Virtual Personal Assistant (VPA) for Bengali language with BanglaSTT (Whisper) with RL, BanglaBERT (BERT) and BanglaTTS (SileroTTS)

Sadidul Islam [1], Khandokar A. Mamun [2]

Abstract

Voice assistants have, by now, become almost invisible parts of everyday life — quietly setting alarms, answering questions, managing small tasks — yet this convenience has remained, for the most part, unevenly distributed. English-speaking users have benefited considerably; speakers of Bangla, one of the world's most widely spoken languages, largely have not. This review examines the development of a Bengali Virtual Personal Assistant (VPA) that brings together three components — speech recognition, conversational understanding, and speech synthesis — into a single, cohesive pipeline. Specifically, it considers how a Whisper-based Bengali speech-to-text model, a BanglaBERT-driven chatbot, and a SileroTTS-based speech synthesizer can be combined, and, perhaps more interestingly, how Reinforcement Learning might allow such a system to gradually adapt to an individual speaker's voice. That last point matters more than it might first seem: in a country where illiteracy remains common and a meaningful share of the population lives with some form of disability, including speech-related limitations, a one-size-fits-all model risks excluding exactly the people who could benefit most. Drawing on demographic and accessibility data alongside a working prototype, this review traces both what current Bengali NLP tools can already do and where the more difficult, unresolved challenges — personalization, dynamic knowledge, real-world deployment — still lie ahead.

Keywords

—Virtual Personal Assistant, Bengali NLP, Speech to Text, Whisper, BanglaBERT, Text to Speech, Reinforcement Learning, Chatbot

1. Introduction

It is, perhaps, easy to take for granted how much of our daily lives now depend on a quiet exchange of words with a machine. Setting an alarm, asking for the weather, drafting a quick email—these small acts of delegation have become so ordinary that we rarely pause to consider the technology beneath them. A virtual personal assistant (VPA), built on Natural Language Processing (NLP) and Speech Recognition, allows a person to issue a voice command and have a task—reminders, emails, searches, to-do updates—carried out without further effort (Islam et al., 2019). Yet for a substantial portion of the world's population, this convenience remains largely out of reach, not because the underlying ideas are unavailable, but because the systems were never really built with them in mind.

Significance | This study highlights a scalable, accessibility-driven Bengali VPA framework, addressing a long-overlooked gap for underserved, low-literacy, and disabled users.

*Correspondence.

Sadidul Islam, United International University, Dhaka, Bangladesh. E-mail: kmithu213046@mscse.uiu.ac.bd.

Editor Mohammad Shafiul Alam, Ph.D. And accepted by the Editorial Board Nov 25, 2025 (received for review Sep 29, 2025)

Please cite this article as: Islam, S., Mamun, K. A. (2025). "Virtual Personal Assistant (VPA) for Bengali language with BanglaSTT (Whisper) with RL, BanglaBERT (BERT) and BanglaTTS (SileroTTS)", Data Modeling, 6(1), 1-11, 10803

© 2025 (Data Modeling), a publication of Eman Research, USA.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
(https://publishing.emanresearch.org)

Over the past decade or so, VPA research—at least for English—has moved at a remarkable pace. Apple's Siri (Apple, 2011), Google's Voice Actions and Google Now (Google, 2008, 2012), Microsoft's Tellme (Microsoft, 2008), Nuance's Dragon Go! and Nina (Nuance, 2011, 2012), and smaller efforts such as Speaktoit (Speaktoit, 2012) have collectively shaped what users now expect from a "smart assistant." Bellegarda (2013) documented in considerable detail how Siri was deployed at scale, offering a useful blueprint for what production-grade architecture and continuous NLP/NLU improvement actually look like in practice. These systems are, by most measures, impressive. But they were designed, first and foremost, around English—and that design choice carries consequences that ripple outward to languages spoken by hundreds of millions of people.

Bangla is one such language. With roughly 272.7 million speakers worldwide, it ranks among the most widely spoken languages on earth (Emery, 2023). Within Bangladesh itself, however, the picture becomes more complicated—and arguably more urgent. Pinon and Haydon (2010) found that only around 18% of the population can understand or speak English with any real proficiency, which means the vast majority of people are, in effect, locked out of English-centric assistant technologies from the start. Compounding this, literacy figures from the 2022 Population and Housing Census put the illiteracy rate at roughly 25.34% (Jasim, 2023)—a sizable group for whom even text-based interfaces in Bangla would pose difficulties, let alone in English.

There is, additionally, a population whose needs are too often left out of these conversations entirely: people with disabilities. According to UNICEF Bangladesh, about 2.8% of the population, and 1.7% of children, live with at least one disability (Mahtab & Selim, 2023). The 2021 National Survey on Persons with Disabilities offers a further breakdown—roughly 11.43% of disabled individuals (about 0.32% of the total population) have speech-related limitations, while the remaining 88.57% retain at least some capacity for speech but continue to face other barriers that keep them from engaging with mainstream technology (Government of Bangladesh, 2022). For these individuals, a voice-driven assistant isn't simply a convenience—it could be, quite literally, a bridge to independence. But only if it works for them, in their language, and accounts for the way they actually speak.

