

*Ширяев В.А.,  
студент магистратуры, 1 курс, Институт радиоэлектроники и  
информационных технологий - РтФ  
Уральский федеральный университет имени первого  
Президента России Б. Н. Ельцина,  
Россия, г. Екатеринбург*

## **СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЯЗЫКОВЫХ МОДЕЛЕЙ NLP ДЛЯ ПОДДЕРЖКИ ЛИТЕРАТУРНОГО ТВОРЧЕСТВА**

***Аннотация:** Статья посвящена сравнительному анализу архитектурных и функциональных особенностей больших языковых моделей (англ. *Large language model, LLM*), используемых в качестве интеллектуальных ассистентов литератора. Рассматриваются ключевые критерии оценки моделей для творческих задач: объём контекста, качество стилизации, способность удерживать сюжетную линию и следование длинным инструкциям. В рамках исследования проведена типология современных LLM таких как *ChatGPT-4o, Claude 3.5 Sonnet, Gemini, Llama 3, YandexGPT, GigaChat* и специализированных писательских моделей (*Sudowrite Muse, NovelAI*). Выявлены наиболее эффективные конфигурации для различных этапов литературного труда: от мозгового штурма до финальной стилистической обработки.*

***Ключевые слова:** большие языковые модели, литературное творчество, обработка естественного языка, сравнительный анализ LLM, креативное письмо, стилистическая генерация, бенчмарк.*

## **COMPARATIVE ANALYSIS OF NLP LANGUAGE MODELS FOR LITERARY CREATIVITY SUPPORT**

**Annotation:** *The article is devoted to a comparative analysis of the architectural and functional features of large language models (LLM) used as intellectual assistants to the writer. The key criteria for evaluating models for creative tasks are considered: the amount of context, the quality of stylization, the ability to keep the storyline and follow long instructions. The study provides a typology of modern LLMs such as ChatGPT-4o, Claude 3.5 Sonnet, Gemini, Llama 3, YandexGPT, GigaChat and specialized writing models (Sudowrite Muse, NovelAI). The most effective configurations for various stages of literary work have been identified: from brainstorming to final stylistic processing.*

**Key words:** *large language models, literary creativity, natural language processing, comparative analysis of LLMs, creative writing, stylistic generation, benchmark.*

## **Введение**

Трансформация литературного творчества под влиянием технологий искусственного интеллекта достигла этапа, на котором литератор уже не просто пользуется инструментами проверки орфографии или анализа стиля, а взаимодействует с полноправным интеллектуальным соавтором. Современные большие языковые модели способны не только завершать фразу, но и выстраивать драматургически целостные сцены, имитировать авторский стиль и генерировать многослойные нарративы на основе сложных промптов. Однако многообразие LLM с их различными архитектурами, объёмами параметров и ориентированностью на ту или иную языковую или жанровую специфику ставит перед литератором и исследователем проблему выбора: какая модель оптимальна для генерации диалогов, какая — для иммерсивного миростроения, а какая — для глубокой стилистической правки?

Актуальность данного исследования обусловлена отсутствием в отечественной и зарубежной литературе комплексного сопоставления LLM именно в разрезе задач художественного, а не технического или делового

письма. Цель статьи — провести сравнительный анализ современных языковых моделей как интеллектуальных ассистентов литератора, выявив их сильные и слабые стороны на разных этапах творческого процесса. Достижение цели предполагает решение задач: классификации моделей по функциональному назначению, оценки их архитектурных особенностей (контекстное окно, метод обучения, специализация), сопоставления качества генерации художественного текста и формулирования практических рекомендаций.

### **Методология и критерии оценки**

В основе работы лежит сравнительно-функциональный анализ, дополненный элементами качественной оценки выходных текстов. Выборка моделей включает как универсальные LLM (ChatGPT-4o, Claude 3.5 Sonnet, Gemini 1.5/2.0, Llama 3.1 70B/405B, YandexGPT 5 Pro, GigaChat 2 Max), так и специализированные писательские решения (Sudowrite Muse, NovelAI, RuGPT). Оценка проводилась по следующим критериям:

1. Объём контекстного окна — способность модели удерживать и обрабатывать длинные фрагменты текста (важно для поддержания сюжетной линии в крупной форме).

