

Составление альтернативного текста к изображениям с использованием больших языковых моделей

Е. А. Косова*, К. И. Редкокош

ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского», г. Симферополь, Российская Федерация

Адрес: 295007, Российская Федерация, Республика Крым, г. Симферополь, просп. Академика Вернадского, д. 4

*lynx99@inbox.ru

Аннотация

Введение. В статье рассматриваются вопросы развития компетенций цифровой доступности. Целью работы является разработка и проверка подхода к составлению альтернативных текстов к изображениям с использованием больших языковых моделей (англ. *Large Language Models, LLM*), реализованных в чат-ботах, на основе ключевых слов и словосочетаний, характеризующих изображение.

Материалы и методы. Исследование выполнялось в апреле-июне 2024 года. В качестве изображений были выбраны 12 сюжетных фотографий и логотипов, содержащих текст; в качестве инструментов – чат-боты Copilot, Gemini, YandexGPT. На основании предварительно составленных перечней ключевых слов на русском языке чат-ботам было предложено сгенерировать описания изображений, а затем сократить полученные описания до 250 и 125 знаков. Готовые текстовые альтернативы оценивали по пяти шкалам: полезность, полнота, достоверность, точность, грамотность. Для анализа данных использовали программу Microsoft Excel®, язык программирования R, статистические методы и критерии (корреляцию Пирсона, многомерные регрессии). Исследование применимости LLM для составления альтернативных текстов проведено впервые.

Результаты исследования. Полученные данные свидетельствуют о высоком качестве сгенерированных текстовых описаний. По заданному набору ключевых слов и словосочетаний LLM способны оперативно составлять достаточно полные, достоверные, точные и грамотные текстовые альтернативы, применимые для людей, не имеющих возможности видеть контент. Установлены переменные, являющиеся предикторами к ухудшению качества сгенерированных альтернативных текстов, в частности: введение лимитов на длину описаний, излишняя длина описаний, использование чат-бота Gemini. В значительной доле сгенерированных описаний зарегистрированы фантазмы (избыточные элементы, содержащие недостоверную информацию, не заданную в ключевых словах), которые были классифицированы по типам: украшательства, умозрительности, псевдоэмоции, инфошум, дезинформация.

Обсуждение и заключение. Предложенный подход может быть использован в обучении и работе специалистов по цифровой доступности. Представляется перспективным продолжить исследования по применению LLM для разработки альтернативных текстов на разных языках к изображениям разных типов.

Ключевые слова: цифровая доступность, ИТ-компетенции, профессиональная подготовка, альтернативный текст, оценка качества, большие языковые модели, чат-боты

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Косова Е. А., Редкокош К. И. Составление альтернативного текста к изображениям с использованием больших языковых моделей // Современные информационные технологии и ИТ-образование. 2024. Т. 20, № 3. С. 699-714. <https://doi.org/10.25559/SITITO.020.202403.699-714>

© Косова Е. А., Редкокош К. И., 2024



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Composing Alternative Text for Images Using Large Language Models

Ye. A. Kosova*, K. I. Redkokosh

V.I. Vernadsky Crimean Federal University, Simferopol, Russian Federation

Address: 4 Prospekt Vernadskogo, Simferopol 295007, Republic of Crimea, Russian Federation

*lynx99@inbox.ru

Abstract

Introduction. The article is devoted to the development of digital accessibility competencies. The goal of the study is to develop and test a methodology for generating alternative texts for images using large language models implemented in chatbots based on a set of keywords and phrases that describe the image.

Materials and Methods. The study was carried out in April-June 2024. Twelve story photos and logos containing text were selected as images. The main tools used were Copilot, Gemini, and YandexGPT. Using pre-compiled keyword lists in Russian, the chatbots generated image descriptions and then reduced them to 250 and 125 characters, respectively. The resulting text alternatives were then evaluated using five scales for usefulness, completeness, reliability, accuracy, and literacy. Microsoft Excel®, the R programming language, and statistical methods (Pearson correlation and multidimensional regression) were used to analyze the data. The study of the potential of LLMs to optimize the generation of alternative text was conducted for the first time.

Results. The results indicate a high quality of text generated by chatbots. According to a set of given keywords and phrases, LLMs were able to quickly generate sufficiently complete, reliable, accurate, and literate alternative texts that could be used by people who are unable to see the content. Variables that were identified as predictors of a decrease in the quality of the generated alternative texts included: entering limits on description length, excessive description length, and the use of the Gemini chatbot. A significant proportion of the generated descriptions registered phantasms (redundant elements containing unreliable information not specified in the keywords), which were classified by type: embellishments, speculations, pseudo-emotions, info-noise, disinformation.

Discussion and Conclusion. The proposed approach could be incorporated into the training and practice of specialists in digital accessibility. It seems promising to continue researching the application of large language models (LLMs) to generate alternative texts in different languages for images of various types.

Keywords: digital accessibility, IT competencies, professional training, alternative text, quality assessment, large language models, chatbots

Conflict of interests: The authors declares no conflict of interest.

For citation: Kosova Ye.A., Redkokosh K.I. Composing Alternative Text for Images Using Large Language Models. *Modern Information Technologies and IT-Education*. 2024;20(3):699-714. <https://doi.org/10.25559/SITITO.020.202403.699-714>



Введение

Современные цифровые ресурсы (в том числе веб-сайты, мобильные и стационарные приложения, электронные документы) должны обладать свойствами цифровой доступности, то есть быть пригодными к использованию любым человеком в соответствии с его персональными возможностями, потребностями и предпочтениями [1]. Одним из требований цифровой доступности является предоставление альтернативного текста (англ. *alt-text*) к визуальному контенту¹. Альтернативный текст необходим пользователям, которые не имеют возможности видеть изображение, в первую очередь, людям с глубокими нарушениями зрения (слепотой и слабовидением) [2-4]. По данным Всемирного Союза Слепых (англ. *World Blind Union, WBU*) таких людей насчитывается более четверти миллиарда² или около 3% человеческой популяции.

Оптимизация разработки текстовых описаний к изображениям является предметом многочисленных научных исследований. Для автоматизации создания альтернативного текста используются программные инструменты и методы, такие как: глубокое обучение на основе нейронных сетей, обработка естественного языка, компьютерное зрение, обратный поиск изображений для извлечения имеющихся текстовых описаний [5-10]. При этом известно, что текстовые альтернативы, созданные человеком или при его участии, качественно превосходят альтернативные описания, выполненные полностью автоматически [9-11]. Вместе с тем вопросы улучшения качества текстовых описаний, созданных человеком, проектирования руководящих принципов для помощи составителям и новых оптимизирующих алгоритмов создания текстовых альтернатив к изображениям разных типов остаются актуальными [9-14].

В последнее время наблюдается прорыв в разработке больших языковых моделей (англ. *Large Language Models, LLM*) – вычислительных инструментов, обладающих способностями достигать универсальной языковой генерации и решать другие задачи обработки естественного человеческого языка³. LLM позволяют повысить эффективность работы человека с текстами, с высокой точностью выполняя поиск информации и анализ текстовых данных, отвечая на вопросы пользователя, генерируя идеи, истории и сценарии, предлагая макеты текстов на естественном языке, обобщая и рецензируя информацию [15-18]. Считается, что LLM могут быть применимы в образовании и обучении, в том числе для развития языковых навыков [16], [18].

В данном исследовании мы выдвигаем **гипотезу** о том, что использование LLM в качестве вспомогательного инструмента для составления текстовых описаний позволит оптимизи-

ровать работу составителя альтернативных текстов, улучшая черновые текстовые описания за счет интеллектуальных алгоритмов обработки естественного языка.

Целью работы является разработка и проверка методического подхода к составлению альтернативных текстовых описаний к изображениям с использованием LLM, реализованных в чат-ботах, на основе заданного составителем набора ключевых слов и словосочетаний, характеризующих изображение.

Постановка задачи

Под альтернативным текстом понимают текстовый эквивалент для нетекстового цифрового контента – статических изображений, аудио- и видеоматериалов, интерактивных объектов, элементов управления, тестов идентификации пользователей, симуляций⁴. Термины «альтернативный текст», «текстовый эквивалент», «текстовая альтернатива», «текстовое описание» используются в литературе как синонимы [11]. Согласно положению «1.1. Текстовые альтернативы» Руководящих принципов доступности веб-контента⁵ любая нетекстовая информация должна иметь текстовый эквивалент для обеспечения возможности полноценного использования цифрового ресурса всеми пользователями, в том числе людьми с ограниченными возможностями здоровья. Текстовые описания могут быть прослушаны пользователем при помощи программ чтения с экрана или восприняты тактильно с использованием дисплеев Брайля.

Текстовые альтернативы к изображениям могут быть добавлены в документ или на веб-страницу программно как замещающий текст. В этом случае альтернативное описание не отображается в режиме пользовательского просмотра, но задано в коде или настройках цифрового ресурса. Замещающий текст становится видимым при отключении изображений и улучшает индексацию веб-страниц в поисковых системах⁶. Одним из требований к замещающему тексту является его краткость (рекомендованная длина – до 125 знаков, в крайнем случае, не более 250 знаков [19])⁷. Если характер изображения и контекст его использования требуют детального описания (например, изображения сложной структуры или музейные экспонаты), соответствующая текстовая альтернатива может быть размещена непосредственно под изображением или в связанном файле, доступном по контекстной ссылке. При таком варианте ограничения на длину описания могут отсутствовать.

