

CHATGPT FOR VISUALLY IMPAIRED AND BLIND

Askat Kuzdeuov¹, Shakhizat Nurgaliyev¹, Huseyin Atakan Varol¹

¹Institute of Smart Systems and Artificial Intelligence, Nazarbayev University, Astana, Kazakhstan

NOTE: Corresponding author: Huseyin Atakan Varol, ahvarol@nu.edu.kz

Abstract – The World Health Organization (WHO) reports that millions of individuals worldwide experience visual disabilities, which can have significant personal implications as more than 80% of our perception, cognition, learning, and daily activities rely on vision. As a result, people with visual impairments face challenges in education and professional settings due to inaccessible technologies. To address this issue, we introduce a speech-based assistive system that allows visually impaired and blind individuals to interact with ChatGPT through natural conversation. The system incorporates automatic speech recognition (ASR), text-to-speech (TTS) synthesis, and a messenger bot (Telegram) to create a seamless interface to ChatGPT. Through preliminary experiments, we have demonstrated the potential of this system as an assistive technology. We have made the source code publicly available at <https://github.com/IS2AI/telegram-bot-chatgpt> to encourage the development of assistive technologies for individuals with visual impairments.

Keywords – Artificial intelligence, assistive technology, automatic speech recognition, ChatGPT, text-to-speech

1. INTRODUCTION

As reported by the World Health Organization (WHO), approximately 1.3 billion individuals, constituting one-sixth of the global population, experience some form of disability [1]. Among these, roughly 253 million people have visual impairments, including around 36 million who are blind. Socioeconomically, individuals with visual impairments encounter numerous challenges related to education, employment, and income. A study by the National Federation of the Blind found that only 10% of blind children in the United States are literate, in contrast to over 90% of their sighted counterparts [2]. Furthermore, research conducted in the United States revealed that the unemployment rate among blind individuals is nearly twice that of the general population [3]. Moreover, individuals with visual impairments are more likely to work in low-wage positions with limited opportunities for career advancement [4], experience higher poverty rates, and have lower homeownership levels.

Individuals with visual impairments obtain information through various assistive technologies, including screen readers, Braille displays, and audiobooks [5]. Braille displays are devices that transform text into Braille, a tactile system of raised dots that can be read through touch. Screen readers are software applications that vocalize text on computers or mobile devices, enabling visually impaired users to access websites, documents, and other digital content. These individuals can employ search engines like Google to locate information by using screen reader software compatible with these engines. Certain screen readers also feature integrated support for Google search, permitting users to conduct searches and navigate results directly within the software.

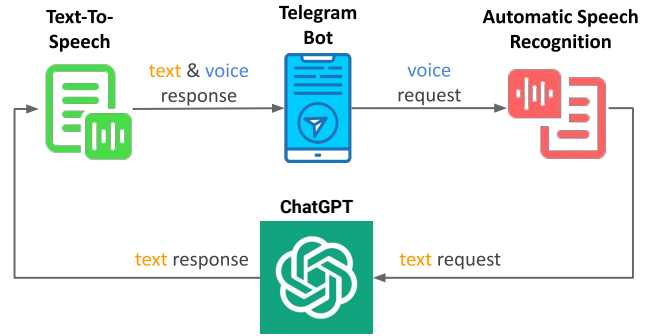


Figure 1 – System architecture of the ChatGPT with the speech-based interface for the visually impaired and blind.

However, information access in the digital realm remains challenging for those with visual impairments. For example, numerous websites and documents lack proper design for screen reader accessibility. Furthermore, interacting with technology, such as smartphones, tablets, and computers, can hinder access to information. Consequently, visually impaired individuals often depend on the assistance of others, limiting their independence and autonomy and negatively impacting their psychological well-being.

Speech has been an essential component of human communication from the earliest times [6]. As human societies evolved and became more complex, language and speech became increasingly important for social interactions, such as negotiation and storytelling. Starting with hieroglyphs and cuneiforms, writing has allowed for the preservation of knowledge and cultural traditions and has boosted the abstract thinking abilities of humankind.

