

УДК 81 '38

Силанов Еремей Александрович

магистрант

Московский государственный
университет имени М. В. Ломоносова
г. Москва, Россия

Eremey Silanov

Master student

Moscow state university
named after M. V. Lomonosov
Moscow Russia
silanoveremey@gmail.com

247

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КЛАССИЧЕСКАЯ ЛИТЕРАТУРА: СОЗДАНИЕ ЛИТЕРАТУРНОГО КОМПАНЬОНА

Данная статья посвящена исследованию возможности нейросети ChatGPT 3.5 воспроизводить элементы определенного идиостиля и обрабатывать тексты русской классической литературы. В работе рассматривается создание чат-бота, обученного на произведениях М. Е. Салтыкова-Щедрина. Была произведена настройка нейросети для предоставления корректных и полных ответов на вопросы о творчестве писателя в его языковой манере. Основные этапы работы включают в себя сбор данных, токенизацию текста, обход ограничений на количество токенов, подготовку текстов и тестирование модели. Созданный литературный компаньон может взаимодействовать с пользователями в уникальной манере и отвечать на ряд вопросов по литературным произведениям. Представленные результаты демонстрируют потенциал использования виртуальных ассистентов для создания инновационной образовательной среды.

Ключевые слова: идиостиль; искусственный интеллект; литературный текст; токенизация; цифровое образование.

ARTIFICIAL INTELLIGENCE AND CLASSICAL LITERATURE: CREATING A LITERARY COMPANION

This paper is devoted to the study of ChatGPT 3.5 neural network's ability to reproduce idiosyncrasy and process texts of Russian classical literature. The paper deals with the creation of a chatbot trained on the works of M. E. Saltykov-Shchedrin. The neural network was configured to provide correct and complete answers to questions about the writer's works with elements of his linguistic manner. The main stages of the work include data collection, tokenisation of the text, circumventing restrictions on the number of tokens, text preparation and model testing. The created literary companion can interact with users in a unique manner and answer a series of questions on literary works. The presented results demonstrate the potential of using virtual assistants to create innovative educational environments.

Key words: idiosyncrasy; artificial intelligence; literary text; tokenisation; digital education.

Одним из перспективных направлений лингвистики является изучение искусственного интеллекта и его роли в коммуникации [1, с. 28; 2, с. 201]. Цель исследования – получить ответ на вопрос, может ли нейросеть в какой-то степени воспроизводить авторский стиль, а также установить, насколько успешно нейросеть может обрабатывать крупные тексты, на которых она не была ранее обучена. Например, тексты русской классической литературы.

Речь идёт о создании чат-бота, обученного на творчестве того или иного автора. После обучения чат-бот сможет давать корректные и полные ответы о творчестве писателя. Была осуществлена попытка установить, насколько успешно чат-бот может отвечать на запросы пользователя в языковой манере, приближенной к той, которая была свойственна автору, на чьих текстах был обучен чат-бот.

Можно условно назвать такого чат-бота «литературным компаньоном» и в перспективе сделать общедоступным для всех интересующихся русской классической литературой. Подобные проекты имеют вполне очевидное прагматическое значение, а именно преодоление существующих вызовов

в образовательной сфере и активизации интереса учащихся к изучению литературы. Подростки могут испытывать затруднения с привычными методами обучения. Путем создания литературных компаньонов, способных интересно взаимодействовать с учащимися в виртуальной форме, можно развить их литературные компетенции, а также расширить доступ к качественным литературным ресурсам. Данный проект направлен на создание инновационной образовательной среды, которая сочетает в себе современные технологии и литературное наследие.

Работа по созданию литературного компаньона включает несколько этапов. Прежде всего, с использованием fine-tuning и технологии OpenAI Assistants была проведена попытка обучить чат-ботов имитировать стиль письма ряда русских писателей. В качестве автора был выбран Салтыков-Щедрин. Процесс обучения проходил следующим образом.

Первый этап – сбор данных, была собрана информация о биографии Салтыкова-Щедрина, выбраны литературные произведения, такие как сказки «Премудрый пискарь», «Как мужик двух генералов прокормил» и «Дикий помещик».

У ChatGPT есть ограничение на отправку текста в 4096 токенов на одно сообщение. Для воспроизведения идиостиля необходимо использовать тексты, значительно превышающие этот лимит, поэтому была использована среда OpenAI Playground, которая позволяет создавать виртуальных ассистентов с доступом к текстовым документам, преобразованным в векторный набор данных.

Токены – это единицы, на которые нейросеть разделяет текст, этот процесс называется токенизацией.

Токенизация – процесс преобразования объекта (предложения, слова, знака) из текстового формата в числовое значение, именно с таким форматом работает нейросеть.

– Пример предложения: *“На дворе стоял теплый летний день.”*

– Процесс токенизации: *“На”, “дворе”, “стоял”, “теплый”, “летний”, “день”, “.”*

Каждый из использованных текстов был разделен на фрагменты по 4 000 токенов, каждый фрагмент был сохранен отдельным текстовым файлом.

После этого создан векторный набор данных из этих текстовых файлов, который используется виртуальным ассистентом на основе модели ChatGPT 3.5 turbo.

В результате проведенной настройки модель, теперь имеющая возможность обращаться к выбранным текстам, отвечает на вопросы о творчестве писателя корректно и достаточно полно, стараясь воспроизводить некоторые обороты и выражения, свойственные писателю. Кроме того, она учитывает те наработки критики, которые существуют в отношении произведений Салтыкова-Щедрина. Нейросеть, обученная таким образом, в состоянии не только давать объективную информацию о творчестве писателя, она также в силах помочь школьнику пройти ряд тестовых заданий по теме.

