

УДК 004.4

## РАЗРАБОТКА ФРЕЙМВОРКА ДЛЯ СОЗДАНИЯ ГОЛОСОВЫХ АССИСТЕНТОВ

Хаяров Э.А. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук, Муромцев Д.И.  
(Университет ИТМО)

**Введение.** В силу развития генеративных моделей в сфере обработки естественного языка, генерации изображений и синтеза речи появились возможности для увеличения степени интерактивности диалоговых систем. В современные голосовые ассистенты уже внедрены большие языковые модели, однако набор инструментов для сравнительно лёгкого создания таких и даже более сложных систем невелик. Существующие фреймворки «JAICF» (язык программирования Kotlin) и «S.T.A.R.K.» (язык программирования Python) позволяют использовать текстовые каналы и аудиоканалы для взаимодействия. В целях повышения степени интерактивности предлагается реализовать собственный фреймворк, включающий возможность использования визуального канала взаимодействия для получения большего количества информации о пользователе в процессе диалога.

**Ключевые слова.** Диалоговая система, голосовой ассистент, средство разработки, фреймворк, синтез речи, распознавание речи, большая языковая модель, генеративный искусственный интеллект, генерация изображений, распознавание эмоций.

**Основная часть.** Фреймворк спроектирован таким образом, чтобы разрабатываемые с его помощью системы имели блочную структуру. Разработаны: система состояний, блоки интерфейсов для инструментов ввода-вывода и механизмов обработки информации.

Блоки ввода-вывода включают в себя механизмы, позволяющие встраивать в систему множество инструментов, таких как модели распознавания и синтеза речи, модели вывода изображений или распознавания эмоций собеседника с помощью технического зрения.

Блоки механизмов обработки информации включают в себя инструменты управления, предназначенные для обработки входных данных, генерации выходной информации и изменения состояний. На практике эти блоки могут по желанию разработчика содержать большую языковую модель с механизмом зависимости промптов от состояния, систем RAG или иных прописанных сценариев. Логика обработки входной информации может быть реализована в рамках нескольких блоков.

Блоки имеют определённые интерфейсы для взаимодействия друг с другом, что позволяет разработчику инкапсулировать реализацию логики каждого блока, обеспечивать их взаимозаменяемость и повторное использование в других проектах. Она может быть реализована как на локальной машине, так и частично на удалённых машинах, взаимодействие с которыми происходит через программные интерфейсы. Данное свойство позволяет создавать голосовые ассистенты даже для компактных устройств в узких рамках ограничений вычислительных возможностей.

На выбор разработчику предлагается базовый набор блоков с различными моделями. Также разработчик может создавать свои реализации блоков для решения специфических задач.

Приложение может собирать данные из  $N$  различных блоков ввода, обрабатывать в цепочке обязанностей из  $K$  блоков обработки информации и возвращать ответ при помощи  $M$  блоков вывода. Каждый блок может содержать переменные состояния, одну или несколько моделей нейронной сети или механизм взаимодействия с API. Между блоками передаются метаданные (состояния или команды), данные, которые необходимо обработать.

**Выводы.** Разрабатываемое решение совмещает в себе простоту использования и высокую степень гибкости для разработки. Фреймворк предназначен для запуска на интерпретаторах языка Python. Возможности представленного решения включают в себя

добавление собственных каналов ввода и вывода, включая визуальный.

Фреймворк можно использовать как для разработки ассистентов автоматизации бизнес-процессов коммерческих организаций, благодаря совмещению функции голосового ассистента с возможностями LLM с RAG системами, так и социально полезные приложения, например, «цифровые сиделки».

**Список использованных источников:**

1. JAICF documentation: [Электронный ресурс] // URL: <https://help.jaicf.com/>. (Дата обращения 05.02.2024).
2. Speech and Text Algorithmic Recognition Kit: [Электронный ресурс] // URL: <https://stark.markparker.me/>. (Дата обращения 05.02.2024).