Until recently, writing has been an activity exclusive to humans. However, in recent years, advances in artificial intelligence led to the emergence of large-scale language models (LLMs) [7]. For instance, GPT-2 and GPT-3, released by OpenAI, have demonstrated state-of-the-art performance on a wide range of natural language processing (NLP) tasks such as language translation, text summarization, and question answering. These models have been trained on massive amounts of text data and have achieved comparable or even superior results to those of human experts in some cases. LLMs with billions of parameters can process and understand natural language at an unprecedented level.

ChatGPT [8], a recent LLM optimized for dialogues using reinforcement learning from human feedback, has the potential to act as an assistive technology for information access among individuals with visual impairments. However, the text prompt-based interface provided by OpenAI poses a barrier for visually impaired users. Furthermore, ChatGPT tends to generate less meaningful responses for prompts in lower-resourced languages, such as Kazakh, due to its training on extensive text datasets procured from the internet, where higher-resourced languages like English predominate. This presents an additional technical challenge for monolingual visually impaired individuals who speak one of these lower-resourced languages.

In this study, we introduce the integration of cloud-based automatic speech recognition (ASR) and text-to-speech (TTS) technologies with ChatGPT, providing individuals with visual impairments a natural interface for information access. ASR allows computers to identify spoken words and transform them into text [9], while TTS technology converts text into natural-sounding speech. Recent advancements in deep learning have resulted in ASR and TTS models achieving human-level performance, particularly in higher-resourced languages such as English, Mandarin, and Japanese. These models have also been made accessible as services through cloud computing.

The utilization of cloud-based ASR and TTS models offers several benefits. For example, they can simultaneously process multiple requests, making them appropriate for large-scale applications. These technologies can be accessed from any location with a stable internet connection, ensuring high portability and accessibility. Regular updates with the latest features and enhancements can be made available. Moreover, cloud-based technologies can be more cost-effective than on-premises solutions, as they negate the need for upfront investments. Nevertheless, cloud-based technologies have certain limitations, including downtime, network latency, elevated costs, and restricted control. An alternative approach involves employing ASR and TTS models offline on local hardware; however, this may necessitate considerable computational power to execute large deep-

learning models. One potential solution involves simpler models, though their accuracy may be inferior to their larger counterparts. In this study, both online and offline options for ASR and TTS models were examined.

The remainder of this paper is structured as follows: Section 2 details the software and hardware architecture of our proposed assistive technology system. Section 3 demonstrates sample prompts and corresponding ChatGPT responses for two languages, namely English and Kazakh. Subsequently, Section 4 delves into the results and limitations of the system, while Section 5 provides a conclusion of our work.

2. METHODOLOGY

As depicted in Fig. 1, the system architecture comprises ChatGPT, a messenger bot (Telegram), ASR, and a TTS module. The messenger bot acts as an interface, enabling users to engage with ChatGPT. Visually impaired and blind individuals submit their requests to ChatGPT through voice messages via the messenger bot. Subsequently, the ASR module transcribes the voice messages into text, as ChatGPT solely processes text-based prompts. The ChatGPT module then analyzes the text inputs and generates responses in textual form. These textual responses are transformed into audible speech through the TTS module. Ultimately, the text and voice responses are delivered to the messenger bot for user interaction. This research investigates the interaction scenario for both English and a lower-resourced language, namely Kazakh.

2.1 Messenger Bot

Telegram messenger, a widely accessible, complimentary, cross-platform, encrypted, cloud-based, and centralized instant messaging service, ranks among the top five downloaded applications globally in 2022, boasting over 700 million monthly active users [10]. In this study, we developed two Telegram bots, catering to English¹ and Kazakh² speakers, respectively. The `pyTelegramBotAPI` [11], a versatile Python implementation for the Telegram Bot Application Programming Interface (API), was utilized to program these bots. Telegram facilitates the exchange of messages in various formats, including photos, videos, voice, and text. However, as our system is designed for visually impaired and blind individuals, our bots exclusively accept voice messages as input, thereby disregarding inadvertently sent text messages or files in other formats.

