

**Аблабекова Чынара Азисовна** к.ф.-м.н.,  
доцент

кафедры программной инженерии Международного  
университета Кыргызской Республики

**Рыскулов Сыймык Аскарбекович**

Международный университет Кыргызстана, кафедра  
программной инженерии

**Жээналиева Назгуль Мидиновна**

ст. препод. кафедры программной инженерии Международного  
университета Кыргызской Республики

**Аблабекова Чынара Азисовна**, Кыргыз

Республикасынын Эл аралык университети программалык  
инженерия кафедрасынын доцент

**Рыскулов Сыймык Аскарбекович**

Кыргыз Республикасынын Эл аралык университети  
программалык инженерия кафедрасы **Назгуль**

**Мидиновна**, улук окутуучу

Кыргыз Республикасынын Эл аралык университети программалык  
инженерия кафедрасы

**Ablabekova Ch. A.**

Ph.D., Associate Professor

Department of Software Engineering from the

International University of the Kyrgyz Republic **Ryskulov S. A.**

Department of Software Engineering from the  
International University of the Kyrgyz Republic

**Jeenaliev N. M.**

Senior Lecturer, Deputy Head

of the department Software Engineering the International  
University of the Kyrgyz Republic

## **РАЗРАБОТКА TELEGRAM-БОТА VOICE AI ДЛЯ РАСПОЗНАВАНИЯ РЕЧИ И ПРЕОБРАЗОВАНИЯ ЕЁ В ТЕКСТ**

## **СҮЙЛӨГӨН ҮНДҮ ТАЛУУ ЖАНА ТЕКСТКЕ АЙЛАНТУУ ҮЧҮН VOICE AI TELEGRAM БОТТУ ИШТЕП ЧЫГУУ**

## **DEVELOPMENT OF THE VOICE AI TELEGRAM BOT FOR SPEECH RECOGNITION AND TEXT CONVERSION**

---

*Аннотация.* В статье исследуются перспективы использования искусственного интеллекта (ИИ) через созданного Telegram-бота для распознавания речи и преобразования ее в текст. Обсуждаются потенциальные преимущества автоматизации процессов, преобразования речи в

текст и улучшения принятия решений с помощью ИИ. А также рассматриваются вопросы, связанные с работой бота и его социальными аспектами.

**Ключевые слова:** Telegram-бот, язык Python, искусственный интеллект, распознавание речи, аудио технологии, обработка текста, автоматизация, социальные аспекты.

**Аннотациясы.** Макалада жасалма интеллекттин (ЖИ) жардамы менен түзүлгөн Telegram-бот аркылуу сүйлөмдү таануу жана аны текстке айландыруунун мүмкүнчүлүктөрүн изилделинет. Процессерди автоматташтыруунун, үндү текстке айлантуунун жана ЖИ аркылуу маалыматтарды иштеп чыгуу боюнча чечимдерди жакшыртуунун потенциалдуу артыкчылыктары талкууланат. Ошол эле учурда Telegram-боттун иштешин жана анын социалдык аспектилери сыяктуу маселелер каралат.

**Негизги сөздөр:** Telegram-бот, Python тили, жасалма интеллект, үндү таануу, аудио технологиялар, текстти иштеп чыгуу, автоматташтыруу, социалдык аспектилер.

**Abstract.** The article explores the prospects of using artificial intelligence (AI) through a Telegram bot for speech recognition and text conversion.

It discusses the potential benefits of process automation, speech-to-text conversion, and improving decision-making through AI. At the same time, the article addresses issues related to the bot's operation and its social aspects.

**Keywords:** Telegram bot, artificial intelligence, speech recognition, text processing, automation, decision-making, security, social aspects.

В последние годы голосовые технологии активно внедряются в повседневную жизнь, облегчая взаимодействие человека с техникой. От голосовых ассистентов в смартфонах до «умных» колонок и автомобильных навигаторов — системы распознавания речи стали неотъемлемой частью цифровой среды. С их помощью пользователи могут управлять устройствами, искать информацию в интернете, оставлять заметки и даже управлять умным домом. Такие технологии особенно ценны в условиях, где нет возможности использовать клавиатуру или экран [2].