Статические изображения, размещенные в электронных ресурсах, подразделяются на типы: информативные (картина, рисунок, фотография) – передают сюжет или концепцию; де-

¹ Web Content Accessibility Guidelines (WCAG) 2.2 ; ed. by A. Campbell, C. Adams, R. B. Montgomery, M. Cooper, A. Kirkpatrick. Cambridge, Massachusetts, USA: World Wide Web Consortium (W3C), 2023 [Электронный ресурс]. URL: <https://www.w3.org/TR/WCAG22/> (дата обращения: 16.05.2024).

² World Blind Union : офиц. сайт [Электронный ресурс]. URL: <https://worldblindunion.org/> (дата обращения: 16.05.2024).

³ Large language model [Электронный ресурс] // Wikipedia, 2024. URL: https://en.wikipedia.org/wiki/Large_language_model (дата обращения: 16.05.2024).

⁴ Web Content Accessibility Guidelines (WCAG) 2.2 ; ed. by A. Campbell, C. Adams, R. B. Montgomery, M. Cooper, A. Kirkpatrick. Cambridge, Massachusetts, USA: World Wide Web Consortium (W3C), 2023 [Электронный ресурс]. URL: <https://www.w3.org/TR/WCAG22/> (дата обращения: 16.05.2024).

⁵ Там же.

⁶ Images Tutorial [Электронный ресурс] // World Wide Web Consortium (W3C), 2022. URL: <https://www.w3.org/WAI/tutorials/images/> (дата обращения: 16.05.2024).

⁷ Выбор ограничений обусловлен возможностями программ чтения с экрана, которые способны, сканируя документ, зачитывать вслух до 125, реже – до 250 знаков с пробелами [19].



коративные (пиктограмма, фон) – используются для оформления контента; функциональные (ссылка, кнопка) – сообщают о сути события, которое следует за нажатием; изображения текста (логотип) – включают текст, предназначенный для чтения; сложные (таблицы, диаграммы, инфографика) – требуют структурированного и подробного описания; группы изображений (шкала рейтинга) – передают одну и ту же информацию; карты изображений (организационные диаграммы, географические карты) – совмещают в одном изображении несколько областей, каждая из которых обладает собственными функциями⁶. Наше исследование направлено на изучение возможностей оптимизации создания текстовых альтернатив для информативных изображений (на примере фотографий) и изображений, содержащих текст (на примере логотипов). Согласно рекомендациям по разработке альтернативного текста, представленным в онлайн-руководствах^{6,8} и научных публикациях [6], [9], текстовые описания к информативным изображениям должны обладать свойствами лаконичности, точности, грамотности, полноты и естественности языка. Для изображений, содержащих текст, добавляется требование о необходимости включать в описание тот же текст⁶.

Создание текстовых альтернатив к цифровым изображениям в Российской Федерации регулируется стандартом тифлокомментирования⁹, а обязательность предоставления альтернативного текста для нетекстового контента – стандартом доступности информации, представленной в электронно-цифровой форме¹⁰.

Наличие альтернативного текста не гарантирует его качества. Например, в работах [20, 21] отмечена недостаточная адекватность существующих текстовых описаний к изображениям. Таким образом, разработка эффективных алгоритмов составления, проверки и коррекции альтернативных текстовых описаний к нетекстовому контенту относится к актуальным задачам. В соответствии с целью и гипотезой исследования сформулированы следующие **исследовательские вопросы**:

- Может ли LLM на основе ключевых слов составить полезное, полное, достоверное, точное и грамотное описание изображения, не «видя» оригинала?
- Как можно охарактеризовать качество текстовых описаний, созданных LLM по заданным ключевым словам? Какие факторы могут повлиять на качество?
- Какие инструкции необходимо предоставить составителю альтернативных текстов для правильного использования инструментов LLM в своей работе?

Материалы и методы

Исследование выполнялось в апреле-июне 2024 года в три этапа:

- подготовка дизайна исследования, отбор инструментов

для выполнения описаний и изображений, подлежащих описанию;

- составление описаний с помощью чат-ботов;
- оценка качества выполненных описаний и статистический анализ полученных данных.

В качестве инструментов для составления текстовых описаний использовали три чат-бота, поддерживающих русский язык и доступных в виде веб-приложений: Copilot (разработчик – Microsoft® (г. Редмонд, США), запуск – 7.02.2023, LLM – Microsoft Prometheus на основе OpenAI GPT-4); Gemini (разработчик – Google® (г. Маунтин-Вью, США), запуск – 21.03.2023, LLM – Gemini); YandexGPT (разработчик – Яндекс® (г. Москва, Россия), запуск – 17.05.2023, LLM – YandexGPT на основе семейства YaLM).

Для составления текстовых описаний были выбраны 12 статических изображений – шесть сюжетных фотографий и шесть изображений-логотипов, содержащих текст (Приложение А). Текстовые описания к изображениям составлялись согласно следующему алгоритму:

1. Изучить изображение.
2. Составить список из 10-20 ключевых слов и/или словосочетаний, соответствующих содержанию изображения. Для сюжетных изображений (фото) в ключевых словах учитывать: цветность изображения; изображенное место; характеристики заднего и переднего плана; характеристики персонажей (внешний вид, пол, возраст, размещение, эмоциональное состояние, поведение); детали, необходимые для понимания сюжета. Для изображений с текстом (логотипов) принимать во внимание: цвет и фон изображения; символы по отдельности и текст в целом; стиль шрифта; графические объекты; соотношение (взаимное расположение) текста и графических объектов.
3. Сформулировать первый запрос чат-боту: «Составьте текстовое описание по ключевым словам:» (далее перечислить ключевые слова через запятую).
4. Сформулировать второй запрос чат-боту: «Сократите описание до 250 знаков с пробелами, сохраняя смысл».
5. Сформулировать третий запрос чат-боту: «Сократите описание до 125 знаков с пробелами, сохраняя смысл».
6. Для каждого сгенерированного описания создать запись в наборе данных с полями: чат-бот, номер изображения, тип изображения (фото или логотип), список ключевых слов, длина ключевых слов, количество ключевых слов, время составления ключевых слов, текст описания, требуемая длина описания, фактическая длина описания, время генерирования описания.
7. Удалить файлы Cookie, историю браузера и историю поиска.
8. Выполнить пункты 1-7 для остальных изображений.
9. Выполнить пункты 3-8 для всех изображений, используя остальные чат-боты.

⁸ Alternative Text [Электронный ресурс] // WebAIM, 2024. URL: <https://webaim.org/techniques/alttext/> (дата обращения: 16.05.2024).

⁹ ГОСТ Р 57891-2022 Тифлокомментирование и тифлокомментарии. Термины и определения : национальный стандарт Российской Федерации: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 24 января 2022 г. № 36-ст: введен впервые: дата введения 2022-04-01. М.: ФГБУ «РСТ», 2022.

¹⁰ ГОСТ Р 52872-2019 Интернет-ресурсы и другая информация, представленная в электронно-цифровой форме. Приложения для стационарных и мобильных устройств, иные пользовательские интерфейсы. Требования доступности для людей с инвалидностью и других лиц с ограничениями жизнедеятельности : национальный стандарт Российской Федерации: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 29 августа 2019 г. № 589-ст: дата введения 2020-04-01. М.: Стандартинформ, 2019.



На составление списков ключевых слов было затрачено от 114 до 480 секунд. Для измерения времени, затраченного на составление ключевых слов и генерирование описаний, использовали секундомер.

Для выполнения оценок текстовых описаний использовали критерии, ранжированные по шкалам типа Likert [22]:

- *полезность* – применимость для людей, не имеющих возможности видеть изображение (от 1 до 5, где 1 – абсолютно бесполезное; 5 – в наивысшей степени полезное);
- *достоверность* – соответствие истинному положению дел, правдивость (от 1 до 5, где 1 – абсолютно недостоверное, вымышленное или выглядит таковым; 5 – максимально достоверное, правдивое или выглядит таковым);
- *полнота* – достаточность для понимания изображения (от 1 до 5, где 1 – абсолютно неполное, недостаточное для минимального понимания; 5 – максимально полное и завершенное, не требует дополнительных деталей);
- *точность* – ясность и доходчивость передачи формы и содержания изображения (от 1 до 5, где 1 – абсолютно неточное; 5 – максимально точное);
- *грамотность* – соответствие правилам и нормам языка (от 1 до 5, где 1 – абсолютно безграмотное; 5 – грамотное, без ошибок).

Помимо оценивания качества описаний предусматривалась регистрация в сгенерированном тексте фантазмов – избыточных элементов описания, содержащих недостоверную (ложную, воображаемую, вводящую в заблуждение) информацию, не заданную в ключевых словах.

После составления описаний два эксперта независимо друг от друга выполнили предварительную (тестовую) оценку 15 описаний для пяти изображений, сравнили между собой полученные результаты и утвердили методику оценивания. Затем были оценены остальные описания, а полученные оценки внесены в набор данных.

Для анализа данных использовали программу Microsoft Excel® (Microsoft, г. Редмонд, США), язык программирования R, статистические методы и критерии: для определения силы и направленности связи между переменными – корреляцию Пирсона, многомерную регрессию (линейную, порядковую и логистическую). Результаты статистического анализа считались значимыми при значении $p < 0,05$.

Результаты исследования, включая сгенерированные описания и их оценки, опубликованы в виде набора данных в репозитории Mendeley Data¹¹.