2.2 Automatic Speech Recognition (ASR)

ChatGPT exclusively processes text-based input, necessitating the use of the `SpeechRecognition` library to transcribe voice messages into text [12]. This li-

¹@issai_chatgpt_en_bot

²@issai_chatgpt_kz_bot

brary is compatible with numerous online and offline speech recognition engines and APIs, such as CMU Sphinx (offline), Google Speech Recognition, Google Cloud Speech API, Wit.ai, Microsoft Azure Speech, Houndify API, IBM Speech to Text, Snowboy Hotword Detection (offline), Tensorflow, Vosk API (offline), and OpenAI Whisper (offline). For testing purposes, we employed the Google Speech Recognition API with the default API key as a cloud-based solution and integrated the Vosk API as an offline speech recognition model into our system. Google Speech Recognition is a free service and supports a multitude of languages, including Kazakh. The Vosk API offers models for thirty languages and also comes with Kazakh language models of varying sizes. The Kazakh Vosk ASR models were trained on the Kazakh Speech Corpus dataset, developed by the Institute of Smart Systems and Artificial Intelligence [13].

When utilizing the English language, the ASR model’s text output is directly forwarded to ChatGPT. Conversely, for the Kazakh language, the text output can be optionally translated into English using the `googletrans` library [14] before sending it to ChatGPT. The `googletrans` library is employed due to its status as a freely accessible and unrestricted Python library, implementing the Google Translate API.

2.3 Interaction With ChatGPT

The OpenAI official API [15] was utilized to engage with ChatGPT, employing the `gpt-3.5-turbo` model. The model operates by accepting a sequence of messages as input and generating a response as output. The primary input parameter is the messages array, which consists of message objects that specify the role (system, user, or assistant) and content of each message. Conversations may vary in length from a single message to numerous pages, and typically follow a structure of alternating user and assistant messages, preceded by a system message that establishes the assistant’s behavior. User messages serve as instructions for the assistant, and may be either user-generated or developer-specified. Assistant messages retain prior responses and may also be developer-authored to provide examples of desired behavior. Including conversation history is useful when user instructions reference previous messages. The maximum input and output token limit for `gpt-3.5-turbo` is 4,096 tokens, with both input and output tokens counting toward this quantity. The more advanced `gpt-4` model is also available via the API, but it costs more and operates slower than `gpt-3.5-turbo`.

2.4 Text-To-Speech (TTS)

To transform the text response from ChatGPT into speech, the `gTTS` (Google Text-to-Speech) library [16] was employed for the English language, while the `Kazakh_TTS` model [17] was used for the Kazakh lan-

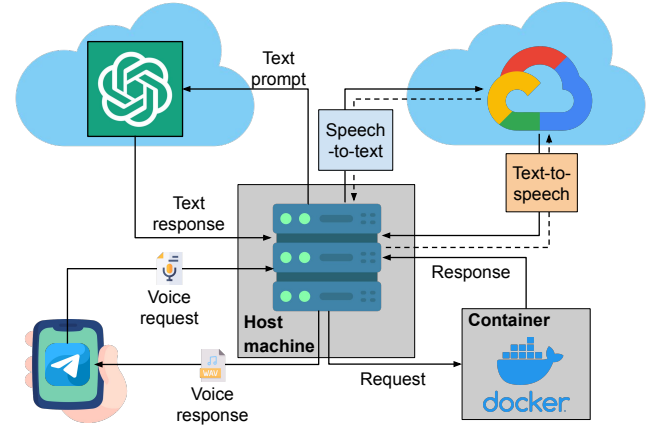


Figure 2 – Software architecture of the ChatGPT-based assistive system for the blind.

guage. The `gTTS` serves as an interface to Google Translate’s Text-to-Speech API and primarily supports high-resourced languages such as English, French, Mandarin, Portuguese, and Spanish. The `Kazakh_TTS` model operates offline and offers the voices of five speakers (two males, three females). When engaging with ChatGPT in the Kazakh language, the response text is translated from English to Kazakh utilizing the `googletrans` library. Subsequently, the Telegram bot delivers the generated voice responses and corresponding text messages to the user. These text messages may prove beneficial in other tasks where visually impaired and blind individuals have access to screen readers.