Распознавание речи открывает новые горизонты в сфере доступности цифровых решений. Для людей с нарушениями зрения или моторики голосовой ввод становится основным способом взаимодействия с устройствами. Кроме того, в условиях многозадачности голосовые команды позволяют экономить время и упрощают повседневные действия — например, отправку сообщений в мессенджерах, создание напоминаний или поиск маршрутов. Особенно актуальной становится поддержка родных языков и диалектов, которые часто остаются за рамками крупных коммерческих решений.

Что повышает значимость работы.

Проект **Voice AI** был разработан с целью создать Telegram-бота, способного распознавать речь пользователей и переводить её в текст. Основной упор был сделан на поддержку **кыргызского языка**, так как существующие решения редко обеспечивают качественное распознавание для малораспространённых языков. Также в проект заложена поддержка **русского и английского** языков, что делает бота универсальным инструментом для пользователей из разных языковых сред.

Voice AI создавался как простое и удобное решение, которое можно использовать без необходимости установки дополнительных программ или приложений. Telegram как платформа был выбран из-за своей популярности, гибкости и широкой пользовательской аудитории.

Созданный бот Voice AI предоставляет пользователю следующие возможности: выбрать язык общения (кыргызский, русский или английский); отправить аудиосообщение в Telegram; получить текстовую расшифровку голоса в ответ; использовать как голосовой блокнот, переводчик или помощник. Все действия происходят в одном окне диалога, без сложных настроек и переключений.

Проект направлен на снижение языкового и технологического барьера — особенно для пользователей, которые не имеют возможности или навыков пользоваться сложными IT-решениями. Он также подойдёт для повседневных задач, связанных с голосовыми заметками, переводами, диктовками и доступностью информации.

Для создания этого проекта были выполнены следующие ключевые задачи:

1. Создание удобного интерфейса взаимодействия с ботом.

Пользователь должен интуитивно понимать, как пользоваться ботом, не прибегая к инструкциям. Для этого реализовано меню выбора языка и простой механизм отправки голосовых сообщений.

2. Интеграция мультязычной системы распознавания речи.

Бот должен распознавать кыргызский, русский и английский языки, адаптируясь под выбор пользователя.

3. Конвертация аудио в формат, пригодный для обработки.

Многие аудиосообщения приходят в формате .ogg, поэтому реализована система автоматической конвертации в формат .wav — необходимый для обработки голосовыми движками.

4. Использование внешних NLP-моделей для распознавания речи.

Для кыргызского языка была выбрана подходящая модель с открытым исходным кодом, которая эффективно справляется с задачей транскрипции.

5. Обработка ошибок и исключений.

Бот корректно реагирует на проблемы с качеством записи, форматами файлов или неполадки связи, отправляя пользователю соответствующее сообщение.

6. Обеспечение минимальной нагрузки на пользователя.

Вся работа происходит прямо в Telegram — без скачивания сторонних приложений или регистрации на сайтах.

7. Обеспечение приватности и временного хранения данных.

Временные аудиофайлы удаляются сразу после обработки, что повышает доверие к использованию бота.

Разработка Telegram-бота Voice AI велась с упором на стабильность, расширяемость и простоту взаимодействия. Чтобы предоставлять пользователю быстрый и точный результат, для реализации проекта были выбраны аудио технологии и библиотеки, позволяющие эффективно обрабатывать данные.

Проект реализован на языке Python, благодаря его читаемому синтаксису, богатому выбору библиотек и активному сообществу. Это позволило быстро разработать рабочий прототип и постепенно масштабировать функциональность.

Для взаимодействия с пользователями использовалась библиотека telebot (PyTelegramBotAPI), которая предоставляет простой интерфейс для: получения голосовых сообщений, отправки текстовых ответов, реализации клавиатуры с кнопками выбора языка, обработки команд, таких как /start. Бот запускается в режиме polling и работает без необходимости развёртывания вебхука, что делает его удобным для локального тестирования.

В зависимости от выбранного пользователем языка, бот применяет разные подходы к распознаванию речи:

- Для кыргызского используется модель машинного обучения, адаптированная под фонетику и лексику кыргызского языка. Она реализована через pipeline из библиотеки Transformers, что позволяет легко интегрировать и использовать предварительно обученные модели.

- Для русского и английского языков применяется библиотека Speech Recognition, в связке с сервисом Google Web Speech API. Это решение обеспечивает достаточно высокую точность распознавания при хорошем качестве аудио.