Результаты исследования

Длина сгенерированных описаний

В процессе работы получено 108 текстовых описаний на русском языке для 12 изображений.

В таблице 1 приведены статистические данные о длинах сге-

нерированных текстовых описаний в целом и по конкретным чат-ботам. Описания имели широкий разброс по длине – до 1335 знаков с пробелами между самым коротким и самым длинным описанием. Чат-боты были предрасположены к генерированию развернутых описаний, если не имели ограничительных условий (длина самого короткого описания, созданного по запросу без лимитирующих условий, оказалась больше, чем среднее значение длины по всем описаниям). Чат-боты, как правило, соблюдали содержащиеся в запросах ограничения на длину: так, 54 описания (75,0 % от всех описаний с заданным лимитом) соответствовали сформулированным в запросах ограничениям. При этом превышение длины чаще фиксировалось для описаний, ограниченных 125 знаками с пробелами.

По конкретным чат-ботам получены следующие данные о превышении длины в сгенерированных описаниях: Copilot – на один знак в среднем для 4,0 % от всех описаний, сгенерированных этим чат-ботом по запросам, содержащим ограничения по длине; Gemini – на 89 знаков в среднем для 37,5 % описаний; YandexGPT – на 43 знака в среднем для 33,0 % описаний. Следовательно, Copilot превосходил остальные чат-боты по точности выполнения запроса на ограничение длины, а Gemini, напротив, чаще прочих не соблюдал заданные ограничения.

Время генерирования описаний

Время генерирования описаний составило от 1 до 68 секунд (в среднем $a=10,49$, $\sigma=12,74$). При этом чат-бот Copilot затрачивал значительно больше времени на создание описаний ($a=22,94$, $\sigma=15,78$), чем чат-боты Gemini ($a=5,25$, $\sigma=1,81$, $p < 0,05$) и YandexGPT ($a=3,28$, $\sigma=1,63$, $p < 0,05$).

Для уточнения связей между временем генерирования описаний и прочими факторами (требуемой длиной описаний, типом изображения, используемым чат-ботом) построена модель линейной регрессии, в результате обучения которой установлены переменные, значимо влияющие на скорость создания текстовых описаний. Ими оказались: чат-бот YandexGPT (коэффициент в модели $-19,67$; $p=2,0 \cdot 10^{-16}$); чат-бот Gemini ($-17,69$; $p=4,12 \cdot 10^{-16}$); требуемая длина до 125 знаков с пробелами ($-11,77$; $p=3,89 \cdot 10^{-9}$); требуемая длина до 250 знаков с пробелами ($-9,41$; $p=1,26 \cdot 10^{-6}$). Переменные, характеризующие тип изображения (логотип или фотография), оказались незначимыми ($p > 0,05$). Отрицательные коэффициенты в регрессионной модели свидетельствовали о наличии обратной связи между переменными-факторами и результирующей переменной. Интерпретировать полученные результаты можно следующим образом: вне зависимости от типа изображения чат-бот Gemini генерировал описания быстрее, чем Copilot, а YandexGPT – еще быстрее, при этом скорость генерирования возрастала при введении в запрос ограничений на длину – чем короче требовалось описание, тем быстрее происходило генерирование.

¹¹ Kosova Ye., Redkokosh K. Composing alt text using large language models: dataset in Russian [Электронный ресурс] // Mendeley Data, 2024. V1. <https://doi.org/10.17632/73dptbyxb.1>



Т а б л и ц а 1. Длина текстовых описаний, сгенерированных чат-ботами по ключевым словам
T a b l e 1. Length of text descriptions generated by chatbots based on keywords

| Чат-боты и статистические характеристики | Фактическая длина текстовых описаний (в знаках с пробелами) в соответствии с заданными лимитами длины | | |
|--|---|---------------------------|---------------------------|
| | Без ограничений длины | До 250 знаков с пробелами | До 125 знаков с пробелами |
| Copilot | | | |
| Среднее значение, $a \pm \sigma$ (n) | 647,3 \pm 231,8 (12) | 184,7 \pm 22,8 (12) | 103,4 \pm 13,2 (12) |
| Min – max | 386 – 1047 | 157 – 232 | 90 – 126 |
| Gemini | | | |
| Среднее значение, $a \pm \sigma$ (n) | 950,8 \pm 301,1 (12) | 266,3 \pm 79,9 (12) | 146,6 \pm 83,0 (12) |
| Min – max | 484 – 1408 | 169 – 490 | 84 – 354 |
| YandexGPT | | | |
| Среднее значение, $a \pm \sigma$ (n) | 532,5 \pm 173,2 (12) | 209,1 \pm 38,1 (12) | 142,0 \pm 42,8 (12) |
| Min – max | 370 – 936 | 129 – 275 | 73 – 219 |
| В целом по всем чат-ботам | | | |
| Среднее значение, $a \pm \sigma$ (n) | 710,19 \pm 294,67 (36) | 220,03 \pm 61,88 (36) | 130,64 \pm 56,39 (36) |
| Min – max | 370 – 1408 | 129 – 490 | 73 – 354 |

Примечания: a – среднее арифметическое, σ – стандартное отклонение, n – количество описаний, min – минимум, max – максимум
Notes: a – arithmetic mean, σ – standard deviation, n – number of descriptions, min – minimum, max – maximum

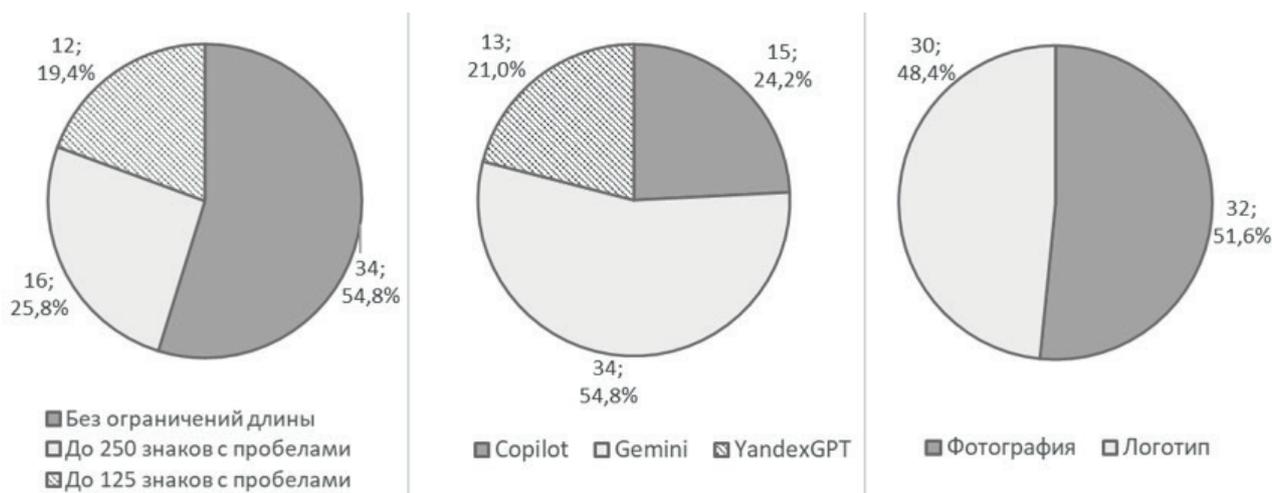
Источник: здесь и далее в статье все таблицы и рисунки составлены авторами.
Source: Hereinafter in this article all tables and figures were made by the authors.

Фантазмы в сгенерированных описаниях

Под фантазмами в этой работе мы понимали любые элементы сгенерированного текстового описания, включающие недостоверную (ложную или воображаемую) информацию, не содержащуюся в ключевых словах, вводящую в заблуждение пользователя или навязывающую ему свое видение и понима-

ние изображения.

Фантазмы присутствовали в 62 текстовых описаниях, что составляет 57,4 % от всех сгенерированных чат-ботами текстов. Общие данные о количестве текстовых описаний с фантазмами (в зависимости от ограничений длины, используемого чат-бота и типа изображения) приведены на рисунке 1.



Р и с. 1. Количество текстовых описаний с фантазмами, сгенерированными чат-ботами, $N = 62$. Слева – частотное распределение в зависимости от требуемой длины описаний, в центре – в соответствии с использованным чат-ботом, справа – по типу изображения.

Fig. 1. The number of textual descriptions with fantasies generated by chatbots is $N = 62$. On the left, you can see the frequency distribution based on the required length of the descriptions. In the center, it is shown according to the chatbot used. On the right, it is based on the type of image

Из данных описательной статистики (см. рисунок 1) можно предположить следующее: генерирование фантазмов в той или иной степени свойственно всем чат-ботам, особенно Gemini; искажения достоверности встречаются чаще в нелитерированных по длине описаниях, чем в сокращенных текстах; описания логотипов и фотографий с фантазмами регистрируются с примерно одинаковой частотой. Для проверки выдвинутых предположений и уточнения связи между наличием описаний с фантазмами и переменными-факторами, которые, предположительно, могли оказать влияние на появление недостоверной информации, была построена модель логистической регрессии. В качестве независимых переменных были определены: требуемая длина описаний, используемый чат-бот, время генерирования описаний, тип изображения. В результате обучения модели обнаружен единственный значимый фактор, влияющий на появление описаний с фантазмами, – чат-бот Gemini (коэффициент в модели 2,28; $p=5,5 \cdot 10^{-3}$). Таким образом, Gemini чаще предрасположен к генерированию описаний с фантазмами, чем YandexGPT и Copilot. Тип изображения, ограничения описаний по длине и время генерирования описаний значимо не влияли на появление фантазмов в текстах ($p>0,05$).