2.5 System Deployment

The implementation of our Python-based application for the system was executed on a high-performance workstation, equipped with an AMD Ryzen Threadripper 3970X 32-core Processor, 128 GB RAM, and an Nvidia GeForce RTX 3070 graphics processing unit (GPU), accompanied by a stable internet connection. To guarantee consistent and reliable performance, we encapsulated the application within a Docker container. This approach offers several advantages over a virtual environment, as it ensures that the application can operate across diverse systems. The architecture of the deployed system is illustrated in Fig. 2. A voice request is transmitted to ChatGPT via the Telegram bot and subsequently forwarded to a processing container on the host machine. Following this, the voice request is transcribed into a text prompt through the Google Speech-To-Text API and forwarded to ChatGPT. Ultimately, the text response generated by ChatGPT is converted into speech utilizing the Google Text-To-Speech API and returned to the Telegram Bot. In instances where offline TTS and ASR are employed, computations are conducted on the host machine’s GPU.

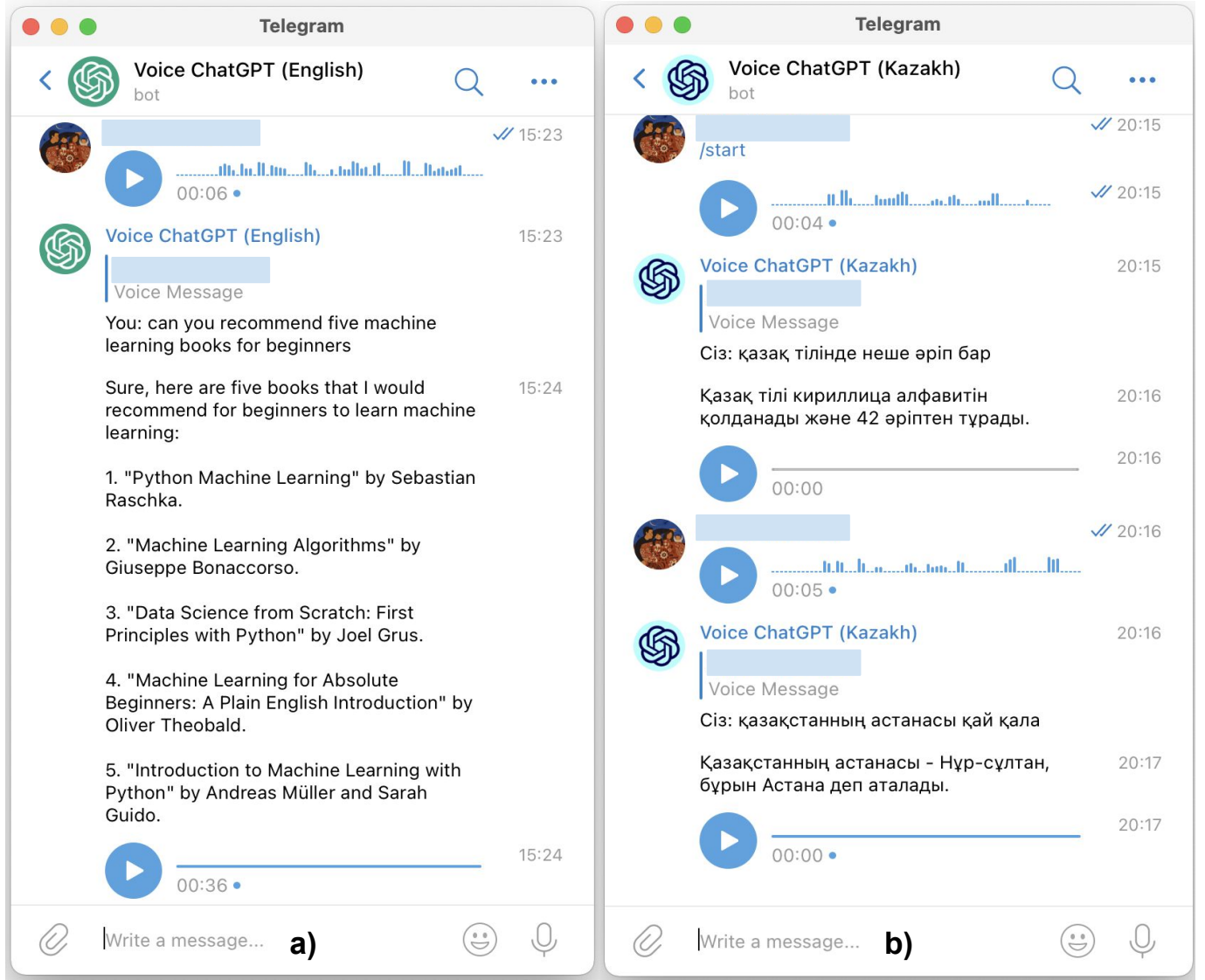


Figure 3 – Screenshots of the Telegram bots for interacting with ChatGPT showing the requests as voice messages, their text equivalents processed by ASR, and the ChatGPT responses in text and voice message formats: a) English and b) Kazakh.

3. RESULTS

The interfaces for the Telegram bots in both English and Kazakh languages are shown in Fig. 3. In Fig. 3a, we submitted a query to ChatGPT, requesting recommendations for five beginner-friendly machine learning books in English. We transmitted our inquiry via a voice message to the bot. The ASR model subsequently converted the voice message into text and returned it to us as a reply message, confirming that our request was acknowledged and in the process of being addressed. Upon successful processing of our prompt by ChatGPT, we obtained a list of five recommended books in both text and voice message formats. In the instance depicted in Fig. 3b, we inquired about the number of letters in the Kazakh language, posed in Kazakh. ChatGPT provided the accurate response, "The Kazakh language uses the Cyrillic alphabet and has 42 letters." The primary distinction between the two bots lies in the translation of

requests and responses to English when the Kazakh language is utilized.

In order to evaluate whether ChatGPT generates more coherent and meaningful responses for prompts in English compared to those in Kazakh, we conducted experiments with the following three cases:

1. We posed questions to ChatGPT in English, encompassing various domains such as history, art, technology, sport, and science.
2. In the second case, we asked the same set of questions, but in the Kazakh language.
3. In the final case, we posed the same set of questions, but translated from Kazakh to English.

We conducted extensive experiments with ChatGPT for these three cases. For brevity, we summarize our

findings in several examples that highlight the most occurred outcomes. The following examples illustrate conversations with ChatGPT in English (en), Kazakh (kk), and translated from Kazakh to English using the `googletrans` (gt). English translations are provided for the responses generated by ChatGPT in Kazakh. We requested ChatGPT to provide a one-sentence response to restrict its answer length.

Example #1

Prompt (en): When and where were the pyramids of Giza built?

Response (en): The Pyramids of Giza were built around 2500 BCE in Giza, Egypt.

Prompt (kk): Гиза пирамидалары қашан және қай жерде салынған?

Response (kk): Гиза пирамидалары екінші мың жылдың басында Мысырда жасалған. The Giza pyramids were built in Egypt at the beginning of the second millennium.

Prompt (gt): When and where were the Giza pyramids built?

Response (en): The Giza Pyramids were built in Giza, Egypt during the Fourth Dynasty of the Old Kingdom Period, approximately from 2580 to 2560 BCE.

Example #2

Prompt (en): What does the International Telecommunication Union do?

Response (en): The International Telecommunication Union is a specialized agency of the United Nations that aims to promote and coordinate international telecommunications and information and communication technologies (ICT) services and policies.

Prompt (kk): Халықаралық телекоммуникация одағы немен айналысады?

Response (kk): Халықаралық телекоммуникация одағы интернет, телефония, радио және телебағдарламалар арқылы айналысады. The International Telecommunication Union deals with Internet, telephony, radio and television programs.

Prompt (gt): What does the International Telecommunication Union do?

Response (en): The International Telecommunication Union is a specialized agency of the United Nations that coordinates global telecommunications networks and services.

Example #3

Prompt (en): What is Newton's first law?

Response (en): Newton's first law states that an object at rest stays at rest and an object in motion stays in motion with a constant speed and in a straight line unless acted upon by an unbalanced force.

