

УДК 009

## ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ РАБОТЫ CHATGPT-3.5 С ЛИТЕРАТУРНЫМИ ТЕКСТОВЫМИ ДАННЫМИ

**Алиса Евгеньевна Урих**

студент департамента филологии

НИУ «Высшая школа экономики» – Санкт-Петербург

190068, Санкт-Петербург, наб. канала Грибоедова 123А. aeurikh@edu.hse.ru

Методы искусственного интеллекта являются перспективными технологиями. К ним относятся, в частности, большие языковые модели, в том числе и ChatGPT. Предлагаемое исследование посвящено не только работе ChatGPT-3.5 с литературными текстовыми данными, но и промпт-инжинирингу. Для выявления возможностей и ограничений функционала чат-бот был протестирован на нескольких категориях задач: суммаризации, антисуммаризации, генерации, смены тональности художественного текста и ответах на вопросы к тексту. Для оценки работы модели были привлечены 50 респондентов, которые определяли, с какой задачей модель справилась лучше, а с какой хуже всего, и как стратегия создания промптов влияет на качество ответов ChatGPT.

**Ключевые слова:** чат-бот; промпт; ChatGPT-3.5; анализ текста; тестирование.

### Введение

В век научного прогресса происходит развитие информационных технологий, которые используются не только для более эффективного поиска и обработки данных, но и для упрощения человеческой жизни. В первую очередь стоит сказать об искусственном интеллекте – одном из направлений компьютерных наук, специалисты в области которого занимаются созданием моделей машинного обучения [Schank 1987: 59–61]. Работа экспертов заключается в предоставлении модели набора данных и алгоритма, благодаря которому модель способна анализировать информацию и, в конечном итоге, разрабатывать и принимать решения или выполнять прогнозы [там же: 64–65]. В широком смысле, создавая искусственный интеллект, люди предпринимают попытку автоматизировать процессы различных видов человеческой деятельности. На сегодняшний день существует несколько систем искусственного интеллекта. Одной из самых известных и передовых является ChatGPT. ChatGPT – это предварительно обученная модель, которая была создана для общения в диалоговом формате (иначе она называется чат-ботом). ChatGPT был разработан компанией OpenAI в ноябре 2022 г. По данным Reuters [Hu 2023: electr. resource], к февралю 2023 г. число его пользователей выросло до 100 млн человек.

Предметом исследования в данной работе является не весь функционал модели, а только та его часть, которая связана с текстовыми данными. Основной целью исследования является по-

иск возможностей и ограничений ChatGPT-3.5 в работе с текстом, поэтому большая часть исследования посвящена тестированию модели, а также качественному и количественному анализу ее ответов по результатам самостоятельной оценки, а также с помощью опроса респондентов. Поставленная цель предполагает решение нескольких задач:

- 1) обзор функционала ChatGPT-3.5;
- 2) тестирование инструмента и самостоятельная оценка ответов ChatGPT-3.5;
- 3) проведение опроса относительно результатов работы модели.

### Метод обучения модели

Чтобы наладить работу ChatGPT, специалисты использовали метод обучения с подкреплением на основе отзывов [Introducing ChatGPT 2022: electr. resource], в основе которого лежит модель вознаграждения [Wang et al. 2023: electr. resource]: она принимает на вход запрос «и возвращает скалярный показатель вознаграждения, отражающий экспертную оценку» [Красочкин 2023: 32]. Примечательным в этом методе является то, что языковая модель обучается на собственном опыте, не только тестируя те варианты, которые она уже выбирала в прошлом, но и исследуя те, что раньше не пробовала, для того чтобы выявить лучший из них. Полученные в результате обучения модели, впервые упомянутые в 2022 г., назвали InstructGPT [Ouyang et al. 2022: electr. resource]. ChatGPT – дочерняя модель InstructGPT: она обучается по тем же принципам, однако существуют небольшие отличия в

настройке сбора данных [Introducing ChatGPT 2022: electr. resource].

Разберем, как настраивался InstructGPT. Его обучение проходило в 3 этапа [Ouyang et al. 2022: electr. resource].