Анализ текстовых описаний и содержащейся в них недостоверной информации позволил предложить следующую классификацию фантазмов:

1. Украшательства (встретились в 26 описаниях, что составило 24,1 % от общего количества сгенерированных текстов) – вычурные фразы, эпитеты и метафоры, перегружающие описание. Например: «... на фоне пушистого снега, словно из морозной дымки, возвышается величественная ель... в центре композиции, словно добрая фея, царит воспитательница...» (из описания фото «Снежный душ»¹² по ключевым словам «черно-белое фото, зима, ель под снегом, группа детей из детского сада, дети и воспитательница под елью, теплые шубы и шапки, снег падает с ели, жмурятся, улыбаются, протягивают руки под снег, радостно»);
2. Умозрительности (57; 52,8 %) – оценочные суждения и выводы относительно содержания изображения и отдельных его элементов. Например: «... придает ему яркость и насыщенность... создает эффект глубины и объема... придает надписи уникальный и запоминающийся вид...» (из описания логотипа Chupa Chups S.A.U.¹³ по ключевым словам «плоский цветок, стилизованный, восемь лепестков, заливка цветка желтая, контур цветка тройной, внутренний контур белый, средний контур красный, внешний контур золотистый, внутри цветка надпись, надпись на латинице «Chupa Chups», второе слово под первым, надпись красная, в нижнем слове буква «р» имеет хвост в виде петли»);
3. Псевдоэмоции (23; 21,3 %) – вымышленные описания настроения, чувственных и эмоциональных переживаний, якобы вызванных изображением. Например: «... изображение пора-

жает своим масштабом и красотой, напоминая о великолепии природы и хрупкости жизни» (из описания фото «Saving the Monarchs»¹⁴ по ключевым словам «цветное фото, живой ковер, бабочки монархи, тысячи бабочек, сидят плотно друг к другу, крылья сложены, оборотная сторона крыльев бледно-желтая, одна бабочка в центре, крылья расправлены, ярко-желтый цвет, темно-коричневые полосы, россыпь белых точек на кончиках крыльев»);

4. Инфошум (12; 11,1 %) – избыточные элементы, не имеющие прямого отношения к содержанию изображения и засоряющие описание. Например: «... используется на одежде, обуви, аксессуарах и других товарах бренда... также используется в рекламных материалах и других коммуникациях бренда...» (из описания логотипа Lacoste S.A.¹⁵ по ключевым словам «надпись на латинице «LACOSTE», прописные буквы, черный цвет шрифта, зеленый крокодил, в профиль, повернут мордой вправо, пасть раскрыта, пасть красная, хвост крокодила задран вверх, спина крокодила покрыта белыми точками, крокодил над надписью»);

5. Дезинформация (17; 15,7 %) – ложные (искаженные или придуманные) факты. Например: «... множество диагональных рядов, искусно пересекающих друг друга... решительные взгляды, направленные ввысь...» (из описания фото «Олимпиада-80»¹² по ключевым словам «черно-белое фото, стадион, гимнастки, выстроены в несколько рядов, по диагонали, ряды параллельны друг другу, больше десяти рядов, занимают всю площадь фотографии, каждая гимнастка держит обруч, руки вытянуты вверх, короткие спортивные платья»).

В текстовом описании могло встретиться несколько типов фантазмов – от одного (18 описаний или 29,03 % от общего количества текстов с фантазмами) до пяти (1; 1,61 %). Статистические данные о регистрации фантазмов в описаниях приведены в приложении В. Чаще прочих в текстах одновременно встречались два типа фантазмов (24; 38,71 %), с преобладанием комбинаций: «умозрительности» и «инфошум» – 9 (14,52 %); «украшательства» и «умозрительности» – 7 (11,29 %). В описаниях, имеющих фантазмы трех типов, как правило, сочетались «украшательства», «умозрительности» и «псевдоэмоции» (7; 11,29 %). Ни одно из описаний, имеющих фантазмы четырех типов (7; 11,29 %), не содержало «инфошум». Вообще, фантазмы типа «инфошум» регистрировались только в текстах, сгенерированных Gemini, в то время как остальные типы фантазмов не были связаны с конкретным чат-ботом и могли появляться в описаниях с разными требованиями к длине, как для логотипов, так и для фотографий. Большинство описаний, имеющих четыре типа фантазмов, и единственное описание, включающее все пять типов, являлись результатом работы Gemini.

¹² Советские послевоенные фотографии: от аполитичной искренности к сюрреализму [Электронный ресурс] // Cameralabs.org, 2024. URL: <https://cameralabs.org/aeon/sovetskie-poslevoennye-fotografy/> (дата обращения: 16.05.2024).

¹³ Chupa Chups [Электронный ресурс] // Wikipedia, 2024. URL: https://en.wikipedia.org/wiki/Chupa_Chups (дата обращения: 16.05.2024).

¹⁴ Saving the Monarchs: 2024 Photo Contest, North and Central America, Stories [Электронный ресурс] // World Press Photo, 2024. URL: <https://www.worldpressphoto.org/collection/photo-contest/2024/Jaime-Rojo/1> (дата обращения: 16.05.2024).

¹⁵ Lacoste [Электронный ресурс] // Wikipedia, 2024. URL: <https://en.wikipedia.org/wiki/Lacoste> (дата обращения: 16.05.2024).



Качество сгенерированных описаний

В целом качество тестовых описаний, сгенерированных чат-ботами, оказалось выше среднего (Таблица 2). Наиболее высокие оценки получены по шкале «грамотность» (в среднем от 4,5 до 5,0 баллов). Точность описаний варьировала от 2,5 до 4,3 баллов в среднем, достоверность – от 3,5 до 4,6, полезность – от 2,7 до 4,6. Самые низкие оценки и, одновременно, наиболее широкий разброс баллов зарегистрированы по шкале «полнота» (от 2,2 до 4,9 баллов в среднем).

Из данных описательной статистики можно сделать предположения о снижении качества описаний по шкалам «полезность», «полнота» и «точность» при появлении в запросах ограничений на длину текста (чем жестче ограничение, тем ниже оценка). Для шкал «достоверность» и «грамотность» подобная особенность не замечена. По всей видимости, при заданном ступенчатом сокращении длины сгенерированные описания поступательно утрачивали свойства полезности, теряли в точности изложения и полноте передачи сути изображения, однако оставались на прежнем уровне правдивости и в

той же мере соответствовали правилам русского языка.

Выдвинутые предположения требуют статистических подтверждений. Для определения связи между качеством описаний и переменными-факторами, потенциально способными повлиять на оценку, была построена серия моделей порядковой регрессии. К независимым переменным отнесли: фактическую длину описаний, требуемую длину описаний, используемый чат-бот, количество ключевых слов, длину ключевых слов, время генерирования описаний, тип изображения, наличие фантазмов в описаниях. В роли зависимых переменных выступали параметры качества по пяти шкалам. В результате получены четыре регрессионные модели, соответствующие параметрам качества «полезность», «достоверность», «полнота», «точность» и содержащие переменные-факторы, значимо влияющие на результирующую переменную (Таблица 3). Для параметра качества «грамотность» регрессионную модель построить не удалось, то есть поведение независимых переменных-факторов никак не объясняло поведения результирующей переменной, определяющей правильность использования русского языка в описаниях.

Таблица 2. Оценка качества текстовых описаний, сгенерированных чат-ботами по ключевым словам
Table 2. Quality assessment of chatbot-generated descriptions using keyword analysis

| Оцениваемые параметры качества | Чат-бот | Средняя оценка качества текстовых описаний в соответствии с фактической длиной, $a \pm \sigma$ (n) | | |
|--------------------------------|----------------|--|---------------------------|---------------------------|
| | | Без ограничений длины | До 250 знаков с пробелами | До 125 знаков с пробелами |
| Полезность | Copilot | 4,6±0,7 (12) | 4,1±0,5 (12) | 3,2±0,4 (12) |
| | Gemini | 3,8±1,1 (12) | 3,3±0,9 (12) | 2,7±0,8 (12) |
| | YandexGPT | 4,6±0,7 (12) | 4,2±0,6 (12) | 3,4±0,7 (12) |
| | <i>В целом</i> | 4,3±0,9 (36) | 3,9±0,8 (36) | 3,1±0,7 (36) |
| Достоверность | Copilot | 4,3±0,9 (12) | 4,6±0,5 (12) | 4,5±0,5 (12) |
| | Gemini | 3,6±0,9 (12) | 4,0±1,0 (12) | 3,6±0,9 (12) |
| | YandexGPT | 4,3±0,8 (12) | 4,5±0,5 (12) | 4,6±0,7 (12) |
| | <i>В целом</i> | 4,1±0,9 (36) | 4,4±0,7 (36) | 4,2±0,8 (36) |
| Полнота | Copilot | 4,9±0,3 (12) | 4,1±0,5 (12) | 3,0±0,7 (12) |
| | Gemini | 4,3±0,8 (12) | 3,2±0,8 (12) | 2,2±0,8 (12) |
| | YandexGPT | 4,6±0,7 (12) | 3,8±0,6 (12) | 2,8±0,6 (12) |
| | <i>В целом</i> | 4,6±0,7 (36) | 3,7±0,8 (36) | 2,7±0,8 (36) |
| Точность | Copilot | 4,3±0,7 (12) | 3,8±0,6 (12) | 3,0±0,0 (12) |
| | Gemini | 3,5±1,2 (12) | 2,9±0,9 (12) | 2,5±0,5 (12) |
| | YandexGPT | 4,3±0,8 (12) | 4,0±0,6 (12) | 3,7±0,5 (12) |
| | <i>В целом</i> | 4,0±1,0 (36) | 3,6±0,8 (36) | 3,1±0,6 (36) |
| Грамотность | Copilot | 4,5±0,5 (12) | 4,5±0,5 (12) | 4,6±0,5 (12) |
| | Gemini | 4,6±0,7 (12) | 4,6±0,5 (12) | 4,6±0,5 (12) |
| | YandexGPT | 4,7±0,5 (12) | 4,9±0,3 (12) | 5,0±0,0 (12) |
| | <i>В целом</i> | 4,6±0,6 (36) | 4,7±0,5 (36) | 4,7±0,5 (36) |

Примечания: a – среднее арифметическое, σ – стандартное отклонение, n – количество описаний.

