



DOI: 10.22363/2313-2299-2026-17-2-427-441

EDN: LIUVUU

Научная статья / Research Article

В поисках алгоритма определения текста, сгенерированного нейросетевой моделью, с опорой на его языковые особенности

Л.Р. Комалова  

Институт научной информации по общественным наукам Российской академии наук,
Москва, Российская Федерация

Московский государственный лингвистический университет, *Москва, Российская
Федерация*

 komalova@inion.ru

Аннотация. С 2023 г. наблюдается устойчивый рост научных публикаций, в которых присутствует текст, сгенерированный нейронными сетями на основе больших языковых моделей. Развитие методов машинного обучения и глубокого обучения применительно к задачам обработки естественного языка приближает результат генерации текста к текстам, написанным человеком. На данном этапе развития лингвистика не располагает разработанным инструментарием, позволяющим разграничить текст, написанный человеком, и текст, сгенерированный нейросетью. Цель проводимого исследования заключается в выработке алгоритма формальной проверки рукописи, поступившей в редакцию научного журнала / научному руководителю, с опорой на доступные научному редактору / научному руководителю инструменты (в основном открытого доступа) детекции сгенерированного нейросетями текста на русском языке. Методология исследования опирается на результаты изысканий в области компьютерной и математической лингвистики по описанию «стиля письма» различных нейросетей при генерации текстов научных жанров на русском языке. Материалом исследования послужили научные публикации за 2021–2025 гг., доступные через базы данных научной информации, авторами которых являются исследователи, аффилированные преимущественно с российскими научными и образовательными организациями, а также с российскими технологическими компаниями. В ходе исследования было выявлено, что практика применения Gen-AI при написании научных текстов намного опережает рефлексию ее результатов и прогнозирование последствий в долгосрочной перспективе. При этом калейдоскоп публикаций, описывающих отличительные черты результата генерации различными нейросетями, позволяет выработать подход к обнаружению подобных текстов, опираясь на ошибки в генерации на различных языковых уровнях.

Ключевые слова: генеративный искусственный интеллект, обнаружение сгенерированного текста

© Комалова Л.Р., 2026



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Финансирование: Исследование выполнено в рамках государственного задания ФГБУН «Институт научной информации по общественным наукам РАН» по теме «Новейшие тенденции в развитии гуманитарной мысли России и Запада: философские течения, развитие культуры и духовная жизнь общества в цифровую эпоху».

Конфликт интересов: автор заявил об отсутствии конфликтов интересов.

История статьи: дата поступления: 23.12.2025; дата приема в печать: 15.02.2026.

Для цитирования: Комалова Л.Р. В поисках алгоритма определения текста, сгенерированного нейросетевой моделью, с опорой на его языковые особенности // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2026. Т. 17. № 2. С. 427–441. <https://doi.org/10.22363/2313-2299-2026-17-2-427-441> EDN LIUVUU

Towards an Algorithm for Identifying Text Generated by Neural Network Models through Linguistic Cues

Liliya R. Komalova  

Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences,
Moscow, Russian Federation
Moscow State Linguistic University, Moscow, Russian Federation
[✉komalova@inion.ru](mailto:komalova@inion.ru)

Abstract. Since 2023, there has been a steady increase in research papers containing text generated by neural networks based on large language models. The development of machine and deep learning methods applied to natural language processing tasks is bringing the results of text generation closer to those of human authors. At this stage of development, linguistics lacks practical tools to distinguish between human-written text and text generated by a neural network. The goal of this study is to develop an algorithm for manuscripts formal verification by editors of scientific journals / academic supervisors, relying on (primarily open access) tools available to the scientific editor / supervisor for detecting text generated by neural networks in Russian language. The study methodology is based on results of computational and mathematical linguistics researches to describe the “writing style” of various neural networks when generating scientific texts in Russian language. We analyzed research papers issued from 2021 to 2025, accessible through scientific information databases, authored by researchers primarily affiliated with Russian scientific and educational organizations, as well as Russian technology companies. The study revealed that the use of Gen-AI in academic writing significantly outpaces the reflection of its results and the prediction of its long-term implications. Furthermore, the plethora of publications describing the distinctive features of the output generated by various neural networks allows for the development of an approach to detecting such texts based on generation errors at various linguistic levels.

Keywords: generative artificial intelligence, generated text detection, linguistic features of generated text

Financing: The research is carried out within the framework of the state assignment to the Federal State Budgetary Institution of Science «Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences», project «The latest trends in the development of humanitarian thought in Russia and the West: philosophical movements, the development of Cultural Studies and the spiritual life of society in the digital era».

Conflict of interests: The author states that there were no conflicts of interest.

Article history: Received: 23.12.2025; Accepted: 15.02.2025.

For citation: Komalova, L.R. (2026). Towards an Algorithm for Identifying Text Generated by Neural Network Models through Linguistic Cues. *RUDN Journal of Language Studies, Semiotics and Semantics*, 17(2), 427–441. (In Russ.). <https://doi.org/10.22363/2313-2299-2026-17-2-427-441> EDN LIUVUU

Генеративный искусственный интеллект — новая текстовая реальность

В статье 5 Национальной стратегии развития искусственного интеллекта на период до 2030 г., утвержденной Указом Президента Российской Федерации от 10 октября 2019 г. № 4901, искусственный интеллект определяется как «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их»¹. В стратегии говорится об ускоренном внедрении технологических решений, разработанных на основе искусственного интеллекта, в различные отрасли экономики и сферы общественных отношений. В документе указывается как на позитивные изменения в ключевых сферах жизнедеятельности, которые сопровождаются использованием технологий искусственного интеллекта (ИИ), так и на негативные последствия, вызванные социальными и технологическими изменениями, которые сопутствуют развитию технологий искусственного интеллекта.