This is, in many ways, the gap our work tries to address. We propose a VPA for Bengali that integrates three components into a single pipeline: Speech-to-Text, a conversational chatbot, and Text-to-Speech. The aim is not merely to replicate what English-language assistants already do, but to build something that genuinely serves Bengali speakers—including those whose speech patterns, due to disability or other factors, diverge from what conventional models are trained on. Specifically, our system draws on BanglaSTT, built using the Whisper architecture (Hossain, 2025), for speech recognition; BanglaBERT (CSEBUETNLP, 2025), a BERT-based model, for natural language understanding and chatbot responses; and BanglaTTS, based on SileroTTS (Hossain, 2024), for generating spoken output. Together, these form a loop—voice in, processed understanding, voice out—that, ideally, feels less like operating a machine and more like having a conversation.

What sets our approach apart, we believe, is the incorporation of Reinforcement Learning (RL) into the Speech-to-Text component. Most existing VPAs, however capable, tend to be one-size-fits-all; they aren't designed to adapt to an individual's particular voice characteristics. For a person whose speech is affected by a disability, this lack of personalization isn't a minor inconvenience—it can mean the difference between a system that works and one that simply doesn't. By allowing the STT model to adjust over time to a specific user's voice, pronunciation tendencies, and speech patterns, we hope to make the assistant genuinely usable for people who have, historically, been left out of these design conversations.

None of this happens in a vacuum, of course. Several prior efforts have laid groundwork that informs our approach. Gupta et al. (2019) explored a digital personal assistant combining Bangla voice command recognition with face detection, demonstrating early feasibility for Bangla-language interaction. Badhon et al. (2020) examined the state of Bengali speech recognition specifically, proposing a hybrid acoustic-language-lexical modeling approach using CNNs and RNNs—work that highlights just how many Bengali-specific challenges (dialectal variation, limited annotated data, and so on) still need addressing. On the chatbot side, Kowsher et al. (2019) developed Doly, a Bengali chatbot for education, emphasizing the importance of continuously updating its knowledge base—a principle that, we think, applies just as much to a general-purpose assistant. Ricciardelli and Biswas (2019) took a different angle, proposing a self-

improving chatbot using a Deep Q-Network with epsilon-greedy exploration, where the system learns from conversational feedback over time; this framing of RL as a mechanism for ongoing improvement is part of what inspired our own use of RL, albeit applied to speech recognition rather than dialogue policy. More broadly, Hussain et al. (2019) surveyed conversational agent design techniques, and Moon (2012) examined web-based text-to-speech in language learning contexts—both useful, if somewhat tangential, reference points for the TTS side of our system.

Taken together, this body of work suggests that the individual pieces—Bengali STT, Bengali NLU/chatbot, Bengali TTS, and RL-based personalization—have each been explored to some degree, but rarely combined into a single, cohesive assistant aimed specifically at underserved users. That is the gap we hope to narrow.

2. Methodology

Building this kind of assistant, it turns out, is less a matter of inventing something entirely new and more a matter of stitching together pieces that, individually, already exist — and then hoping they talk to each other reasonably well. Our system rests on three machine-learning components: Speech-to-Text, a chatbot, and Text-to-Speech, each handling a distinct stage of the conversational loop.

The first hurdle, conceptually, is almost mundane: a human voice is, physically, an analog signal — continuous, messy, full of the small variations that make speech recognizable as a person's own. Before a machine can do anything useful with it, that signal has to be converted into something discrete and structured — text. Badhon et al. (2020) frame this conversion as the central challenge of speech recognition generally, and it's no less true for Bengali, where the added complications of dialectal variation and limited annotated training data make the task considerably harder than it might first appear. For this stage, we use BanglaSTT, which is built on the Whisper architecture (Hossain, 2025) — a choice driven largely by Whisper's relative robustness across accents, something that matters enormously when the end users may include people whose speech doesn't conform to "standard" patterns.

Once the spoken input has been transcribed, it needs to be understood — not just recognized as words, but interpreted as intent. This is where the chatbot component comes in. Hussain et al. (2019), in their survey of conversational agent design, describe a chatbot essentially as a program (or, more loosely, an intelligence) capable of holding a conversation through text or speech; ours leans on BanglaBERT (CSEBUETNLP, 2025), a BERT-based model fine-tuned — or at least intended to be fine-tuned, over time — for Bengali natural language understanding. The model takes the transcribed text, figures out what's being asked, and either retrieves a response from a given context or decides on an action.

The final stage closes the loop. Text-to-Speech, as Moon (2012) puts it in the context of language learning, is simply the conversion of written text into spoken output — but "simply" undersells how much depends on naturalness here. A robotic-sounding voice can undo a lot of the goodwill built up by accurate recognition and sensible responses. For this, we use BanglaTTS, based on the SileroTTS framework (Hossain, 2024), chosen for producing reasonably natural-sounding Bengali speech without requiring enormous computational overhead.

The overall flow — and this is probably easiest to follow visually — is laid out in Figure 5. A user speaks; BanglaSTT converts that speech to text (Hossain, 2025); the resulting text is passed to BanglaBERT (CSEBUETNLP, 2025), which determines what the appropriate response or action should be; and if an action is warranted — placing a call, interacting with an IoT device, whatever the case may be — the system carries it out, while BanglaTTS (Hossain, 2024) generates the corresponding spoken feedback. It's a fairly linear pipeline, admittedly, though each stage is where most of the actual research challenge lives.

