**Разработка генератора анекдотов на русском языке**

**Техническое задание**

**Листов 10**

**Содержание**

[**1. Введение** 3](#_Toc153232693)

[**1.1 Наименование программы** 3](#_Toc153232694)

[**1.2 Краткая характеристика области применения** 3](#_Toc153232695)

[**2. Основания для разработки.** 4](#_Toc153232696)

[**3. Назначение разработки** 5](#_Toc153232697)

[**3.1 Функциональное назначение** 5](#_Toc153232698)

[**3.2 Эксплуатационное назначение** 5](#_Toc153232699)

[**4. Требования к программе или программному изделию** 5](#_Toc153232700)

[**4.1 Требования к функциональным характеристикам** 6](#_Toc153232701)

[4.1.1 Требования к составу выполняемых функций 6](#_Toc153232702)

[4.1.2 Требования к организации входных и выходных данных 7](#_Toc153232703)

[4.1.3 Требования к временным характеристикам 8](#_Toc153232704)

[**4.2 Требования к надежности** 8](#_Toc153232705)

[4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы 8](#_Toc153232706)

[4.2.2 Время восстановления после отказа 8](#_Toc153232707)

[**4.3 Условия эксплуатации** 8](#_Toc153232708)

[4.3.1 Климатические условия эксплуатации 8](#_Toc153232709)

[4.3.2 Требования к видам облуживания 9](#_Toc153232710)

[4.3.3 Требования к численности и квалификации персонала 9](#_Toc153232711)

[**4.4 Требования к составу и параметрам технических средств** 9](#_Toc153232712)

[**4.5 Требования к информационной и программной совместимости** 9](#_Toc153232713)

[4.5.1. Требования к исходным кодам и языкам программирования 9](#_Toc153232714)

[4.5.2. Требования к программным средствам, используемым программой 10](#_Toc153232715)

[**4.6 Требование к маркировке и упаковке** 10](#_Toc153232716)

[**4.7 Требования к транспортированию и хранению** 10](#_Toc153232717)

[**4.8 Специальные требования** 10](#_Toc153232718)

# **1. Введение**

## **1.1 Наименование программы**

Наименование программы – «Генератор анекдотов на русском языке»

## **1.2 Краткая характеристика области применения**

«Генератор анекдотов на русском языке» - программа, способная принимать на вход заранее собранную коллекцию существующих анекдотов и генерировать новые анекдоты на основе входных данных. Другими словами, функционал генератора позволяет принимать, обрабатывать, анализировать входящую коллекцию текстов, обучать на ее основе модели машинного обучения и в результате генерировать новые тексты. Финальный продукт может быть использован преимущественно в развлекательных целях. Целевую аудиторию составляют как отдельные пользователи, так и организации, желающие приобрести подобный продукт в целях повышения лояльности сотрудников и клиентов, улучшения атмосферы в рабочем коллективе. Однако возможно его применение и при разработке других проектов, основанных на идее генерации текста и/или связанных с методами машинного обучения и нейронными сетями.

Рынок развлекательных услуг имеет существенное значение в социальной и экономической сферах жизнедеятельности общества. Однако, на сегодняшний день в России он пока существенно отстает в развитии по количеству, уровню и разнообразию, в том числе и интернет-среда развлекательного сегмента услуг. Данный проект, в свою очередь, направлен на дополнение и развитие данной сферы, и поэтому основная цель разрабатываемой программы – создание уникального инструмента, способного генерировать качественный развлекательный контент.

# **2. Основания для разработки.**

Основанием для разработки является Договор №111 от 01.11.2023. Договор утвержден Директором ООО «Хахаряшки» Михайловым Петром Игоревичем, именуемым в дальнейшем Заказчиком, и Егоровым Егором Егоровичем (самозанятый), именуемым в дальнейшем исполнителем, 01.11.2023.

Согласно Договору, Исполнитель обязан разработать и установить программу «Генератор анекдотов на русском языке» на оборудовании Заказчика не позднее 13.12.2023, предоставить исходные коды и документацию к разработанной системе не позднее 13.12.2023.

Договором предусмотрены этапы разработки программного продукта, по окончании срока реализации которых Исполнитель обязан предоставлять отчетность о выполненной работе Заказчику:

Этап 1. Срок реализации: 01.11.2023 – 08.11.2023

1. Формирование перечня требований по функционалу программы.
2. Проведение мини-исследования с целью выбора инструментов для генерации текста.

Этап 2. Срок реализации: 08.11.2023 – 29.11.2023

1. Поиск и сбор входного текстового корпуса с анекдотами.
2. Создание и обучение собственной модели.
3. Работа с моделью, предоставленной в открытом доступе.

Этап 3. Срок реализации: 29.11.2023 – 13.12.2023

1. Разработка архитектуры программы.
2. Создание интерфейса для генератора.
3. Сравнение результатов и технических возможностей моделей.

Наименование темы разработки – «Разработка программы «Генератор анекдотов на русском языке».  
Условное обозначение темы разработки (шифр темы) – «Анекдоты-01».

# **3. Назначение разработки**

## **3.1 Функциональное назначение**

Функционал программы предоставляет возможность загрузки и автоматической обработки коллекции анекдотов на русском языке любого размера и возможность дальнейшей генерации новых анекдотов на основе входных текстов с применением двух моделей машинного обучения и нейронных сетей. Пользователю предоставляется выбор модели генерации, тематики анекдота и его длины.

