ФОНД СОДЕЙСТВИЯ ИННОВАЦИЯМ

Заявка №: C1-100335 **Подана:** 20.01.2021

ИНФОРМАЦИЯ О ПРОЕКТЕ

Тематика проекта

Название проекта:
Система естественного клонирования голоса и озвучки текста

Название проекта на английском языке:
Natural voice and text to speech cloning system

Описание конечного продукта:
"Система естественного клонирования голоса и озвучки текста" - программный продукт для клонирования человеческого голоса с последующим его использованием для озвучки текста и синтеза речи.

Данная система характеризуется синтезом естественного голоса, соответствующего живому человеческому голосу с учетом тембра, произношения и интонации, а также возможностью голосового воспроизведения текста в режиме реального времени.

Требуется ли выполнение 2-го этапа (года) НИОКР?
Да

Обоснование необходимости проведения НИОКР 2-го этапа (года)
Модификация модели для использования разных голосов
Порт модели на мобильные устройства. Написание приложения

Требуется ли выполнение 3-го этапа (года) НИОКР?
Нет

Обоснование необходимости проведения НИОКР 3-го этапа (года)

Основное направление программы СТАРТ:
Н1. Цифровые технологии

Поднаправления:
14. Искусственный интеллект. Нейрокомпьютерные технологии и эволюционные алгоритмы.

Фокусная тематика:
Распознавание образов и речи

Приоритетные направления:
Информационно-телекоммуникационные системы

Ключевые слова:
Синтез речи, Синтез фонем, Text To Speech, Генеративные трансформеры, GPT, Обработка естественного языка, Глубинное обучение

Осуществление НИОКР в сфере спорта, городской среды, экологии, социального предпринимательства:
Нет

Описание соответствия НИОКР сферам спорта, городской среды, экологии, социального предпринимательства:

Направление в рамках Стратегии научно-технологического развития Российской Федерации:

Запрашиваемая сумма гранта (рублей):
2 000 000

Срок выполнения работ по 1-ому этапу проекта:
12

ИНФОРМАЦИЯ О ЗАЯВИТЕЛЕ И УЧАСТНИКАХ ПРОЕКТА

Основные сведения

Тип заявителя:
Физическое лицо

Руководитель (потенциальный) предприятия:
Рыжиков Артём Сергеевич

Научный руководитель проекта:
Василенко Владислав Юрьевич

Участие в конкурсном отборе:
Готовы приехать в Москву

Члены проектной команды:

Сотрудник	Должность	Роль в проекте	Опыт и квалификация
Василенко Владислав Юрьевич	Руководитель группы по разработке проекта	Развитие бизнес модели, организация процессов тестирования и разработки	

Планы по привлечению новых специалистов:
НИОКР 1-ого года (этапа) реализации проекта: нет необходимости в привлечении специалистов
НИОКР 2-ого года (этапа) реализации проекта: 1) Специалист по разработке веб-сервиса платформы 2) Специалист по анализу данных

Для исполнителей по программе УМНИК

Подача заявки в рамках обязательств по программе «УМНИК»:
Нет

Номер контракта и тема проекта по программе «УМНИК» :

Роль исполнителя по программе «УМНИК» в заявке по программе «Старт»:

Заполняется если выбранно «Иное» в поле «Роль исполнителя по программе «УМНИК» в заявке по программе «Старт»:

Информация о заявителе

Заявитель:
Рыжиков Артём Сергеевич

Дата регистрации предприятия:

Наличие в Едином реестре субъектов МСП:

Регион заявителя:
Москва

Выручка от реализации товаров (работ, услуг) за последний календарный год (рублей):
0

Среднесписочная численность сотрудников за последний календарный год, человек:
0

Профиль деятельности предприятия:

Заполняется если выбранно «Иное» в поле «Профиль деятельности предприятия»:

Участник проекта «Сколково»:

Учредители

№ п/п	Учредитель	Доля

Создано в соответствии с 217-ФЗ:
Нет

Учредитель 217-ФЗ:

СОДЕРЖАНИЕ ПРОЕКТА

Аннотация проекта

Многие люди редко находят в современном плотном графике всё меньше времени на чтение полезных ресурсов, таких как книги, статьи или иные информационные сводки. Иные же не против, чтобы кто-нибудь просто читал им перед сном приятным голосом.