Prompt (kk): Ньютонның бірінші заңы дегеніміз не?

Response (kk): Бірінші заң - кепілдік заңы, соны

қауіпсіздік сейімділігі туралы. The first law is the law of guarantee, and that is about the capacity of security.

Prompt (gt): What is Newton's first law?

Response (en): Newton's first law states that an object will remain at rest or in uniform motion in a straight line unless acted upon by an external force.

Example #4

Prompt (en): Who was Abai Qunanbaiuly?

Response (en): Abai Qunanbaiuly was a Kazakh philosopher, poet, writer, and composer considered to be one of the greatest cultural figures in Kazakh history.

Prompt (kk): Абай Құнанбайұлы кім болған?

Response (kk): Абай Құнанбайұлы Қазақ әдебиетінің жоғары дағдысы, әлемге танымал мәдени адамдарының бірі болып табылады. Abai Kunanbayuly is one of the world-renowned cultural figures of Kazakh literature.

Prompt (gt): Who was Abai Kunanbayuly?

Response (en): Abai Kunanbayuly was a prominent Kazakh poet, composer, and philosopher who lived in the 19th century.

In the first example, we posed a common historical query (i.e., "When and where were the pyramids of Giza built?"). When the question was presented in English, ChatGPT provided the answer with the accurate date and place. When the same inquiry was made in Kazakh, the response was correct, albeit less comprehensive. However, when translating the prompt into English, ChatGPT offered a more in-depth answer. In the subsequent example, we inquired about the functions of the International Telecommunication Union (ITU). ChatGPT delivered a suitable response when asked in English. Yet, when the question was posed in Kazakh, ChatGPT delivered a seemingly credible but factually inaccurate reply. This issue is referred to as "hallucination" [18] within the context of LLMs. However, when the question was translated into English, ChatGPT provided the correct answer. In the third example, we asked about Newton's first law. ChatGPT delivered an accurate explanation when presented with the English prompt. Nevertheless, the response to the Kazakh prompt was entirely unrelated. This issue was mitigated when the prompt was rendered in English. In the final example, we queried ChatGPT about a renowned Kazakh poet, Abai Qunanbaiuly. ChatGPT offered pertinent information in both English and Kazakh. However, the Kazakh response exhibited grammatical inaccuracies.

4. DISCUSSION

We demonstrated that our system is capable of facilitating interaction with ChatGPT through voice messages and can potentially serve as an assistive technology for

visually impaired and blind individuals to access information. The system offers several benefits, including an intuitive interface that does not necessitate additional user training. Moreover, the utilized libraries, APIs, and engines are freely accessible, with the exception of ChatGPT. As of the time this paper was written, the `gpt-3.5-turbo` model incurred a cost of \$0.002 per 1,000 tokens. Currently, our bots share a single key to access the OpenAI API, leading to shared usage among all users and a consequent accumulation of costs. To address this issue, our future endeavors will focus on the development of a mobile application designed to provide individual users with access to ChatGPT through their personal accounts.

The findings indicate that ChatGPT performs significantly better in the English language. This is presumably due to the fact that ChatGPT was trained on an extensive text corpus gathered from the internet, where English serves as the lingua franca. Furthermore, the training process for ChatGPT utilized human feedback to guide its learning, which may have primarily focused on higher-resourced languages, resulting in enhanced performance for those languages. To address this issue, we translated prompts from Kazakh to English. However, the quality of the answers provided by ChatGPT in this scenario is heavily reliant on the accuracy of the translation tool employed. In this project, we utilize the publicly available `googletrans`, which may not be as precise as commercial translation services such as Google Cloud Translator, Bing Microsoft Translator, or DeepL Translator. Consequently, translated responses might occasionally lack precision. Fine-tuning ChatGPT for lower-resourced languages is currently unfeasible as the model’s weights are not open source. However, Meta has recently introduced open and efficient foundation language models called Llama [19], which can be fine-tuned on custom instruction-following data [20]. In our future work, we plan to fine-tune Llama for the Kazakh language.