Многие эксперты и практики отмечают, что применение ИИ-инструментов имеет потенциал для преобразования различных сфер жизни, включая образование и научные исследования [1; 2]. Всплеск активности на уровне рядовых пользователей произошел после введения в массовое пользование предназначенных для генерации текста в письменной форме чат-ботов на основе предобученных нейронных сетевых языковых моделей, основанных на архитектуре типа «трансформеры» (generative pre-trained transformer, GPT), на механизме самовнимания (self-attention) [3] и на парадигме самообучения (self-supervised learning) на огромном объеме текстовых данных. В отличие от традиционных задач классификации и регрессии генеративный ИИ (Gen-AI) обладает способностью автономно генерировать новые данные, включая изображения, музыку и текст [4; цит. по: 5].

Генерация текста реализуется с помощью двух базовых подходов [6. С. 283]: (1) использование предопределенного шаблона для генерации текста на основе различных входных данных (например, используя предопределенный шаблон, искусственный интеллект может сгенерировать определенное описание продукта на основе типа продукта, его характеристик и преимуществ); (2) использование метода неконтролируемого (глубоком)

¹ См.: Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» URL: <https://base.garant.ru/72838946/?ysclid=ml9m8kv11q126675642> (дата обращения: 05.12.2025).

обучения, который изучает нюансы языковых структур и функционирует с условием наличия больших объемов входных данных — по сравнению с подходом на основе шаблонов результат генерации текста на основе метода глубокого обучения приближается к тексту, написанному человеком.

«Генеративные нейросети настроены на создание наиболее предсказуемых, статистически чаще встречающихся последовательностей слов, вызывающих наименьшее недоумение у читателя и с большей вероятностью воспринимаемых как осмысленные» [7. С. 91]. В основе работы моделей лежат статистические пермутации (перестановки элементов), «пропущенные через миллионные обучающие выборки на примерах» [8. С. 470]. Развитие методов машинного глубокого обучения применительно к задачам обработки естественного языка «позволяет получать искусственно созданные тексты, степень сходства которых с написанными человеком с годами становится лишь выше» [9. С. 308]. Однако статистическое «понимание» языка, которым обладают генеративные нейросетевые модели, отлично от семантического понимания языка человеком [10. С. 235].

Сгенерированный текст как лингвистическая проблема

Начиная с 2020 г. мир захлестнула волна публикаций, искусственно сгенерированных с помощью технологии GPT под видом статей, написанных человеком [11]. К настоящему времени «сгенерированные нейросетями тексты уже стали частью нашей коммуникативной среды и культуры» [12. С. 790]. В частности, в образовательной среде возникла коллизия между широким функционалом Gen-AI (системами аннотирования, реферирования, перефразирования для подготовки и написания рефератов, обзорных и исследовательских статей, выпускных квалификационных работ) и развитием творческих исследовательских способностей у обучающихся вузов, а также возможностями преподавателей оценить интеллектуальные компетенции обучающихся в сложившейся ситуации [13. С. 58—59]. Опасность появления большого количества синтетических данных заключается в том, что они становятся (пусть и формально) частью научной онтологии, материалом для обучения следующего поколения генеративных нейросетевых языковых моделей, в то время как «программы-детекторы работают пока недостаточно надежно» [8. С. 476].

Можно выделить несколько типов недобросовестного использования генеративных нейросетей [9; 14]:

- из-за огромного объема данных для обучения модель генерирует ответ, который не всегда соответствует действительности, при этом пользователи не всегда проверяют информацию, полученную в результате генерации;
- генерация научного или учебного текста, гарантированно проходящего проверку на наличие плагиата;
- перефразирование текста для того, чтобы затруднить системам обнаружения заимствований выявление того, что текст был ранее опубликован.

Задача «проверяющего» усложняется тем, что текст чаще всего не генерируется моделями целиком: как правило, это умелая комбинация машинно сгенерированных и написанных человеком текстовых фрагментов [15; 16].

Следует зафиксировать, что на данном этапе развития лингвистика не располагает разработанным инструментарием (устойчивым комплексом признаков), позволяющим разграничить текст, написанный человеком, и текст, сгенерированный нейросетью [17. С. 165]. При этом «тексты одной жанровой принадлежности, но сгенерированные разными языковыми моделями, существенно различаются» как по языковым параметрам, так и по ошибкам, которые делают эти генеративные нейросети [18; цит. по: 12. С. 789]. Результат генерации зависит также и от задачи, которую пользователь формулирует в промпте (сбор, реферирование, аннотирование текстового материала, его перефразирование и т.д.), и от субъективных показателей автора промпта [13. С. 62].

Цель проводимого нами исследования заключается в выработке алгоритма формальной проверки рукописи, поступившей в редакцию научного журнала / научному руководителю, с опорой на доступные научному редактору / научному руководителю инструменты (в основном открытого доступа) детекции сгенерированного нейросетями текста на русском языке.

Методология исследования опирается на результаты изысканий в области компьютерной лингвистики по описанию «стиля письма» различных нейросетей при генерации текстов научных жанров на русском языке.

Материалом для исследования послужили научные публикации за 2021–2025 г., доступные через базы данных научной информации, авторами которых являются исследователи, аффилированные преимущественно с российскими научными и образовательными организациями, а также с российскими технологическими компаниями.

Автоматическая детекция сгенерированного текста

Как отмечается в ряде работ в области математической лингвистики, существующие методы не решают проблемы детекции текстов большого объема [9], так как длина фрагментов, на которые разбивается анализируемый текст, оказывает влияние на результат детекции [19; 20]. Для разных языков оптимальная длина варьируется, но в каждом существует плато, позволяющее зафиксировать качество классификации [9]. Для русского языка это порядка 4000 символов.

Высокий уровень полноты детекции и большая обобщающая способность модели достигаются при обучении детектирующей модели на мультидоменных текстах: чем больше разнородных по тематикам текстов будет на стадии обучения, тем выше будут метрики качества классификации при использовании детектора в реальных задачах поиска искусственно сгенерированных фрагментов текста [21; цит. по: 9. С. 312].