2. Качество стилизации — точность имитации заданного авторского стиля или жанровой тональности.

3. Способность удерживать сюжетную связность — внутренняя когерентность повествования на протяжении десятков тысяч токенов.

4. Следование длинным инструкциям — точность исполнения многошаговых творческих заданий.

5. Работа с русским языком — для оценки использованы данные MERA-бенчмарка и специализированных исследований.

В качестве дополнительного параметра учитывалась специфика обучения: дообучение на корпусах художественной литературы повышает литературные качества модели [9, с. 31–34], тогда как универсальные модели

демонстрируют большую гибкость, но требуют более тщательной ручной доработки [5, с. 7–10].

### **Архитектурные и функциональные особенности моделей**

ChatGPT-4o (OpenAI). Универсальный мультимодальный ассистент с одним из самых быстрых откликов — среднее время ответа составляет 1,9 секунды, что заметно превосходит конкурентов по скорости [1, с. 31]. Обладая значительным стилистическим диапазоном, GPT-4o успешно справляется с задачами структурирования, создания планов, мозгового штурма и написания диалогов [7, с. 13–16]. Исследование MIT продемонстрировало, что модель, дообученная на 50 известных писателях, генерирует тексты, которые большинство читателей оценивают выше оригинальных авторских произведений [9, с. 31–35]. СтилOMETрический анализ, проведённый G. Mikros (2025), подтверждает способность GPT-4o к точной имитации контрастных литературных стилей — например, Эрнеста Хемингуэя и Мэри Шелли [9, с. 4–11].

Однако существуют и ограничения. GPT-4o использует пассивные конструкции вдвое реже, чем человек, и склонна к информационно насыщенному, перегруженному существительными стилю [9, с. 42–45], что обедняет вариативность повествования. Кроме того, фиксируются случаи регрессии в нарративной глубине [9, с. 23–29], требующие от литератора усиленного контроля.

Claude 3.5 Sonnet (Anthropic). Широко признаётся одной из лучших моделей для художественной литературы. Её ключевое преимущество — способность писать «человечным» языком и строго следовать длинным инструкциям [7, с. 11–15]. В многоязычном бенчмарке RWS Claude 3.5 Sonnet занял первое место по качеству в шести из восьми тестируемых языков, стабильно получая оценку выше 4,0 по 5-балльной шкале во всех языках, кроме самого сложного (Киньяруанда) [4, с. 36–40].

Модель демонстрирует исключительное мастерство в стихоподражании и поэтической стилизации [9, с. 5–8]. При этом скорость Claude 3.5 заметно уступает GPT-4o: среднее время ответа составляет 5,8 секунды против 1,9 [1, с. 31]. Для литератора это означает необходимость балансировать между качеством стиля и оперативностью интеракции.

Gemini 1.5 Pro / 2.0 (Google). Основное преимущество — сверхдлинное контекстное окно, достигающее десятков тысяч токенов. Это делает модель идеальным инструментом для анализа «лора» сложных фэнтези-вселенных, проверки согласованности деталей на всём протяжении романа и ведения многотурных диалогов [7, с. 20–22]. В рейтинге RWS Gemini 1.5 Pro занял третье место после Claude 3.5 Sonnet и GPT-4o, демонстрируя высокое качество многоязычной генерации [4, с. 41–44]. Интеграция с экосистемой Google (Документы, Почта, Поиск) предоставляет литератору удобный инструментарий для сбора материалов и первичного структурирования [5, с. 32–35].

Llama 3.1 (Meta) и её производные. 70B-параметрическая модель Llama 3.1 представляет собой основу для множества специализированных дообучений. На её базе созданы модели Storywriter, ArliAI-RPMax, Spellbound-StoryWriter и др. [5, с. 4–44], оптимизированные для ролевых игр и творческого письма с контекстным окном до 128K токенов. Ключевой недостаток — скорость: Llama 3.1 70B демонстрирует 6,6 секунды отклика, что заметно медленнее GPT-4o [1, с. 31]. При этом для русскоязычной аудитории Llama 3.3 70B в ряде лингвистических тестов уступает как Mistral Large, так и локальным русскоязычным моделям [7, с. 28–31].

