

УДК 81`255

Doi 10.26456/vtfilol/2026.2.223

БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ В РАБОТЕ ПЕРЕВОДЧИКА ПИСЬМЕННЫХ ТЕКСТОВ

Е. Д. Малёнова, Д. П. Винский

Омский государственный университет им. Ф. М. Достоевского, г. Омск

Статья посвящена изучению вопроса применения больших языковых моделей (далее – БЯМ) в процессе письменного перевода. Проводится обзор некоторых существующих БЯМ, рассматривается специфика применения БЯМ в письменном переводе в сопоставлении с системами нейронного машинного перевода, разрабатываются рекомендации по оценке качества переводов, выполненных с применением БЯМ, с учётом специфики взаимодействия пользователя с системами искусственного интеллекта.

Ключевые слова: перевод, генеративный искусственный интеллект, большие языковые модели, БЯМ, нейронный машинный перевод, оценка качества перевода.

Использование компьютерных технологий в целях повышения производительности труда давно закрепилось в переводческой практике. Со времен Джорджтаунского эксперимента (1954 г.) и до появления нейронного машинного перевода информационные технологии в переводе сделали огромный скачок вперед. В наши дни существует множество сервисов, предоставляющих услуги машинного перевода, такие как Google Translate, Yandex Translate, DeepL, Prompt и т.д.. Вместе с тем в течение последних нескольких лет наблюдается прирост популярности использования генеративного искусственного интеллекта, в частности больших языковых моделей (далее – БЯМ). Процесс взаимодействия с БЯМ, в связи с применением в них уникальной технологии обработки больших массивов текстовой информации, существенно отличается от использования систем машинного перевода. В связи с этим возникает потребность в поиске новых способов оценки качества переводов, выполняемых генеративным искусственным интеллектом.

В переводческой практике оценка качества машинного перевода отличается отказом от критериев адекватности и эквивалентности, применимых к переводам, которые были выполнены человеком. Вместо этого оценивается уровень потребности в последующем редактировании полученного перевода. Так, бюро переводов АКМ-Вест предлагает следующий ряд критериев оценки качества машинного перевода [1]:

1. Смысловые ошибки. Предполагается, что качественный машинный перевод будет как можно более точно передавать смысл оригинального текста, не будет содержать грубых искажений смысла.

2. Немотивированные добавления/опущения. В полученном переводе не должно быть опущений тех или иных лексических единиц или фрагментов текста. В то же время такие лексические единицы или фрагменты текста не должны оставаться на языке оригинала.

3. Единообразие терминологии. Необходимо сохранять последовательность перевода терминов в тексте, не допускается использование нескольких переводов одного и того же термина. Данный критерий особенно важен для переводов технических, медицинских, юридических текстов.

4. Соответствие нормам ПЯ. В машинном переводе не должны проследиваться грубые отклонения от норм языка, на который осуществляется перевод текста, например, не допускается сохранение синтаксиса языка оригинала.

5. Целостность полученного перевода. В качественном переводе должно проследиваться сохранение связей, присутствующих в оригинале, сохранение стиля и регистра. Так, например, недопустимо, чтобы в переводе диалога одновременно встречались обращения «вы» и «ты» от одного и того же лица.

Искусственный интеллект (ИИ) уже успел закрепиться в сфере машинного перевода, постепенно вытесняя используемый ранее статистический машинный перевод, в основе которого лежало сравнение языковых пар или корпусов текстов. На замену ему некоторые компании, например, Google и Yandex, представили нейронный машинный перевод (далее – НМП) – технологию, в которой искусственная нейронная сеть обучается на большом массиве текстов с целью повышения эффективности решения переводческих задач, подобно тому, как человек может целенаправленно осваивать определённый род деятельности. Однако, как отмечает К. Уолк, всё же есть ощутимая разница между умственной деятельностью человека и работой искусственной нейронной сети: «Человеческая память не вечна, множество различных факторов может привести к отмиранию мозговых клеток с последующей потерей хранящейся в них информации, в то время как информация, которая была закреплена в нейронной сети посредством тренировок, никогда не будет забыта» [14: 4]. Помимо этого, К. Уолк отмечает, что точность выполненных искусственной нейронной сетью переводов сохраняется, в то время как качество переводов, выполненных человеком, варьируется [там же].