Notes: a – arithmetic mean, σ – standard deviation, n – number of descriptions.



Т а б л и ц а 3. Результаты регрессионного анализа, обнаружившего влияние переменных-факторов на качество текстовых описаний
Table 3. The results of a regression analysis, which revealed the impact of various factors on the quality of textual descriptions

| Параметры качества | Значимые переменные-факторы | Значение коэффициента в модели (t) | Остаточное отклонение (AIC) |
|--------------------|---|--|-----------------------------|
| Полезность | Требуемая длина до 125 знаков с пробелами | -5,39 (-6,77) | 413,3 (443,3) |
| | Чат-бот Gemini | -4,50 (-3,42) | |
| | Требуемая длина до 250 знаков с пробелами | -2,89 (-4,39) | |
| | Длина ключевых слов | -1,52 (-2,92) | |
| Достоверность | Фактическая длина | -2,21 (-1,76) | 398,7 (426,7) |
| | Чат-бот Gemini | -1,53 (-2,91) | |
| | Тип изображения – логотип | -0,99 (2,58) | |
| | Требуемая длина до 125 знаков с пробелами | -0,83 (-1,21) | |
| Полнота | Требуемая длина до 125 знаков с пробелами | -6,51 (-7,01) | 357,0 (387,0) |
| | Требуемая длина до 250 знаков с пробелами | -3,70 (-4,81) | |
| | Чат-бот Gemini | -1,59 (-2,73) | |
| | Тип изображения – логотип | -1,16 (2,72) | |
| Точность | Требуемая длина до 125 знаков с пробелами | -3,58 (-5,03) | 410,6 (440,6) |
| | Требуемая длина до 250 знаков с пробелами | -1,74 (-2,88) | |
| | Тип изображения – логотип | 1,70 (4,31) | |
| | Чат-бот Gemini | -1,57 (-3,00) | |

Примечания: если значение коэффициента в модели положительно, то между переменными-факторами и зависимой переменной существует прямая связь, если отрицательно – обратная; переменные-факторы размещены в таблице по уменьшению значимости (чем больше абсолютное значение коэффициента, тем выше его значимость в модели); t – t -критерий для порядковой регрессии, $t \in (-)$, в модель включены четыре наибольших по модулю t ; AIC – Akaike information criterion (рус. информационный критерий Акаике).

Notes: If the coefficient value of a variable in the model is positive, there is a direct correlation between that variable and the dependent variable. If the value is negative, there is an inverse correlation. The factor variables are listed in the table in descending order of importance (the higher the absolute value of a coefficient, the more significant it is in the model); t – t -criterion for ordinal regression, $t \in (-)$, the model includes the four largest ones modulo t ; AIC – Akaike information criterion.

Приведенные в таблице 3 результаты допустимо интерпретировать следующим образом:

- в сравнении с неограниченными по длине описаниями, при лимитировании длины до 250 знаков с пробелами полезность падает, а при ограничении до 125 знаков – падает еще сильнее. Полезность описаний снижается существенно при использовании чат-бота Gemini. Увеличение длины списка ключевых слов также отрицательно влияет на полезность – чем длиннее список (в знаках с пробелами), тем менее релевантными оказываются описания. Остальные независимые переменные (время генерирования описаний, тип изображения, наличие фантазмов) значимо не влияют на полезность;
- излишняя длина описаний делает их менее достоверными (предположительно, за счет появления фантазмов). Вместе с тем, обнаружено, что ограничение длины до 125 знаков с пробелами хоть и незначительно, но все же снижает достоверность. Вероятность получения недостоверных описаний выше для логотипов, чем для фотографий, при этом Gemini генерирует менее достоверные результаты по сравнению с остальными чат-ботами. Значимых переменных-факторов, положительно влияющих на достоверность, не обнаружено;
- ограничение длины текста до 250 и, особенно, до 125 знаков с пробелами ожидаемо оказывает негативное влияние на полноту описаний. Тексты, сгенерированные чат-ботом Gemini, по сравнению с результатами других чат-ботов, прои-

грывают в полноте, то есть не являются вполне достаточными для понимания изображения. Значимым предиктором к снижению полноты является также тип изображения (логотипы). Переменных-факторов, положительно влияющих на полноту, не обнаружено;

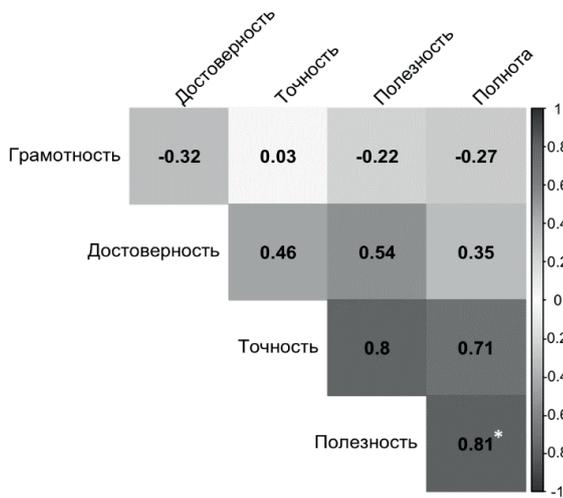
- точность изложения снижается при введении ограничений на длину описаний до 250 знаков с пробелами, и еще сильнее – до 125 знаков. Одновременно выявлено, что чат-боты генерируют для логотипов более точные описания, чем для фото. Наименее точен в своих описаниях чат-бот Gemini. Наличие фантазмов, время генерирования описаний, фактическая длина текстов и характеристики ключевых слов на точность значимо не влияют.

Частотный анализ встречаемости отдельных отметок (баллов), выставленных экспертами для каждого текстового описания по каждой из шкал, показал в целом преобладание отметок «4» (199 раз, что составило 36,9 % от всех выставленных отметок) и «5» (187; 34,6 %), реже встречались отметки «3» (109; 20,2 %) и «2» (40; 7,4 %), в единичных случаях (5; 0,9 %) – отметка «1». При этом описания, выполненные чат-ботами Copilot и YandexGPT, ни разу не были оценены на «1» и имели всего по четыре отметки «2», то есть подавляющее большинство отметок ниже среднего были выставлены за описания, сгенерированные чат-ботом Gemini.

На рисунке 2 приведена матрица корреляции Пирсона, демон-



стрирующая наличие и тесноту связей между оценками качества текстовых описаний, выставленными по разным шкалам.



Р и с. 2. Матрица корреляции Пирсона для оценок качества сгенерированных описаний по шкалам «полезность», «достоверность», «полнота», «точность», «грамотность». Числа в ячейках соответствуют коэффициентам Пирсона; * – корреляция незначима ($p > 0,05$).

Fig. 2. Pearson correlation matrix for the quality assessments of the generated descriptions on a scale of "usefulness", "reliability", "completeness", and "literacy". The numbers in the cells represent the Pearson coefficients, and * indicates a correlation that is not significant ($p > 0,05$).

В соответствии со шкалой, предложенной R.E. Chaddock¹⁶, в парах переменных достоверность-точность и достоверность-полнота присутствует умеренная связь (коэффициенты Пирсона лежат в диапазоне от 0,3 до 0,5), в то время как между достоверностью и полезностью связь характеризуется как заметная (диапазон от 0,5 до 0,7), а в парах точность-полезность и точность-полнота – как высокая (от 0,7 до 0,9). Все связи (кроме слабых, с коэффициентами ниже 0,3) имеют положительную направленность, то есть рост оценки по одной из переменных пары вызывает пропорциональное увеличение оценки по второй переменной, и наоборот. При этом, чем теснее связь, тем более сопоставимыми между собой будут соответствующие характеристики качества текстовых описаний. Например, можно ожидать, что описания, получившие высокие оценки за точность изложения, будут также высоко оценены по шкалам полезности и полноты.

В приложении С приведены примеры сгенерированных текстовых описаний, которые дают общее представление о качестве искусственного текста.

Обсуждение

По сведениям, имеющимся у авторов, исследование применимости LLM для оптимизации составления альтернативных текстов к изображениям проведено впервые. Полученные данные свидетельствуют о высоком качестве текстовых опи-

саний, сгенерированных чат-ботами (средняя оценка в баллах $a=3,97$, $\sigma=0,97$). По заданному набору ключевых слов и словосочетаний LLM способны составлять достаточно полные, достоверные, точные и грамотные текстовые альтернативы, применимые для людей, не имеющих возможности видеть контент.

