

УДК 81'42;004.8

<https://doi.org/10.25587/2222-5404-2025-22-4-234-249>

Оригинальная научная статья



Искусственный интеллект в медиадискурсе: классификация лингвопрагматических признаков сгенерированного текста

E. A. Юрова

Кубанский государственный университет, г. Краснодар, Российская Федерация

✉ elizabeth1999fedorova@mail.ru

Аннотация

Целью работы являются исследование отличительных признаков сгенерированных нейросетью информационно-аналитических текстов в специальном выпуске ежедневной деловой газеты РБК (выпуск заявлен как частично сгенерированный с помощью нейросетей GigaChat и Kandinsky) и обоснование этих признаков в классификацию. В работе был применен сопоставительный метод (автор статьи предпринял попытку выявить и описать не только общие признаки (маркеры) сгенерированного текста, но и особенности генерации в различных типах, видах и жанрах медиадискурса, также описана разница между лингвопрагматическими признаками в различных языках). Среди прочих использован квантитативный метод (контент-анализ специально выбранных единиц текста – признаков текстовой нейросетевой генерации). По результатам исследования была подтверждена ранее выявленная классификация лингвопрагматических маркеров сгенерированных текстов в медиадискурсе; приведены статистические данные, иллюстрирующие наличие тех или иных признаков генерации в текстах и доказывающие их искусственное происхождение: пунктуационные ошибки, морфологические ошибки, синтаксические ошибки, лексические ошибки, лексический повтор, орфографические ошибки, графические ошибки, логические ошибки, фактические ошибки, клише, редundантность, шаблонная структура. В тексте статьи приводится пример анализа одной из информационно-аналитических статей специального выпуска РБК, содержащей наибольшее количество лингвопрагматических признаков генерации нейросетью. Теоретическая значимость результатов исследования состоит в необходимости обоснования новой отдельной области лингвистического знания – медиалингвистики сгенерированных нейросетью текстов. Практическая значимость данного исследования заключается в том, что выделенную автором классификацию лингвопрагматических признаков и особенностей нейросетевого медиатекста возможно применять для экспертизы контента на предмет его искусственного происхождения и интеллектуального мошенничества.

Ключевые слова: газетно-публицистический медиадискурс, генерация текста, ежедневная деловая газета РБК, искусственный интеллект (ИИ), медиаконтент, лингвопрагматические признаки сгенерированного текста, нейросеть, маркеры генерации, медиатекст, синтетический текст

Финансирование. Исследование не имело финансовой поддержки

Для цитирования: Юрова Е. А. Искусственный интеллект в медиадискурсе: классификация лингвопрагматических признаков сгенерированного текста. *Вестник СВФУ*. 2025, Т. 22, № 4. С. 234–249. DOI: 10.25587/2222-5404-2025-22-4-234-249

Original article

Artificial Intelligence in media discourse: the classification of linguistic and pragmatic features of the generated text

Elizaveta A. Yurova

Kuban State University, Krasnodar, Russian Federation

✉ elizabeth1999fedorova@mail.ru

Abstract

The purpose of this scientific article is to study the distinctive features of information and analytical texts generated by a neural network in a special issue of the daily business newspaper RBC (РБК) (the edition is claimed to be partially generated using GigaChat and Kandinsky neural networks) and to put these features into a classification. Procedure and methods. A comparative method was used in the work (the author of the article attempted to identify and describe not only the common features (markers) of the generated text, but also the features of generation in various types and genres of media discourse, the difference between linguistic and pragmatic features in different languages is also described). Among others, a quantitative method was used (content analysis of specially selected text units – features of textual neural network generation). According to the results of the study, the previously identified classification of linguistic and pragmatic features of generated texts in media discourse was confirmed; statistical data were provided illustrating the presence of certain features of generation in the text (160) and proving their artificial origin: punctuation errors (10); morphological errors (7); syntactic errors (7); lexical errors (8); lexical repetition (13); spelling errors (7); graphic errors (8); logical errors (7); factual errors (11); clichés (71); redundancy (6); template structure (6). The text of the article provides an example of the analysis of one of the data analytics articles of the special issue of RBC (РБК), containing the largest number of linguistic and pragmatic features of neural network generation (55 items). Theoretical significance of the research results lies in the necessity to isolate a new separate area of linguistic knowledge – media linguistics of texts generated by a neural network. The practical significance contains the idea that the classification of linguistic and pragmatic features of a neural network media text identified by the author can be used to examine content for its artificial origin and intellectual fraud.

Keywords: newspaper and journalistic media discourse, text generation, the daily business newspaper RBK (РБК), artificial intelligence (AI), media content, linguistic and pragmatic features of the generated text, neural network, generation markers, media text, synthetic text

Funding. No funding was received for writing this manuscript

For citation: Yurova E. A. Artificial Intelligence in media discourse: the classification of linguistic and pragmatic features of the generated text. *Vestnik of NEFU*. 2025, Vol. 22, No. 4. Pp. 234–249.
DOI: 10.25587/2222-5404-2025-22-4-234-249

Введение

Классификация текстов медиапространства активно разрабатывается учеными-лингвистами и филологами (Н. Н. Оломской [1], Т. Г. Добросклонской [2], Н. И. Клушиной [3]). Ввиду взаимопроникновения типов и видов дискурса и медиадискурса друг в друга возникают новые единицы, поэтому выделение и обособление «чистых» типов не представляется более возможным.

В последние десятилетия классические типы дискурса (публицистический, политический, педагогический и т. д.) становятся типами медиадискурса, то есть для их реализации и продвижения контента используются средства массовой информации, таким образом, реализация политического дискурса с помощью медиакоммуникаций приобретает статус политического медиадискурса.

Т. Г. Добросклонская разделяет медиатексты по степени производства на авторские и коллегиальные (написанные несколькими авторами) [4]. В 2022 г., после релиза генеративной нейросети ChatGPT, медиатексты приобретают также и авторство нейросетей, так как стремительное развитие технологий искусственного интеллекта стимулировало появление новой реальности – искусственно сгенерированных текстов [5].

Примером авторского медиатекста может являться любой материал с указанием имени журналиста или корреспондента. Коллегиальные медиатексты – в основном

публикации крупных медиаиздательств, например, «Российской газеты», «Интерфакса», RT, РБК [4].