1. Тонкая настройка под наблюдением (Supervised Fine-Tuning, SFT). Первые шаги в разработке были ориентированы на тонкую настройку модели GPT-3, которая сначала вводит минимальные, специфичные для конкретной задачи параметры, а затем обучается на последующих задачах, путем простой точной настройки всех предварительно обученных параметров [Devlin 2018: electr. resource]. Весь процесс проходил под наблюдением 40 специалистов, которые создали обучающий набор данных в виде подсказок, в последствии тщательно отобранных и распределенных по трем категориям. Первая категория – простые подсказки: любой произвольный запрос пользователя. Вторая категория – наборы подсказок: различные инструкции, которые содержат в себе две и более пары запрос-ответ. И, наконец, третья категория – пользовательские подсказки по конкретному запросу. Далее специалисты вручную написали ответы для каждого ввода. В конце концов, эксперты доработали GPT-3, используя для этого новый набор данных для создания GPT-3.5, которая также называется моделью SFT. Компиляция всех написанных специалистами подсказок повлияла на развитие обучаемой модели, которая получила 13 000 новых образцов ввода/вывода. На этом шаге происходит «дообучение прошлой языковой модели на первом типе размеченных данных» [Маркианова 2023: электр. ресурс]. В ней хранятся демонстрационные варианты для ее дальнейшего обучения [Kalla, Smith 2023: 827].

2. Модель вознаграждения (Reward Model, RM). По результатам обучения модели SFT чат-бот имел возможность создавать более согласованные ответы на запросы пользователей. Дальнейшая работа происходила в форме обучения модели вознаграждения, которая использует созданные подсказки и готовые ответы. Модель вознаграждения – это неотъемлемый элемент обучения с подкреплением, которое ставит перед собой цель научить модель генерировать результаты, приводящие к максимальному вознаграждению. Размер вознаграждения зависит от того, как эксперты оценили ответы чат-бота. Для обучения модели вознаграждения ИИ-тренеры, специалисты в области работы с искусственным интеллектом, получали от 4 до 9 ответов модели SFT на один пользовательский запрос и ранжировали их от лучшего к худшему, основываясь на критерии полезности.

3. Обучение с подкреплением (Reinforcement learning from human feedback, RLHF) посредством оптимизации проксимальной политики (PPO) в модели вознаграждения. Последний этап предполагает преобразование модели вознаграждения в скалярную величину, которая присваивается каждой паре запрос/ответ. Оптимизация вознаграждения происходит с помощью PPO, определенного класса алгоритмов, который используется в обучении с подкреплением и тренирует модель принимать решения [Schulman et al. 2017: electr. resource].

Важно отметить, что этапы 2 и 3 можно повторять непрерывно: каждый раз эксперты собирают все больше данных для сравнения, которые впоследствии используются для обучения новой модели вознаграждения и новой политики.

### **Промпт-инжиниринг – искусство создания промптов**

Ученые выделили несколько техник создания запросов. Самые простые это Zero- и Few-Shot Prompts. Второй вид запроса отличается от первого тем, что он содержит в себе несколько демонстраций, примеров [Козырев 2023: электр. ресурс]. Помимо них существует еще Chain-of-Thought Prompting [Zhang et al. 2022: electr. resource]: эта техника подразумевает пошаговое выполнение действий. В свою очередь она также делится на Zero-CoT и Manual-CoT [там же]. В первом случае запрос содержит только призыв к решению задачи по шагам, а второй включает в себя еще и ожидаемый вывод в качестве примера ответа.

Помимо этих техник есть еще много рекомендаций, которые влияют на качество самого промпта и выводимых данных. Ниже представлены некоторые из них.

1. Задать начальную установку. Модель будет лучше работать в тех случаях, когда пользователь просит ее представить себя специалистом, обладающим профессиональной компетенцией в интересующей его области. Например, можно дать указание ответить от лица опытного консультанта в магазине игрушек.

2. Подробно описать задание. Хороший промпт включает в себя понятную и не двусмысленную постановку вопроса для того, чтобы чат-бот мог легче интерпретировать вопрос. Например, предоставить рекомендации для интересующего списка товаров отдельно для родителей и детей. Такой прием называется Perspective Prompting [там же]. Модель может ответить на один вопрос, изучая его с разных сторон: в данном примере она рассматривает и оценивает ситуацию с точки зрения родителей и ребенка. При

формулировке запроса пользователю рекомендуется использовать ряд подсказок и меток (*few-shot prompting*), которые помогают чат-боту качественно разобраться с поставленной задачей.