Автоматическая детекция сгенерированного текста строится на различных статистических метриках [6; 7; 22–24]: частоте совместной встречаемости n -грамм, вычислении коэффициентов перекрытия слов, длины предложений, распределения слов в предложении, значений «бурности» (burstiness) и «недоумения» (perplexity), вычислении различных метрик сложности текста. Если после генерации текста нейросетью проводится постредактирование, его идентификация становится еще более сложной. Известны следующие подходы к детекции такого типа искусственно сгенерированных текстов [25. С. 963]:

- обнаружение неконтекстуальных токенов (non-contextual tokens): например, ChatGPT-модель часто вставляет специальные токены или маркеры в свои ответы для указания начала или конца ответа;

- анализ статистических свойств сгенерированного текста: распределение слов, грамматические или стилистические особенности (проверка использования редких или специфичных слов и фраз, обнаружение неправильной грамматики или «плохого» стиля);

- выявление метаданных: для обнаружения применения, например, ChatGPT-модели используется дополнительная информация, такая как сведения о запросах, метаданные сеанса и временные отметки;

- анализ контекста вопросов и ответов: если в ответах модели отсутствуют устойчивые или явные связи с предыдущими вопросами либо контекст быстро теряется, это может быть признаком автоматической генерации;

- сравнение следующих токенов при совпадающих предыдущих (идентичном левом контексте) в проверяемом тексте и в тексте, генерируемом GPT-подобной моделью;

- проверка фрагментов текста путем их поиска в интернет-архивах.

Системы, детектирующие искусственно сгенерированный текст, широко представлены в сети Интернет². Результат детекции подается либо в виде ответа типа «текст сгенерирован / текст написан человеком», либо в виде указания доли предположительно искусственно сгенерированного текста.

Экспертное обнаружение сгенерированного текста

В результате анализа научной литературы было принято решение объединить выявленные маркеры сгенерированности текста в три группы. Макроуровень охватывает всю рукопись и также касается взаимодействия структурных частей рукописи между собой, мезоуровень относится к отдельным абзацам и взаимодействию между ними, микроуровень характеризует особенности текста на уровне предложения и взаимодействия между предложениями в рамках одного абзаца.

² См., например: GigaCheck. URL: <https://developers.sber.ru/portal/products/gigacheck> (accessed: 05.12.2025); GPTZero. URL: <https://gptzero.me> (accessed: 05.12.2025); ZeroGPT. URL: <https://www.zerogpt.com> (accessed: 05.12.2025); Copyleaks. URL: <https://copyleaks.com> (accessed: 05.12.2025); Contentatscale. URL: <https://ai.tenorshare.com/products/contentatscale-ai-detector> (accessed: 05.12.2025) и др.

Макроуровень

Наиболее частотной ошибкой, упоминаемой в научной литературе, которая отличает сгенерированный текст от текста, написанного человеком, является некорректная цитация и составление списка литературы. Внутритекстовые отсылки к списку литературы в сгенерированных текстах либо не соответствуют списку литературы, либо являются вымышленными, т.е. таких источников в научной онтологии не существует, а указанные в списке идентификаторы (например DOI) связаны с иными публикациями [24; 26–28]. Частота данной особенности сгенерированных текстов вынуждает экспертов (научных редакторов и руководителей) проверять каждую цитату и каждую позицию в списке литературы.

Второй и более серьезной частотной погрешностью сгенерированных текстов является формулировка недостоверной информации и искажение фактов [5; 12; 25; 29], что требует высокой компетенции проверяющего в предметной области и тематике проверяемой рукописи. Фактчекинг — серьезный вызов, т.к. в случае включения недостоверной информации в научный оборот, впоследствии она может лечь в основу методологии исследования неопытного исследователя, что повлечет за собой дальнейшее искажение научных данных.

Исследователи отмечают, что сгенерированные нейросетью тексты носят поверхностный характер (на основе поверхностного анализа), отличаются неясностью формулировок, в таких текстах взаимосвязь между его частями не всегда прослеживается, исследуемый вопрос в полной мере не раскрывается, такие тексты содержат общую информацию и имеют более простую структуру [5; 7; 29–32]. При генерации текста обзорной части рукописи выявляются многочисленные ошибки при перефразировании, а также явный плагиат [33]. Такие тексты тяготеют к публицистическому, а иногда и обиходно-разговорному функциональному стилю письма. При этом «сгенерированные тексты отличает более четкая организация, соответствующая правилам IMRaD» [24. С. 1002].

Отмечается «неспособность языковых моделей выстраивать долгосрочные семантические зависимости в тексте, что иногда приводит к „вырождению“ текста» [34; цит. по: 17. С. 166]. В сгенерированных нейросетью текстах отмечается единообразный стиль изложения, в то время как текст, написанный человеком, «может иметь изменения стиля и тона по всему тексту» [25. С. 962]. В силу программных особенностей больших языковых моделей в порождаемых нейросетями текстах высока лексическая плотность³ при низком лексическом разнообразии⁴ [5].

³ Количественная характеристика обширности словарного запаса — соотношение содержательных слов (самостоятельных частей речи: имен существительных, имен прилагательных, глаголов и наречий) в тексте к общему количеству слов в тексте.

⁴ Количественная характеристика текста, отражающая степень разнообразия словаря при построении текста заданной длины — соотношение отдельных лексических единиц (лемм) и количества их употребления в тексте (словоформ).

Сгенерированный нейросетью текст отличает «отсутствие идеи, объективно или субъективно характеризующей наличие нового знания» [7. С. 92]. Такой текст не содержит научной новизны, не имеет теоретической и практической значимости [29]. За сгенерированным текстом не «проглядывается» личность автора [24]. Эти особенности может обнаружить только человек — эксперт в предметной области, затронутой в тексте.