На примере медиатекстов из различных газет возможно проследить те или иные признаки сгенерированного текста, т. к. некоторые СМИ начинают активно применять нейросети в производстве контента. В данной статье лингвистические признаки генерации рассмотрены на материале информационно-аналитических текстов специального выпуска деловой газеты РБК от 28 апреля 2023 года.

Проблема лингвистики сгенерированных текстов находится на стадии разработки с 2022 г. после релиза текстовой нейросети ChatGPT, исследованию проблемы посвящаются научные труды отечественных и зарубежных лингвистов и ИТ-специалистов [5–13].

Предполагается, что теоретическая значимость результатов исследования состоит в необходимости обосновления новой отдельной области лингвистического знания – медиалингвистики сгенерированных нейросетью текстов. Практическая значимость данного исследования заключается в том, что выделенную автором классификацию лингвопрагматических признаков и особенностей нейросетевого медиатекста возможно применять для экспертизы контента на предмет его искусственного происхождения и интеллектуального мошенничества.

Место ежедневной деловой газеты РБК в классификации дискурса СМИ и медиадискурса (по каналам реализации)

Понятие «медиадискурс» коррелирует с дискурсом в целом, но ограничено сферой употребления – средствами массовой информации (СМИ) [14]. Согласно Закону РФ от 27.12.1991 №2124-1 (ред. от 23.11.2024) «О средствах массовой информации», к СМИ относятся: печатные и сетевые издания, теле- и радиопрограммы, кинохроника и другие жанры под постоянным названием, воздействующие на массовую аудиторию [15]. Газета РБК в данном случае относится к сетевому изданию, публикуемому на официальном сайте [16] и в официальном Телеграм-канале [17].

РБК является жанром медиадискурса, находящимся на стыке нескольких его типов, среди которых газетно-публицистический и компьютерный, т. к. газета выходит на Интернет-портале и не является печатной продукцией.

Согласно авторской классификации типов и видов медиадискурса, разработанной на основе исследования Н. Н. Оломской [1], Интернет-издание РБК относится к компьютерному дискурсу по каналам реализации, далее в данном типе выделяются Интернет-медиадискурс, а затем медиадискурс веб-сайтов. Иерархия сверху вниз отображает уровни реализации контента с помощью компьютера. По коммуникативным функциям выделяется газетно-публицистический дискурс. Таким образом, положение газеты РБК в типологии медиадискурса отражает следующую ядерную структуру от большего к меньшему (рис. 1).



Рис. 1. Место газетно-публицистического дискурса

в типологии медиадискурса по каналам реализации

Fig. 1. The place of newspaper and journalistic discourse
in the typology of media discourse by channels of implementation



Рис. 2. Место газетно-публицистического дискурса
в типологии медиадискурса по каналам реализации
Fig. 2. The place of newspaper and journalistic discourse
in the typology of media discourse by channels of implementation

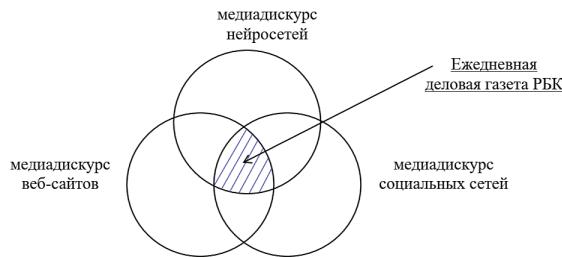


Рис. 3. Жанр ежедневной деловой газеты РБК
в типологии медиадискурса по каналам реализации
Fig. 3. The genre of the daily business newspaper RBC
in the typology of media discourse by distribution channels

Стоит однако отличать вид публицистического медиадискурса внутри компьютерного от самостоятельного газетно-публицистического типа медиадискурса, выделяемого по каналам реализации, так как в первом публикация и распространение контента происходят через цепочку веб-сайт – Интернет – компьютер. Также звено «веб-сайт» может быть заменено «социальной сетью», так как, согласно типологии автора, Интернет-медиадискурс подразделяется на медиадискурс веб-сайтов и медиадискурс социальных сетей (рис. 2). Ежедневная деловая газета РБК также размещает свои статьи в официальном Телеграм-канале.

Интернет-портал РБК как один из жанров современных деловых медиа (классификация медиадискурса по коммуникативным функциям)

В эпоху искусственного интеллекта к инструментам и ресурсам создания и распространения контента стали также относить и нейросети, поэтому наравне с медиадискурсом веб-сайтов и медиадискурсом социальных сетей, выделенным автором статьи в классификацию типов медиадискурса по каналам реализации, можно обосновать таким образом и медиадискурс нейросетей.

Ежедневная деловая газета РБК находится на пересечении всех трех подвидов компьютерного и Интернет-медиадискурса (рис. 3).

Газета РБК, согласно представленной выше иерархии медиадискурса, считается одним из его жанров [18], в частности компьютерного, в то время как веб-сайт, социальная сеть или нейросеть – это инструменты [18] создания или распространения контента.

Стремительное появление СМИ на просторах Интернета явилось предпосылкой для обоснования новых синтетических типов медиа [19] – Интернет-издательств или официальных каналов газет в социальных сетях.

Согласно классификации Н. Н. Оломской, по коммуникативным функциям выделяются рекламный, публицистический и PR-дискурсы [1]. Н. И. Клушина выделяет информационный/новостной, публицистический медиадискурс и дискурс развлечения [3]. По критерию основных тем публикуемого контента газета РБК относится к деловым медиа, публикующим также периодически новости [19]. Таким образом, автор статьи выделяет новый тип медиадискурса (на стыке газетно-публицистического и новостного) – публицистико-новостной.

В классификации текстов СМИ выделяются следующие функционально-жанровые типы: новости, информационно-аналитические тексты, features (статьи на разные популярные темы) и рекламные тексты [2]. В РБК можно встретить первые три типа: новости, информационно-аналитические тексты и features – при этом анализируемый в данном исследовании специальный выпуск газеты содержит только информационно-аналитические тексты о внедрении искусственного интеллекта в журналистику.