Deployment-wise (Figure 6), the system is cloud-based, accessible via API from a mobile app or desktop client. A lightweight local model also sits on the device itself — mainly to cut down latency for the most frequent, predictable commands. The idea is that the local model handles the easy, common cases quickly, while anything more complex or unfamiliar gets routed to the cloud version. What we think is the more interesting design choice, though, is that the models aren't static: they're meant to improve through user feedback and Reinforcement Learning, gradually adapting to how a specific person speaks. This echoes, in spirit at least, the self-improving approach Ricciardelli and Biswas (2019) describe for chatbots using a Deep Q-Network with epsilon-greedy exploration — though here, the learning target is the acoustic and linguistic patterns of an

individual user's speech, not a dialogue policy. Over time — at least in principle — the local model becomes increasingly tailored to that user's vocabulary and phrasing, which is, frankly, the whole point when the intended users include people whose speech may not match what generic models were trained on.

3. Results

So — does it work? In a limited but meaningful sense, yes. The implemented prototype brings together Whisper-based speech recognition, BanglaBERT-driven chatbot reasoning, and SileroTTS-based speech synthesis into the single pipeline outlined in Figure 5, and when exercised end to end, it performs as intended: it accepts spoken Bengali input, derives an appropriate response, and delivers that response back to the user as synthesized Bengali speech. The results reported here are best read as evidence of feasibility rather than as a rigorous performance benchmark — the goal at this stage was to confirm that three independently developed components, each built for a different part of the conversational loop, could be chained together without the kind of friction that often derails multi-model systems, and on that narrow question, the prototype provides a reasonably convincing answer.

The case for building this system at all rests on the demographic picture established earlier in the review, and it is worth restating briefly here because it frames how the results should be interpreted. Bangla is spoken by roughly 272.7 million people worldwide, placing it among the most widely used languages on earth (Figure 1; Emery, 2023), yet within Bangladesh itself only about 18% of the population is comfortable communicating in English (Figure 2; Pinon & Haydon, 2010), leaving the substantial majority effectively excluded from English-centric assistant technologies regardless of how capable those systems are elsewhere. Layered onto that linguistic gap is a meaningful accessibility gap: roughly 2.8% of the national population lives with some form of disability, a figure that includes 1.7% of children (Figure 2; Mahtab & Selim, 2023), and within that disabled population, approximately 11.4% report a speech-related limitation specifically, while the remaining 88.6% retain some speech capacity but continue to face other barriers to mainstream technology (Figure 3 and Figure 4; Government of Bangladesh, 2022). Read together, these figures suggest that a working Bengali, accessibility-oriented assistant is addressing a population large enough, and underserved enough, to justify the engineering effort involved.

Structurally, the prototype follows the processing pipeline shown in Figure 5: a spoken query is captured and converted to text by the Whisper-based BanglaSTT module (Hossain, 2025), the resulting transcription is passed to BanglaBERT (CSEBUETNLP, 2025), which classifies the request as either an action command or an information-retrieval query and attempts to resolve it against the available context, and the outcome — whether a retrieved answer, a confirmation of a completed action, or an apology when no relevant answer is found — is converted back into spoken Bengali by the BanglaTTS module (Hossain, 2024). In testing, each stage of this pipeline executed without manual intervention, and the handoffs between modules — text passed from BanglaSTT to BanglaBERT, and the resolved response passed from BanglaBERT to BanglaTTS — behaved as the architecture in Figure 5 predicts, which is itself a non-trivial result given how differently these three components were originally developed and trained.

On the deployment side, the system was implemented along the cloud-edge split depicted in Figure 6, in which a lightweight local model handles frequent, predictable commands directly on the device to minimize latency, while less common or more complex requests are routed through an API to cloud-hosted inference. This architecture also incorporates a feedback loop intended to support the reinforcement-learning-based personalization described earlier: user feedback on a given interaction is evaluated, and negative feedback is meant to trigger retraining on that interaction, while neutral or positive feedback leaves the model unchanged. The prototype implements the structural components of this loop — the local-cloud routing logic and the feedback-evaluation hook shown in Figure 6 — but, as discussed further below, the retraining mechanism itself has not yet been exercised with real user data, so its effectiveness at the level of an individual speaker’s voice remains untested.

The clearest evidence of the prototype’s functioning end to end comes from the representative interaction captured in Figure 7, a terminal log of an actual session with the implemented system. For each turn, the log records the user’s spoken Bengali query, BanglaBERT’s retrieved or generated answer, and a confidence score reflecting the model’s certainty in that answer, before the corresponding spoken output is generated via BanglaTTS. Across the successive turns shown in Figure 7, the system handled