Мезоуровень

К особенностям, позволяющим обнаружить сгенерированные фрагменты текста, можно отнести [29. С. 155; 35. С. 50–51]:

- «закольцованность» темы (т.е. повтор темы без ввода рематического компонента);
- наличие множества рем, ни одна из которых в дальнейшем не преобразуется в тему и не продолжает свое текстовое развитие;
- недостаточный учет факторов автора и адресата;
- маркированность текста показателями модуса предположения, некатегоричности, сигнализирующими о наличии других, альтернативных вариантов ответа;
- повторение одной и той же мысли в разных предложениях;
- отсутствие связи сгенерированного фрагмента с остальным текстом.

Микроуровень

На уровне предложений сгенерированные тексты обладают меньшей степенью связности, чем тексты, написанные человеком. В сгенерированных текстах обнаруживается «доставление нарратива причинно-следственными операторами *поэтому, так что, вот и, действительно* и т.п., что устраняет „нарушения“ в оригинальном тексте-источнике» [12. С. 789]. Генеративные «нейросети создают не только предсказуемый порядок слов, но и строят предложения с несложной структурой со средней длиной порядка 10–20 слов» [7. С. 91], в то время как текст, написанный человеком, более насыщен причастными и деепричастными оборотами.

В сгенерированных текстах встречаются нарушения синтаксической сочетаемости; избыточность клише, штампов, канцеляризмов и слов с более высоким уровнем формальности; неестественная сочетаемость и частота специфической терминологии, которая используется для создания более авторитетного или научного стиля текста; неуместные уточнения; формальные повторы словосочетаний и отдельных слов; аббревиатуры терминов на английском языке, которые приводятся в дополнении к терминам на русском; ошибки или искажения в лексической сочетаемости; нарушение сочетаемости слов по линии предсказуемости / непредсказуемости, обязательности / факультативности синтаксической связи; нарушение правил использования степени сравнения прилагательного, лексической валентности слов, грамматической

сочетаемости при использовании разговорных слов и жаргонизмов, правил употребления предлогов перед согласными; наличие значительного (неоправданного или избыточного) количества вводных слов и конструкций [5; 17; 24; 25; 29; 30; 35]. В тексте, сгенерированном нейросетью, отсутствует выражение собственного мнения, субъективные высказывания (авторская позиция) [7].

В сгенерированных текстах слова, включенные в заголовок, встречаются по тексту значительно чаще, чем в текстах, написанных человеком [36]. В тексте, сгенерированном нейросетью, очень редко встречаются опечатки и орфографические ошибки, не встречаются разные типы тире (-, — и —) или кавычек (“” и «») [25; 37–39].

Заключение

На основе проведенного анализа научной литературы можно заключить, что практика применения Gen-AI при написании научных текстов намного опережает рефлексию ее результатов и прогнозирование последствий в долгосрочной перспективе. Вероятно, в ближайшие годы этот разрыв сократится⁵, так как в силу недостаточности обучающего материала генеративные нейронные сети придут к морфологически более стройному и стабильному набору характеристик (например, стереотипных ошибок на разных уровнях языка) результата генерации, который позволит разработать автоматические системы — детекторы генерации текста. До тех пор научный редактор / руководитель / рецензент вынуждены подкреплять интуитивное обнаружение сгенерированного текста разношерстными данными из научной литературы и из личных проб работы с системами-детекторами. Наряду с этим обобщение результатов экспертной оценки позволяет сформулировать следующие выводы.

1. Необходимо сочетать автоматические и ручные способы подтверждения наличия в рукописи сгенерированного текста. Следует использовать параллельно несколько систем автоматической проверки, так как каждая из них «натренирована» на материале конкретных Gen-AI-моделей и на специфическом языковом материале (например, на массиве текстов, написанных в публицистическом стиле). При ручной проверке можно опереться на предложенный в этой статье подход поуровневого анализа ошибок и погрешностей, обнаруживаемых в тексте.

2. Проверку (особенно автоматическую) следует проводить автономно относительно каждого раздела рукописи (для статей это аннотация, введение, обзор литературы / история вопроса, описание методики исследования, описание результатов исследования, обсуждение, выводы / заключение, список литературы). Запрещать использование инструментов Gen-AI, вероятно, непродуктивно, но обозначить границы приемлемого использования необходимо. По нашему мнению, недопустимо передавать на откуп пока еще

⁵ См., например: Коммерсантъ. URL: <https://www.kommersant.ru/doc/6849411> (дата обращения: 05.12.2025).

«сырой» технологии задачу по аналитике практического материала, описанию и обобщению полученных данных. Особенно значимо требование самостоятельности выполнения научной работы и ее описания для начинающих исследователей, которые находятся в процессе освоения профессиональных навыков, погружения в предметную область исследования, становления исследовательской позиции, выработки индивидуального стиля письма.

3. При использовании AI-инструментов на любом этапе подготовки и написания научного текста для авторов должно стать правилом декларирование того, какие именно инструменты, в каком объеме и для генерации какой конкретно части текста применялись.