Возможности генерации контента нейросетями и их роль в современной публицистической медиапрактике

Ввиду того что газета РБК находится на стыке трех типов Интернет-медиадискурса по каналам реализации, можно прийти к выводу, что публикуемый контент может создаваться с помощью генеративных нейросетей и публиковаться на официальном веб-сайте и в официальных сообществах в социальных сетях.

28 апреля 2023 года издательством был опубликован специальный выпуск, посвященный искусственноому интеллекту и частично, по словам главного редактора Петра Канаева, сгенерированный нейросетью GigaChat от Сбера [20].

Эволюция технологий искусственного интеллекта вывела на передний план автоматизированные вычислительные приложения, упростив извлечение речи и языковых показателей [13]. В связи с часто скрытым фактом создания синтетического контента человеческое общество сталкивается с проблемой поиска достоверных и подлинных материалов, не содержащих искаженные факты [8].

Например, новости могут генерироваться специальными профильными нейросетями «Нейропушкин» и «Порфириевич» [6]. Однако другие нейросети способны обнаруживать синтетический текст на основе одних и тех же языковых моделей, тем самым борясь с распространением дезинформирующих сгенерированных фейков [12].

Общие и частные лингвистические, прагматические и стилистические признаки генерации контента в газетно-публицистическом / публицистико-новостном медиадискурсе (на примере газеты РБК)

При выделении признаков сгенерированного текста в медиадискурсе следует обращать внимание на тип и жанр публикуемого материала. Существует множество различных видов нейросетей, которые адаптируют сгенерированный текст на выходе под запрос пользователя: например, в промпте (промте) возможно указать «Сгенерируй пост для социальной сети», «Сгенерируй новость» или «Сгенерируй информационно-аналитическую статью» и т. д. В таком случае массив сгенерированных текстов будет различаться не только лингвистически, но и стилистически.

С точки зрения прагматики сгенерированный текст содержит ненужные определения, слишком общую информацию, избыточные повторы одних и тех же смыслов, что нарушает прагматическую концепцию Г. П. Грайса (количество, качество, уместность и манера преподнесения информации) [10].

Синтетический нейросетевой текст отличается простыми поверхностными примерами, множественными повторами лексики и идей в сопутствующих предложениях и искажением фактов [11].

Также лингвисты предполагают, что признаками сгенерированного текста являются стереотипный «ход мыслей», ограниченный набор структур и их повторяемость («закольцованность» темы [7]), нейросеть работает только с объектом с явными признаками, не имея возможности работать со скрытыми связями [5].

Контент, сгенерированный с помощью ИИ, содержит шаблонные формулировки и избыточность, отличая текст стилистически от созданного человеком [9].

По результатам нескольких проведенных ранее исследований автор статьи предпринял попытку выявить и описать не только общие признаки (маркеры) сгенерированного текста, но и особенности различных типов, видов и жанров медиадискурса. Стоит отметить еще раз, что лингвопрагматические признаки генерации изменяются в зависимости от конкретного вида и жанра медиадискурса, для публикации в котором созданы. Например, одним из маркеров является избыточная терминология: для онлайн-энциклопедий она является нормой [21], а для политических текстов необходимо ее отсутствие и деспециализация понятий с целью упрощения понимания посланий политикам электорату [22].

Таким образом, текст, подготовленный нейросетью для поста в социальной сети, будет визуально отличаться от текста, подготовленного для онлайн-энциклопедии. Те или иные типографические и стилистические признаки также учитываются нейросетью при написании промпта.

Например, энциклопедическая статья в Википедии отличается научным стилем и списком литературы в конце, заметки для путешественников на тематическом веб-сайте, как правило, имеют множество сопроводительных изображений, привлекательные заголовки и «зазывающие» высказывания.

Также признаки сгенерированных текстов могут быть общими вне зависимости от языка генерации или могут быть свойственными только конкретным типам языков. Например, в русском языке к маркерам сгенерированных текстов относятся морфологические и синтаксические ошибки. Русский язык входит в число синтетических (также флексивных) языков, в которых грамматические отношения выражаются с помощью морфем, в отличие от, например, английского, который является аналитическим. Таким образом, категорию рода в английском языке невозможно отследить. По наблюдениям автора, в режиме работы на русском языке нейросети часто допускают ошибки в определении рода, согласовании подлежащего и сказуемого (также несклоняемые существительные единственного числа согласуются с глаголами в форме множественного числа).

Шаблонная или кольцевая структура свойственна постам для социальных сетей. Это связано, в первую очередь, с потребностью личного обращения «автора» текста к читателям. Наиболее часто к генерации постов обращаются политики, которым необходимо вести предвыборную кампанию и непосредственно контактировать с избирателем. В таком случае шаблонная структура может выглядеть так: «Здравствуйте, дорогие друзья!» – перечисление актуальных проблем – «Безусловно, эти проблемы требуют решения!».

Генерация списка литературы появляется в том случае, если в запросе для нейросети указать «сгенерируй научную/энциклопедическую статью». Стоит отметить, что, согласно классификации фактических ошибок, также разработанной автором статьи, данный признак (наличие списка источников) будет относиться к нейросетевой «галлюцинации» и не будет представлять собой реально существующие работы.

Ранее автор статьи предпринял попытку определить универсальную классификацию, подходящую практически любому жанру медиадискурса, на основе корпуса текстов, который выделен в медиадискурсе социальных сетей и состоит из сгенерированных постов в телеграм-канале кандидата на пост

губернатора Нижегородской области 2023 Ильи Померанцева [23]. Таким образом, лингвопрагматические признаки сгенерированного текста, выявленные на материалах 38 постов в Телеграм (423 маркера), были подтверждены на примере 6 статей из полного специального выпуска РБК. Оба корпуса анализируемых текстов были заявлены их авторами как сгенерированные. Таким образом, можно с большой степенью вероятности полагать, что идентичные маркеры, обнаруженные в обоих текстовых блоках, могут быть выделены и обособлены в классификацию, проиллюстрированные признаки которой можно наблюдать и в других сгенерированных текстах.