К сожалению, доступные на сегодняшний день массовым пользователям продукты не удовлетворяют современным критериям качества и остаются ещё весьма далекими от естественной человеческой речи.

Мы предлагаем качественно новую технологию для озвучки текста и синтеза речи. Данная технология позволит не только естественным образом читать разного рода тексты, но и, например, поможет эффективно заменять телефонных операторов, учителей, актеров озвучки, открывая тем самым принципиально новый рынок и возможности.

Результатом проведения НИОКР будет приложение на основе обученной и протестированной модели, которое будет производить синтез речи на основе текста, используя загруженный паттерн голоса.

Научно-техническая часть проекта

Новизна предлагаемых в инновационном проекте решений:

Text-to-speech (TTS) трансформация текста имеет большие перспективы развития и широкие области применения. Однако существующие на рынке подходы обладают рядом недостатков:

Плохое качество генерируемой речи на длинных последовательностях текста

Отсутствие качественных моделей на русском языке

Невозможность качественно адаптировать речь под произвольный голос

В данной работе предлагается создать концептуально новый продукт, основанный на новый нейросетевых архитектурах типа Transformer, лежащих в основе GPT-3, для качественной озвучки текстов произвольной длины

Способы и методы решения поставленных задач НИОКР:

1. Написание и тестирование модели для синтеза фонем и речи для длинных последовательностей (текста).

Достигаем воспроизводимости результатов на имеющихся моделях без функции клонирования голоса. За основу берём модели текстовых кодировщиков для работы с длинными текстами (статья https://arxiv.org/pdf/1809.08895.pdf, имплементация https://github.com/soobinseo/Transformer-TTS)

Улучшаем имеющиеся подходы синтеза речи: заменяем авторегрессионный WaveNet в Tacotron 2 на Transformer, обучаем его на российском корпусе текста, используем обновлённые и улучшенные архитектуры Transformer (в п.1 в частности)

2. Модификация модели для использования разных голосов

Как только модель с фиксированными голосами покажет хорошие результаты, будет произведена модификация модели для клонирования произвольных голосов.

3. Порт модели на мобильные устройства. Написание приложения

Для эффективного использования обученной модели на мобильных устройствах требуются современные методы ускорения и разреживания нейросетей. Так как в основе предлагаемой архитектуры будут Transformer'ы, для ускорения будут опробованы следующие подходы: https://arxiv.org/abs/1906.00532, https://arxiv.org/abs/1910.10485. Также будут исследованы подходы с байесовским разреживанием (https://arxiv.org/abs/1701.05369, https://github.com/HolyBayes/pytorch_ard - автор проекта Артем Сергеевич)

Задел по тематике проекта:

Что сделано:

Проведен анализ существующих технологий

Исследовано качество существующих открытых решений в области клонирования голоса и машинного чтения текста

Проведен анализ рынка в сфере Text to Speech

Публикации соучредителя проекта:

https://arxiv.org/abs/1906.06096

https://arxiv.org/abs/1912.09323

https://arxiv.org/abs/2001.07493

https://iopscience.iop.org/article/10.1088/1742-6596/1085/4/042018

Также автором заявки предложен новый подход к обучению нейронных сетей https://github.com/HolyBayes/pytorch_ard, которые дает до 300 раз рост скорости обучения как сверхточные так и глубинные разреженные полносвязные сети.

Перспективы коммерциализации

Конкурентные преимущества создаваемого продукта, сравнение технико-экономических характеристик с основными аналогами, в том числе мировыми:

Основное преимущество данного продукта:

1. иммитация естественного голоса
2. обучение сети для иммитации произвольного голоса
3. выразительное прочтение с поддержкой фонем русского языка

Технико-экономические характеристики:

1. Продолжительность клонируемой речи - 5-7 секунд
2. Более 20 тысяч часов русской речи и 16 млн высказываний в датасете
3. 3-этапное глубинное обучение для создания числового представления голоса

Коммерческие аналоги:

Yandex SpeechKit - https://cloud.yandex.ru/docs/speechkit/tts/

Microsoft Text-To-Speech https://azure.microsoft.com/ru-ru/services/cognitive-services/text-to-speech/

Google Text-To-Speech https://cloud.google.com/text-to-speech

Целевые потребительские сегменты (рынки) создаваемого продукта, их объемы, динамика и потенциал развития:

Потребительские рынки продукта:

1. Рынок речевых технологий в России около 3 млрд.руб., с ростом до 25% в год
2. Рынок аудиокниг, до 6.5 млрд в России, с ростом до 30% в год
3. Международный рынок голосовых помощников 1.7 млрд $ на 2019 год, с драйвером роста в колл-центрах до 30% в год

Описание бизнес-модели проекта и стратегии продвижения продукта на рынок:
Бизнес модель проекта подразумевает включает в себя возможность продавать лицензию на использование технологии, и монетизацию через собственное мобильное приложение с внедренной технологией.

Описание бизнес-модель включает в себя:

Стратегия коммерциализации:
1. Разработка демо технологии для демонстрации потенциальным инвесторам на этапе НИОКР 1-го этапа.
2. Привлечение средств внешних инвесторов после разработки демо технологии.
3. Коммерциализации через мобильное приложение, используя модель SaaS
доставки сервиса. Пользователь платит за набор обрабатываемых аудиозаписей, изначально даётся
возможность ознакомится с демонстрационными примерами обработки.
4. Продажа лицензии на использование технологии в сторонних сервисах.

Маркетинговая стратегия по выводу собственного продукта на массовый рынок:
1. Создание портрета клиента
2. Поиск мест где могут находится наши клиенты
3. Глубинные интервью для развития клиента
4. На основе собранных интервью, формирование коммерческого предложения
5. Создание рекламных креативов и проведение тестовых рекламных компаний в соц.сетях

Инструменты маркетинга включают в себя:
использование инструментов SMM (реклама в социальных сетях), таргетированную и контекстную рекламу. Большую роль для вывода массового продукта (мобильного приложения) играют блоггеры (инфльюенсеры). Таким образом маркетинговая стратегия будет включать закупку и взаимодействие с популярными блоггерами социальных сетей.

ТЕХНИЧЕСКОЕ ЗАДАНИЕ НА ВЫПОЛНЕНИЕ НИОКР

Техническое задание на выполнение НИОКР

Цель выполнения НИОКР

Цель выполнения НИОКР, поиск наиболее эффективных способов эмбендинга голоса для превращения в цифровой вид.
Повышение качества синтеза голоса с использованием большего числа размерностей характеристик анализа речи.

Назначение научно-технического продукта (изделия и т.п.)

Основные области применения технологии Natural Voice Cloning Text-To-Speech:

1. Аудирование книг, статей, текстов (в том числе для помощи глухим)
2. Применение в дубляже видеофильмов
3. Голосовые помощники в колл-центрах
4. Голосовое управление бытовыми приборами (телефоны, автомобили, бытовая техника, умный дом)
5. Подмена голоса для приложений на основе DeepFake

Технические требования к научно-техническому продукту (макету, прототипу, лабораторному образцу, опытному образцу), который должен быть разработан в рамках текущего этапа выполнения НИОКР

Основные технические параметры, определяющие функциональные, количественные (числовые) и качественные характеристики научно-технического продукта, полученного в результате выполнения текущего этап НИОКР

Функции, выполнение которых должен обеспечивать разрабатываемый научно-технический продукт

Функция считывания голоса для создания цифрового представления
а. Данную функцию выполняет Speaker encoder на основе многослойной нейронной сети, который создает векторное представление фиксированной размерности из звука

Функция синтеза спектограммы звучания с учетом цифрового представления голоса

Функция синтеза речи на основе синтезированных спектограмм

Количественные параметры, определяющие выполнение научно-техническим продуктом своих функций

1. MOS (Mean Opinion Score) модели не менее 4.40 (или MUSHRA со значениями больше, чем у TransformerTTS (https://arxiv.org/pdf/1809.08895.pdf) и Flowtron (https://arxiv.org/pdf/2005.05957.pdf)) на текстах с длиной более 500 символов. Сравнение метрик будет производиться с помощью критерия Стьюдента при уровне значимости 5%.
2. Склонированные и оригинальные голоса в 95% случаях неразличимы. Для сравнения голосов будут использованы нейросетевые классификаторы с отрытым исходным кодом (наподобие https://github.com/jurgenarias/Portfolio/tree/master/Voice%20Classification). Измерения будут производиться по голосовым примерам длиной от 1 секунды до 5-10 минут.