Русскоязычные модели: YandexGPT 5 Pro, GigaChat 2 Max. Эти модели ориентированы прежде всего на работу с русским языком и демонстрируют высокую точность в лингвистических задачах — например, в градуальной оценке приемлемости синтаксических конструкций YandexGPT превосходит GigaChat [7, с. 29–31]. Общий лидер российского рынка LLM по версии MERA

— GLM-5.1, с общим результатом 0,804 и результатами выше человеческих по целому ряду задач [8, с. 2–16].

Специализированные писательские модели: Sudowrite Muse, NovelAI. Muse разработана создателями Sudowrite как модель, целенаправленно обученная художественной прозе. Её авторы характеризуют возможности Muse как соответствующие уровню студента или начинающего писателя [6, с. 8–14]. В отличие от универсальных моделей, Muse специализируется на сенсорных «гипертекстовых» дополнениях — расширении описаний с акцентом на визуальные, обонятельные, осязательные впечатления, а также на метафорическое развитие текста [6, с. 24–38]. NovelAI, в свою очередь, предоставляет возможность выбора стиля повествования и генерации историй «с нуля», ориентируясь на жанровую специфику [7, с. 18–19].

### **Сравнительный анализ и обсуждение результатов**

Проведённое сопоставление позволяет сформулировать несколько ключевых закономерностей. Во-первых, универсальные модели типа ChatGPT-4o и Claude 3.5 Sonnet лидируют по креативности и гибкости, однако требуют более активной ручной доработки; специализированные инструменты (Sudowrite, NovelAI) лучше справляются с академическим или литературоведческим стилем, но предлагают меньшую универсальность [5, с. 7–10]. Во-вторых, конкуренция между мультязычными LLM и русскоязычными моделями на русскоязычном материале остаётся открытой — как показало исследование вариативного согласования, лучшие результаты демонстрирует мультязычная Mistral Large, в то время как YandexGPT превосходит GigaChat [7, с. 28–31].

Наиболее сложной остаётся задача объективной оценки качества сгенерированного художественного текста. Традиционные метрики (BLEU, ROUGE) измеряют лишь совпадение n-грамм с эталонным текстом и непригодны для оценки художественной ценности [13, с. 24–28]. Метрики на основе контекстуальных эмбеддингов (BERTScore) учитывают

семантическую близость, но не креативность или стилистическую целостность [14, с. 29–33]. Подход «LLM-как-судья» набирает популярность, но рискует воспроизводить системные смещения самих моделей [13, с. 14–18]. MAUVE измеряет «разрыв» между распределением сгенерированных текстов и человеческих, но не оценивает содержательную сторону [12, с. 4–8].

Особую сложность представляет субъективный характер литературного качества: как показывают исследования, читатели с различным уровнем экспертизы ценят разные аспекты текста [10, с. 7–17]. «Поверхностные» читатели отдадут приоритет читаемости и лексической насыщенности, тогда как эксперты-литераторы акцентируют тематическое развитие, риторическое разнообразие и динамику эмоций [10, с. 17–22].

Для литератора, работающего с русскоязычными текстами, рекомендуется комбинированная стратегия: на этапе генерации идей и черновиков — использование Claude 3.5 Sonnet или GPT-4o (с учётом их сильных стилистических качеств), для длинных аналитических задач — Gemini благодаря её сверхдлинному контекстному окну, а для финальной доработки и устранения «англицизмов» — дообученные модели на базе Llama 3.1 или специализированные русскоязычные модели (YandexGPT 5 Pro, GLM-5.1), показавшие лучшие результаты в российских бенчмарках.

### **Заключение**

Современные языковые модели представляют собой спектр инструментов с различными архитектурными компромиссами. Claude 3.5 Sonnet лидирует по «человечности» стиля и способности следовать сложным инструкциям, GPT-4o — по скорости и универсальности, Gemini — по объёму контекста, Llama 3.1 — как основа для специализированных писательских дообучений. Русскоязычные YandexGPT 5 Pro, GigaChat 2 Max и GLM-5.1 достигают высокой точности, но ещё не всегда превосходят мультиязычных конкурентов на всех типах задач.