Среди признаков выделяются:

- редундантность («закольцованность» темы, повторение идентичных идей и смыслов в следующих друг за другом предложениях);
 - оформление текста по шаблону;
 - клише (типичные для нейросети слова и фразы);
 - чрезмерное «насыщение» текста терминологическими единицами;
 - фактические ошибки:
 - 1) неверные источники и ссылки на несуществующие материалы;
 - 2) частично неправильные данные, например, ошибки в числовых показателях или неверный хронотоп (указание времени и места);
 - 3) неверные, перефразированные цитаты;
 - 4) полностью ошибочные и придуманные факты;
 - 5) намеренно запущенная дезинформация, фейки;
 - ошибки следующего характера:
 - 1) нарушения логики повествования;
 - 2) неверное употребление лексических единиц;
 - 3) лексические повторы;
 - 4) неправильное использование грамматических единиц (включая морфологические и синтаксические);
 - 5) неточность в написании слов (орфография);
 - 6) некорректное использование знаков пунктуации [23].

В рамках данной научной работы была предпринята попытка классифицировать признаки генерации применительно именно к газетно-публицистическому или выделенному новому публицистико-новостному медиадискурсу. В качестве эмпирического материала использованы информационно-аналитические тексты специального выпуска газеты РБК (рис. 4) [20].



Рис. 4. Специальный выпуск ежедневной деловой газеты РБК

Источник: <https://www.rbc.ru/society/29/04/2023/644cc8fe9a79476874dd9f28>

Fig. 4. A special edition of the daily business newspaper RBC

Link: <https://www.rbc.ru/society/29/04/2023/644cc8fe9a79476874dd9f28>

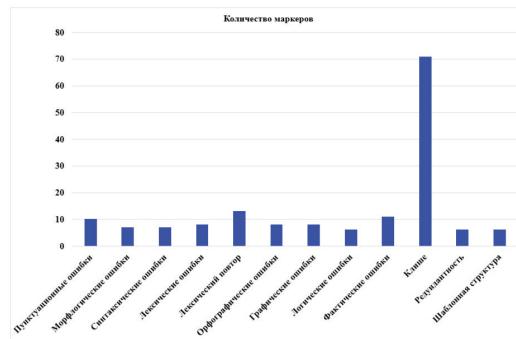


Рис. 5. Количество содержание каждого из типов лингвистических маркеров генерации в информационно-аналитических статьях РБК
Fig. 5. The quantitative content of each type of linguistic marker of generation in RBC information and analytical articles

Предполагается, что влияние человеческого фактора на публикацию данного выпуска РБК с многочисленными языковыми ошибками может быть сведено к минимуму, так как анализируемый в данной статье корпус текстов заявлен главным редактором как частично сгенерированный с помощью нейросетей GigaChat и Kandinsky и опубликованный под заголовком «Как выглядит первая газета РБК, созданная в соавторстве с нейросетями» [20]. По результатам исследования оказалось, что тексты, отмеченные авторскими, также являются сгенерированными и содержащими идентичные признаки «присутствия» нейросетей. Не вызывает сомнения тот факт, что крупные серьезные издательства, такие как РБК, не выпускают заведомо некачественный и непроверенный контент, содержащий множество редакторских «опечаток» и фактических ошибок, так как существующие должности редактора и корректора обеспечивают проверку информации на подлинность, аспекты стилистики и лингвистические ошибки. Также упомянутые выше признаки не могут носить человеческий характер в связи с тем, что анализируемый в данной статье выпуск был заявлен как сгенерированный.

Практический аспект исследования по выделению признаков сгенерированных текстов в специальном выпуске газеты РБК

По результатам исследования был обнаружен новый признак (маркер) генерации: графические ошибки (выделение некоторых слов в предложении функцией Caps Lock, двойной пробел).

Количественные показатели каждого из типов лингвистических маркеров (160) сгенерированного текста включают:

- пунктуационные ошибки (10);
- морфологические ошибки (7);
- синтаксические ошибки (7);
- лексические ошибки (8);
- лексический повтор (13);
- орфографические ошибки (7);
- графические ошибки (8);
- логические ошибки (7);
- фактические ошибки (11);
- клише (71);
- редундантность (6);
- шаблонная структура (6) (рис. 5).

Таблица

**Количественное содержание различных клише
в информационно-аналитических статьях РБК**

Table

Quantitative content of various clichés in RBC information and analytical articles

	Статья 1	Статья 2	Статья 3	Статья 4	Статья 5	Статья 6	СУММА
«решение»	7	10		5		1	
«задача»	14	2		1		2	
«кроме того»	1	5		1	1		
«ключевой»		1	3				
«эффективный»		2	2				
«инструмент»		1	3				
«в первую очередь»			1			3	
«один из»		1	2				
«инициатива»			1				
«сложный»	1						
СУММА	23	22	12	7		6	

Наибольшее количество (71) занимают типичные, повторяющиеся слова, известные также в зарубежных источниках как buzzwords [24], AI words [25], redundant vocabulary [26]. Автор настоящего исследования предлагает использовать в данной ситуации термин «клише». Среди выявленных клише наиболее частотными являются: «решение» (23 маркера), «задача» (19), «кроме того» (7), «эффективный» (4), «ключевой» (4), «в первую очередь» (4), «один из» (3), «сложный» (1), «инициатива» (1) (рис. 6). Клише, употребленные в информационно-аналитических статьях специального выпуска РБК единожды, были выявлены автором ранее как типичные на основе другого сгенерированного материала [23]. Данные лексемы могут предположительно быть использованы в профессиональном публицистическом дискурсе, в деловой и технологической журналистике, однако, учитывая наличие в крупных издательствах должности редактора и корректора, факт использования данных слов (в данном корпусе приведены 64 лексемы в 6 статьях) сводится к минимуму.

Постатейные количественные показатели обнаруженных лингвопрагматических признаков генерации содержат:

1. «Нейрообъятия для человечества» – 41 единица генерации;
2. «От дипфейков до пиратства» – 53 единицы генерации;
3. «Быстрее, бюджетнее, практичнее» – 29 единиц генерации;
4. «Венчуры фокусируются на конкретном» – 15 единиц генерации;
5. «Карма электрокара» – 13 единиц генерации;
6. «Расставание с рутиной» – 9 единиц генерации.