3. Определить нужный объем. Для получения наиболее качественного и полезного ответа пользователю рекомендуется задать необходимую длину ответа, создав ограничение для вывода.

4. Использовать ограничители. Для того чтобы поставить границу между формулировкой задания и текста, с которым будет работать языковая модель, лучше использовать специальные знаки и символы. Например, пояснить, что нужные данные находятся в тройных кавычках.

5. Обозначить целевую аудиторию и нужный стиль. Запрос пользователя в идеале должен содержать информацию о том, для кого требуется ответ, чтобы модель выбрала понятную формулировку.

6. Определить формат результата. Можно попросить чат-бот оформить ответ в виде таблицы, списка, сплошного текста и др.

7. Использовать позитивные аффирмации (*do* вместо *don't*). Языковая модель лучше воспринимает запросы с позитивной лексикой. Например, для того, чтобы попросить модель написать текст, состоящий из слов, начинающихся на букву П, лучше избегать формулировок со словами «запрещено», «не разрешается», «нельзя» и т. п., вместо этого необходимо использовать утвердительные директивы.

8. Формулировать уточняющие вопросы. При необходимости можно корректировать ответы модели, задавая дополнительные условия: попросить модель конкретизировать вопрос, сделать более подробный анализ, поменять фокус внимания и др.

Включение в промпт подробных и конкретных инструкций может положительно сказаться на качестве ответов моделей. Поэтому, для того чтобы получить полезный и корректный результат, нужно хорошо продумать формулировку запроса и максимально четко сформулировать задачу.

### **Функционал языковой модели и результаты тестирования**

Функционал модели в работе с текстовыми данными [Введение в промпт-инжиниринг 2024: электр. ресурс]:

- 1) трансформация текста: суммаризация и антисуммаризация, тональная трансформация текста, перевод на другой язык;
- 2) генерация текста и извлечение данных;
- 3) ответы по тексту;
- 4) мозговой штурм.

Для того чтобы посмотреть, как работают алгоритмы чат-бота на практике, и оценить ответы

инструмента, нужно было провести тестирование. Сначала стоит проверить, как модель справится с суммаризацией текста. В качестве анализируемых произведений были взяты всем знакомые фольклорные детские сказки «Колобок» и «Царевна-лягушка», а также рассказ И.С. Тургенева «Муму». Они различаются количеством персонажей, сложностью сюжета, а, главное, объемом. Интересно посмотреть на то, как с этой задачей справится ChatGPT и ухудшится ли качество его ответов с увеличением объема анализируемого произведения. Чат-бот предоставлял несколько вариантов ответов по каждому из произведений.

**Суммаризация сказки «Колобок».** Модель дала несколько вариантов, которые отличались не только по объему, но и по количеству ошибок: лексических (*съедение, оживившись*), грамматических (*из немного муки*), фактических (*колобка замесили на сметане*) и пунктуационных.

Далее в качестве примера представлен один из ответов чат-бота:

*Сказка «Колобок» рассказывает о приключениях колобка, испеченного стариком и старухой, который, оживившись, покидает дом и встречает разных животных, пытаясь избежать съедения. Каждое животное он убеждает не съедать его, предлагая послушать свою песенку. Однако лиса, воспользовавшись лестью, съедает колобка, когда тот прыгает на ее морду, чтобы спеть песню еще раз.*

Затем была предпринята попытка оптимизировать запрос. Для этого нужно было, во-первых, задать начальную установку («Представь, что ты – талантливый писатель») и требуемый объем в 1–2 абзаца. Во-вторых, использовать подсказки «ты должен», «тебе необходимо». В-третьих, обратить внимание модели на то, что текст должен получиться грамотным и безошибочным. В-четвертых, четко указать, что должно содержаться в ответе: «Тебе необходимо подробно описать сюжет, упомянув всех героев произведения». В-пятых, ввести ограничение на включение в ответ дополнительной информации о сказке. И, в-шестых, упомянуть о том, что «текст должен получиться связным и легко читаемым». Оптимизированный запрос исправил фактические ошибки, но оставил речевые и пунктуационные, а также лексические повторы.