Telegram присылает голосовые сообщения в формате .ogg, что несовместимо напрямую с большинством движков распознавания речи. Для этого используется: библиотека pydub, преобразует входной .ogg или .mp3 файл в формат .wav, который затем Подается в модели [3,4].

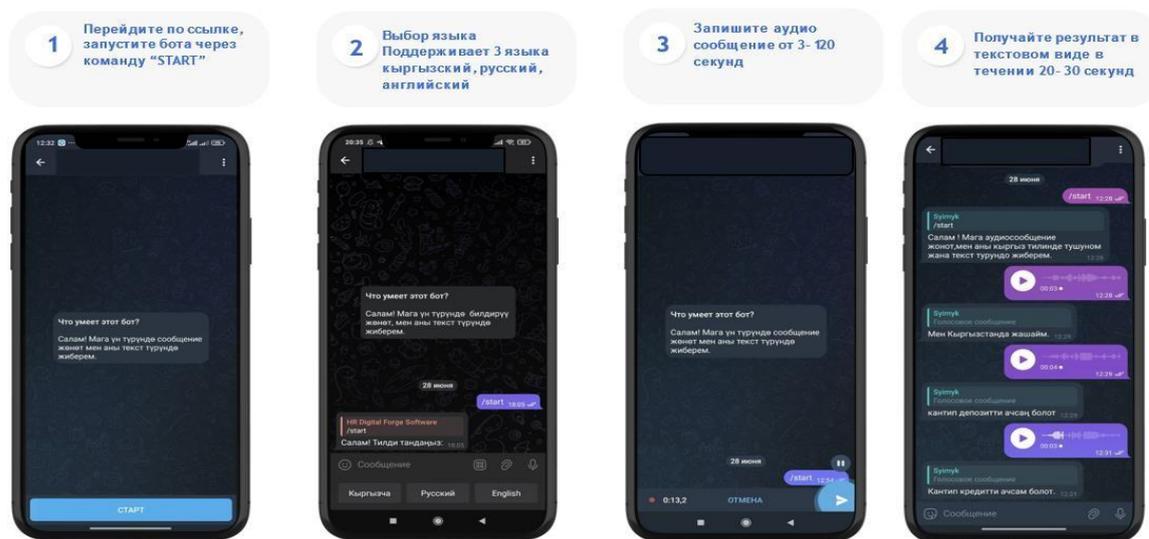
Пример: python

```
Копировать-Редактировать
audio = AudioSegment.from_file(input_path), audio.export(output_path,
format="wav")
```

Все аудиофайлы удаляются сразу после обработки, чтобы экономить дисковое пространство и повысить безопасность [8].

Для определения предпочтительного языка каждого пользователя применяется словарь user\_languages, где по ключу (ID пользователя) хранится выбранный язык. Это позволяет боту "помнить" выбор пользователя между сессиями и автоматически обрабатывать его голосовые сообщения в нужной языковой модели.

Выделим основные преимущества и области применения Voice AI: **Поддержка нескольких языков.** (Наш бот поддерживает три языка: кыргызский, русский и английский. Это делает его удобным и доступным для пользователей из разных регионов, особенно для стран Центральной Азии и СНГ). **Быстрая обработка аудио.** (Благодаря использованию современных моделей распознавания речи и оптимизации работы с аудиофайлами, время ожидания сокращается до минимума. Результат предоставляется пользователю в течение нескольких секунд). **Простота использования.** (Интерфейс реализован в виде Telegram-бота, не требующего отдельной установки или регистрации. Весь процесс работы происходит в пределах привычного мессенджера). **Персонализированная обработка.** (Бот запоминает выбранный язык конкретного пользователя и автоматически использует его при дальнейшем взаимодействии, что делает опыт использования). **Безопасность и очистка данных.** (Все аудиофайлы автоматически удаляются после завершения обработки, что исключает риск утечки данных и уменьшает нагрузку на систему).