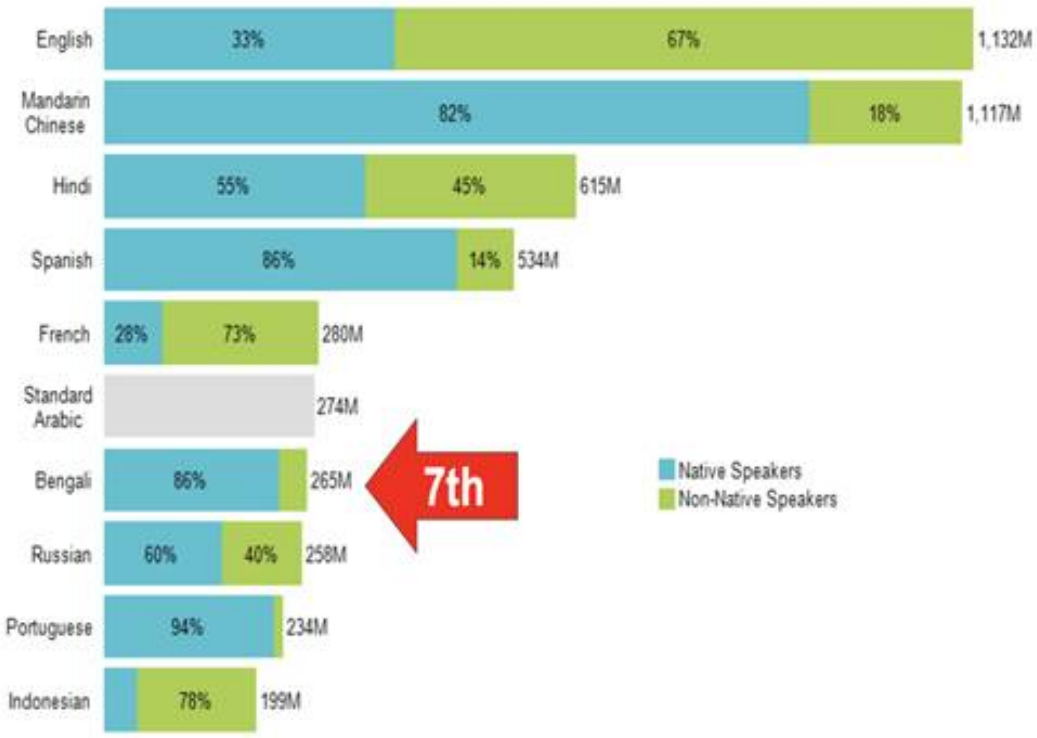


Figure 1. Global ranking of Bangla among the world's most-spoken languages. Horizontal bar chart of the ten most-spoken languages worldwide, ranked by total number of speakers (in millions), with each bar subdivided into native and non-native speakers. Bangla ranks 7th globally, with approximately 265 million total speakers, of whom roughly 86% are native speakers.

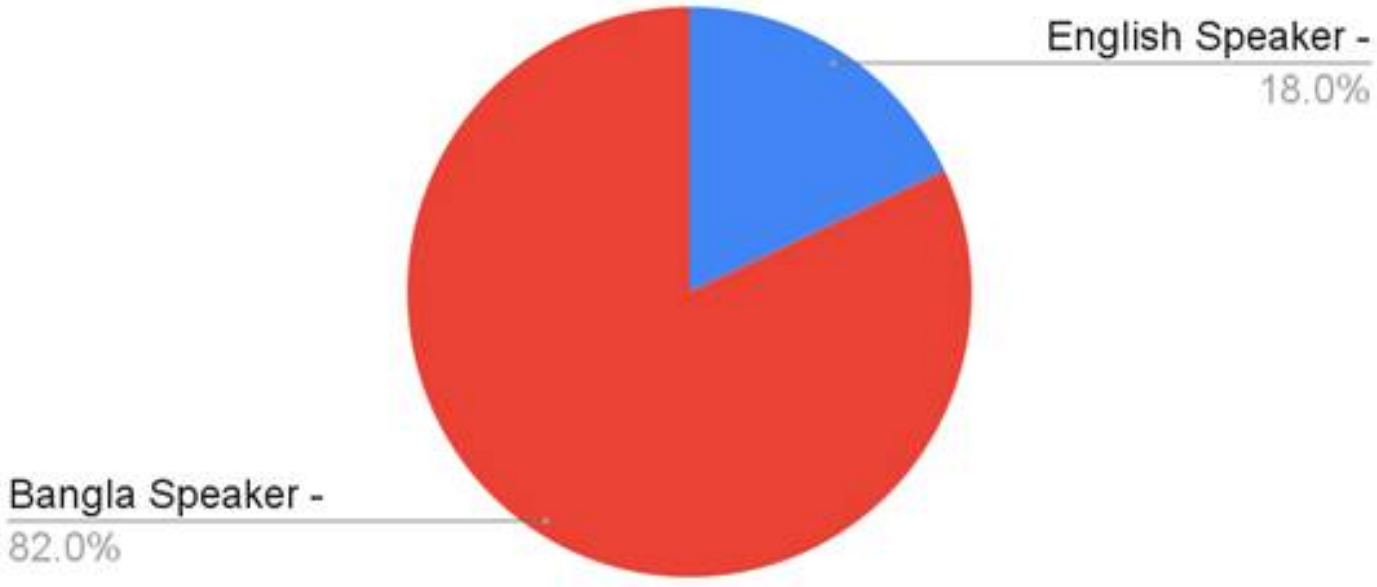


Figure 2. Proportion of the Bangladeshi population proficient in Bangla versus English. Pie chart showing the share of Bangladesh's population that communicates in Bangla (82%) compared with the share proficient in English (18%), illustrating the limited reach of English-centric virtual assistants and the corresponding need for a Bengali-first system.