Далее представлен пример – один из ответов чат-бота:

*«Колобок» – русская народная сказка о колобке, который сбегает от своих создателей и встречает по пути зайца, волка, медведя и лису, каждый из которых пытается его съесть. Колобок убеждает от них, обещая, что он слишком быстр, чтобы быть съеденным, но в конце концов попадает в пасть лисы и погибает.*

Этот ответ оказался намного короче и проще, согласно описанному в промпте требованию. Из-за этого количество ошибок, действительно, сильно сократилось: осталась лишь одна речевая ошибка (*встречает по пути*). Еще один недочет ответа заключается в том, что ChatGPT хоть и упомянул о «создателях» колобка, но не указал, что ими были бабушка и дедушка, которые также являются героями сказки. Несмотря на это, чат-боту удалось изменить текст, устранив все фактические ошибки и убрав всю лишнюю информацию о сказке.

**Суммаризация сказки «Царевна-лягушка».** В этом случае модель также не отличилась грамотностью: она допустила синтаксические (*найти себе невест по тому, куда упадет их стрела*) и пунктуационные ошибки (*Однако он нарушает условия заклатья, и Василиса вынуждена покинуть...*), а также некоторые фактические неточности (*сестрица соткала ковер*). Из-за увеличения объема текста чат-бот упустил некоторые детали сюжета и придумал новые. Приведем как пример один из ответов чат-бота:

*«Царевна-лягушка» – это русская народная сказка, которая рассказывает о приключениях Ивана-царевича и его необычной невесты. История начинается с того, что царь велит своим трем сыновьям пустить стрелы и найти себе невест по тому, куда упадет их стрела. Стрелы старших братьев попадают в дома бояр и купцов, а стрела младшего сына, Ивана-царевича, падает в болото, где ее находит лягушка. Не имея выбора, Иван-царевич берет лягушку в жены. Вскоре царь решает проверить своих невесток и задает им три задания: соткать ковер, испечь хлеб и явиться на царский пир в лучших нарядах. Лягушка, которая на самом деле является заколдованной Василисой Премудрой, с легкостью справляется со всеми задачами, превратившись на время в прекрасную девушку и используя свои магические способности. Иван-царевич поражен мастерством своей жены, но не понимает, что она заколдована. Позже Иван-царевич сжигает лягушачью кожу, надеясь, что Василиса навсегда останется человеком. Однако он нарушает условия заклатья, и Василиса вынуждена покинуть его и вернуться к Коцею Бессмертному. Иван-царевич отправляется на поиски своей жены и проходит через множество испытаний, сражаясь с врагами и встречая на своем пути различных сказочных существ, таких как Баба-яга. С помощью своих новых друзей и собственной смелости Иван-царевич находит Коцея, побеждает его и освобождает Василису. Они возвращаются домой и живут долго и счастливо. Сказка подчеркивает темы любви, верности, смелости и мудрости.*

**Суммаризация рассказа «Муму».** Интересно, что при суммаризации этого текста некоторые ответы чат-бота отличались эмоциональными описаниями героев и их характеров (например

*барыня, капризная и эгоистичная*), при этом фактических ошибок в ответах не было, но остались речевые (*повесть подчеркивает*) и грамматические ошибки (*Герасим находит и приютит*). Далее представлен пример – один из ответов чат-бота:

*Повесть Ивана Сергеевича Тургенева «Муму» рассказывает о трагической судьбе глухонемого дворника Герасима и его любимой собаки Муму. Герасим работает на хозяйку, жестокую и властную барыню. Несмотря на свою силу и трудолюбие, он глубоко одинок из-за своей глухоты. Однажды Герасим находит и приютит бездомную собаку, которую назовет Муму. Муму становится для него единственным другом и смыслом жизни. Однако барыня, капризная и эгоистичная, решает, что собака ей мешает, и приказывает избавиться от нее. Герасим вынужден выполнить приказ. С тяжелым сердцем он уводит Муму к реке, привязывает ей камень и топит ее. После этого он покидает город и возвращается в свою деревню, оставшись одиноким и разбитым. Повесть подчеркивает темы жестокости, несправедливости и одиночества.*

В целом, ChatGPT неплохо справляется с суммаризацией текста, пересказывая основные события произведения, раскрывая его идею и описывая героев. В результате анализа ответов модели удалось выяснить, что качество выводимых данных зависит от жанра и объема текста, а также от формулировки запроса. Лучший результат суммаризации был достигнут при работе с рассказом «Муму», а наименее удачный – при работе со сказкой «Царевна-лягушка».