Список литературы

1. Лебедев А.Н. Константа М.Н. Ливанова и психофизиологические закономерности работы мозга // Психологический журнал. 2008. Т. 29. № 1. С. 133–137. EDN: INMIWB
2. Кузьминов Я.И., Кручинская Е.В., Груздев И.А., Наумов А.А. Отстающие и опережающие: как студенты используют генеративный искусственный интеллект в образовательных целях // Высшее образование в России. 2025. Т. 34. № 6. С. 9–35. <https://doi.org/10.31992/0869-3617-2025-34-6-9-35> EDN: RXDTXQ
3. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems 30. Long Beach : Neural Information Processing Systems Foundation. New Orleans : IBM Research, 2017. P. 6000–6010.
4. Yu H., Guo Y. Generative artificial intelligence empowers educational reform: current status, issues, and prospects // Frontiers in Education. 2023. Vol. 11. Iss. 7. <https://doi.org/10.31992/0869-3617-2025-34-6-9-35> EDN: RXDTXQ
5. Телицына А.Ю. Оптимизация научной деятельности через интеграцию ИИ: нейронные сети как инструмент в работе с академической литературой // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 218–236. <https://doi.org/10.14515/monitoring.2024.5.2623> EDN: FEHOZO
6. Голубинский А.Н., Толстых А.А., Толстых М.Ю. Автоматическая генерация аннотаций научных статей на основе больших языковых моделей // Informatics and Automation. 2025. Т. 24. № 1. С. 275–301. <https://doi.org/10.15622/ia.24.1.10> EDN: ICXSIC
7. Осипенко Л.Е., Коротков А.В. Текстовые генеративные нейросети в исследовательской деятельности студентов // Мир науки, культуры, образования. 2024. № 4 (107). С. 90–93. <https://doi.org/10.24412/1991-5497-2024-4107-90-93> EDN: COOBKW
8. Гаврилова Т.А. «Стохастический попугай» на службе бизнеса: успехи и проблемы генеративного искусственного интеллекта // Российский журнал менеджмента. 2024. Т. 2. № 3. С. 461–482. <https://doi.org/10.21638/spbu18.2024.305> EDN: VZIVZB
9. Грицай Г.М., Грабовой А.В., Кильдяков А.С., Чехович Ю.В. Поиск искусственно сгенерированных текстовых фрагментов в научных документах // Доклады Российской академии наук. Математика, информатика, процессы управления. 2023. Т. 514. № 2. С. 308–317. <https://doi.org/10.31857/S2686954323601677> EDN: GQRWLF
10. Зашихина И.М. Научные публикации и большие языковые модели: поймет ли нейросеть российскую науку? // Научный редактор и издатель. 2024. Т. 9. № 52. С. 31–46. <https://doi.org/10.24069/SEP-24-11> EDN: KOZFIO
11. Павлов А.А. Созданные с использованием искусственного интеллекта «научные тексты»: текущее состояние и перспективы развития // Методологические проблемы цивилистических исследований. 2024. № 6. С. 111–127. <https://doi.org/10.33397/2619-0559-2024-6-6-111-127> EDN: LNFESF

12. Гусаренко С.В., Гусаренко М.К. Функционально-семантические, макроструктурные и пресуппозиционально-прагматические параметры сгенерированных русскоязычных текстов в лингвистических нейросетях GigaChat, ChatGPT Марти и Яндекс Алиса // Гуманитарные и юридические исследования. 2024. Т. 11. № 4. С. 788–800. <https://doi.org/10.37493/24091030.2024.4.23> EDN: JXCRGC
13. Фомин В.В., Фомина И.К., Тарануха С.Н. Анализ системы автоматической проверки текстовых документов на наличие заимствований при использовании обучающимися инструментом генеративного искусственного интеллекта // Автоматизация и управление технологическими процессами и производствами. 2024. № 3 (63). С. 57–67.
14. Николаев В.В., Рахконен М.Е. Применение различных инструментов и использование чат-бота «ChatGpt» при написании научных работ, проверяемых в программе «Антиплагиат» // Профессиональное юридическое образование и наука. 2023. № 1 (9). С. 78–81. EDN: RJDDBY
15. Liu Y., Zhang Z., Zhang W., Yue S., Zhao X., Cheng X., Zhang Y., Hu H. ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models // arXiv. 2023. <https://doi.org/10.48550/arXiv.2304.07666>
16. Ma Y., Liu J., Yi F., Cheng Q., Huang Y., Lu W., Liu X. AI vs. human – differentiation analysis of scientific content generation // Computational Linguistics. 2023. Vol. 1. Iss. 1. <https://doi.org/arxiv.org/abs/2301.10416>
17. Огорелков И.В., Махаева Е.В. Искусственная нейросетевая генерация текста как новый вид маскировки письменной речи автора // Виртуальная коммуникация и социальные сети. 2025. Т. 4. № 2. С. 163–171. <https://doi.org/10.21603/2782-4799-2025-4-2-163-171> EDN: ENIMQP
18. Гусаренко С.В. Дефекты когнитивно-семантических структур как причина высокой энтропии актуального дискурса // Известия Южного федерального университета. Филологические науки. 2009. № 2. С. 68–74. EDN: KVLHWD
19. Gritsay G., Grabovoy A., Chekhovich Yu. Automatic detection of machine generated texts: need more tokens // Ivannikov Memorial Workshop (IVMEM). Moscow, 2022. P. 20–26. <https://doi.org/10.1109/IVMEM57067.2022.9983964>
20. Ye J., Chen X., Xu N., Zu C., Shao Z., Liu S., Cui Y., Zhou Z., Gong C., Shen Y., Zhou J., Chen S., Gui T., Zhang Q., Huang X. A comprehensive capability analysis of GPT-3 and GPT-3.5 Series Models // arXiv. 2023. <https://doi.org/10.48550/arXiv.2405.10913>
21. Rodriguez J.D., Hay T., Gros D., Shamsi Z., Srinivasan R. Cross-domain detection of GPT-2-generated technical text // Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics. Seattle : Washington. 2022. P. 1213–1233. <https://doi.org/10.18653/v1/2022.naacl-main.88>
22. Herbold S., Hautli-Janisz A., Heuer U., Kikteva Z., Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays // Scientific Reports. 2023. Vol. 13. Iss. 1. P. 18617–18627. <https://doi.org/10.1038/s41598-023-45644-9> EDN: MFECWA
23. Прохоров А.И., Асадчая К.В. Инструментальные средства определения текста, сгенерированного при помощи нейросети // Научный вектор: сборник научных трудов / под ред. Е.Н. Макаренко. Т. 9. Ростов-на-Дону : Ростовский государственный экономический университет «РИНХ», 2023. С. 250–253. EDN: MGTGIE
24. Хохлова М.В., Корышев М.В. Сравнительный анализ сгенерированных и оригинальных аннотаций научных статей по филологической тематике // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24. № 6. С. 999–1006. <https://doi.org/10.17586/2226-1494-2024-24-6-999-1006> EDN: KDMIJL
25. Кулешов С.В., Зайцева А.А., Аксенов А.Ю. Анализ статистических характеристик искусственно сгенерированных текстов // Известия высших учебных заведений. Приборостроение. 2024. Т. 67. № 11. С. 958–968. <https://doi.org/10.17586/0021-3454-2024-67-11-958-968> EDN: HIYQAL