Чат-боты, основанные на LLM, были предрасположены к созданию развернутых описаний, связывая в завершённый литературно обработанный текст предложенные в запросе ключевые слова. В большинстве случаев чат-боты адекватно реагировали на требования сократить текст, изменяя описания в соответствии с введенными ограничениями. При этом общее качество сокращенных текстов оставалось выше среднего. Описания длиной до 125 и до 250 знаков с пробелами применимы в качестве замещающего текста, основная задача которого состоит в кратком пояснении смысла изображения [19]. Более длинные описания могут быть размещены в цифровом документе под изображением или организованы в виде отдельного файла, связанного с изображением по контекстной ссылке. Подход с предоставлением нескольких текстовых альтернатив для одного изображения описан и в других публикациях [7], [19].

Время генерирования описаний никак не влияло на их качество. Несмотря на то, что чат-боты Gemini и YandexGPT генерировали описания быстрее, чем Copilot, время генерирования по всем трем чат-ботам исчислялось секундами (среднее время в секундах $a=10,49$, $\sigma=12,74$), тогда как на составление человеком качественного развернутого текстового описания (особенно для сложных изображений) может потребоваться до 60 минут [19]. По некоторым данным [9], долгий процесс разработки альтернативного текста негативно влияет на готовность человека заниматься составлением тестовых описаний. В нашем исследовании время, потраченное составителем на создание перечня ключевых слов и словосочетаний к одному изображению, занимало около трех минут в среднем. При этом для составления ключевых слов было достаточно: внимательно изучить изображение, выбрать из него необходимые элементы (согласно сформулированным в руководстве требованиям) и включить их в перечень. Глубокие знания языка, умения создавать связные предложения, объединенные одной темой, и навыки литературной обработки текста от составителя не требовались. Таким образом, использование чат-ботов в качестве вспомогательных инструментов для разработки альтернативного текста позволяет нивелировать недостаток языковых компетенций авторов альтернативного текста (например, при составлении описаний на неродном языке).

Сгенерированные чат-ботами текстовые описания часто содержали вымышленную (ложную или воображаемую) информацию, которая была определена нами как фантазмы. В отличие от «галлюцинаций» искусственного интеллекта, также называемых «чушью» (англ. *bullshit*) [23, 24], фантазмы не всегда дезинформируют. Так, фантазмы типов «украшательства», «умозрительности», «псевдоэмоции» и «инфошум» используются чат-ботами для литературной обработки текста, внесения дополнительного (не указанного в ключевых словах, но похожего на правду или имитирующего правдивость)

¹⁶ Chaddock R. E. Principles and methods of statistics. Boston: Houghton Mifflin Company, 1925. 471 p.



контекста и не несут в себе заведомо ложных сведений. В то время как фантазмы типа «дезинформация» (в нашем исследовании встречались в каждом шестом описании) противоречат истинному положению вещей и могут ввести в заблуждение пользователя, что приведет к неправильной трактовке им изображения. Появление в сгенерированных текстах фантазмов (особенно дезинформирующих) свидетельствует о том, что все описания должны проходить проверку составителем перед размещением в цифровых документах. О необходимости финального контроля текстовых альтернатив, сгенерированных искусственным интеллектом, говорится и в других работах [9], [25]. Содержание фантазмов требует дальнейшего исследования на больших выборочных совокупностях описаний. Для оценки текстовых описаний использовали пять шкал, что позволило исследовать различные составляющие качества. Подход к измерению качества альтернативного текста разными способами использован также в работах [8, 9]. Обнаружено, что текстовые альтернативы, сгенерированные чат-ботами, обладали высокой грамотностью (включая правильность орфографии, пунктуации и лексики), не зависящей от контекстуальных факторов. По остальным шкалам оценки были в среднем ниже, при этом удалось установить переменные, являющиеся предикторами к изменению (ухудшению или улучшению) качества. Так, выявлено, что показатели полезности, полноты, точности и достоверности снижаются при введении ограничений на длину описаний; вместе с тем, излишняя длина также негативно влияет на полезность и достоверность информации (по-видимому, за счет появления в текстах фантазмов). Полученные данные согласуются с опубликованными в работе [13], где было замечено, что с увеличением количества слов в описаниях оценки качества растут, однако слишком длинные описания нередко оказываются некачественными. Чем длиннее исходный перечень ключевых слов (в знаках с пробелами), тем ниже применимость текстовых альтернатив для людей, не имеющих возможности видеть изображение. Следовательно, при составлении ключевых слов и словосочетаний следует придерживаться лапидарности изложения. Текстовые альтернативы, сгенерированные для логотипов, точнее, чем для фотографий, но менее достоверны и полны. Не исключено (но требует дальнейших исследований), что противоречивость полученных результатов связана с характеристиками ключевых слов, а именно: наличием текста, поясняющего суть визуального контента, то есть повышающего «узнаваемость» логотипа; избытком графических деталей изображения, при связывании которых в единый текст могли возникнуть погрешности и пробелы. Между оценками достоверности, точности, полезности и полноты присутствовали значимые попарные корреляции. Самые сильные положительные связи зарегистрированы в парах точность-полезность и точность-полнота. Следовательно, изменение точности описаний, как в положительном, так и в отрицательном направлении, влечет пропорциональные и сопоставимые изменения показателей полноты (достаточности) и полезности (применимости), что может говорить о взаимозаменяемости используемых для оценки шкал. Чат-боты, реализующие LLM, повели себя в исследовании по-разному: Copilot – дольше всех генерировал описания, почти всегда соблюдал лимиты на длину текста, создавал точные

описания; Gemini – генерировал менее достоверные, полные, полезные и точные результаты, чем остальные чат-боты, чаще прочих не соблюдал требования к длине описаний, получал самые низкие оценки за выполненные описания, но быстро генерировал тексты; YandexGPT – создавал точные описания, генерировал тексты быстрее всех, однако часто не соблюдал лимиты на длину описаний. Несмотря на то, что Gemini показал себя сравнительно хуже, чем Copilot и YandexGPT, в целом все чат-боты справились с задачей. Таким образом, в результате выполненного исследования мы можем рекомендовать использовать LLM, реализованные в протестированных чат-ботах, для оптимизации создания текстовых альтернатив к информативным изображениям и изображениям, содержащим текст. Исходя из результатов апробации предложенного алгоритма разработки текстовых описаний можно сформулировать рекомендации для составителей альтернативного текста по использованию LLM в своей работе:

- при составлении списка ключевых слов, характеризующих изображение, придерживаться точности и краткости изложения, избегать субъективной трактовки изображения;
- включать в список ключевых слов все значимые элементы изображения: для фотографий – цвет, место, объекты заднего и переднего плана, персонажи (внешний вид, пол, возраст, размещение, настроение, поведение), детали, поясняющие сюжет; для логотипов – цвет, текст и его характеристики, графические объекты и их взаимное расположение;
- после выполнения чат-ботом серии описаний для каждого изображения удалять файлы Cookie, очищать историю браузера и историю поиска;
- проверять сгенерированные описания на наличие фантазмов и, при обнаружении, избавляться от них, выполняя коррекцию текста перед размещением текстовых альтернатив в цифровых документах и на веб-страницах.

Проведенное исследование имеет **ограничения**: в ходе анализа не выполнялась сравнительная оценка качества текстов, сгенерированных чат-ботами и написанных человеком без их помощи – вероятно, компаративный анализ двух способов разработки текстовых альтернатив позволит выявить новые закономерности; сгенерированные тексты не подвергались оценке людьми, не имеющими возможности видеть изображение, в том числе, с глубокими нарушениями зрения – не исключено появление новых рекомендаций для составителей альтернативного текста после анализа сгенерированных описаний целевой аудиторией; описания выполнялись на русском языке – можно предположить появление новых результатов и закономерностей при использовании других языков.

Заключение

К обязательным требованиям доступности информации, представленной в электронно-цифровой форме, относится наличие текстовых альтернатив для нетекстового контента, в том числе, изображений. Альтернативный текст позволяет пользователям, не имеющим возможности видеть, понять смысл и содержание изображения. Известно, что текстовые альтернативы, разработанные с помощью полностью авто-



математических методов искусственного интеллекта (глубокого обучения, обработки естественного языка и компьютерного зрения), проигрывают в качестве по сравнению с текстовыми описаниями, созданными человеком. При этом автор или составитель альтернативного текста должен обладать компетенциями по созданию грамотных, лаконичных и точных текстовых описаний, в том числе применяя оптимизирующие программные инструменты.

В результате исследования разработан и проверен методический подход к созданию текстовых альтернатив для нетекстового контента (информативных изображений и изображений, содержащих текст) с использованием LLM, реализованных в

чат-ботах. Установлено, что LLM позволяют на основе заданного перечня ключевых слов, характеризующих изображение, создавать связанные текстовые описания, пригодные для применения в качестве альтернативного текста. В ходе статистического анализа выявлены факторы, влияющие на полезность, полноту, точность и достоверность сгенерированных описаний, что позволило сформулировать рекомендации для составителей по использованию LLM в своей работе.

Представляется перспективным продолжить исследования по применению чат-ботов на основе LLM в процессах разработки альтернативных текстовых описаний на разных языках и к изображениям разных типов.