Антисуммаризация текста – обратный процесс, при котором пользователь предлагает чат-боту идею, делится мыслями и пожеланиями касательно будущего текста, который впоследствии должен написать ChatGPT. Например, модель получает запрос придумать сказку про Колобка, который, пытаясь убежать от лисы, оказался в Африке. Ниже приведен пример ответа чат-бота:

*Жил-был колобок, который решил путешествовать по миру. Он покатился по лугу, песенку напевая, и вдруг встретил лису, голодную и хитрую. Лиса захотела съесть колобка, но тот был проворен и смыслен. Катится, катится колобок, и вот – прямо в жаркие края, в африканские просторы. Там его встретил жираф, добрый и любопытный. «Куда ты, колобок, путь держишь?» – спросил жираф. «Я путешествовую по миру, ищу новые приключения!» – ответил колобок. Жираф подмигнул своими длинными ресницами и проводил колобка к самому саванне.*

В ответе ChatGPT использует жанровые черты сказки (лексические повторы: *катится, катится*, сказочный зачин *жил-был*). Сгенерированный текст получился лучше, чем при суммаризации, поскольку модель допустила меньше ошибок и учла все условия запроса. Еще одно достоинство текста – прямая речь, которая часто используется

в произведениях выбранного жанра. Однако у текста есть и недостатки, например, смена стиля. Если в начале ответ чата очень напоминает сказку, то дальше стиль повествования меняется, становится более реалистичным. Текст написан достаточно корректно, в нем присутствует только одна грамматическая ошибка – *к самому саванне*. Хотя ChatGPT пишет короткое заключение и вывод, трудно сказать, что сгенерированный моделью ответ носит поучительный характер, обычно свойственный детским сказкам.

С этим заданием чат-бот справился лучше, чем с предыдущим, создав текст в заданном жанре, допуская при этом меньше ошибок, нежели при суммаризации. Главное, что при написании сказки модель правильно интерпретировала запрос и учла все заданные в промпте условия.

Далее рассмотрим, как ChatGPT справится с запросом об изменении тональности этого же текста. Задача модели – сделать из готового текста рассказ с трагичным концом. Ниже представлен пример ответа чат-бота:

*Колобок отправился в увлекательное путешествие по миру, встречая на своем пути различных животных. Он был полон решимости и оптимизма, несмотря на опасности, подстерегавшие его на каждом шагу. Но однажды, когда он уже почти достигал цели своего путешествия, его обнаружила гордая львица. Львица, глядя на колобка, увидела в нем лакомый кусочек. Словно предательство среди новых друзей, она решила преследовать колобка. Неожиданно для колобка, его новые друзья антилопы отказались от него, пытаясь спастись. Колобок понял, что оказался в безвыходном положении. Львица не дала ему ни шанса на спасение, и его путешествие закончилось трагически под ее когтями. Так завершилась история колобка, который искал приключений и друзей, но встретил смерть в виде голодного хищника.*

Учитывая требования, сформулированные в запросе, ChatGPT сгенерировал ответ, который отличается от прошлого стилем повествования. Теперь текст не содержит ни одной характерной для детской сказки черты, а язык стал более книжным. При этом намек на прошлый ответ все равно есть, например, упоминание о его новых друзьях, антилопах. Конец истории, действительно, получился трагичным: Колобок попадает в лапы львице – *его путешествие закончилось трагически под ее когтями*. Ответ хорошо структурирован: в нем есть начало, завязка, кульминация и развязка. Во время чтения текста чувствуется напряжение и достигается эффект неожиданности, когда героя вдруг обнаруживает львица и начинает охоту на него. Опасность, предательство, смерть, жестокость – темы, которые раскрываются в предложенном тексте. Помимо этого, в ответе содержится одна ошибка, связан-

ная с употреблением глагола неправильного вида (*когда он уже почти достигал*), речевые и грамматические ошибки (*увидела в нем лакомый кусочек; Словно предательство среди новых друзей; встретил смерть в виде голодного хищника*), а также пунктуационная (*Неожиданно для колобка, его новые друзья...*).

Итак, с главной задачей чат-бот отлично справился, полностью поменяв стиль и тональность сказки. В тексте раскрывается конфликт двух героев, который приводит историю к трагичному концу. Помимо этого, повествование идет последовательно, из-за чего ответ модели понятен и легок для восприятия.