The performance of our system can also be influenced by the efficacy of the employed ASR and TTS models. The accuracy of the ASR model is of greater importance, as it generates the input prompt for ChatGPT. Consequently, a low word error rate ASR engine that is robust to background noise is essential for ensuring high-quality interaction. Latency presents another limitation, as each prompt-response must pass through the ASR, ChatGPT, and TTS components. The response time may be further exacerbated by a poor internet connection when utilizing cloud-based ASR and TTS models. This issue can be mitigated by employing offline ASR and TTS models; however, these offline models may not be as accurate as their cloud-based counterparts.

5. CONCLUSION

In this study, we introduced a Telegram bot-based assistive technology that enables visually impaired and blind individuals to interact with ChatGPT via voice messages. The Telegram bot is capable of utilizing both cloud-based and offline ASR and TTS technologies. Furthermore, we demonstrated that ChatGPT exhibits greater accuracy with high-resourced languages (e.g., English) compared to low-resourced languages (e.g., Kazakh). To address this issue, we translated prompts from Kazakh to English using a cloud-based translator prior to sending them to ChatGPT. We have made the source code³ publicly accessible to encourage the development of assistive systems for visually impaired and blind individuals.

REFERENCES

- [1] Rupert R A Bourne, Seth R Flaxman, Tasanee Braithwaite, Maria V Cicinelli, Aditi Das, Jost B Jonas, et al. “Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis.” In: *The Lancet Global Health* 5.9 (2017), e888–e897. ISSN: 2214-109X. DOI: [https://doi.org/10.1016/S2214-109X\(17\)30293-0](https://doi.org/10.1016/S2214-109X(17)30293-0). URL: <https://www.sciencedirect.com/science/article/pii/S2214109X17302930>.
- [2] *The Braille Literacy Crisis in America*. Tech. rep. National Federation of the Blind Jernigan Institute, 2009.
- [3] Michele C. McDonnall and Zhen Sui. “Employment and Unemployment Rates of People Who Are Blind or Visually Impaired: Estimates from Multiple Sources.” In: *Journal of Visual Impairment & Blindness* 113.6 (2019), pp. 481–492. DOI: [10.1177/0145482X19887620](https://doi.org/10.1177/0145482X19887620). eprint: <https://doi.org/10.1177/0145482X19887620>. URL: <https://doi.org/10.1177/0145482X19887620>.
- [4] Michele C. McDonnall, Jennifer L. Cmar, and Zhen S. McKnight. “Beyond Employment Rates: Earnings of People With Visual Impairments.” In: *Journal of Visual Impairment & Blindness* 116.4 (2022), pp. 526–532. DOI: [10.1177/0145482X221121830](https://doi.org/10.1177/0145482X221121830). eprint: <https://doi.org/10.1177/0145482X221121830>. URL: <https://doi.org/10.1177/0145482X221121830>.
- [5] Laura Marshall and Jeanne-Louise Moys. “Readers’ Experiences of Braille in an Evolving Technological World.” In: *Visible Language* 54.1-2 (2020), pp. 9–29.