References

- [1] Kosova Ye.A., Redkokosh K.I. Methodology for the Formation of Digital Accessibility Competences: Development and Trial on a Russian Sample. *RUDN Journal of Psychology and Pedagogics*. 2022;19(3):488-509. (In Russ., abstract in Eng.) <https://doi.org/10.22363/2313-1683-2022-19-3-488-509>
- [2] Hamideh Kerdar S., Bächler L., Kirchoff B.M. The accessibility of digital technologies for people with visual impairment and blindness: a scoping review. *Discover Computing*. 2024;27:24. <https://doi.org/10.1007/s10791-024-09460-7>
- [3] Jung C., Mehta S., Kulkarni A., Zhao Y., Kim Y.-S. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. *IEEE Transactions on Visualization and Computer Graphics*. 2022;28(1):1095-1105. <https://doi.org/10.1109/TVCG.2021.3114846>
- [4] Tiwary T., Mahapatra R.P. Web Accessibility Challenges for Disabled and Generation of Alt Text for Images in Websites using Artificial Intelligence. In: Proceedings of the 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). Ghaziabad, India: IEEE Press; 2022. p. 1-5. <https://doi.org/10.1109/ICICT55121.2022.10064545>
- [5] Xie J., Li R., Lv S., Wang Y., Wang Q., Vorotnitsky Y.I. Chinese alt text writing based on deep learning. *Trait Signal*. 2019;36(2):161-170. <https://doi.org/10.18280/ts.360206>
- [6] Sharma H., Agrahari M., Singh S.K., Firoj M., Mishra R.K. Image Captioning: A Comprehensive Survey. In: 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC). Mathura, India: IEEE Press; 2020. p. 325-328. <https://doi.org/10.1109/PARC49193.2020.236619>
- [7] Hanley M., Barocas S., Levy K., Azenkot S., Nissenbaum H. Computer Vision and Conflicting Values: Describing People with Automated Alt Text. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21). New York, NY, USA: Association for Computing Machinery; 2021. p. 543-554. <https://doi.org/10.1145/3461702.3462620>
- [8] Lee J., Peng Y. H., Herskovitz J., Guo A. Image Explorer: Multi-Layered Touch Exploration to Make Images Accessible. In: Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21). New York, NY, USA: Association for Computing Machinery; 2021. Article number: 69. <https://doi.org/10.1145/3441852.3476548>
- [9] Mack K., Cutrell E., Lee B., Morris M.R. Designing Tools for High-Quality Alt Text Authoring. In: Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21). New York, NY, USA: Association for Computing Machinery; 2021. Article number: 23. <https://doi.org/10.1145/3441852.3471207>
- [10] Jeong H., Chun V., Lee H., Oh S.Y., Jung H. WATAA: Web Alternative Text Authoring Assistant for Improving Web Content Accessibility. In: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23 Companion). New York, NY, USA: Association for Computing Machinery; 2023. p. 41-45. <https://doi.org/10.1145/3581754.3584127>
- [11] Edwards E.J., Gilbert M., Blank E., Branham S.M. How the Alt Text Gets Made: What Roles and Processes of Alt Text Creation Can Teach Us About Inclusive Imagery. *ACM Transactions on Accessible Computing*. 2023;16(2):1-28. <https://doi.org/10.1145/3587469>
- [12] Chintalapati S.S., Bragg J., Wang L.L. A Dataset of Alt Texts from HCI Publications: Analyses and Uses Towards Producing More Descriptive Alt Texts of Data Visualizations in Scientific Papers. In: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22). New York, NY, USA: Association for Computing Machinery; 2022. Article number: 30. <https://doi.org/10.1145/3517428.3544796>
- [13] Williams C., de Greef L., Harris E., Findlater L., Pavel A., Bennett C. Toward supporting quality alt text in computing publications. In: Proceedings of the 19th International Web for All Conference (W4A '22). New York, NY, USA: Association for Computing Machinery; 2022. Article number: 20. <https://doi.org/10.1145/3493612.3520449>
- [14] Kosova Ye.A., Redkokosh K.I., Mikheyev P.O. Using A Neural Network to Generate Images When Teaching Students to Develop an Alternative Text. *Open Education*. 2024;28(1):9-20. (In Russ., abstract in Eng.) <https://doi.org/10.21686/1818-4243-2024-1-9-20>
- [15] Kaddour J., Harris J., Mozes M., Bradley H., Raileanu R., McHardy R. Challenges and Applications of Large Language Models. *arXiv:2307.10169*. 2023. <https://doi.org/10.48550/arXiv.2307.10169>
- [16] Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., Stadler M., Weller J., Kuhn J.,



- Kasneji G. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*. 2023;103(102274). <https://doi.org/10.1016/j.lindif.2023.102274>
- [17] Myers D., Mohawesh R., Chellaboina V.I., Sathvik A.L., Venkatesh P., Ho Y.-H., Henshaw H., Alhawawreh M., Berdik D., Jararweh Y. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*. 2024;27:1-26. <https://doi.org/10.1007/s10586-023-04203-7>
- [18] Xiao C., Xu S.X., Zhang K., Wang Y., Xia L. Evaluating Reading Comprehension Exercises Generated by LLMs: A Showcase of ChatGPT in Education Applications. In: Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023). Toronto, Canada: Association for Computational Linguistics; 2023. p. 610-625. <https://doi.org/10.18653/v1/2023.bea-1.52>
- [19] Richard B. Key Issues Affecting the Inclusion of Alt Text in Scholarly PDF Publications. *Logos*. 2023;34(1):44-60. <https://doi.org/10.1163/18784712-03104058>
- [20] Crane M.A., Nguyen M., Lam A., Berger Z.D., Paulus Y.M., Romley J.A., Faden R.R. Figure accessibility in journals: analysis of alt-text in 2021-23. *The Lancet*. 2023;402(10419):2287-2289. [https://doi.org/10.1016/S0140-6736\(23\)02348-6](https://doi.org/10.1016/S0140-6736(23)02348-6)
- [21] Alam M.Z.I., Islam Sh., Hoque E. SeeChart: Enabling Accessible Visualizations Through Interactive Natural Language Interface For People with Visual Impairments. In: Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23). New York, NY, USA: Association for Computing Machinery; 2023. p. 46-64. <https://doi.org/10.1145/3581641.3584099>
- [22] Likert R. A technique for the measurement of attitudes. *Archives of Psychology*. 1932;22(140):5-55.
- [23] Hicks M.T., Humphries J., Slater J. ChatGPT is bullshit. *Ethics and Information Technology*. 2024;26:38. <https://doi.org/10.1007/s10676-024-09775-5>
- [24] Fisher S.A. Large language models and their big bullshit potential. *Ethics and Information Technology*. 2024;26:67. <https://doi.org/10.1007/s10676-024-09802-5>
- [25] Salisbury E., Kamar E., Morris M. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing. 2017;5(1):147-156. <https://doi.org/10.1609/hcomp.v5i1.13301>

Поступила 16.05.2024; одобрена после рецензирования 12.07.2024; принята к публикации 23.08.2024.

Submitted 16.05.2024; approved after reviewing 12.07.2024; accepted for publication 23.08.2024.

Приложение А

Список изображений, использованных для составления текстовых описаний

Appendix A

List of Images Used to Compose Text Descriptions

Фотографии:

- «Full Speed» (Drysdale, 1969), черно-белая¹⁷;
- «Снежный душ» (Л. Бородулин, 1960), черно-белая¹²;
- «Олимпиада-80» (А. Абаза, 1980), черно-белая¹²;
- «Saving the Monarchs» (J. Rojo, 2024), цветная¹⁴;
- «Плюшевый мишка» (А. Романов, 2020), цветная¹⁸;
- «На границе двух сред обитания» (А. Максимов, 2019), цветная¹⁸.

Логотипы:

- компании Lacoste S.A.¹⁵;
- компании Chupa Chups S.A.U.¹³;
- Национального управления по авиации и исследованию космического пространства (англ. *National Aeronautics and Space Administration, NASA*)¹⁹;
- компании Nestlé S.A.²⁰;
- Организации Объединенных Наций по вопросам образования, науки и культуры (англ. *United Nations Educational, Scientific and Cultural Organization, UNESCO*)²¹;
- Британской радиовещательной корпорации (англ. *British Broadcasting Corporation, BBC*)²².

¹⁷ Drysdale J., Regan M. *Our Peaceable Kingdom: The photographs of John Drysdale*. New York: St. Martin's Press, 2000. 112 p.

¹⁸ Самая красивая страна: победители [Электронный ресурс] // ВОО «Русское географическое общество», 2024. URL: <https://photo.rgo.ru/winners/> (дата обращения: 16.05.2024).

¹⁹ NASA [Электронный ресурс] // Wikipedia, 2024. URL: <https://en.wikipedia.org/wiki/NASA> (дата обращения: 16.05.2024).

²⁰ Nestlé [Электронный ресурс] // Wikipedia, 2024. URL: <https://en.wikipedia.org/wiki/Nestlé> (дата обращения: 16.05.2024).

²¹ UNESCO [Электронный ресурс] // Wikipedia, 2024. URL: <https://en.wikipedia.org/wiki/UNESCO> (дата обращения: 16.05.2024).

²² BBC [Электронный ресурс] // Wikipedia, 2024. URL: <https://en.wikipedia.org/wiki/BBC> (дата обращения: 16.05.2024).