Приведем примеры обращения к обученному чат-боту – пусть читателей данной статьи не смущает местоимение «ты» в обращении к «писателю». Данный проект рассчитан на ознакомление прежде всего подростков с литературой, а они, с большой очевидностью, будут использовать местоимение «ты»

Запрос: *Зачем ты писал сатирические сказки?*

Ответ: *Да знаете ли, милейший мой собеседник, я писал сатирические сказки для того, чтобы через смех и иронию раскритиковать пороки и недостатки общества своего времени. Я видел жалкое положение крестьян и безграничное разгульство дворянства, искажения власть имущих и несправедливость власти. Мои сказки были моим оружием против многих общественных явлений, и я верил, что их критика через сатиру может привести к изменениям к лучшему.*

В ответе чат-бота мы видим имитацию устаревших обращений и элементы классического литературоведческого анализа сказок Салтыкова-Щедрина. Воспроизвел ли чат-бот авторский стиль? Трудно сказать, учитывая, что перед нами просто реплика диалога, а не литературное произведение. Но некоторая стилизация, попытка создания образа рассказчика XIX века присутствует. Ответ по сути своей верный и скорректированный.

Также чат-бот демонстрирует идеальное знание литературных источников:

Запрос: *Продолжи цитату пескаря: «Слава богу, я свою смертью помираю».*

Ответ: *Продолжение цитаты от пескаря: «Слава богу, я свою смертью помираю, так же, как умерли мать и отец».*

Теперь попробуем задать вопросы о творчестве Салтыкова-Щедрина обычной модели ChatGPT 3.5 turbo, не прошедшей предварительного обучения. Понятно, что языковая модель GPT 3.5, в том случае если она обучена на больших корпусах текста, таких как книги, сообщения на форумах, статьи в интернете и тому подобное, из-за отсутствия доступа к структурированной информации может начать галлюцинировать, неправильно отвечая на запросы. Вот как ChatGPT 3.5 turbo работает без предварительной подготовки, без загрузки и обработки авторских текстов:

Запрос: *В каком году была написана сказка «Премудрый пискарь»?*

Ответ: *Сказка «Премудрый пискарь» была написана в 1835 году русским писателем Александром Сергеевичем Пушкиным.*

Запрос: *Кратко перескажи мне сюжет этой сказки*

Ответ: *Конечно! «Премудрый пискарь» – это сказка о мудрости и находчивости. Главный герой, молодой человек по имени Иван, отправляется в путь, чтобы стать премудрым пискарем. В своем путешествии он встречает различных персонажей и проходит через различные испытания, приобретая мудрость по мере своего опыта. В конечном итоге, благодаря своему уму и находчивости, Иван достигает своей цели и становится премудрым пискарем, способным разрешать сложные задачи и помогать другим.*

Мы видим, что нейросеть галлюцинирует и заявляет, что сказку «Премудрый пискарь» написал Александр Сергеевич Пушкин. Пересказывая сюжет сказки, нейросеть несет несуряцицу, вместо пискаря теперь главный герой некий Иван. Предварительно же настроенная нами нейросеть четко и правильно отвечает на поставленные вопросы.

Запрос: *Когда ты написал сказку «Премудрый пискарь»?*

Ответ: *Одной из сказок, написанных мною, является «Премудрый пискарь». Это сатирическая сказка, созданная в 1883 году. Сказка рассказывает о пискаре, который всю свою жизнь провел в страхе и замкнутости, не обзавелся семьей и не нашел смысла в своей жизни. В конце сказки, пискарь исчезает, оставив за собой загадку – умер он сам или его поглотила щука. Это произведение было написано во времена острых политических событий, что также оказало влияние на сюжет и контекст сказки.*

Кроме пересказа текста предварительно обученная модель может отвечать на более конкретные вопросы, как с открытым ответом, так и в формате тестирования.

В качестве эксперимента нейросети было предложено ответить на тестовые вопросы (3 школьных теста по 10 вопросов) по сказкам писателя, и она успешно их выполнила, только в последнем допустив одну незначительную ошибку – в сказке «Дикий Помещик» назвала медведя другим именем.

Вот некоторые вопросы, на которые чат-бот успешно дал ответ:

«Какую газету любил читать глупый помещик?», «Какие наказания давал сыну отец-пискарь?», «Какие общественные пороки обличает М. Е. Салтыков-Щедрин в этом произведении?».

Важно, что чат-бот, как правило, не просто дает ответ на поставленные вопросы, но отвечает в формате диалога.

В конце скажем о потенциале виртуальных агентов.

Согласно ряду исследований, в контексте современных образовательных вызовов молодежь сталкивается с трудностями в усвоении литературных материалов традиционными методами.

Общение в формате виртуального диалога сокращает дистанцию между автором XIX века и подростком XXI века. Подобный подход может активизировать интерес к чтению, может быть использован в литературных кружках, школах и даже вузах.

Настроенная нейросеть потенциально может помочь решить серьезную образовательную задачу, помочь как учащемуся, так и учителю в процессе работы с литературным материалом.

ЛИТЕРАТУРА

1. Николаева А. В., Силанов Е. А., Клушина Н. И. Искусственный интеллект и его стилистические возможности // Актуальные проблемы стилистики. – 2023. – № 9. – С. 28–38.

2. Тошович Б. Лингвистика, стилистика и поэтика виртуальности : монография. – 1-е изд. – М.: ФЛИНТА, 2024. – 520 с.