³<https://github.com/IS2AI/telegram-bot-chatgpt>

- [6] Robert C. Berwick, Angela D. Friederici, Noam Chomsky, and Johan J. Bolhuis. "Evolution, brain, and the nature of language." In: *Trends in Cognitive Sciences* 17.2 (2013), pp. 89–98. ISSN: 1364-6613. DOI: <https://doi.org/10.1016/j.tics.2012.12.002>. URL: <https://www.sciencedirect.com/science/article/pii/S1364661312002823>.
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, et al. "Language Models are Few-Shot Learners." In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901.
- [8] OpenAI. *ChatGPT: Optimizing Language Models for Dialogue*. 2022. URL: <https://openai.com/blog/chatgpt/> (visited on 11/30/2022).
- [9] Jinyu Li. "Recent Advances in End-to-End Automatic Speech Recognition." In: *APSIPA Transactions on Signal and Information Processing* (Apr. 2022).
- [10] The Telegram Team. *700 Million Users and Telegram Premium*. 2022. URL: <https://telegram.org/blog/700-million-and-premium> (visited on 06/19/2022).
- [11] Wang Frank. *pyTelegramBotAPI*. Version 4.10.0. 2023. URL: <https://github.com/eternnoir/pyTelegramBotAPI>.
- [12] Zhang Anthony. *SpeechRecognition*. Version 3.9.0. 2022. URL: https://github.com/UberI/speech%5C_recognition.
- [13] Saida Mussakhoyayeva, Yerbolat Khassanov, and Huseyin Atakan Varol. "KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus." In: *Proc. Interspeech 2022*. 2022, pp. 1367–1371. DOI: 10.21437/Interspeech.2022-421.
- [14] Han Suhun. *Googletrans*. Version 3.0.0. 2020. URL: <https://github.com/ssut/py-googletrans>.
- [15] OpenAI. *Chat completions*. Version gpt-3.5-turbo. 2023. URL: <https://platform.openai.com/docs/guides/chat>.
- [16] Durette Pierre Nicolas. *gTTS*. Version 2.3.1. 2023. URL: <https://github.com/pndurette/gTTS>.
- [17] Saida Mussakhoyayeva, Aigerim Janaliyeva, Almas Mirzakhmetov, Yerbolat Khassanov, and Huseyin Atakan Varol. "KazakhTTS: An Open-Source Kazakh Text-to-Speech Synthesis Dataset." In: *Proc. Interspeech 2021*. 2021, pp. 2786–2790. DOI: 10.21437/Interspeech.2021-2124.
- [18] Ashish Agarwal, Clara Wong-Fillnjiang, David Sussillo, Katherine Lee, and Orhan Firat. "Hallucinations in Neural Machine Translation." In: 2018.
- [19] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, et al. "LLaMA: Open and Efficient Foundation Language Models." In: *arXiv preprint arXiv:2302.13971* (2023).
- [20] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, et al. *Stanford Alpaca: An Instruction-following LLaMA model*. https://github.com/tatsu-lab/stanford_alpaca. 2023.

AUTHORS



Askat Kuzdeuov received his B.S. degree in automation and control engineering from the Almaty University of Power Engineering and Telecommunication, Almaty, Kazakhstan, in 2014, and the M.S. degree in Robotics from Nazarbayev University, Astana, Kazakhstan, in 2019. He is currently a Senior Data Scientist with the Institute of Smart Systems and Artificial Intelligence at Nazarbayev University. Prior to that, he was a Software Engineer in the oil & gas industry, from 2014 to 2017, participating to more than 30 local and international projects. His current research interests include computer vision, machine learning, and robotics.



Shakhizat Nurgaliyev received his B.S. degree in automation and control from the Kazakhstan-British Technical University, Almaty, Kazakhstan, in 2014, and the M.S. degree in in robotics, autonomous and interactive systems with a specialism in embedded robotics from the Heriot-Watt University, Edinburgh, UK, in 2016. His research interests include robotics, the internet of things, computer vision, sensor fusion, and machine learning.



Huseyin Atakan Varol is the department chair and full professor of robotics at Nazarbayev University's School of Engineering and Digital Sciences. He received his BS degree in mechatronics engineering from Sabanci University, Istanbul, Turkey, in 2005. He received the MS and PhD degrees both in electrical engineering from Vanderbilt University, Nashville, TN, USA, in 2007 and 2009, respectively. From August 2009 to August 2011, he was first a post-

603 doctoral research associate and then a research assistant
604 professor with the Center for Intelligent Mechatronics,
605 Department of Mechanical Engineering, Vanderbilt Uni-
606 versity, Nashville, TN, USA.
607 In 2011, he joined the faculty of Nazarbayev University,
608 Astana, Kazakhstan, as a founding associate professor
609 of robotics, where he currently chairs the department
610 and directs the Institute of Smart Systems and Artifi-
611 cial Intelligence (ISSAI). His research interests include
612 artificial intelligence, soft robotics, machine learning,
613 sensor fusion and tensegrity. He has published more
614 than 100 technical papers on these topics in reputable
615 international journals and conferences. He holds multi-
616 ple international patents in the area of bionics. He has
617 attracted multiple competitively-funded research grants
618 from Ministry of Education and Science of Kazakhstan,
619 World Bank and Nazarbayev University Research Fund.
620