Генеративный ИИ отличается от машинного перевода более широкой сферой применения. Данные системы проводят анализ и синтез информации в соответствии с запросом, составленным пользователем. На сегодняшний день многие системы генеративного ИИ позволяют вести работу не только с текстом, но и с изображениями, звуком и видео. Наиболее распространёнными системами генеративного ИИ являются БЯМ, проводящие обработку текстов на естественном языке (NLP).

Одной из главных целей языковых моделей является создание органично звучащих текстов посредством серии предположений, сформированных в результате обработки множества текстов на стадии обучения модели. Многие БЯМ используют интерфейс чат-бота, позволяя пользователю составлять запросы в форме беседы с искусственным интеллектом.

К началу 2026 года на глобальные рынки было выведено множество БЯМ, разработанных как зарубежными, так и отечественными компаниями. Среди них можно выделить следующие:

1. GPT-3.5. Третье поколение моделей GPT, выпущенное компанией OpenAI в 2020 году. GPT-3.5 – одна из наиболее популярных моделей. С помощью веб-интерфейса ChatGPT данная модель позволяет вести работу с текстом (перевод, краткий пересказ, составление новых текстов) и компьютерным кодом. ChatGPT также позволяет параллельно использовать нейронную сеть DALL-E для синтеза изображений. По платной подписке пользователи ChatGPT могут получить доступ к новым версиям модели: GPT-4o и GPT-5. ChatGPT обладает способностью обрабатывать большой объём контекста взаимодействий с пользователем, что позволяет языковой модели выполнять такие задачи как объяснение упомянутых ранее явлений или предметов, синтез творческих произведений и решение представленных пользователем проблем [11: 1].

2. BLOOM. Альтернатива GPT-3, которая была разработана в рамках проекта BigScience с целью создания БЯМ, доступной для академических исследований [6]. Модель BLOOM способна составлять и анализировать текст на 46 языках, а также позволяет составлять компьютерный код на 13 языках программирования [8]. Разработчики проекта BigScience заявляют, что данная модель не нацелена на применение в сферах медицины, экономики, политики и юриспруденции, а также не должна быть использована для получения фактически корректной информации по запросам пользователя.

3. DeepSeek. БЯМ, разработанная одноимённой китайской компанией. Данная модель опирается на ансамблевое обучение искусственного интеллекта, что позволяет значительно понизить вычислительные затраты в сравнении с моделями OpenAI, в то время как процесс дистилляции знаний позволяет модели отсеивать избыточные данные [11: 1-2].

4. Claude. БЯМ, разработанная компанией Anthropic. Основными сферами применения данной модели являются программирование и техническая поддержка, хотя Claude также может проводить и различные операции над текстом, будь то перевод, составление краткого пересказа или ответов на заданные пользователем вопросы. На данный момент модель доступна в трёх разновидностях: Naïve, которая позволяет быстро решать простые задачи; Opus, которая жертвует скоростью ради качества

выполнения трудных расчётов; и Sonnet, в которой достигается баланс между скоростью и качеством выполнения запросов пользователей [5].

5. Gemini. Улучшенная версия модели LaMDA, разработанной компанией Google и доступной через чат-бот программу Bard. При создании модели Gemini особое внимание было уделено мультимодальному функционалу, позволяя данной модели ИИ обрабатывать текст, компьютерный код, изображения, а также видео- и аудиоматериалы. Одной из особенностей данной модели является встроенная система проверки качества, позволяющая пользователям оценить, насколько корректно ИИ отвечает на заданные вопросы [10: 860].

6. Alice AI. Усовершенствованная версия модели YandexGPT, разработанной компанией Яндекс в качестве отечественной альтернативы ChatGPT. Данная модель была интегрирована в виртуального помощника «Алиса», а также во многие другие сервисы компании, включая поисковую систему и платформу электронной торговли. Система расчёта токенов позволяет данной модели более экономично обрабатывать представляемую пользователем информацию [4]. В первую очередь, данная модель рассчитана на обработку текстов на русском языке.