Для понимания общей логики работы Voice AI полезно рассмотреть алгоритм действий, начиная с первого взаимодействия пользователя и заканчивая выдачей текста на основе голосового сообщения. Этот процесс можно разделить на несколько ключевых этапов:

# VoiceAI

Распознавание кыргызской речи в текст

**START**

КЫРГЫЗСКИЙ » РУССКИЙ

Голосового ввода пока для этого языка нет

1. Запуск и выбор языка. Когда пользователь впервые запускает бота с помощью команды /start, бот:

- Приветствует пользователя.
- Предлагает выбрать один из трёх языков: **Кыргызча**,

**Русский**, или **English**.

- Сохраняет выбранный язык в словаре user\_languages по идентификатору пользователя (user\_id), что позволяет обрабатывать будущие сообщения на нужном языке.

*Пример:* Пользователь выбирает "Русский" — теперь все голосовые сообщения от него будут обрабатываться через распознавание речи на русском языке.

2. Получение голосового сообщения. Когда пользователь отправляет аудиосообщение (будь то voice или audio), бот:

- Скачивает файл с серверов Telegram с помощью метода get\_file().
- Сохраняет его локально как input\_audio.

Это необходимо, чтобы затем конвертировать файл в формат, пригодный для распознавания речи.

3. Конвертация аудио. Так как большинство сообщений в Telegram приходят в формате .ogg, используется библиотека pydub для преобразования в .wav. Формат .wav лучше всего воспринимается как специализированными моделями распознавания, так и Google API.

```
audio.export(output_path,
              format="wav")
```

4. Распознавание речи. На этом этапе происходит основная "магия" — бот обрабатывает аудиофайл и

преобразует его в текст. Подход зависит от выбранного языка: Кыргызский язык: используется модель глубокого обучения, обученная на кыргызской речи. Если речь не распознана или возник сбой соединения, пользователю отправляется понятное сообщение об ошибке.

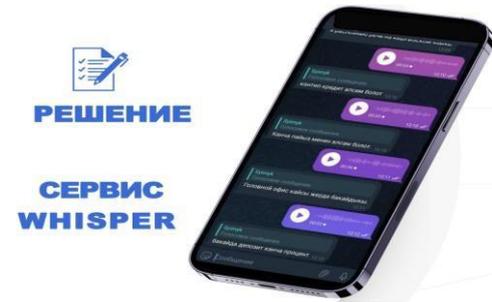
Через библиотеку transformers и pipeline API бот извлекает текст непосредственно из аудио.

- Русский / Английский язык: Применяется Google Web Speech API через библиотеку speech\_recognition, которая передаёт звуковую дорожку на сервера Google, а затем возвращает расшифровку

5. Ответ пользователю. После успешного распознавания, бот отправляет пользователю **Вы:** "Привет, как дела?" полученный текст. Если язык был кыргызский — **Voice AI:** "Привет, как дела?" ответ будет на кыргызском, если русский — на русском, и так далее

6. Очистка временных файлов. В завершение процесса бот удаляет все временные файлы (input\_audio, audio.wav) из локальной директории. Это делается для экономии ресурсов и обеспечения безопасности.

```
os.remove(audio_file)
```



Данный чат - бот может быть полезен преподавателям и студентам: для преобразования устной речи в текст во время лекций, диктантов, подготовки к экзаменам или конспектирования аудио заметок; для бизнеса при деловой переписке используется для быстрой записи голосовых напоминаний, составления деловой переписки и перевода аудио фрагментов совещаний в текст для хранения и анализа. Созданный чат бот может быть доступным для людей с особыми

потребностями. Подходит для пользователей с нарушениями речи, письма или зрения. Бот может выполнять функцию голосового помощника, обеспечивая доступ к информации в удобной текстовой форме. А также с нашим ботом можно вести работу в социальных сетях и блогах. Голосовые записи могут быть быстро преобразованы в текст для публикации или обработки, что особенно полезно для блогеров, журналистов и редакторов. Telegram-бот Voice AI помогает вести научную и исследовательскую деятельность академическому сообществу. Применяется для расшифровки интервью, подкастов, лекций, встреч и других форм аудиоинформации, облегчая последующую работу с материалами для научных публикаций.

Для разработки Telegram-бота Voice AI были использованы различные библиотеки и инструменты, обеспечивающие работу с аудиофайлами, распознавание речи и взаимодействие с пользователями. Каждый из этих компонентов играет ключевую роль в реализации проекта. Основная библиотека – Telebot, использованная для создания Telegram-бота.