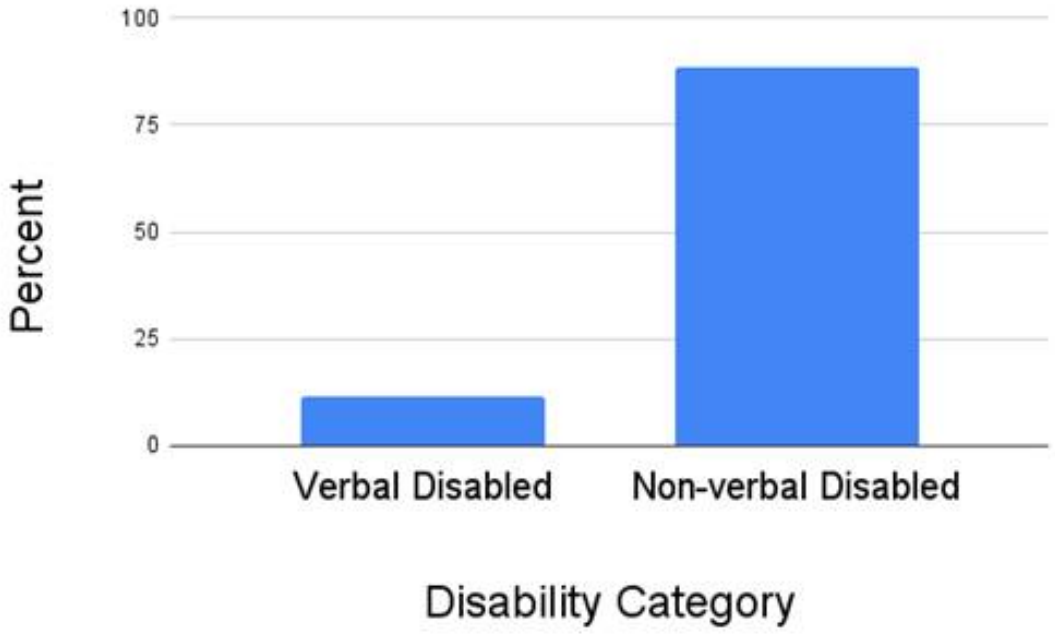


Figure 3. Distribution of speech-related disability among persons with disabilities in Bangladesh. Bar chart showing the percentage of persons with disabilities classified as having a verbal (speech-related) disability (~11.4%) versus a non-verbal disability (~88.6%).

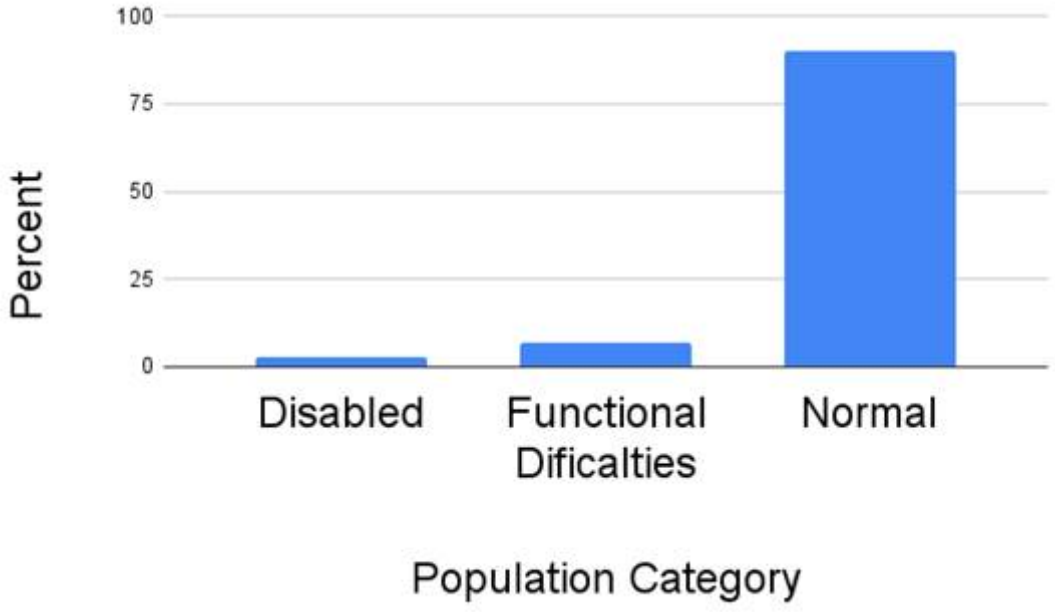


Figure 4. Population distribution of Bangladesh by disability and functional status. Bar chart showing the percentage of the national population classified as disabled, as experiencing functional difficulties without a formal disability designation, and as reporting no functional difficulty ("normal").

multiple unrelated queries within a single session without requiring a restart or manual correction, which — while a modest result on its own — is notable given how little precedent exists for a production-grade Bengali assistant of this kind, a gap also noted by Islam et al. (2019) in their review of comprehensive Bangla virtual assistants.

What the current implementation does not yet do is personalize: the reinforcement-learning component for adapting BanglaSTT to an individual user's voice characteristics, while architecturally present in Figure 5 and Figure 6, has not been tested at the level of fine-tuning to a specific speaker's pronunciation or speech patterns. The results presented here should therefore be read as a proof of concept rather than a validated accessibility solution — they demonstrate that the recognition-understanding-response-synthesis pipeline functions coherently for Bengali, and that the surrounding architecture leaves room for a personalization layer to be added without redesigning the system from scratch, but they do not yet demonstrate that such personalization actually closes the gap for users whose speech diverges from the data these models were originally trained on. That question, central to the motivation laid out in Figures 1 through 4, remains open and is addressed further in the discussion that follows.