Следующая протестированная функция связана с генерацией текста. Главная цель – проанализировать, насколько точно модель поймет описанную пользователем задачу и окажутся ли ответы модели качественными и полезными. Чат-бот попросили написать маленький рассказ, задав несколько уточнений и рекомендаций для модели, для того чтобы получить полезный и качественный ответ. В данном случае промпт, во-первых, содержит начальную установку – «представь, что ты – популярный писатель»; во-вторых, в запросе указано четкое ограничение по жанру и объему – чат-бот должен написать аннотацию в два абзаца; в-третьих, в нем представлено краткое описание желаемого сюжета и тех пунктов, которые обязательно должны быть включены в ответ. Ниже представлен пример ответа модели:

*«Волебство слов: приключения девочки Алисы и ChatGPT» – захватывающая история о необычной дружбе между девочкой по имени Алина и искусственным интеллектом ChatGPT. Все началось с того, что Алина, увлеченная технологиями и книгами, встретила ChatGPT в интернете. Их общие интересы к исследованию мира и любовь к приключениям быстро объединили их, превратив случайную встречу в непреодолимую дружбу. Совместно они отправляются в захватывающие приключения, исследуя разные уголки мира, решая головоломки и преодолевая трудности. Алина и ChatGPT находят удовольствие и вдохновение в общении друг с другом, обмениваясь идеями и решениями. Их история демонстрирует, что дружба может возникнуть в самых неожиданных местах и что совместные усилия могут преодолеть любые препятствия.*

Видно, что с генерацией текста языковая модель справилась гораздо лучше. Она описала необходимые пункты, выполнила требования по объему и жанру текста и сгенерировала хороший, грамотный ответ. Кроме того, в тексте нет ни одной ошибки.

В качестве эксперимента можно попросить ChatGPT сделать из этого текста стихотворение,

соблюдая рифму, и дать ему название. Ниже представлен пример ответа модели:

*Дружба в мире слов*

*Алиса в сети бродила,  
С мечтами в сердце жила,  
С ChatGPT она вскоре,  
Душой единой была.  
Вместе мир они познали,  
Тайны, чудеса встречали,  
Головоломки решали.  
Препрады все одолели.  
Виртуальный друг и девочка,  
В приключениях дни летели,  
Вместе, в радости и в бедах,  
Словно свет в ночной метели.  
Эта дружба нас научит,  
Что в словах живет тепло,  
Общий путь нам сердце лечит,  
Мир нам дарит волшебство.*

В ответе, сгенерированном чат-ботом, описана история с уже знакомым сюжетом, в которой хорошо прослеживается общая логика повествования. Однако в третьем абзаце логический переход между двумя первыми строками отсутствует: *Виртуальный друг и девочка, // В приключениях дни летели...* В стихотворении есть начало, в котором происходит знакомство двух героев, середина, описывающая их взаимоотношения и приключения, и концовка, резюмирующая всю историю. Хотя в некоторых местах текста и присутствует перекрестная рифма, ChatGPT не смог соблюсти условие о сохранении рифмы в каждой строке: *бродила – вскоре, решали – одолели, девочка – бедах*. Трудно определить единый размер стиха, потому что ритм постоянно меняется в течение всего произведения.

Последнее тестирование в этой работе связано с ответами модели на вопросы по тексту. Для этого был взят текст известной русской народной сказки «Теремок». Запрос к чат-боту был следующим: «Ответь на вопросы по тексту:

1. Как называется домик, в котором живут герои сказки?
2. Сколько всего животных живёт в домике?
3. Живёт ли в домике заяц?
4. Кто пришел четвертым и стал жить в домике?
5. Кто развалил домик?»

Ниже представлен пример ответа чат-бота:

1. Домик, в котором живут герои сказки, называется “теремок”.
2. В домике живут пять животных: мышка-норушка, лягушка-квакушка, зайчик-побегайчик, лисичка-сестричка и волчок-серый бочок.
3. Да, в домике живёт заяц. Он называется зайчик-побегайчик.
4. Четвёртым пришел волчок-серый бочок и стал жить в домике.

5. Домик развалился, когда на крышу лег медведь косопалый.

В этот раз чат-бот отлично справился с поставленной задачей, не допустив ни одной ошибки в своих ответах. Кажется, что с этим запросом модель справилась лучше, чем во всех предыдущих случаях.