Дальнейшие исследования должны быть направлены на три магистральных направления: разработку релевантных метрик оценки художественного качества генерации, не сводимых к формальным совпадениям; изучение возможностей «механистического» управления стилем (на основе анализа индивидуальных нейронов, как продемонстрировано на GPT-2 [11, с. 11–21]); а также создание открытых бенчмарков для творческого письма на русском языке по аналогии с MERA. Только при соблюдении этих условий LLM станут не просто инструментом, а полноценным расширением творческих возможностей литератора, сохраняющим его авторский суверенитет.

#### **Использованные источники:**

1. Сравнение скорости LLM: GPT-4o (1.9 с), Claude 3.5 (5.8 с), Llama 3.1 (6.6 с) // DotData. 2025. [Электронный ресурс]. URL: <https://jp.dotdata.com> (дата обращения: 14.05.2026).
2. Какие нейросети лучше всего подходят для написания книг // Author.Today. 2026. [Электронный ресурс]. URL: <https://author.today/post/846262> (дата обращения: 14.05.2026).
3. ТОП-10 лучших нейросетей для написания книги // vc.ru. 2025. [Электронный ресурс]. URL: <https://vc.ru/id5463987/2668231-top-10-neurosetey-dlya-napisaniya-knigi> (дата обращения: 14.05.2026).
4. LLM benchmarking leaderboard: Languages, creativity and tasks // RWS. 2025. [Электронный ресурс]. URL: <https://www.rws.com/blog/llm-benchmarking-leaderboard-languages-creativity-tasks/> (дата обращения: 14.05.2026).
5. Популярные LLM в 2025–2026: что выбрать новичку (ChatGPT, Claude, Gemini, Llama) // AI-Nov. 2025. [Электронный ресурс]. URL: <https://ai-nov.ru/2025/06/02/iskusstvennyj-intellekt-dlya-novichkov-42/> (дата обращения: 14.05.2026).

6. Sudowrite: Muse AI model for prose writing // Techsty. 2025. [Электронный ресурс]. URL: [https://techsty.art.pl/elektroblog/Sudowrite\\_AI\\_dla\\_pisarzy.html](https://techsty.art.pl/elektroblog/Sudowrite_AI_dla_pisarzy.html) (дата обращения: 14.05.2026).
7. Студеникина К.А., Лютикова Е.А., Герасимова А.А. Оценка лингвистической компетенции больших языковых моделей на материале корпуса согласовательной вариативности // Computational Linguistics and Cognitive Computing. 2025. Т. 9. С. 100–119.
8. GLM-5.1 MERA Leaderboard // MERA — Альянс в сфере ИИ. 2026. [Электронный ресурс]. URL: <https://mera.a-ai.ru/ru/text/submits/16044> (дата обращения: 14.05.2026).
9. Beyond the surface: stylometric analysis of GPT-4o’s capacity for literary style imitation // Oxford Academic. 2025. [Электронный ресурс]. URL: <https://preview.academic.oup.com> (дата обращения: 14.05.2026).
10. Marco G., Gonzalo J., Fresno V. The Reader is the Metric: How Textual Features and Reader Profiles Explain Conflicting Evaluations of AI Creative Writing // arXiv:2506.03310. 2025.
11. Enkhbayar Т.-О. Atomic Literary Styling: Mechanistic Manipulation of Prose Generation in Neural Language Models // arXiv:2510.17909. 2025.
12. MAUVE — автоматическая метрика для оценки качества текста // Systems-Analysis.ru. 2025. [Электронный ресурс]. URL: <https://systems-analysis.ru/mauve> (дата обращения: 14.05.2026).
13. Метрики для генеративных моделей: LLM-as-a-Judge // HSE. 2025. [Электронный ресурс]. URL: <https://www.hse.ru> (дата обращения: 14.05.2026).
14. BERTScore для русского языка // CyberLeninka. 2025. [Электронный ресурс]. URL: <https://cyberleninka.ru> (дата обращения: 14.05.2026).