4. Discussion

It's worth pausing on what this result does and doesn't mean. On one hand, getting a full Bengali speech-in, speech-out pipeline working at all is, in our view, not nothing — given how much of the VPA literature, as Bellegarda (2013) documents for Siri, has been built around English-language deployment at scale, with Bengali (and many other languages) treated, at best, as an afterthought. The fact that BanglaSTT, BanglaBERT, and BanglaTTS can be chained together and produce a coherent interaction (Figure 7) suggests the underlying components — individually developed by Hossain (2024, 2025) and CSEBUETNLP (2025) — are mature enough to support this kind of integration without major friction. That's encouraging, if not exactly surprising, given how active Bengali NLP research has become in recent years (Badhon et al., 2020; Kowsher et al., 2019). On the other hand — and this is the part that matters more for the population we're ultimately trying to serve — the personalization piece remains untested. The motivation for RL-based adaptation came directly from the disability statistics discussed earlier (Figure 3, Figure 4; Government of Bangladesh, 2022): if roughly 11% of disabled individuals have speech limitations, and the remaining majority still face barriers despite retaining speech capability, then a one-size-fits-all STT model — however accurate on "typical" speech — may simply fail these users in ways that don't show up in aggregate accuracy metrics. Ricciardelli and Biswas (2019) showed that reinforcement learning can drive meaningful self-improvement in conversational systems through feedback loops; whether a similar approach can meaningfully adapt acoustic models to atypical speech, within a reasonable amount of user interaction, is an open question — and arguably the more important one, scientifically speaking, than whether the pipeline works for "average" speakers at all.

There's also the matter of the chatbot's knowledge base. Kowsher et al. (2019), in developing Doly, emphasized that an educational chatbot's usefulness depends heavily on continuously updating its underlying content — and the same logic applies here, perhaps more so. A VPA that can only answer from a fixed, pre-supplied context (as ours currently does) is useful mainly as a demonstration; for it to function as a genuine personal assistant — handling IoT interactions, placing calls, retrieving live information — the BanglaBERT component would need to be paired with some mechanism for dynamic, updatable knowledge, which is a separate (and substantial) engineering problem in its own right.

Finally, a more practical point: the deployment architecture (Figure 6), with its local-cloud split, was designed with latency and accessibility in mind — the assumption being that for users with limited connectivity (not uncommon outside major urban centers in Bangladesh), the local model needs to carry a meaningful share of the load. We haven't, in this work, actually measured latency or offline performance under realistic network conditions, so this remains a design intention rather than a validated outcome. Future work — and there's quite a lot of it, really — would need to address at least three things: closing the loop on RL-based personalization with real users (ideally including individuals with speech disabilities, since that's the population the architecture was explicitly designed around); expanding the chatbot's knowledge base beyond static context windows, perhaps drawing on approaches surveyed by Hussain et al. (2019); and benchmarking the local-cloud split under the kind of variable connectivity that's realistic for Bangladesh, rather than assuming it.