7. GigaChat. Отечественная БЯМ, разработанная Сбербанком России. Данная модель может выполнять большое количество операций, ассоциируемых с различными моделями генеративного искусственного интеллекта, будь то анализ и синтез мультимедийных данных или «отыгрывание» определённых персонажей в ходе коммуникации с пользователем.

Как было отмечено ранее, в отличие от систем НМП, БЯМ обучены выполнению множества различных операций над текстом, будь то корректура, перевод на другой язык, краткий пересказ содержания или составление нового текста в соответствии с запросом пользователя. Подобное разнообразие функций, выполняемых одной системой, позволяет проводить более гибкую работу над переводом текста. Однако стоит ли ожидать, что БЯМ заменят машинный перевод в будущем или же будут существовать в качестве альтернативного подхода к использованию компьютерных технологий в процессе перевода? Несмотря на то, что в основе каждого из данных подходов лежит использование искусственных нейронных систем, существуют определённые различия в том, как системы НМП и БЯМ справляются с поставленной переводческой задачей. Пользователям систем НМП известны недостатки машинного перевода и типичные ошибки, которые допускают такие системы, в связи с чем становится легко предугадать, в каких ситуациях потребуются постредактирование полученного перевода. Переводы, предлагаемые БЯМ, демонстрируют более низкий уровень предсказуемости. Помимо этого, разница в особенностях тренировки ИИ приводит к тому, что БЯМ демонстрируют более

высокий уровень владения языком, но не достигают точности перевода, выполняемого системами машинного перевода. Л. Варга и Й. Чищельски, рассуждая о различиях между данными подходами к автоматизации переводческой деятельности, отмечают, что БЯМ не следует рассматривать как замену машинному переводу: «Будущее лежит за поиском синергии между языковыми моделями и машинным переводом, взвешиванием сильных сторон каждого из этих подходов с целью получения точного, контекстуально правильного и натурально звучащего перевода» [13].

Отдельное внимание следует уделить особенностям работы с генеративным ИИ. Одним из наиболее часто упоминаемых недостатков БЯМ является склонность данных моделей к созданию ложной информации, получившей название «галлюцинации». Как отмечают Аменицкий и др., БЯМ «закодированы таким образом, чтобы придумывать ответ» [2: 227], иными словами, при нехватке информации ИИ может предложить лишь искусственно сгенерированный текст, который структурно похож на ответ на вопрос. Это представляет особую опасность для использования генеративного ИИ в условиях, где подобная ошибка может повлечь за собой тяжелые последствия. С точки зрения письменного перевода, подобные галлюцинации могут стать причиной затруднений в случае необходимости объяснения культурных реалий, не имеющих эквивалента в языке перевода. Вместе с тем следует обращать внимание на размеры контекстного окна, с которым может работать выбранная модель. При синтезе ответов БЯМ могут учитывать ограниченное количество новой информации, полученной в ходе взаимодействия пользователя с моделью. Текстовая информация, которую обрабатывает модель, разбивается на единицы, называемые «токенами». Один токен может являться словом, символом или знаком препинания. При превышении допустимого размера контекстного окна модель будет забывать некоторые части ранее обработанных фрагментов текста, в связи с чем могут возникнуть трудности со связностью текста или единством терминологии в полученном переводе. На примере перевода имён зарубежных монархов Е.К. Толстикова продемонстрировала необходимость ссылок на контекст, так как исследуемая модель YandexGPT не учитывала содержание предыдущих запросов [3: 157]. В настоящее время разработчики БЯМ активно расширяют допустимые размеры контекстных окон моделей, однако данное ограничение всё ещё может создавать сложности при продолжительной работе с большим объёмом переводимого текста.