В качестве эмпирического материала целесообразно продемонстрировать некоторые элементы анализа информационно-аналитической статьи с наибольшим количеством обнаруженных лингвопрагматических признаков генерации («2. От дипфейков до пиратства»). Ниже приводятся цитаты из сгенерированного специального выпуска ежедневной деловой газеты РБК [20]:

1. Графическая ошибка (выделение некоторых слов в предложении прописными буквами): *«Есть ли в условиях кризиса ВОЗМОЖНОСТИ ДЛЯ РАЗВИТИЯ ТЕХНОЛОГИЙ и каким в 2023 году станет отечественный рынок инноваций?*

Самое главное из исследования РАЭК, посвященного технологическим прогнозам РАЗВИТИЯ РУНЕТА, в материале РБК» [27].

2. Лексическая ошибка (существительное «реальность» заменено паронимом «реалистичность»): «Сейчас пользователи присутствуют в виртуальных пространствах в виде аватаров, которые пока далеки от реалистичности» [27].

3. Логическая ошибка (из материалов информационно-аналитической статьи не ясно, о каких авторах исследования идет речь): «По мнению авторов исследования, с помощью технологии дипфейков разработчикам метавселенных удастся полностью воссоздать образы людей в виртуальном мире» [27].

4. Клише (типичный для нейросети производный предлог «кроме того»): «Кроме того, биометрия нужна для дистанционной сдачи экзаменов, удаленного получения электронной подписи, подключения eSIM или оплаты проезда в московском метро» [27].

5. Клише (существительное «решение»): «Компаниям предлагаются льготные условия кредитования, для того чтобы они оперативно и с минимальными издержками смогли внедрить российские ИТ-решения» [27].

6, 7. Логическая ошибка (из материалов информационно-аналитической статьи не ясно, о каких экспертах идет речь): «Эксперты надеются, что эти меры ускорят появление в стране цифровых сервисов, которые станут достойной заменой зарубежным решениям» [27]. Также в данном предложении присутствует клише (существительное «решение»).

8. Лексический повтор («стандарты ... в рамках стандартизации»): «Стандарты разработаны в рамках перспективной программы стандартизации по приоритетному направлению «Искусственный интеллект» на период 2021–2024 годов» [27].

9. Логическая ошибка (из материалов информационно-аналитической статьи не ясно, о каких аналитиках идет речь): «В условиях geopolитической неопределенности, по мнению аналитиков, стоит ожидать дальнейших шагов по оптимизации бизнеса со стороны ИТ-компаний» [27].

10, 11. Логическая ошибка (из материалов информационно-аналитической статьи не ясно, о каких авторах исследования идет речь): «По мнению авторов исследования, российские решения и сервисы обладают высоким экспортным потенциалом» [27]. Также используется клише (существительное «решение»).

12. Клише (существительное «решение» и его формы являются типичными для генерированного текста): «К тому же перспектива выйти на рынки дружественных стран стимулирует компании совершенствовать свои продукты и решения» [27].

13. Клише (прилагательное «эффективный»): «Экосистемная модель бизнеса эффективна не только для самих компаний (это помогает диверсифицировать риски), но и для пользователей (получение разнообразных услуг в «одном окне»)» [27].

14, 15. Клише (производный предлог «кроме того»): «Кроме того, платформы помогают развитию малого и среднего бизнеса: сотрудничество с ними позволяет сократить издержки на ведение бизнеса и открывает доступ к новым рынкам и категориям покупателей» [27]. Также употреблен лексический повтор существительного «бизнес».

16. Клише (существительное «эффективность»): «Платформенная экономика «выстрелила» в 2020 году с началом пандемии и за это время показала свою эффективность» [27].

17. Клише (числительное с предлогом «один из»): «Нехватка оборудования и комплектующих – один из основных факторов, который будет оказывать

негативное влияние на рост IT-отрасли наравне с экономическим общим спадом и санкционным давлением» [27].

18. Синтаксическая ошибка (необоснованное начало предложения со сложно-подчинительного союза «поэтому»): *«В 2022 году многие зарубежные игроки в области информационной безопасности покинули российский рынок. Поэтому в ближайшее время ожидается рост количества отечественных решений» [27].*

19. Клише (существительное «решение»): *«Поэтому в ближайшее время ожидается рост количества отечественных решений» [27].*

20. Клише (производный предлог «кроме того»): *«Кроме того, с 2025 года государственным органам будет запрещено использовать средства защиты из недружественных стран» [27].*

21. Клише (существительное «решение»): *«После введения санкций были приняты дополнительные меры по поддержке развития отечественных решений» [27].*

22, 23. Клише (существительное «решение» и существительное «задача»): *«Проект под названием «Госдата.хаб» будет содержать обезличенные данные, которыми смогут пользоваться не только государственные органы, но и компании для решения бизнес-задач» [27].*

24. Лексическая ошибка (некорректно расшифрованная аббревиатура «ЦФО» – «цифровые финансовые активы»): *«Цифровые финансовые активы – это цифровые права, выпуск, учет и обращение которых возможны только путем внесения (изменения) записей в информационную систему на основе блокчейна. Геополитическая ситуация заставила государство по-новому взглянуть на ЦФО» [27].*

25, 26 Морфологическая ошибка (существительное «оплата» сочетается с другими существительными в родительном падеже без предлога: «оплата (чего?) товаров, работы, услуг, интеллектуальной деятельности и т.д.»): *«В условиях санкционных ограничений на трансграничные платежи цифровые финансовые активы могут быть использованы для оплаты за товары, работу, услуги, интеллектуальную деятельность и т.д» [27].* Также употреблено клише (существительное «решение»).