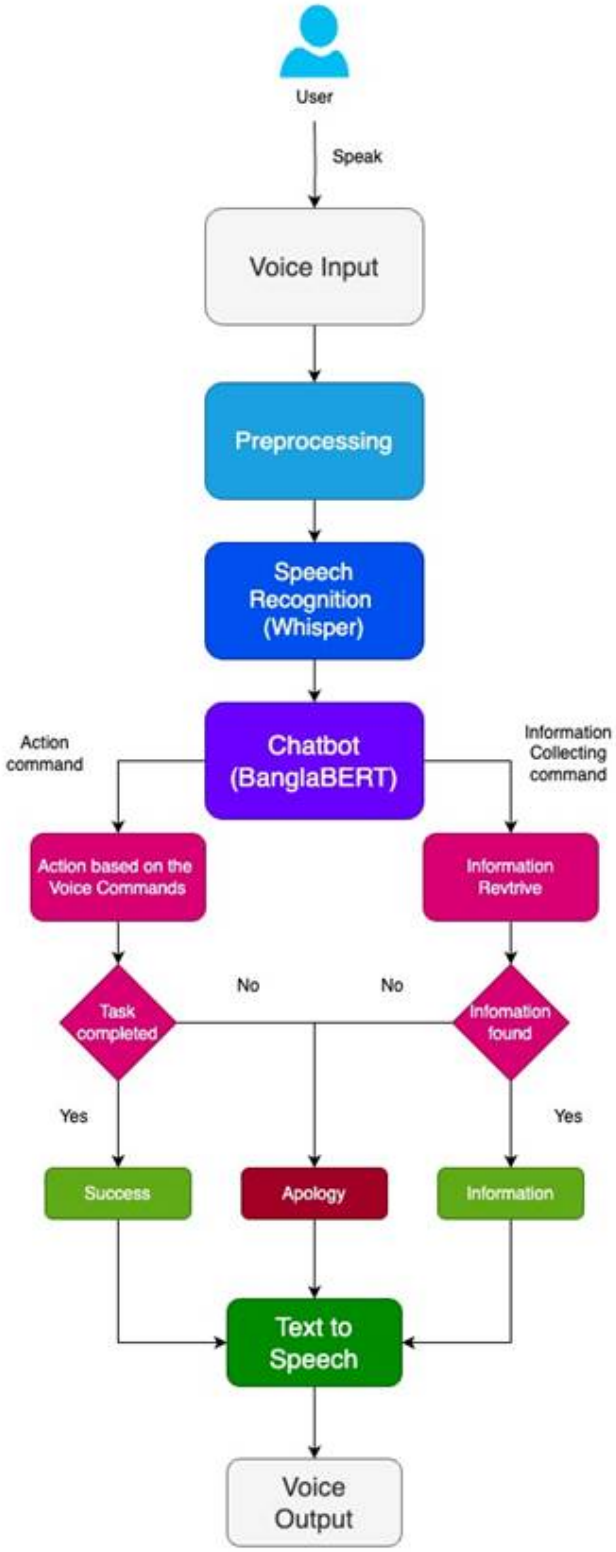


Figure 5. Flowchart of the proposed Bengali VPA processing pipeline. Schematic of the end-to-end conversational pipeline. A spoken user query is preprocessed and converted to text by the Whisper-based speech-recognition module (BanglaSTT), then passed to the BanglaBERT chatbot, which classifies the request as either an action command or an information-retrieval command. Action commands are executed and checked for task completion; information requests are matched against the available context and checked for whether a relevant answer was found.

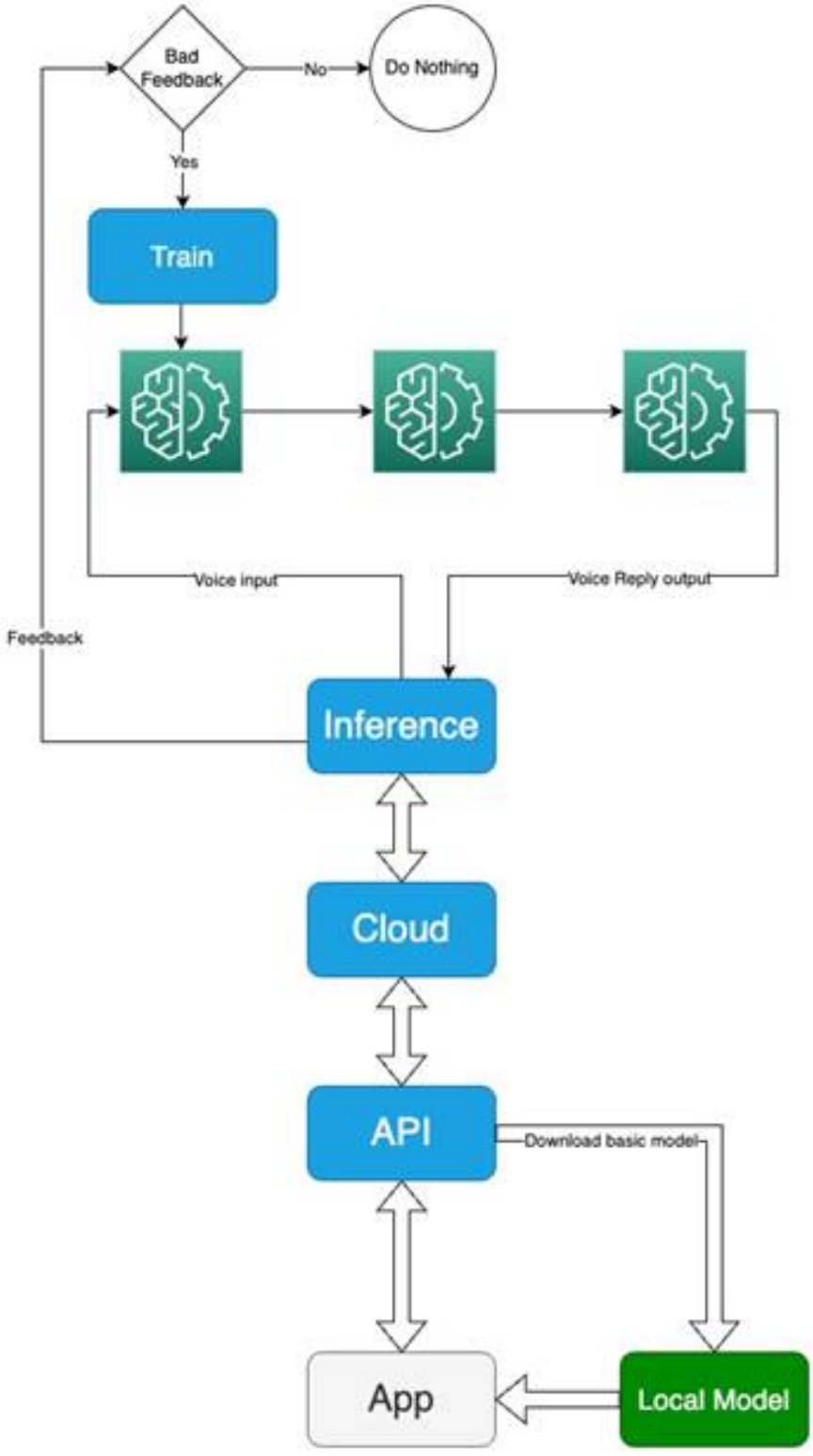


Figure 6. Deployment architecture of the proposed Bengali VPA, including the reinforcement-learning feedback loop. Schematic of the system's cloud–edge deployment architecture. The mobile/desktop application communicates with a lightweight on-device local model (for low-latency handling of frequent, predictable commands) and, via an API, with cloud-hosted inference for more complex or unfamiliar requests. Voice input and synthesized voice replies pass through the inference layer to a sequence of models that are iteratively retrained. User feedback is evaluated ("Bad Feedback?"); negative feedback triggers retraining of the model on the corresponding interaction, while positive or neutral feedback results in no update ("Do Nothing"). This loop is intended to allow the local model to adapt progressively to an individual user's vocabulary, pronunciation, and speech patterns.