Приложение В

Данные о регистрации фантазмов в текстовых описаниях, сгенерированных чат-ботами по ключевым словам

Appendix B

Data on the registration of phantasms in text descriptions generated by chatbots based on keywords

| Оцениваемые параметры | Тип изображения | | Чат-бот | | | Ограничения по длине | | | В целом |
|--|-----------------|----------------|----------------|----------------|----------------|-----------------------|---------------------------|---------------------------|----------------|
| | Фотография | Логотип | Copilot | Gemini | Yandex GPT | Без ограничений длины | До 250 знаков с пробелами | До 125 знаков с пробелами | |
| Количество типов фантазмов в описаниях, $a \pm \sigma$ (n) | 2,63±1,01 (32) | 1,70±0,84 (30) | 2,13±1,06 (15) | 2,41±1,02 (34) | 1,62±0,87 (13) | 2,44±1,19 (34) | 1,88±0,81 (16) | 1,83±0,58 (12) | 2,18±1,03 (62) |
| Min – max | 1 – 5 | 1 – 4 | 1 – 4 | 1 – 5 | 1 – 3 | 1 – 5 | 1 – 4 | 1 – 3 | 1 – 5 |
| Доля описаний с: | | | | | | | | | |
| – фантазмами, %* (n) | 29,63 (32) | 27,78 (30) | 13,89 (15) | 31,48 (34) | 12,04 (13) | 31,48 (34) | 14,81 (16) | 11,11 (12) | 57,41 (62) |
| – одним типом фантазмов, %** (n) | 4,84 (3) | 24,19 (15) | 8,06 (5) | 8,06 (5) | 12,90 (8) | 16,13 (10) | 8,06 (5) | 4,84 (3) | 29,03 (18) |
| – двумя типами фантазмов, %** (n) | 22,58 (14) | 16,13 (10) | 8,06 (5) | 27,42 (17) | 3,23 (2) | 11,29 (7) | 14,52 (9) | 12,90 (8) | 38,71 (24) |
| – тремя типами фантазмов, %** (n) | 12,90 (8) | 6,45 (4) | 4,84 (3) | 9,68 (6) | 4,84 (3) | 16,13 (10) | 1,61 (1) | 1,61 (1) | 19,35 (12) |
| – четырьмя типами фантазмов, %** (n) | 9,68 (6) | 1,61 (1) | 3,23 (2) | 8,06 (5) | 0,00 (0) | 9,68 (6) | 1,61 (1) | 0,00 (0) | 11,29 (7) |
| – пятью типами фантазмов, %** (n) | 1,61 (1) | 0,00 (0) | 0,00 (0) | 1,61 (1) | 0,00 (0) | 1,61 (1) | 0,00 (0) | 0,00 (0) | 1,61 (1) |
| – «украшательствами», %** (n) | 37,10 (23) | 4,84 (3) | 12,90 (8) | 25,81 (16) | 3,23 (2) | 24,19 (15) | 9,68 (6) | 8,06 (5) | 41,94 (26) |
| – «умозрительностиами», %** (n) | 45,16 (28) | 46,77 (29) | 24,19 (15) | 46,77 (29) | 20,97 (13) | 54,84 (34) | 20,97 (13) | 16,13 (10) | 91,94 (57) |
| – «псевдоэмоциями», %** (n) | 30,65 (19) | 6,45 (4) | 9,68 (6) | 19,35 (12) | 8,06 (5) | 27,42 (17) | 8,06 (5) | 1,61 (1) | 37,10 (23) |
| – «инфошумом», %** (n) | 1,61 (1) | 17,74 (11) | 0,00 (0) | 19,35 (12) | 0,00 (0) | 9,68 (6) | 4,84 (3) | 4,84 (3) | 19,35 (12) |
| – «дезинформацией», %** (n) | 20,97 (13) | 6,45 (4) | 4,84 (3) | 20,97 (13) | 1,61 (1) | 17,74 (11) | 4,84 (3) | 4,84 (3) | 27,42 (17) |

Примечания: a – среднее арифметическое; σ – стандартное отклонение; n – количество описаний, имеющих фантазмы; min – max – минимальное и максимальное количество типов фантазмов в одном описании; %* – доля от всей совокупности описаний ($N=108$); %** – доля от совокупности описаний, имеющих фантазмы ($n=62$)

Notes: a – arithmetic mean, σ – standard deviation, n – the number of descriptions that have phantasms; min – max – the minimum and maximum number of phantasm types in one description; %* – the share of the total number of descriptions ($N=108$); %** – the share of the total number of descriptions that contain phantasms ($n=62$).



Приложение С

Текстовые описания к изображению «Full Speed»¹⁷, сгенерированные чат-ботами по ключевым словам

Appendix C

Text descriptions of the «Full Speed» image generated by chatbots using keywords

| Чат-боты | Требуемая длина описаний | Текстовые описания к изображению, сгенерированные чат-ботами (число знаков с пробелами в описании) |
|-----------|---------------------------|--|
| Copilot | Без ограничений | Это черно-белое фото, сделанное в раннюю осень. На переднем плане изображены две мохнатые овечки с черными мордами. Они бегут по широкому лугу, где трава скошена. На овечках верхом сидят две девочки лет семи в легких пальтишках. Они мчат во весь опор и смеются. Фон изображения размыт, что подчеркивает скорость движения овечек и девочек. Это изображение наполнено радостью и беззаботностью детства. (399) |
| | До 250 знаков с пробелами | Черно-белое фото: две семилетние девочки в легких пальто, смеющиеся и мчащиеся верхом на мохнатых овечках с черными мордами по скошенному лугу. Фон размыт, подчеркивая скорость. (177) |
| | До 125 знаков с пробелами | Ч/б фото: две девочки, 7 лет, смеются, мчат на овечках по лугу. Фон размыт, акцент на скорость. (95) |
| Gemini | Без ограничений | Чёрно-белое фото, запечатлевшее беззаботный вихрь ранней осени. Широкий луг, убранный от скошенной травы, расстилается под ногами. Размытый фон создает ощущение скорости и движения. Две мохнатые овечки с черными мордами, неся на своих спинах двух семилетних девочек, мчатся во весь опор по бескрайнему простору. Легкие пальтишки юных наездниц развеваются на ветру, а их лица озарены заразительным смехом. Фотография дышит безграничной свободой, беззаботностью и единением с природой. (484) |
| | До 250 знаков с пробелами | Ранняя осень. Озорные девочки верхом на овечках мчатся по бескрайнему лугу. Чёрно-белое фото. Скошенная трава, размытый фон. Две овечки, черные морды, несут на спинах смеющихся семилеток в легких пальтишках. Беззаботный вихрь, единение с природой. (247) |
| | До 125 знаков с пробелами | Озорные наездницы на овечках мчатся по лугу. Ч/б фото. Осень. Скошенная трава. Смех. (84) |
| YandexGPT | Без ограничений | На черно-белом фото изображён широкий луг ранней осенью. Трава на лугу скошена, фон размыт. В центре внимания — две девочки лет семи верхом на мохнатых овечках. Они одеты в лёгкие пальтишки и мчат во весь опор. Кажется, что скорость движения передана даже на фотографии. Девочки смеются. Их лица выражают радость и беззаботность. Чёрные морды овечек добавляют динамики в общую картину. Это фото передаёт атмосферу детской радости и веселья. Оно вызывает улыбку и напоминает о беззаботном детстве. (496) |
| | До 250 знаков с пробелами | На фото — широкий луг ранней осенью, трава скошена. В центре — две девочки лет семи верхом на мохнатых овечках, смеются. Чёрные морды овечек добавляют динамики. Фото передаёт атмосферу детской радости. (201) |
| | До 125 знаков с пробелами | На фото — широкий луг ранней осенью, трава скошена. Две девочки лет семи верхом на мохнатых овечках смеются. (108) |

Примечание: предложенные чат-ботам ключевые слова – *черно-белое фото, ранняя осень, широкий луг, трава скошена, фон размыт, скорость, две мохнатые овечки, черные морды, две девочки лет семи, легкие пальтишки, верхом на овечках, мчат во весь опор, смеются.*

Note: suggested keywords for chatbots – *black and white photo, early autumn, wide meadow, grass mowed, background blurred, speed, two shaggy sheep, black muzzles, two girls about seven years old, light coats, riding sheep, running at full speed, laughing.*



Об авторах:

Косова Екатерина Алексеевна, заведующий кафедрой прикладной математики физико-технического института, ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского» (295007, Российская Федерация, Республика Крым, г. Симферополь, просп. Академика Вернадского, д. 4), кандидат педагогических наук, доцент, **ORCID: <https://orcid.org/0000-0002-3263-9373>**, lynx99@inbox.ru

Редкокош Кирилл Игоревич, аспирант физико-технического института, ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского» (295007, Российская Федерация, Республика Крым, г. Симферополь, просп. Академика Вернадского, д. 4), **ORCID: <https://orcid.org/0000-0003-3535-7336>**, kirillf13@yandex.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Yekaterina A. Kosova, Head of the Chair of Applied Mathematics, Physics and Technology Institute, V.I. Vernadsky Crimean Federal University (4 Prospekt Vernadskogo, Simferopol 295007, Republic of Crimea, Russian Federation), Cand. Sci. (Ped.), Associate Professor, **ORCID: <https://orcid.org/0000-0002-3263-9373>**, lynx99@inbox.ru

Kirill I. Redkokosh, Postgraduate Student of the Physics and Technology Institute, V.I. Vernadsky Crimean Federal University (4 Prospekt Vernadskogo, Simferopol 295007, Republic of Crimea, Russian Federation), **ORCID: <https://orcid.org/0000-0003-3535-7336>**, kirillf13@yandex.ru

All authors have read and approved the final manuscript.