26. Zhao B., Zhang S., Xu C., Sun Y., Deng C. Deep fake geography? When geospatial data encounter artificial intelligence // *Cartography and Geographic Information Science*. 2021. Vol. 48. Iss. 4. P. 338–352. <https://doi.org/10.1080/15230406.2021.1910075> EDN: MBHDAY
27. Misra D.P., Chandwar K. ChatGPT, artificial intelligence and scientific writing: what authors, peer reviewers and editors should know // *Journal of the Royal College of Physicians of Edinburgh*. 2023. Vol. 53. Iss. 2. P. 90–93. <https://doi.org/10.1177/14782715231181023> EDN: DGPSHF
28. Kacena M.A., Plotkin L.I., Fehrenbacher J.C. The use of artificial intelligence in writing scientific review articles // *Current Osteoporosis Reports*. 2024. Vol. 22. P. 115–121. <https://doi.org/10.1007/s11914023008520> EDN: PDAEEH
29. Айдагулова А.Р. Особенности текстов, сгенерированных искусственным интеллектом // Вестник Башкирского государственного педагогического университета им. М. Акмуллы. 2023. № 4 (72). С. 154–156. EDN: ROEFNK
30. Козловский А.В., Мельник Я.Э., Волощук В.И. О подходе для автоматической генерации сюжетно-связного текста // *Известия ТулГУ. Технические науки*. 2022. Вып. 9. С. 160–167. <https://doi.org/10.24412/2071-6168-2022-9-160-168> EDN: DIRFGA
31. Черкасова М.Н., Тактарова А.В. Признаки сгенерированного текста в академическом дискурсе: проблема идентификации // *Филологические науки. Вопросы теории и практики*. 2024. Т. 17. № 7. С. 2226–2232. <https://doi.org/10.30853/phil20240307> EDN: BEJDOU
32. Kolmogorova A.V., Margolina A.V. Written vs generated text: “naturalness” as a textual and psycholinguistic category // *Научный результат. Вопросы теоретической и прикладной лингвистики*. 2024. Т. 10. № 2. С. 71–99. <https://doi.org/10.18413/2313-8912-2024-10-2-0-4> EDN: CCESHE
33. Aydın Ö., Karaarslan E. OpenAI ChatGPT generated literature review: digital twin in healthcare // *Emerging Computer Technologies 2. İzmir : İzmir Akademi Dernegi*, 2022. P. 22–31. <https://doi.org/10.2139/ssrn.4308687>
34. Mikhaylovskiy N., Churilov I. Autocorrelations decay in texts and applicability limits of language models // *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”*. 2023. Vol. 22. P. 350–360. <https://doi.org/10.48550/arXiv.2305.06615>
35. Осетрова Е.В., Седова А.В. Характеристики сгенерированного текста: языковой и социально-коммуникативный анализ // *Сибирский филологический форум*. 2025. Т. 2. № 31. С. 45–55. <https://doi.org/10.24412/2587-7844-2025-2-45-55> EDN: WYGEFZ
36. Тельнов Р.Е., Ларцина С.В. Типовые различия естественных и сгенерированных нейронной сетью текстов в квантитативном аспекте // *Научный диалог*. 2023. Т. 12. № 7. С. 47–65. <https://doi.org/10.24224/2227-1295-2023-12-7-47-65> EDN: ZZRRKN
37. Chaka C. Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: the case of five AI content detection tools // *Journal of Applied Learning and Teaching*. 2023. Vol. 6. Iss. 2. P. 94–104. <https://doi.org/10.37074/jalt.2023.6.2.12> EDN: BJHFHQ
38. Weber-Wulff D., Anohina-Naumeca A., Bjelobaba S., Foltýnek T., Guerrero-Dib J., Popoola O., Šigut P., Waddington L. Testing of detection tools for AI-generated text // *International Journal for Educational Integration*. 2023. Vol. 19. P. 26. <https://doi.org/10.1007/s40979-023-00146-z> EDN: VCYINN
39. Yang X., Cheng W., Petzold L., Wang W.Y., Chen H. DNA-GPT: divergent n-gram analysis for training-free detection of GPT-generated Text // *ArXiv*. 2023. <https://doi.org/10.48550/arXiv.2305.17359>

References

1. Lebedev, A.N., & Konstanta, M.N. (2008). Livanov and Psychophysiological Patterns of Brain Function. *Psychological Journal*, 29(1), 133–137. (In Russ.) EDN: INMIWB