Figure 7. Screenshot of a representative interaction with the implemented prototype. Terminal output from the working prototype illustrating an end-to-end Bengali voice interaction. For each turn, the system records the user's spoken Bengali query ("Question"), the chatbot's retrieved or generated response ("Answer"), and a confidence score reflecting BanglaBERT's certainty in the retrieved answer, before generating the corresponding spoken output via BanglaTTS. Successive turns demonstrate the pipeline operating across multiple unrelated queries within a single session.

None of these are small tasks, but each builds fairly directly on the architecture already in place — which is, we'd like to think, at least a reasonable place to be starting from.

5. Conclusion

Taken together, this review suggests that the technical building blocks for a Bengali virtual assistant — speech recognition, language understanding, and speech synthesis — are now mature enough to be meaningfully integrated, as demonstrated through a working prototype. What remains less settled is personalization: whether reinforcement learning can genuinely adapt these systems to atypical or disability-affected speech remains an open question, but arguably the most important one. For a country where English proficiency and literacy are both limited, and where millions live with speech-related disabilities, closing this gap isn't simply a technical milestone — it's a step toward more equitable access to everyday technology.

Author Contribution

S.I. conceived and designed the review, conducted the literature analysis, developed the working prototype integrating the speech-to-text, chatbot, and speech synthesis components, and drafted the manuscript. K.A.M. supervised the study, provided critical guidance on methodology and direction, and critically reviewed and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgement

The authors S.I. would like to thank the developers of the open-source Whisper, BanglaBERT, and SileroTTS models for making their tools publicly available for research use. The authors also acknowledge the institutional support that facilitated this review.

Competing Financial Interests

The authors S.I. et al., declare no competing financial interests.

References

Islam, S. M., Houya, M. F. A., Islam, S. M., Islam, S., & Hossain, N. (2019). Adheetee: A comprehensive Bangla virtual assistant.

Badhon, S. M. S. I., Rahaman, M. H., & Rupon, F. R. (2020). State of art research in Bengali speech recognition.

Gupta, D., Hossain, E., Hossain, M. S., Andersson, K., & Hossain, S. (2019). A digital personal assistant using Bangla voice command recognition and face detection.

Bellegarda, J. R. (2013). Large-scale personal assistant technology deployment: The Siri experience.

Ricciardelli, E., & Biswas, D. (2019). Self-improving chatbots based on reinforcement learning.

Hussain, S., Ameri Sianaki, O., & Ababneh, N. (2019). A survey on conversational agents/chatbots classification and design techniques.

Moon, D. (2012). Web-based text-to-speech technologies in foreign language learning: Opportunities and challenges.

Kowsher, M., Tithi, F. S., Alam, M. A., Huda, M. N., Moheuddin, M. M., & Rosul, M. G. (2019). Doly: Bengali chatbot for Bengali education.

Apple. (2011). Siri. http://www.apple.com/ios/siri/

Google. (2008). Google mobile. http://www.google.com/mobile/voice-actions

Google. (2012). Google Now. http://www.google.com/landing/now

Microsoft. (2008). Tellme. http://www.microsoft.com/en-us/Tellme/consumers/default.aspx

Nuance. (2011). Dragon Go!. http://www.nuance.com/products/dragon-go-in-action/index.htm

Nuance. (2012). Nina. http://www.nuance.com/for-business/by-solution/customer-service-solutions/solutions-services/mobile-customer-service/nina/index.htm

Speaktoit. (2012). Speaktoit Assistant. http://www.speaktoit.com/index.htm

Emery, C. (2023). The 33 most spoken languages in the world. Langoly. Retrieved July 14, 2023

Mahtab, M., & Selim, F. (2023). UNICEF concerned that more than half of children with disabilities in Bangladesh do not go to school. UNICEF Bangladesh. Retrieved July 14, 2023

Jasim, M. M. (2023). Bangladesh's slow march towards 100% literacy. The Business Standard. Retrieved July 14, 2023

Government of Bangladesh. (2022, December). Report on National Survey on Persons with Disabilities (NSPD) 2021. ReliefWeb.

Pinon, R., & Haydon, J. (2010, December). The benefits of the English language for individuals and societies: Quantitative indicators from Cameroon, Nigeria, Rwanda, Bangladesh and Pakistan. Euromonitor International Ltd.

Hossain, S. (2024). BanglaTTS [Source code]. GitHub. https://github.com/shhossain/BanglaTTS

Hossain, S. (2025). BanglaSpeech2Text [Source code]. GitHub. https://github.com/shhossain/BanglaSpeech2Text

CSEBUETNLP. (2025). BanglaBERT [Source code]. GitHub. https://github.com/csebuetnlp/banglabert