В то же время, одна из особенностей использования БЯМ, отличающая их от машинного перевода – возможность тонкой настройки получаемого результата. Так, например, при формулировании запроса на перевод фрагмента текста делового письма с английского языка на русский в него можно добавить такие фразы как «с учётом особенностей

официально-делового стиля русского письма», «точный перевод», «с сохранением смысла» и др. для задачи дополнительных требований, которым должен будет соответствовать результирующий перевод. Данный подход получил название «prompt engineering» или «техника подсказок». Возможность включения переводческой задачи в запрос для модели рассматривается Х. Шаркас. Посредством указания в запросе целевой аудитории, цели перевода и так далее оценивался потенциал искусственного интеллекта к достижению результатов, схожих с работой человека. В ходе исследования было установлено, что для повышения качества выполняемого БЯМ перевода необходимо составлять более подробные запросы, например, «включая примеры переводов других текстов того же жанра, соответствующие требованиям» [12: 149].

На сегодняшний день активно проводятся исследования, в ходе которых оценивается способность тех или иных моделей к качественному выполнению перевода текстов. Однако с учётом указанных выше факторов возникает вопрос: позволяют ли системы, предназначенный для оценки качества машинного перевода, адекватно оценить качество переводов, выполненных БЯМ? М.Н. Кайела заявляет, что, несмотря на различия между НМТ и БЯМ, использование алгоритмов оценки качества машинных переводов является «отправной точкой для сравнения результатов, демонстрируемых НМТ и БЯМ» [7] и представляет собой не столько идеальное решение проблемы, сколько практически оправданный шаг. В свою очередь, особенности работы БЯМ вызывают необходимость обратить внимание на следующие условия:

1. Переводы, выполняемые генеративным ИИ, будут в определённой степени отличаться от запроса к запросу. Возможно, для более адекватной оценки качества выполняемой работы следует получить несколько вариантов перевода одного предложения, представленных одной БЯМ и вычислить средние показатели качества перевода.

2. Определённую ценность для анализа имеет техника подсказок. Как показало исследование Х. Шаркас, даже в пределах изучения одной модели можно получить различные результаты посредством подбора инструкций для модели. Сравнение результатов перевода с использованием техники подсказок и без позволит оценить потенциал модели к выполнению качественного перевода.

3. Сопоставительное сравнение качества переводов, выполненных системами НМП и БЯМ может оказаться излишне генерализованным и представить трудности в полноценном раскрытии возможностей того или иного подхода. Полученные таким образом результаты могут дать лишь поверхностное представление о том насколько один подход к использованию технологий ИИ в переводе превосходит альтернативы.

4. Так как ИИ обучается посредством обработки существующих текстов, использование в процессе анализа текстов без перевода может позволить лучше оценить способности ИИ к выполнению перевода.

Алгоритмы, позволяющие оценивать качество перевода, выполненного непосредственно БЯМ, находятся в активной разработке. Одним из таких специализированных алгоритмов является GEMBA-MQM, представленный Т. Кокми и К. Федерманном. Однако разработчики алгоритма утверждают, что, несмотря на удовлетворительные показатели точности оценки, применение его в академических исследованиях не рекомендуется, ввиду зависимости GEMBA-MQM от проприетарной БЯМ GPT-4 [8].

Поиск подходящих алгоритмов и метрик оценки качества переводов с учётом указанных выше условий может стать отправной точкой для адекватной оценки качества выполняемых БЯМ переводов и поспособствовать дальнейшему развитию способностей данных моделей к выполнению переводов требуемого качества.

Список литературы

1. АКМ-Вест. Как оценивается качество машинного перевода? URL: https://wiki.akmw.ru/index.php?title=Как_оценивается_качество_машинного_перевода%3F (дата обращения 07.12.2025).
2. Аменицкий А.В., Рухович И.В., Аменицкая Л.А., Аменицкий Д.А. Побочные эффекты галлюцинаций искусственного интеллекта // Наука, инновации, образование: актуальные вопросы и современные аспекты. Пенза: Наука и Просвещение, 2024. С. 224–235.
3. Толстикова Е.К. Особенности перевода имен собственных с применением ИИ-технологии Yandex GPT (на примере имен английских монархов) // В многомерном пространстве современной лингвистики: Сборник материалов Международной молодежной конференции, Москва, 10 декабря 2024 года. Москва: Р.Валент, 2025. С. 154–158.
4. ООО «Яндекс.Облако». Обзор AI-моделей сервиса Yandex AI Studio. URL: <https://yandex.cloud/ru/docs/ai-studio/concepts/generation/#yandex> (дата обращения 04.12.2025).
5. Anthropic PBC. Claude Models Overview. URL: <https://platform.claude.com/docs/en/about-claude/models/overview> (accessed at 04.12.2025).
6. BigScience. Introducing The World's Largest Open Multilingual Language Model: BLOOM. URL: <https://bigscience.huggingface.co/blog/bloom> (accessed at 04.12.2025).
7. Cayuela M.N. Mirror, mirror, which is the fairest metric of them all? URL: <https://www.aiothinktank.com/post/mirror-mirror-which-is-the-fairest-metric-of-them-all> (accessed at 06.12.2025).
8. Hugging Face Inc. BigScience Large Open-science Open-access Multilingual Language Model. URL: <https://huggingface.co/bigscience/bloom> (accessed at 03.12.2025).
9. Kocmi T., Federmann, C.. GEMBA-MQM: Detecting translation quality error