2. Kuzminov, Ya.I., Kruchinskaia, E.V., Gruzdev, I.A., & Naumov, A.A. (2025). Falling Behind and Getting Ahead: Student Use of Generative AI in Education. *Higher Education in Russia*, 34(6), 9–35. (In Russ.). <https://doi.org/10.31992/0869-3617-2025-34-6-9-35> EDN: RXDTXQ
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. In: Guyon, I., et al. (eds.). *Advances in Neural Information Processing Systems 30* (pp. 6000–6010). Long Beach: Neural Information Processing Systems Foundation.
4. Yu, H., & Guo, Y. (2023). Generative Artificial Intelligence Empowers Educational Reform: Current Status, Issues, and Prospects. *Frontiers in Education*, 11(7). <https://doi.org/10.31992/0869-3617-2025-34-6-9-35> EDN: RXDTXQ
5. Telitsyna, A.Yu. (2024). Optimization of Scientific Activities Through AI Integration: Neural Networks as a Tool in Working with Academic Literature. *Monitoring of Public Opinion: Economic and Social Changes*, 5, 218–236. (In Russ.). <https://doi.org/10.14515/monitoring.2024.5.2623> EDN: FEHOZO
6. Golubinskiy, A., Tolstykh, A., & Tolstykh, M. (2025). Automatic Generation of Scientific Articles Abstracts Based on Large Language Models. *Informatics and Automation*, 24(1), 275–301. (In Russ.). <https://doi.org/10.15622/ia.24.1.10>
7. Osipenko, L.E., & Korotkov, A.V. (2024). Text Generative Neural Networks in Students' Research Activities. *Mir Nauki, Kultury, Obrazovaniya*, 4 (107), 90–93. (In Russ.). <https://doi.org/10.24412/1991-5497-2024-4107-90-93> EDN: COOBKW
8. Gavrilova, T.A. (2024). “Stochastic parrot” as the servant of Business: Achievements and Challenges of Generative Artificial Intelligence. *Russian Management Journal*, 2(3), 461–482. (In Russ.). <https://doi.org/10.21638/spbu18.2024.305> EDN: VZIVZB
9. Gritsay, G.M., Grabovoy, A.V., Kildyakov, A.S., & Checkovich, Yu.V. (2023). Artificially Generated Text Fragments Search in Academic Documents. *Doklady Rossijskoj Akademii Nauk. Matematika, Informatika, Processy Upravleniya*, 514(2), 308–317. (In Russ.). <https://doi.org/10.31857/S2686954323601677> EDN: GQRWLF
10. Zashikhina, I.M. (2024). Scientific Publications and Large Language Models: will Neural Networks Understand Russian Science? *Science Editor and Publisher*, 9(52), 31–46. (In Russ.). <https://doi.org/10.24069/SEP-24-11> EDN: KOZFIO
11. Pavlov, A.A. (2024). Artificially Generated “Scientific” Texts: Current Status and Future Perspectives. *Methodological Problems of the Civil Law Researches*, 6, 111–127. (In Russ.). <https://doi.org/10.33397/2619-0559-2024-6-6-111-127> EDN: LNFESF
12. Gusarenko, S.V., & Gusarenko, M.K. (2024). Functionalsemantic, Macrostructural and Presuppositional-Pragmatic Parameters of Generated Russian-Language Texts in the Linguistic Neural Networks Gigachat, Chatgpt Marty and Yandex Alice. *Humanities and Law Research*, 11(4), 788–800. (In Russ.). <https://doi.org/10.37493/24091030.2024.4.23> EDN: JXCRGC
13. Fomin, V.V., Fomina, I.K., & Taranukha, S.N. (2024). Analysis of the System of Automatic Verification of Text Documents for the Presence of Borrowings when Students Use Generative Artificial Intelligence Tools. *Informatization of Education and Science*, 3 (63), 57–67. (In Russ.).
14. Nikolaev, V.V., & Rakhkonen, M.E. (2023). The Use of Various Tools and the Use of the Chatgpt Chatbot When Writing Scientific Papers Tested in the Antiplagiate Program. *Professional Legal Education and Science*, 1(9), 78–81 (In Russ.).
15. Liu, Y., Zhang, Z., Zhang, W., Yue, S., Zhao, X., Cheng, X., Zhang, Y., & Hu, H. (2023). ArguGPT: Evaluating, Understanding and Identifying Argumentative Essays Generated by GPT Models. *arXiv*. <https://doi.org/10.48550/arXiv.2304.07666>
16. Ma, Y., Liu, J., Yi, F., Cheng, Q., Huang, Y., Lu, W., & Liu, X. (2023). AI vs. Human – Differentiation Analysis of Scientific Content Generation. *arXiv*, 1(1). <https://doi.org/10.48550/arXiv.2301.10416>

