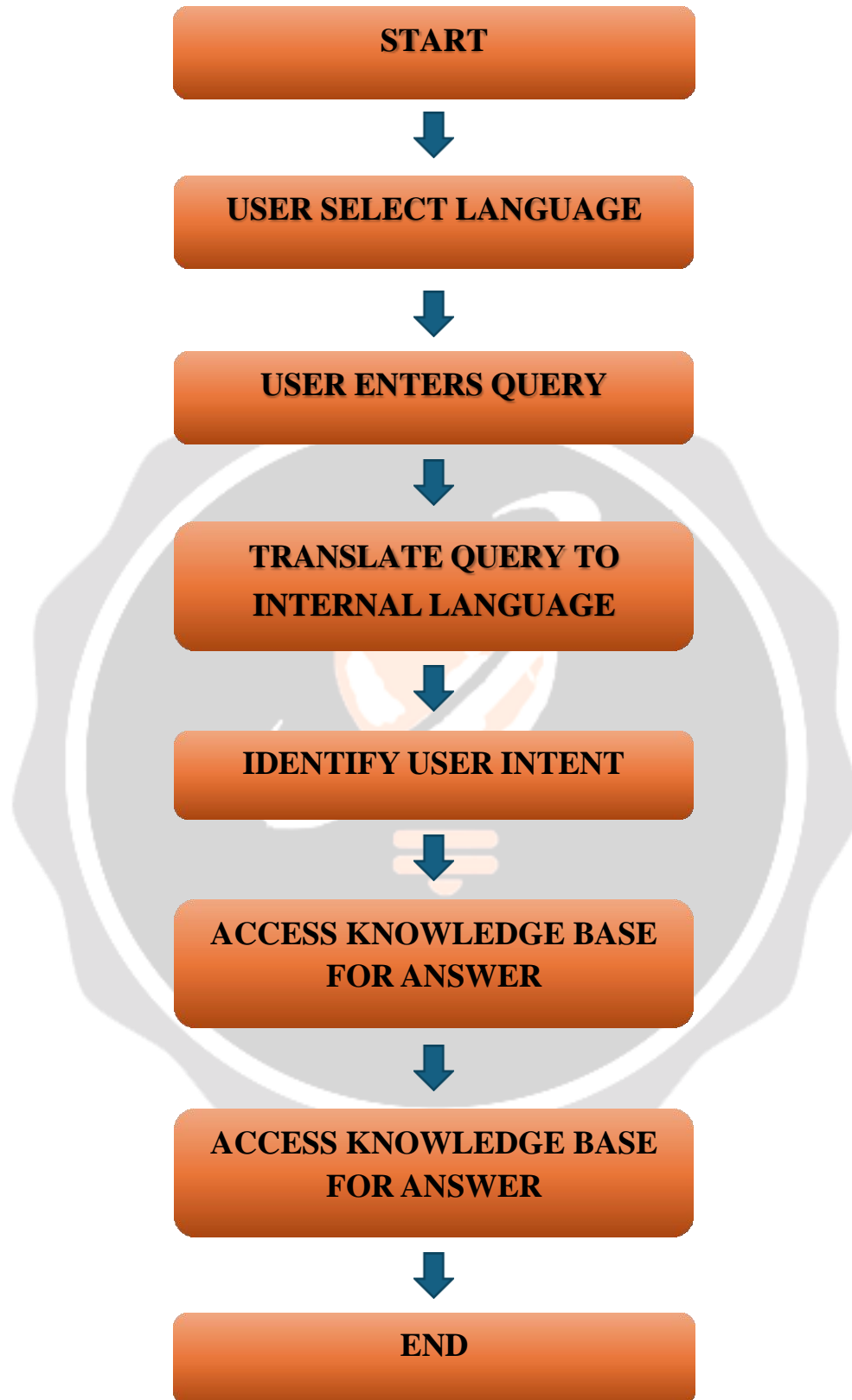


Figure 3: This flowchart illustrates how a Multilingual College Chatbot can translate a student query and provide an answer in their preferred language.

### Additional Considerations:

**Data Privacy:** Ensure compliance with data privacy regulations when collecting and storing student data. Anonymize data whenever possible.

**Data Quality:** Continuously monitor and improve data quality to ensure the chatbot provides accurate and up-to-date information.

NLP is used to analyze user queries, detect keywords and phrases, and map them to appropriate responses based on previously processed data.

Personalization (Optional): For more advanced chatbots, techniques like user profiling and collaborative filtering can be employed. User profiles are built based on past interactions, and the chatbot can leverage this information to personalize recommendations and responses. Collaborative filtering analyses the interactions of comparable users to provide resources or support alternatives that are relevant to the present user's needs. Data processing for multilingual college chatbots is not a one-time event. As students interact with the chatbot, new data is generated. This data can be anonymized and given back into the system for ongoing learning and development. This iterative method enables the chatbot to fine-tune its responses, adjust to changing demands, and provide a more personalized and successful experience for students.

In conclusion, data processing plays a vital role behind the scenes in making multilingual college chatbots function effectively. By acquiring relevant data from various sources, pre-processing it for accuracy and usability, and utilizing machine learning techniques, these chatbots can navigate diverse languages.

### 4. RESULTS AND OUTCOMES:







### Language preferences – Telugu:



## 5. CONCLUSION

Multilingual College Chatbots are an invaluable resource that may help students in a variety of ways. They can employ Natural Language Processing (NLP) to interpret and reply to user inquiries in real time on several topics, including academics, placements, and HoD staff. This is especially useful for students who are not native speakers of the language of teaching. Furthermore, multilingual chatbots can help to widen a college's communication platforms by providing support in a larger range of languages. One significant enhancement for multilingual college chatbots is the incorporation of advanced conversation management mechanisms. This will enable chatbots to better comprehend the topic of a conversation and respond in more useful and informative ways. Furthermore, chatbots can relate to various.

## 6. REFERENCES

- [1] Rahul Jindal, Rohit Kumar, Ritvik Sahajpal, Sanjeev Sofat, and Shailendra Singh. Implementing a natural language conversational interface for Indian language computing. IETE Technical Review, 2022.
- [2] Sagar Badlani, Tanvi Aditya, Meet Dave, and Sheetal Chaudhari. Multilingual healthcare chatbot using machine learning. In 2021 2<sup>nd</sup> International Conferences for Engineering Technology (INCET), pages 1-6. IEEE, 2021.
- [3] Jon Ander Campos, Arantxa Otegi, Aitor Soroa, Jan Deriu, Mark Cieliebak, and Eneko Agirre. Doqa—accessing domain-specific faqs via conversational qa. arXiv preprint arXiv:2005.01328, 2020.
- [4] Arantxa Otegi, Aitor Agirre, Jon Ander Campos, Aitor Soroa, and Eneko Agirre. Conversational question answering in low resource scenarios: A dataset and case study for basque. In Proceedings of the 12th Language Resources and Evaluation Conference, 2020.
- [5] Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. IndicNLP Suite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In Findings of EMNLP, 2020.
- [6] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, System Demonstrations, pages 38–45, Online, October 2020. Association for Computational Linguistics.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv: 1810.04805, 2018.

[8] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzman, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov, Unsupervised cross-lingual representation learning at scale. CoRR, abs/1911.02116, 2019.

[9] Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukar, Muril: Multilingual representations for Indian Languages, 2021. Bus. Intel, 4:19-24, 2019.