Результатом выполнения данного проекта должен стать разработанный генератор анекдотов на русском языке, который обладает всем необходимым функционалом, описанным в ТЗ, и соответствующий заявленным требованиям к интерфейсу, интеграции, регулировке входящих параметров, обработке ошибок, особенностям тестирования, безопасности и масштабируемости.

## **3.2 Эксплуатационное назначение**

Программа является самостоятельным продуктом для генерации развлекательного контента. Она может быть использована как в личных пользовательских целях, так и в корпоративном масштабе.

# **4. Требования к программе или программному изделию**

## **4.1 Требования к функциональным характеристикам**

### 4.1.1 Требования к составу выполняемых функций

При запуске программы для пользователя должно открываться интерфейсное приложение генератора анекдотов, содержащее 2 опции выбора: выбор используемой модели машинного обучения и выбор длины генерируемого текста.

Всего на выбор должно даваться минимум 2 модели, которые должны именоваться, как “Наша” – для модели, разработанной специально для данного проекта и “Чужая” – для модели, взятой для сравнения с разработанной.

При выборе модели “Наша” пользователю также должна предоставляться возможность указать тег сгенерированного анекдота (то есть его тему). В генераторе должно содержаться минимум 10 тем на выбор, среди которых обязательно должны быть следующие:

1. Еда
2. Политика
3. Кошки
4. Вульгарный
5. Работа
6. Компьютеры
7. Дети
8. Штирлиц
9. Студенты
10. Соседи

В данный список Заказчиком в любой момент могут быть добавлены и другие категории анекдотов. Темы также могут предлагаться Исполнителем, но по предварительному согласованию с Заказчиком.

Кроме того, пользователю предоставляется возможность указать необходимую длину сгенерированного анекдота. Минимальная длины должна составлять 30 символов, максимальная – 100 символов.

Опция выбора модели, категории и длины текста должна быть оформлена в виде выпадающего списка.

Над выпадающими списками должно быть поле, в котором пользователь может написать затравку для анекдота. Справа от этого поля - кнопка "Отправить".

**Запуск**

Чтобы получить сгенерированный анекдот, пользователю нужно ввести затравку в поле, выбрать все категории и нажать "Отправить" или нажать Enter. Затем появится затравка и сгенерированный текст.

**Сохранение**

В программе должна быть автоматическая функция сохранения шуток локально. При первом запуске генератора создается специальный текстовый файл, в котором пользователь может просмотреть все сгенерированные ранее результаты.

Интерфейс программы должен быть лаконичным, удобным и понятным для неподготовленного пользователя.

После разработки непосредственно самой программы, должен быть разработан набор тестов для проверки корректности генерации анекдотов в различных сценариях.

### 4.1.2 Требования к организации входных и выходных данных

Программа должна обеспечивать проверку корректности входных данных. Они должны представлять собой именно тексты, а не что-либо иное.

Входные данные подаются пользователем – получаемые программой тексты хранятся в текстовом формате .txt. Итоговые выходные данные должны представлять собой текст в интерфейсном приложении программы.

Также должно обеспечиваться фильтрация входного и выходного контента, чтобы избежать генерации оскорбительных или нецензурных шуток.

### 4.1.3 Требования к временным характеристикам

Время выполнения программы не должно быть более 5 минут.

## **4.2 Требования к надежности**

### 4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы

Пользователю, работающему с программой через веб-браузер, должен быть предоставлен непрерывный доступ к веб-приложению, расположенному по определённому url-адресу. Веб-сервис не должен непредвиденно прерывать свою работу.

### 4.2.2 Время восстановления после отказа

В случае отказа и последующей недоступности веб-приложения, время восстановления не должно превышать одни рабочие сутки.

## **4.3 Условия эксплуатации**

### 4.3.1 Климатические условия эксплуатации

Требований к климатическим условиям эксплуатации не предъявляется

### 4.3.2 Требования к видам облуживания

Обслуживание не требуется

### 4.3.3 Требования к численности и квалификации персонала

Для управления системой достаточно одного человека, способного запустить на сервере систему управления. Требуемая квалификация пользователя – специальные требования не предъявляются

## **4.4 Требования к составу и параметрам технических средств**

Состав технических средств:

Компьютер пользователя, включающий в себя:

* процессор x86 с тактовой частотой, не менее 1 ГГц;
* оперативную память объемом, не менее 1 Гб;
* видеокарту, монитор, мышь.

## **4.5 Требования к информационной и программной совместимости**

### 4.5.1. Требования к исходным кодам и языкам программирования

Исходные коды программы должны быть написаны на языке Python.

Рекомендуемые модули и библиотеки на языке Python для разработки:

* pathlib – работа с файлами (путями к ним)
* pylint – анализ кода
* gdown – загрузка больших файлов с Google Drive
* torch – модуль для машинного обучения
* transformers – загрузка и обучение моделей
* BeautifulSoup4 - поиск информации на веб-страницах
* pandas – работа с таблицами
* requests – загрузки веб-страниц
* tqdm – отслеживание прогресса

Программа должна обучаться на двух моделях, которые основаны на одной и той же архитектуре — RuGPT3.

### 4.5.2. Требования к программным средствам, используемыми программой

Системные программные средства, используемые программой, должны быть представлены лицензионной локализованной версией операционной системы не ниже Windows 7.

## **4.6 Требование к маркировке и упаковке**

Специальных требований не предъявляется.

## **4.7 Требования к транспортированию и хранению**

Специальных требований не предъявляется.

## **4.8 Специальные требования**

Программа должна обеспечивать взаимодействие с пользователем посредством графического пользовательского интерфейса, разработанного согласно рекомендациям компании-производителя операционной системы.