*Telebot* — это популярная библиотека Python для работы с Telegram API. Она упрощает создание ботов, позволяя взаимодействовать с пользователями, обрабатывать команды и сообщения, а также отправлять текстовые и мультимедийные ответы [5].

В данном проекте эта библиотека используется для: обработки входящих сообщений, включая текстовые и голосовые сообщения; управления интерфейсом бота, включая создание кнопок и клавиатур; отправки ответов пользователям.

Библиотека *Telebot* является удобным и эффективным инструментом для разработки ботов, так как она предоставляет мощный и простой интерфейс для работы с API Telegram.

Для преобразования аудиофайлов в текст в проекте используется модель для распознавания речи на основе трансформеров. Эти модели обучены распознавать речь на разных языках, включая кыргызский, русский и английский. Модели на базе архитектуры Whisper от OpenAI показали хорошие результаты в обработке аудиофайлов в реальном времени. В качестве альтернатива использовалась другая модель для транскрибации речи на русском и английском языках, которая позволяет распознавать и интерпретировать речь в этих языках [6].

*SpeechRecognition* — это библиотека Python, которая предоставляет простой интерфейс для работы с различными сервисами распознавания речи, такими как Google Speech API, Sphinx, и другие. В этом проекте она используется для распознавания речи на русском и английском языках, предоставляя доступ к мощным инструментам для транскрибации. Ее основные функции: обработка аудиофайлов в формате WAV; использование облачных сервисов, таких как Google Speech Recognition, для преобразования речи в текст; поддержка различных форматов аудиофайлов для дальнейшей обработки [4].

*Pydub* - это библиотека для обработки аудиофайлов, которая позволяет легко конвертировать аудио из одного формата в другой, а также изменять параметры звука (например, битрейт или частоту). В проекте она используется для конвертации загруженных аудиофайлов в формат WAV, который необходим для дальнейшего анализа с использованием модели распознавания речи. Также прибегли к таким инструментам:

- *Os* - стандартная библиотека Python, используемая для работы с файловой системой, удалением временных файлов и управления путями.
- *Transformers* - библиотека от Hugging Face, которая предоставляет доступ к множеству предобученных моделей для обработки естественного языка и речи. В проекте она используется для загрузки и работы с моделью Whisper, предназначенной для распознавания речи [5].

Использование этих библиотек и технологий позволил создать мощный инструмент для преобразования речи в текст, который эффективно работает с разными языками и аудиоформатами. Эти компоненты вместе образуют надежную основу для создания функционального и удобного Telegram-бота для распознавания речи.

Современные технологии голосового ввода играют важную роль в условиях стремительной цифровизации [9]. Telegram-бот Voice AI представляет собой универсальное решение для преобразования речи в текст, обладая рядом преимуществ и широким спектром применений.

### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Демиденко А. «Создание чат-ботов для начинающих: Telegram и Python» [Электронный ресурс]. – Режим доступа: <https://www.litres.ru/>
2. Дворянкин О. А. «Голосовой помощник в интернете. Куда ведут нас информационные технологии»// Молодой ученый. - № 18(360). 2021. -С10-17.
3. Как работают чат-боты: ML, Big Data, NLU и NLP. – URL: [Электронный ресурс]. – Режим доступа: <https://www.bigdataschool.ru/blog/how-chat-bot-is-made.html>
4. Что такое чат-бот? URL: [Электронный ресурс]. – Режим доступа: <https://www.oracle.com/cis/chatbots/what-is-a-chatbot/>
5. Python. URL: [Электронный ресурс]. – Режим доступа: <https://www.python.org>
6. Andrew Volpert. Развитие ИИ Чат-Ботов: Ключевой Инструмент для Бизнеса в 2024 году – [Phttps://vc.ru/future/1022837-razvitie-ii-chat-botov-klyuchevoj](https://vc.ru/future/1022837-razvitie-ii-chat-botov-klyuchevoj)
7. Аблабекова Ч.А., Жээналиева Н.М., Досболова Ф.К. Использование цифровых технологий для транспортной задачи. Вестник Международного университета Кыргызстана. № 4 (56) 2024. С. 3-9 [https://drive.google.com/file/d/1JqUKT\\_7tbzo3pI2TcVSnGOdPjCVnJ5Br/view](https://drive.google.com/file/d/1JqUKT_7tbzo3pI2TcVSnGOdPjCVnJ5Br/view)