27. Клише (производный предлог «кроме того»): *«Кроме того, власти хотят обеспечить айтишникам и программистам комфортные условия для работы в стране» [27].*

28, 29. Клише (производный предлог «кроме того», существительное «инструмент»): *«Кроме того, это хороший инструмент для поддержки сотрудников в период вынужденного простоя, поиска новой работы или временного снижения дохода на основном месте работы» [27].*

30. Логическая ошибка (из материалов информационно-аналитической статьи не ясно, о каких авторах исследования идет речь): *«По мнению авторов исследования, объемы нелегального контента будут расти» [27].*

31. Пунктуационная ошибка (необоснованное использование кавычек в названиях некоторых организаций): *«Эксперты выделили компании, которые будут определять развитие интернет-экономики в 2023 году: «Сбер», «Яндекс», VK, «Авто», Telegram, маркетплейсы» [27].*

32. Фактическая ошибка (РАЭК организована в 2006 году): *«Российская ассоциация электронных коммуникаций (РАЭК) — некоммерческая организация, занимающаяся развитием интернет-отрасли в России. Она была создана в 2002 году и на сегодняшний день является одной из крупнейших ассоциаций, представляющих интересы участников российского сегмента интернета» [27].*

33. Клише (существительное «задача»): *«Основные задачи РАЭК: ...» [27].*

34. Лексическая ошибка («защита» ... «защиты»): *«...защита интересов участников рынка в части регулирования и защиты прав потребителей» [27].*

35. Графическая ошибка (отсутствие пробела в «домен .RU»): «*В 1991 году был зарегистрирован домен.RU и началась активная работа по созданию национальных доменов для других стран*» [27].

36. Редундантность (повтор одной и той же идеи развития Рунета в двух следующих друг за другом предложениях): «*В это время также начал развиваться и русскоязычный сектор интернета, который стал важным элементом глобальной Сети. Сегодня Рунет продолжает расти и развиваться и является одной из крупнейших интернет-индустрий в мире*» [27].

37. Лексический повтор (необоснованный повтор существительного «интернет»): «*Количество пользователей интернета в России превышает более 80 млн человек, а русский язык является одним из самых распространенных языков в интернете*» [27].

38. Орфографическая ошибка (существительное «рунет» написано в тексте с прописной буквы в связи с тем, что нейросеть позиционирует данное слово как имя собственное, при этом существительное «интернет» написано со строчной, что не является логичным): «*Сегодня Рунет продолжает расти и развиваться и является одной из крупнейших интернет-индустрий в мире. Количество пользователей интернета в России превышает более 80 млн человек, а русский язык является одним из самых распространенных языков в интернете*». Таким образом, либо оба имени собственных должны быть написаны со строчных букв, либо оба с прописных [27].

39–47. Фактические ошибки:

«1990 год Создание Советского национального сервера» [27] (Советский национальный сервер так и не был создан).

«1991 год Регистрация домена.RU» [27] (домен .ru был зарегистрирован в 1993 году).

«1993 год Принятие закона «О национальной компьютеризированной сети России» [27] (закона «О национальной компьютеризированной сети России» не существует).

«1994 год Начало работы первых поисковых систем» [27] (первой поисковой системой в России стал Рамблер, запущенный в 1996 году).

«1995 год Появление социальных сетей» [27] (первая социальная сеть в России – E-Xecutive – появилась в 2001 году).

«1997 год Создание национальной поисковой системы Nigma» [27] (поисковая система Nigma была создана в 2008 году).

«2003 год Создание национальной поисковой системы «Яндекс»» [27] (поисковая система «Яндекс» была создана в 1997 году).

«2012 год Введение национальной поисковой системы «Спутник»» [27] (поисковая система «Спутник» вышла в 2014 году).

«*В 2022 году правительство приняло государственную программу по развитию искусственного интеллекта*» [27] (возможно, речь идет об Указе Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»)

48, 49. Лексический повтор (словосочетание «российский интернет» повторяется в трех следующих друг за другом предложениях): «*Этот текст является описанием технологических трендов и возможностей, которые будут влиять на развитие российского интернета в 2023 году. Ожидается, что технологии продолжат изменять ландшафт российского интернета, включая развитие искусственного интеллекта, блокчейна, биотехнологий и других областей. Рынок инноваций останется ключевым направлением развития российского интернета, так как инновации будут необходимы для улучшения качества жизни людей и*

развития новых отраслей экономики» [27]. Также встречается редундантность (повтор одной и той же идеи развития российского интернета в трех следующих друг за другом предложениях).

50. Клише (прилагательное «ключевой»): «Рынок инноваций останется ключевым направлением развития российского интернета, так как инновации будут необходимы для улучшения качества жизни людей и развития новых отраслей экономики» [27].

51, 52. Клише (существительное «решение» в следующих двух предложениях): «Компании предлагают льготы для ускорения внедрения российских ИТ-решений» [27], «Эксперты надеются, что эти меры помогут ускорить появление новых цифровых сервисов, которые станут альтернативой зарубежным решениям» [27].

53. Шаблонная структура (тиографика сохранена):

– Самое главное из исследования РАЭК, посвященного технологическим прогнозам РАЗВИТИЯ РУНЕТА, в материале РБК.

– ТRENДЫ НАПРАВЛЕНИЯ «ТЕХНОЛОГИИ»

– ТRENДЫ НАПРАВЛЕНИЯ «ЭКОНОМИКА»

– ТRENДЫ НАПРАВЛЕНИЯ «ГОСУДАРСТВО»

– ТRENДЫ НАПРАВЛЕНИЯ «ОБЩЕСТВО»

– НАПРАВЛЕНИЕ «ОТРАСЛЬ»

– ЧТО ПРЕДСТАВЛЯЕТ СОБОЙ РАЭК (рубрика с «комментарием» нейросети)

– ОСНОВНЫЕ ВЕХИ РУНЕТА (нейросетевая аналитика текста)

– О ЧЕМ ЭТА СТАТЬЯ (также нейросетевая аналитика текста) [27].

Заключение

Лингвистический анализ информационно-аналитических статей специального выпуска РБК показал, что тексты, имеющие авторские подписи, также были сгенерированы нейросетью.

Результаты данного исследования вносят вклад в понимание механизмов работы текстовых нейросетей и особенностей обработки естественного языка. Генеративные нейросети активно используются в медиаиндустрии, специалисты которой не всегда маркируют контент как созданный нейросетью. В связи с этим разработанная автором классификация лингвопрагматических признаков может использоваться потребителями медиа-материала для отслеживания факта синтетического происхождения контента.