17. Ogorelkov, I.V., & Makhaeva, E.V. (2025). Artificial Neural Network Text Generation as a New Type of Masking the Authorship. *Virtual Communication and Social Networks*, 4(2), 163–171. (In Russ.). <https://doi.org/10.21603/2782-4799-2025-4-2-163-171> EDN: ENIMQP
18. Gusarenko, S.V. (2009). Defects in Cognitive-Semantic Structures as a Cause of High Entropy in Contemporary Discourse. *Proceedings of Southern Federal University. Philology*, 2, 68–74. (In Russ.). EDN: KVLHWD
19. Gritsay, G., Grabovoy, A., & Chekhovich, Yu. (2022). Automatic Detection of Machine Generated Texts: Need More Tokens. In: *Ivannikov Memorial Workshop (IVMEM)* (pp. 20–26). Moscow. <https://doi.org/10.1109/IVMEM57067.2022.9983964>
20. Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q., & Huang, X. (2023). A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models. *arXiv*. <https://doi.org/10.48550/arXiv.2405.10913>
21. Rodriguez, J.D., Hay, T., Gros, D., Shamsi, Z., & Srinivasan, R. (2022). Cross-Domain Detection of GPT-2-Generated Technical Text. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1213–1233). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.naacl-main.88>
22. Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A Large-Scale Comparison of Human-Written Versus ChatGPT-Generated Essays. *Scientific Reports*, 13(1), 18617–18627. <https://doi.org/10.1038/s41598-023-45644-9>
23. Prokhorov, A.I., & Asadchaya, K.V. (2023). Tools for Determining the Text Generated Using a Neural Network. *Nauchny Vektor*, 9, 250–253. (In Russ.). EDN: MGTGIE
24. Khokhlova, M.V., & Koryshev, M.V. (2024). Comparative Analysis of AI-Generated and Original Abstracts of Academic Articles on Philology. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 24(6), 999–1006. (In Russ.). <https://doi.org/10.17586/2226-1494-2024-24-6-999-1006> EDN: KDMIJI
25. Kuleshov, S.V., Zaytseva, A.A., & Aksenov, A.Yu. (2024). Analysis of Statistical Characteristics of Artificially Generated Texts. *Journal of Instrument Engineering*, 67(11), 958–968. (In Russ.). <https://doi.org/10.17586/0021-3454-2024-67-11-958-968> EDN: HIYQAL
26. Zhao, B., Zhang, S., Xu, C., Sun, Y., & Deng, C. (2021). Deep Fake Geography? When Geospatial Data Encounter Artificial Intelligence. *Cartography and Geographic Information Science*, 48(4), 338–352. <https://doi.org/10.1080/15230406.2021.1910075> EDN: MBHDAY
27. Misra, D.P., & Chandwar, K. (2023). ChatGPT, Artificial Intelligence and Scientific Writing: What Authors, Peer Reviewers and Editors Should Know. *Journal of the Royal College of Physicians of Edinburgh*, 53(2), 90–93. <https://doi.org/10.1177/14782715231181023>
28. Kacena, M.A., Plotkin, L.I., & Fehrenbacher, J.C. (2024). The Use of Artificial Intelligence in Writing Scientific Review Articles. *Current Osteoporosis Reports*, 22, 115–121. <https://doi.org/10.1007/s11914023008520> EDN: PDAAEH
29. Aidagulova, A.R. (2023). Features of Texts Generated by Artificial Intelligence. *Vestnik Bashkirskogo Gosudarstvennogo Pedagogicheskogo Universiteta imeni M. Akmully*, 4, 154–156. (In Russ.). EDN: ROEFNK
30. Kozlovsky, A.V., Melnik, Ya.E., & Voloshchuk, V.I. (2022). On The Approach for Automatic Generation of Narrative-Linked Text. *Izvestiya TulGU. Tekhnicheskie Nauki*, 9, 160–167. (In Russ.). <https://doi.org/10.24412/2071-6168-2022-9-160-168> EDN: DIRFGA
31. Cherkasova, M.N., & Taktarova, A.V. (2024). Attributes Of Generated Text in Academic Discourse: The Problem of Identification. *Philology. Theory & Practice*, 17(7), 2226–2232. (In Russ.). <https://doi.org/10.30853/phil20240307> EDN: BEJDOW
32. Kolmogorova, A.V., & Margolina, A.V. (2024). Written vs Generated Text: “Naturalness” as a Textual and Psycholinguistic Category. *Research Result. Theoretical and Applied Linguistics*, 10(2), 71–99. <https://doi.org/10.18413/2313-8912-2024-10-2-0-4> EDN: CCESHE

33. Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In: *Emerging Computer Technologies 2* (pp. 22–31). İzmir: İzmir Akademik Derneği. <https://doi.org/10.2139/ssrn.4308687>
34. Mikhaylovskiy, N., & Churilov, I. (2023). Autocorrelations Decay in Texts and Applicability Limits of Language Models. In: *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"* (Vol. 22, pp. 350–360). <https://doi.org/10.48550/arXiv.2305.06615>
35. Osetrova, E.V., & Sedova, A.V. (2025). Characteristics of the generated Text: Linguistic and Social-Communicative Analysis. *Siberian Philological Forum*, 2(31), 45–55. (In Russ.). <https://doi.org/10.24412/2587-7844-2025-2-45-55> EDN: WYGEFZ
36. Telpov, R.E., & Lartsina, S.V. (2023). Typological Differences of Natural and Neural Network-Generated Texts in a Quantitative Aspect. *Nauchnyi dialog*, 12(7), 47–65. (In Russ.). <https://doi.org/10.24224/2227-1295-2023-12-7-47-65> EDN: ZZRRKN
37. Chaka, C. (2023). Detecting AI Content in Responses Generated by ChatGPT, YouChat, and Chatsonic: The Case of Five AI Content Detection Tools. *Journal of Applied Learning and Teaching*, 6(2), 94–104. DOI: 10.37074/jalt.2023.6.2.12
38. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of Detection Tools for AI-Generated Text. *International Journal for Educational Integration*, 19, 26. <https://doi.org/10.1007/s40979-023-00146-z> EDN: VCYINN
39. Yang, X., Cheng, W., Petzold, L., Wang, W.Y., & Chen, H. (2023). DNA-GPT: divergent N-Gram Analysis for Training-Free Detection of GPT-Generated Text. *ArXiv*. <https://doi.org/10.48550/arXiv.2305.17359>

Сведения об авторе:

Комалова Лилия Рязиштовна, доктор филологических наук, доцент, руководитель и ведущий научный сотрудник Центра эмерджентных практик, Институт научной информации по общественным наукам Российской академии наук (117418, Российская Федерация, г. Москва, Нахимовский проспект, д. 51/21); профессор кафедры прикладной и экспериментальной лингвистики, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); *сфера научных интересов*: корпусная лингвистика, экспериментальная лингвистика, речевая деятельность, кинодискурс; *e-mail*: komalova@inion.ru

ORCID: 0000-0002-0955-5315, Scopus Author ID: 55873286400, ResearcherID: O-2238-2019, SPIN-код: 9330-0170

Information about the author:

Liliya R. Komalova, Dr.Sc. (Philology), Associate Professor, Leading Research Fellow and Head of the Centre for Emerging Practices, Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences (51/21 Nakhimovsky Prospekt, Moscow, Russian Federation, 117418); Professor at the Department of Applied and Experimental Linguistics, Moscow State Linguistic University (38 build 1 Ostozhenka Str., Moscow, Russian Federation, 119034); *Research interests*: corpus linguistics, experimental linguistics, speech activity, cinematic discourse; *e-mail*: komalova@inion.ru

ORCID: 0000-0002-0955-5315, Scopus Author ID: 55873286400, ResearcherID: O-2238-2019, SPIN-code: 9330-0170