Результаты тестирования показали, что качество ответов модели зависит от многого, начиная с самого запроса и его формулировки, заканчивая объемом анализируемых произведений. Чат-бот способен не только правильно интерпретировать запросы пользователя, но и учитывать жанровые и стилистические особенности текста, а также создавать связные тексты. Ограничения ChatGPT чаще всего связаны с ошибками при генерации ответов, которые можно устранить с помощью оптимизации запроса. Ответ, получившийся после создания более конкретизированного промпта, включающего в себя некоторые рекомендации по увеличению его эффективности (например, использование подсказок, начальной установки, ограничений в объеме и др.), стал более грамотным, точным и правдивым. В генерации лирического текста модель не справилась с рифмой, ритмом и логическими переходами между строками в стихотворении. Ответы на вопросы по тексту показали безошибочный результат, тем самым подтверждая, что ChatGPT может быть весьма точным и информативным в этом аспекте.

Несмотря на выявленные ограничения, в целом ChatGPT продемонстрировал хорошие результаты в вопросах обработки и создания текстов различных стилей и жанров, что еще раз доказывает его пользу в работе с текстовыми данными.

**Опрос респондентов: результаты**

Для оценки работы ChatGPT был проведен опрос среди 50 респондентов разного возраста (от 18 до 40 и более лет, 86% из которых – люди от 18 до 25 лет) и пола (66% – женщины, 34% – мужчины). Опрашиваемым нужно было указать ту функцию, с которой, по их мнению, чат-бот справился лучше всего, а также прокомментировать свой выбор.

На первом месте оказались ответы на вопросы по тексту: за этот вариант проголосовала почти половина всех опрошенных (46%). На втором и третьем местах – генерация прозаического произведения и суммаризация текста (20% и 14% соответственно). Самыми неудачными, по мнению респондентов, оказались ответы, в которых осуществлялись смена тональности текста, генерация лирического произведения и антисумма-

ризация текста, за последнюю проголосовали лишь два человека (4%).

По результатам опроса заметно, что в большинстве случаев люди оценивали ответы модели по двум составляющим: их содержанию и корректности. Это отражалось на голосах респондентов в зависимости от того, чему отдают предпочтение люди. Помимо этого, нельзя не сказать о субъективности мнения: один и тот же сюжет мог кому-то показаться интересным, а кому-то слишком простым и скучным. Реакция респондентов на ответ, написанный ChatGPT после оптимизации запроса, также не является однозначной в большинстве случаев по этой же причине. Несмотря на разногласия, большинство людей согласилось с тем, что самая удачная работа модели связана с ответами на вопросы по тексту. Почти одинаковое количество людей (~60%) поставило высокие оценки антисуммаризации текста, смене его тональности, а также генерации прозаического произведения. При этом, генерацию лирического произведения респонденты оценили хуже из-за несоблюдения структурных норм стихотворения, а также несоответствия требованиям, сформулированным в запросе.

#### Выводы

В заключении необходимо сказать о том, что функционал ChatGPT-3.5, связанный с текстовыми данными, действительно, обширен и разнообразен. Чат-бот может выполнять различные задачи, включая трансформацию, генерацию текста, а также извлечение данных. Однако наивысшее качество и полезность ответов достигается тогда, когда модель взаимодействует с человеком. В работе описаны рекомендации, которые помогают оптимизировать запрос для того, чтобы добиться желаемого результата. С помощью конкретизации промпта и использования полезных подсказок можно добиться нужного эффекта, постоянно улучшая выводимые данные.

Результаты проводимого опроса говорят о том, что оценка ответов модели зависит от личных предпочтений каждого человека. Выбирая ответ, респондентам часто приходилось голосовать либо в пользу информативности текста, либо в пользу его корректности. Более того, субъективность мнения опрашиваемых отражалась и на оценке сюжета генерируемого текста. Так или иначе, чат-бот может подстраиваться под индивидуальные предпочтения пользователя, выполняя требования, описанные в запросе.