Л и т е р а т у р а

1. Оломская Н.Н. К вопросу о жанровой классификации медиадискурса. *Научный диалог*. 2013;5(17):250-259.
2. Добросклонская Т.Г. *Язык средств массовой информации*: учебное пособие, электронное издание сетевого распространения. Москва: Добросвет; 2018. ISBN 978-5-7913-1034-7
3. Клушина Н.И. Интенциональная конфигурация медийного пространства. *Политическая лингвистика*. 2013;2(44):40-45.
4. Добросклонская Т.Г. *Медиалингвистика: системный подход к изучению языка СМИ*. URL: <http://www.flf.msu.ru/research/publications/dobrosklonskaya/dobrosklonskaya-medialingvistika.pdf> (Дата обращения: 30.04.2025).
5. Kolmogorova AV, Margolina AV. Written vs generated text: “naturalness” as a textual and psycholinguistic category. *Research Result. Theoretical and Applied Linguistics*. 2024;(10-2):71-99. DOI: 10.18413/2313 8912-2024-10-2-0-4.
6. Крашенинникова М.А., Барабанова М.А. Автоматизированный текст в современной российской медиапрактике. *МедиаАльманах*. 2022;(5):56-62.
7. Осетрова А.В., Седова А.В. Характеристики сгенерированного текста: языковой и социально-коммуникативный анализ. *Сибирский филологический форум*. 2025;2(31):45-55.
8. Напреенко Г.В., Ковалева П.К. Сгенерированный текст в новостном дискурсе:

особенности восприятия носителями языка. *Гуманитарные науки и образование*. 2025;16(1):152-156. DOI: 10.51609/2079-3499_2025_16_01_152

9. Rakhmanova L. Linguistic evolution in digital discourse: the impact of memetics and AI-generated content on syntax and vocabulary. *Cold Science*. 2025;(17):96-107.

10. Bazarbayeva GA, Karbozova GK. Markers of linguistic artificiality in Internet communication (based on bot materials). *Международный научный журнал «ВЕСТИК НАУКИ»*. 2025;2;6(87):1351-1360.

11. Айдагулова А.Р. Особенности текстов, сгенерированных искусственным интеллектом. *Вестник Башкирского государственного педагогического университета им. М. Акмуллы*. 2023;(4-72):154-156.

12. Flores Vivar JM. Artificial intelligence and journalism: diluting the impact of disinformation and fake news through bots. *Doxa Comunicación*. 2019;(29):197-212. DOI: 10.31921/doxacom.n29a10.

13. Georgiou GP. Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. URL: https://www.researchgate.net/publication/382064960_Differentiating_between_human-written_and_AI-generated_texts_using_linguistic_features Automatically_extracted_from_an_online_computational_tool (Дата обращения: 25.08.2025).

14. Калташкина Е.Ю. Прагматические аспекты изучения политического медиадискурса. *Известия Саратовского университета. Новая серия. Серия Филология. Журналистика*. 2012;2(12):27-32.

15. Закон РФ от 27.12.1991 N 2124-1 (ред. от 23.11.2024) «О средствах массовой информации». URL: https://www.consultant.ru/document/cons_doc_LAW_1511/f977773d5130bdc4b8aa5c541d1fa7ca381b18f4/ (Дата обращения: 25.03.2025).

16. Ежедневная деловая газета РБК. URL: <https://www.rbc.ru/> (Дата обращения: 15.03.2025).

17. Ежедневная деловая газета РБК в Телеграм-канале URL: https://t.me/rbc_news (Дата обращения: 15.03.2025).

18. Пескова Е.Н. Дискурс веб-сайта: взаимодействие с другими типами дискурса, жанровые особенности. *Ученые записки Забайкальского государственного университета. Серия: Филология, история, востоковедение*. 2015;2(61):111-116.

19. Дронова М.В. РБК как тип новых медиа в сфере деловой журналистики. *Вестник Волгоградского государственного университета. Серия 9: Исследования молодых ученых*. 2017;(15):96-98.

20. Как выглядит первая газета РБК, созданная в соавторстве с нейросетями. URL: <https://www.rbc.ru/society/29/04/2023/644cc8fe9a79476874dd9f28> (Дата обращения: 15.03.2025).

21. Юрова Е.А. Выявление сгенерированных медиатекстов в Википедии по лингвистическим признакам. *Филология будущего, которое мы создаем: тенденции, смыслы, приоритеты: Материалы I Международной научно-практической конференции*, г. Москва, 21 марта 2025 г. Москва: Московский международный университет; 2025:167-171.

22. Шейгал Е.И. Семиотика политического дискурса: Дис. докт. филол. наук. Волгоград: 2000:440.

23. Оломская Н.Н., Юрова Е.А. Лингвопрагматические особенности сгенерированного текста в медиадискурсе социальных сетей (на примере текстов о выборах губернатора Нижегородской области – 2023 в России). *Российский социально-гуманитарный журнал*. 2025;(2):1-23. URL: <https://www.evestnik-mgou.ru/jour/article/view/1649/1359> (Дата обращения: 15.08.2025).

24. Decoding AI Language: Common Words and Phrases in AI-Generated Content. URL: <https://www.grammarly.com%2Fblog%2Fai%2Fcommon-ai-words%2F&utf=1> (Дата обращения: 13.05.2025).

25. Common AI Words – What to Look Out For in Your Writing. URL: <https://textero.io/guides/common-ai-words> (Дата обращения 12.05.2025).

26. Muñoz-Ortiz A., Gómez-Rodríguez C., Vilares D. Contrasting Linguistic Patterns in Human and LLM-Generated News Text. *Artificial Intelligence Review*. 2024;57(265):1–28. DOI: 10.1007/s10462-024-10903-2.

27. От дипфейков до пиратства. URL: https://www.rbc.ru/technology_and_media/28/04/2023/644b83309a794774db5e6a67 (Дата обращения: 15.03.2025).