Входные воздействия, необходимые для выполнения научно-техническим продуктом заданных функций

1. Текст (.pdf, .txt, .docx)
2. Голос, которым текст требуется озвучить (в формате .wav, .mp3, .flac)

Выходные реакции, обеспечиваемые научно-техническим продуктом в результате выполнения своих функций

1. Озвученный текст (в формате .wav, .mp3, .flac)

Конструктивные требования к научно-техническому продукту, который должен быть получен в результате выполнения текущего этап НИОКР

Требования к конструкции и составным частям научно-технического продукта

1. Необходима облачная инфраструктура для размещения нейронных сетей, использования процессорного времени вычислительных модулей в процессе обработки входных данных.
2. Необходимо использование технологий облачных лабораторий для проведения лабораторных исследований нейронных сетей.
3. Хранилище для накопления размеченных данных более 1Тб SSD, со скоростью доступа к данным более 3Гб/с, и скоростью записи более 1Гб/с.
4. Оперативная память облачной инфраструктуры должна быть не менее 16Гб.

Требования к массогабаритным характеристикам научно-технического продукта

Требования к массогабаритным харакрестикам научно-технического продукта отсутствуют.

Вид исполнения, товарные формы

Данный продукт будет храниться в цифровом виде, исходный код продукта будет загружен в облачный репозоторий кода с возможностью удаленного доступа по ссылке.

Требования к мощностным характеристикам научно-технического продукта – по потребляемой/производимой энергии

Требования к удельным характеристикам научно-технического продукта – на единицу производимой продукции – для машин и аппаратов

Требования к аппаратной части программных комплексов

Условия эксплуатации, использования научно-технического продукта

Иные требования к научно-техническому продукту (макету, прототипу, лабораторному образцу, опытному образцу), который должен быть разработан в рамках текущего этапа выполнения НИОКР

Требования по патентной охране

В ходе выполнения работ будут проведены мероприятия, обеспечивающие защиту прав на интеллектуальную собственность в соответствии с ч.4 ГК РФ.
Государственная регистрация объекта интеллектуальной собственности, а именно программы для электронно-вычислительных машин, базы данных и математической модели в Роспатенте будет осуществлена в 3-м квартале 2020 года.
Предлагаемая к патентованию в юрисдикции РФ интеллектуальная собственность:
«Математическая модель данных нейронной сети»;
«Алгоритм нейронной сети с набором параметров по результатам обучения».

Перечень основных категорий комплектующих и материалов (входящих в состав разрабатываемого продукта (изделия) или используемых в процессе его разработки и изготовления)

Основные комлектующие и ресурсы входящие в состав этот проекта:

1. Портативная вычислительная техника с возможностью проведения GPU вычислений
2. Облачные сервисы вычислений на GPU (такие как Paperspace.com)
3. Сервисы планирования и командной работы, такие как Miro, Trello, Jira
4. Сервисы ведения репозитория кодовой базы, такие как Github
5. Сервисы анализа рынка для привлечения внебюджетного финсирования, такие как Crunchbase"

Отчетность по НИОКР (перечень технической документации, разрабатываемой в процессе выполнения текущего этапа НИОКР)

- научно-технические отчет
- алгоритм работы программы
- описание программы
- программы и методики испытаний (тестирования) программы

КАЛЕНДАРНЫЙ ПЛАН И СМЕТА

Календарный план

Календарный план выполнения НИОКР. 1-й годовой этап проекта:

№ этапа	Название этапа календарного плана	Длительность этапа, мес	Стоимость, руб.
1	Написание и тестирование модели для синтеза фонем и речи для длинных последовательностей (текста). Перечень работ: 1. Достигаем воспроизводимости результатов на имеющихся моделях без функции клонирования голоса. За основу берём модели текстовых кодировщиков для работы с длинными текстами (статья https://arxiv.org/pdf/1809.08895.pdf, имплементация https://github.com/soobinseo/Transformer-TTS) 2. Заменяем авторегрессионный WaveNet в Tacotron 2 на Transformer, обучаем его на российском корпусе текста, используем обновлённые и улучшенные архитектуры Transformer (в п.1 в частности) 3. Улучшаем имеющиеся подходы синтеза речи. MOS (Mean Opinion Score) модели должен быть не менее 4.40 (или MUSHRA со значениями больше, чем у TransformerTTS (https://arxiv.org/pdf/1809.08895.pdf) и Flowtron (https://arxiv.org/pdf/2005.05957.pdf)) на текстах с длиной более 500 символов. Сравнение метрик будет производиться с помощью критерия Стьюдента при уровне значимости 5%.	6,00	1 000 000,00

2	Модификация модели для использования разных голосов. Перечень работ: 1. Пробуем подход Offline voice cloning (в качестве альтернативного, запасного, подхода) - клонирование голоса посредством индивидуального дообучения модели под каждый голос. 2. Пробуем подход Real time voice cloning (в приоритете, за основу берётся статья https://arxiv.org/pdf/1806.04558.pdf) - клонирование голоса в реальном времени посредством вычленения эмбедингов из клонируемого голоса. Для этого используется отдельная нейросеть - Voice Encoder. 3. Достигаем MOS (Mean Opinion Score) модели не менее 4.40 (или MUSHRA со значениями больше, чем у TransformerTTS (https://arxiv.org/pdf/1809.08895.pdf) и Flowtron (https://arxiv.org/pdf/2005.05957.pdf)) на текстах с длиной более 500 символов. Сравнение метрик будет производиться с помощью критерия Стьюдента при уровне значимости 5%. Склонированные и оригинальные голоса должны быть в 95% случаях неразличимы. Для сравнения голосов будут использованы нейросетевые классификаторы с отрытым исходным кодом (наподобие https://github.com/jurgenarias/Portfolio/tree/master/Voice%20Classification). Измерения будут производиться по голосовым примерам длиной от 1 секунды до 5-10 минут. 4. Порт модели на мобильные устройства. Написание приложения	6,00	1 000 000,00
	ИТОГО:		2 000 000

Смета

Смета затрат на 1-ый год реализации проекта:

№ п/п	Наименование статей расходов:	Сумма (руб.):
1	Начисление на заработную плату	1 122 000,00
2	Материалы, сырье, комплектующие	378 000,00
3	Оплата работ соисполнителей и сторонних организаций	500 000,00
Итого:		**2 000 000,00**

Приложение к заявке № С1-100335

от 10 января 2021 г.

Показатели реализации инновационного проекта

ИП Рыжиков А. С.

Мы, нижеподписавшиеся, заверяем правильность всех данных, указанных в таблице и обязуемся предоставлять необходимую документацию, подтверждающую указанные данные, при мониторинге финансово-производственной деятельности МИП или по требованию сотрудников Фонда. Мы предупреждены о том, что в случае предоставления недостоверных данных Фонд может прекратить финансирование проекта.

Индивидуальный предприниматель _____ (подпись) Рыжиков Артём Сергеевич (ФИО)

М. П.

Главный бухгалтер _____ (подпись) _____ (ФИО)

М. П.

Код	Показатель развития МИП	За 2020 г.	За 2021 г.	За 2022 г.	За 2023 г.	За 2024 г.	За 2025 г.	За 2026 г.	За 2027 г.
Коллектив предприятия									
КЧ*	Среднесписочная численность сотрудников МИП	2	2	3	5	5	5	0	0
Финансы									
ФР1/Ф	Объем израсходованных бюджетных средств на реализацию проекта, представленных Фондом	0	2 000 000	3 000 000	10 000 000	0	0	0	0
Интеллектуальная собственность									
И1	Общее количество объектов интеллектуальной собственности, полученных МИП в рамках реализации проекта	0	1	1	1	1	1	0	0
И1Зр	Количество поданных заявок на регистрацию результатов интеллектуальной деятельности в Российской Федерации, созданных МИП в рамках реализации проекта	0	1	1	1	1	1	0	0