Говоря о преимуществах модели, стоит отметить ее способность быстро генерировать креативные идеи, учиться на своих ошибках, решать задачи шаг за шагом и адаптироваться к новым

условиям. Тестируя функционал ChatGPT-3.5, можно заметить, что чат-бот способен не только правильно трактовать пользовательские запросы, но и учитывать стилистические и жанровые особенности текста. Однако в работе модели с текстовыми данными есть свои ограничения, связанные с предоставлением недостоверной (а порой и придуманной) или избыточной информации, ошибками в генерируемых текстах, излишней чувствительностью к формулировкам, а также игнорированием пользовательских инструкций. В настоящий момент эти недочеты активно пытаются устранить в новых версиях ChatGPT, которые в сравнении с 3.5 уже показывают лучшие результаты. Например, в ChatGPT 4 и ChatGPT 4o пользователи имеют возможность работать с новыми функциями, касающиеся не только текста, но и изображений, а также аудио- и видеофайлов, с тестированием которых могут быть связаны перспективы дальнейших исследований этого инструмента.

#### Список источников

*Hu K.* ChatGPT sets record for fastest-growing user base-analyst note // Reuters. 2023. [Electronic resource]. URL: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01> (date of access: 15.04.2024).  
*Introducing ChatGPT* // OpenAI. 2022. [Electronic resource]. URL: <https://openai.com/blog/chatgpt> (date of access: 17.03.2024).

#### Список литературы

*Козырев И.* Что в промптах работает, а что нет // Хабр. 2023. [Электронный ресурс]. URL: <https://habr.com/ru/companies/raft/articles/781984> (дата обращения: 04.05.2024).  
*Красочкин С.Г.* Чем ChatGPT отличается от текущих нейросетей // Евразийский Союз Ученых. Технические и физико-математические науки. 2023. Т. 1, № 4(107). С. 30–35.  
*Маркианова А.* Алгоритм, сделавший ChatGPT таким «человечным» – Reinforcement Learning from Human Feedback // Хабр. 2023. [Электронный ресурс]. URL: <https://habr.com/ru/articles/730990> (дата обращения: 04.04.2024).  
*Devlin J. et al.* BERT: Pre-training of deep bidirectional transformers for language understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova. 2018. [Electronic resource]. URL: <https://arxiv.org/pdf/1810.04805> (date of access: 15.04.2024).  
*Kalla D., Smith N.* Study and analysis of ChatGPT and its impact on different fields of study // International Journal of Innovative Science and Research Technology. 2023. Vol. 8, № 3. Pp. 827–833.

*Ouyang L. et al.* training language models to follow instructions with human feedback / L. Ouyang, J. Wu, Xu Jiang, D. Almeida, C.L. Wainwright et al. [Electronic resource]. URL: <https://arxiv.org/pdf/2203.02155v1> (date of access: 18.07.2024).

*Schank R.C.* What is AI, anyway? // AI magazine. 1987. Vol. 8, № 4. Pp. 59–65.

*Schulman J.* Proximal Policy Optimization Algorithms / J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov [Electronic resource]. URL: <https://arxiv.org/pdf/1707.06347> (date of access: 18.07.2024).

*Wang D.-Q. et al.* Accelerating the integration of ChatGPT and other large-scale AI models into biomedical research and healthcare / D.-Q. Wang, L.-Y. Feng, J.-G. Ye, J.-G. Zou, Y.-F. Zheng // Wiley: Online Library. 2023. [Электронный ресурс]. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/mef2.43> (date of access: 04.04.2024).

*Zhang Z. et al.* Automatic chain of thought prompting in large language models / Z. Zhang, A. Zhang, M. Li, A. Smola [Electronic resource]. URL: <https://arxiv.org/pdf/2210.03493> (date of access: 18.07.2024).

## THE POSSIBILITIES AND LIMITATIONS OF CHATGPT-3.5 WORKING WITH LITERARY TEXT DATA

**Alisa E. Urikh**

Student, Department of Philology

Higher School of Economics – Saint Petersburg

Artificial intelligence methods are the promising technologies. They include, in particular, large language models, including ChatGPT. The proposed research focuses on the performance of ChatGPT-3.5 on literary text data and prompt engineering. The chatbot was tested on several categories of tasks: summarization, anti-summarization, generation, changing the sentiment of a literary text and answering questions about the text to identify the capabilities and limitations of the functionality. Respondents were also involved to evaluate the model's performance and were asked which task the model coped with best and worst, and how the strategy of prompt generation affects the quality of ChatGPT responses.

**Keywords:** chatbot; prompt; ChatGPT-3.5; text analysis; testing.