spans with GPT-4 // arXiv preprint. 2023. arXiv:2310.13988. URL: <https://arxiv.org/abs/2310.13988> (accessed at 06.12.2025).

10. Pande A., Patil R., Mukkemwar R., Panchal R., Bhoite, S. Comprehensive Study of Google Gemini and Text Generating Models: Understanding Capabilities and Performance // 15th International Conference on Advances in Computing, Control, and Telecommunication Technologies. Hyderabad, India, 2024. Pp. 856–863.

11. Rahman M., Shiplu A.I., Watanobe Y. ChatGPT and DeepSeek: Strengths, Limitations, and the Future of Generative AI // Journal of LaTeX class files. Vol. 18. № 9. Pp. 1–20.

12. Sharkas H. Exploring the Role of Translation Brief Elements in Prompts to Large Language Models // AWEJ for Translation & Literary Studies. 2025 Vol. 9. № 2. Pp. 139–153.

13. Varga L., Ciesielski J. ChatGPT & LLMs – Separating Fact from Fiction for Localization. URL: <https://www.nimdzi.com/chatgpt-llm-separating-fact-from-fiction-for-localization/> (accessed at 03.12.2025).

14. Wołk K., Marasek K. Neural-based Machine Translation for Medical Text Domain. Based on European Medicines Agency Leaflet Texts // Procedia Computer Science. 2015. Vol. 64. Pp. 2–9.

Об авторах:

МАЛЁНОВА Евгения Дмитриевна – к. филол. н., доцент, декан факультета филологии, переводоведения и медиакоммуникаций Омский государственный университет им. Ф.М. Достоевского (644053, г. Омск, ул. Нефтезаводская д. 11); e-mail: malenovae@mail.ru

ВИНСКИЙ Дмитрий Павлович – аспирант, 1 курс, Омский государственный университет им. Ф.М. Достоевского (644053, г. Омск, ул. Нефтезаводская д. 11); e-mail: winski16@gmail.com

LARGE LANGUAGE MODELS IN TRANSLATION PRACTICE

E.D. Malenova, D.P. Winskii

Dostoevsky Omsk State University, Omsk

In this article authors provide a look into specifics of using large language models (LLM) in translation. An overview of several language models is provided along with a description of how LLMs may be used by translators in comparison with neural machine translation systems. Authors also provide recommendations on how to approach assessing quality of translations made with use of AI models with the specifics of user-AI interaction in mind.

Keywords: *translation, generative artificial intelligence, large language models, LLM, neural machine translation, machine translation quality assessment.*

About authors:

MALENOVA Evgeniya Dmitrievna – PhD in Philology, Associate Professor, Dean of the Faculty of Philology, Translation Studies and Media Communications, Dostoevsky Omsk State University (Neftozavodskaya Ulitsa, 11, Omsk, Omsk Oblast, 644053); e-mail: malenovae@mail.ru

WINSKII Dmitrii Pavlovitch – post-graduate, first year Dostoevsky Omsk State University (Neftezavodskaya Ulitsa, 11, Omsk, Omsk Oblast, 644053); e-mail:winski16@gmail.com

Статья поступила в редакцию 19.03.26
Подписана в печать 03.04.26

© Малёнова Е.Д.,
Винский Д.П., 2026