R e f e r e n c e s

1. Olomskaya NN. On Genre Classification of Media Discourse. *Nauchnyi dialog*. 2013;5(17):250-259 (in Russian).
2. Dobrosklonskaya TG. *The language of mass media*: study guide, an electronic publication for online distribution. – Moscow: Publishing House “KDU”, “Dobrosvet”; 2018. ISBN 978-5-7913-1034-7 (in Russian).
3. Klushina NI. Intentional configuration of mediaspace. *Political linguistics*. 2013;2(44):40-45 (in Russian).
4. Dobrosklonskaya TG. *Media Linguistics: a systematic approach to learning the language of the media*. Available at: <http://www.flf.msu.ru/research/publications/dobrosklonskaya/dobrosklonskaya-medialingvistika.pdf> [Accessed 30 April 2025] (in Russian).
5. Kolmogorova AV, Margolina AV. Written vs generated text: “naturalness” as a textual and psycholinguistic category. *Research Result. Theoretical and Applied Linguistics. Research Result. Theoretical and Applied Linguistics*. 2024;(10-2):71-99 (English). DOI: 10.18413/2313 8912-2024-10-2-0-4 (in English).
6. Krasheninnikova MA, Barabanova MA. Automated text in modern Russian media practice. *Media Almanach*. 2022;(5):56-62 (in Russian).
7. Osetrova EV, Sedova AV. Characteristics of the generated text: linguistic and social-communicative analysis. *Siberian Philological Forum*. 2025;2(31):45-55 (in Russian).
8. Napreenko GV, Kovaleva PK. Generated text in news discourse: features of perception by native speakers. *The Humanities and Education*. 2025;16;1(61):152-156. DOI: 10.51609/2079-3499_2025_16_01_152 (in Russian).
9. Rakhmanova L. Linguistic evolution in digital discourse: the impact of memetics and AI-generated content on syntax and vocabulary. *Cold Science*. 2025;(17):96-107 (in English).
10. Bazarbayeva GA, Karbozova GK. Markers of linguistic artificiality in Internet communication (based on bot materials). *Vestnik Nauki*. 2025;2;6(87):1351-1360 (in English).
11. Aidagulova AR. Features of texts generated by artificial intelligence. *Bulletin of the Bashkir State Pedagogical University named after M. Akmulla*. 2023;(4-72):154-156 (in Russian).
12. Flores Vivar JM. Artificial intelligence and journalism: diluting the impact of disinformation and fake news through bots. *Doxa Comunicación*. 2019;(29):197-212 (in English). DOI: 10.31921/doxacom.n29a10 (in English).
13. Georgiou GP. Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. Available at: https://www.researchgate.net/publication/382064960_Differentiating_between_human-written_and_AI-generated_texts_using_linguistic_features Automatically_extracted_from_an_online_computational_tool [Accessed: 25 August 2025] (in English).
14. Kaltashkina EYu. Pragmatic aspects of political media discourse. *Proceedings of the Saratov University. A new series. Philology series. Journalism*. 2012;2(12):27-32 (in Russian).
15. The Law of the Russian Federation dated December 27, 1991 № 2124-1 (as amended on 11.23.2024) “On mass media”. Available at: https://www.consultant.ru/document/cons_doc_LAW_1511/f977773d5130bdc4b8aa5c541d1fa7ca381b18f4/ [Accessed: 25 March 2025] (in Russian).
16. The daily business newspaper “RBC”. Available at: https://t.me/rbc_news [Accessed: 15 March 2025] (in Russian).
17. The daily business newspaper “RBC” in Telegram. Available at: https://t.me/rbc_news [Accessed 15 March 2025] (in Russian).
18. Peskova EN. The website discourse: interaction with other types of discourse, genre features. *Scientific notes of the Transbaikal State University. Series: Philology, History, Oriental Studies*. 2015;2(61):111-116 (in Russian).

19. Dronova MV. RBC as a type of new media in business journalism. *Bulletin of the Volgograd State University. Episode 9: Young scientists' research.* 2017;(15):96-98 (in Russian).
20. What does the first RBC newspaper, created in collaboration with neural networks, look like? Available at: <https://www.rbc.ru/society/29/04/2023/644cc8fe9a79476874dd9f28> [Accessed: 15 March 2025] (in Russian).
21. Yurova EA. The detection of generated media texts in Wikipedia according to the linguistic features. *Philology of the future that we are creating: trends, meanings, priorities:* Proceedings of the First International Scientific and Practical Conference, Moscow, 21 March 2025: Moscow International University; 2025:167-171 (in Russian).
22. Shaigal EI. The semiotics of political discourse: Doctor's dissertation (Philology). Volgograd: 2000:440 (in Russian).
23. Olomskaya NN, Yurova EA. Linguopragmatic features of AI-generated text in the media discourse of social networks (on the example of texts devoted to the governor election in the Nizhny Novgorod region, Russian Federation, 2023). *Russian Social and Humanitarian Journal.* 2025;(2):1-23. Available at: <https://www.evestnik-mgou.ru/jour/article/view/1649/1359> [Accessed: 15 August 2025] (in Russian).
24. Decoding AI Language: Common Words and Phrases in AI-Generated Content. Available at: <https://www.grammarly.com%2Fblog%2Fai%2Fcommon-ai-words%2F&utf=1> [Accessed 13 May 2025] (in English).
25. Common AI Words – What to Look Out For in Your Writing. Available at: <https://textero.io/guides/common-ai-words> [Accessed: 12 May 2025] (in English).
26. Muñoz-Ortiz A., Gómez-Rodríguez C., Vilares D. Contrasting Linguistic Patterns in Human and LLM-Generated News Text). *Artificial Intelligence Review.* 2024;57(265):1–28 (in English). DOI: 10.1007/s10462-024-10903-2 (in English).
27. From deepfakes to piracy. Available at: https://www.rbc.ru/technology_and_media/28/04/2023/644b83309a794774db5e6a67 [Accessed: 15 March 2025] (in Russian).

Сведения об авторе

ЮРОВА Елизавета Андреевна – ассистент кафедры английской филологии, ФГБОУ ВО «Кубанский государственный университет», г. Краснодар, Российская Федерация, <https://orcid.org/0009-0003-1343-7359>, SPIN: 2901-7970, e-mail: elizabeth1999fedorova@mail.ru

About the author

Elizaveta A. YUROVA – Teaching assistant at the Department of English Philology, Kuban State University, Krasnodar, Russian Federation, <https://orcid.org/0009-0003-1343-7359>, SPIN: 2901-7970, e-mail: elizabeth1999fedorova@mail.ru

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов

Conflict of interests

The author declares no relevant conflict of interests

Поступила в редакцию / Received 06.10.2025

Поступила после рецензирования / Revised 28.10.2025

Принята к публикации / Accepted 13.11.2025