

УДК 81'42:004.738.5(470.341)“2023”

# ЛИНГВОПРАГМАТИЧЕСКИЕ ОСОБЕННОСТИ СГЕНЕРИРОВАННОГО ТЕКСТА В МЕДИАДИСКУРСЕ СОЦИАЛЬНЫХ СЕТЕЙ (НА ПРИМЕРЕ ТЕКСТОВ О ВЫБОРАХ ГУБЕРНАТОРА В НИЖЕГОРОДСКОЙ ОБЛАСТИ – 2023 В РОССИИ)

**Оломская Наталья Николаевна**

- e-mail: olomnat@mail.ru;  
доктор филологических наук, доцент, профессор кафедры  
английской филологии;  
Кубанский государственный университет  
350040, г. Краснодар, ул. Ставропольская, д. 149,  
Российская Федерация

**Юрова Елизавета Андреевна**

- e-mail: elizabeth1999fedorova@gmail.com;  
ассистент кафедры английской филологии;  
Кубанский государственный университет  
350040, г. Краснодар, ул. Ставропольская, д. 149,  
Российская Федерация

## Для цитирования

Оломская Н. Н., Юрова Е. А. Лингвопрагматические особенности сгенерированного текста в медиадискурсе социальных сетей (на примере текстов о выборах губернатора в Нижегородской области – 2023 в России) // Российский социально-гуманитарный журнал. 2025. №2. URL: [www.evestnik-mgou.ru](http://www.evestnik-mgou.ru)

- Статья поступила в редакцию 03.03.2025
- Статья размещена на сайте 25.06.2025

## СТРУКТУРА СТАТЬИ

Аннотация

Ключевые слова

Введение

Использование сгенерированного контента в политическом медиадискурсе (посты в социальных сетях, политическая реклама, новости в СМИ, дипфейки)

Практический аспект исследования: классификация лингвостилистических маркеров

Заключение

Литература

Article information

## АННОТАЦИЯ

**Цель.** Исследование отличительных признаков сгенерированных нейросетью текстов в политических постах в телеграм-каналах и обособление их в классификацию для выявления маркеров сгенерированного текста и подтверждения гипотезы о наличии подобных искусственно созданных текстов в медиадискурсе.

**Процедура и методы.** При анализе эмпирического материала был использован ряд методов и приёмов для анализа текстов, сгенерированных нейросетями, таких как описательный (сбор языкового материала, классификация найденных лингвистических единиц, описание их особенностей применительно к медиадискурсу), метод контент-анализа (качественный и количественный анализ маркеров сгенерированных текстов), метод обобщения и систематизации.

**Результаты.** По результатам исследования в 38 постах Ильи Померанцева было выявлено 423 лингвистических маркера сгенерированного текста. Установлено количественное и процентное содержание каждого из видов.

**Теоретическая и практическая значимость.** Результаты научного исследования вносят вклад в более глубокое понимание принципов работы нейросетей, процесса генерации ими текстов и обработки естественного языка. Эксперимент наглядно демонстрирует, на какой стадии лингвистического развития находятся генеративные нейросети. Практическая значимость данного исследования заключается в том, что предложенная классификация лингвистических маркеров поможет

специалистам проводить анализ текста с последующим выявлением элементов сгенерированности с помощью искусственного интеллекта.

## КЛЮЧЕВЫЕ СЛОВА

*генерация текста, губернаторские выборы – 2023, искусственный интеллект (ИИ), маркеры сгенерированного текста, нейросеть, политический медиадискурс*

## ВВЕДЕНИЕ

В последние годы наблюдается рост применения нейросетевых технологий для генерации текстов с целью манипулятивного воздействия или пиар-кампании в электоральном медиадискурсе. С помощью нейросетей генерируются посты в социальных сетях, политическая агитационная реклама, новости в СМИ и дипфейки.

Нейросетевая генерация политических текстов в медиадискурсе может стать мощным инструментом манипуляции мнением избирателей, для обнаружения данного феномена мы выдвинули гипотетические признаки сгенерированных текстов.

Гипотеза исследования состоит в следующем: мы считаем, что сгенерированный текст обладает тождественным набором признаков. Представленная в исследовании классификация позволяет выявить, был ли исследуемый текст полностью или частично сгенерирован с помощью нейросетей. Авторы исследования не утверждают, что каждый пост в социальной сети «Телеграм», опубликованный в период выборов, подвергался генерации с помощью искусственного интеллекта. Однако авторы считают возможным выделение признаков сгенерированности на основе представленной гипотезы и предлагают классификацию признаков, которые характеризуют большинство проанализированных сгенерированных текстов.

Методологической базой исследования являются основные положения теории политического дискурса и медиадискурса (Е. И. Шейгал [14], А. П. Чудинов [13]), материалы по компьютерной лингвистике, NLP (обработке естественного языка) и генерации текста нейросетями (Д. Картер [5], Й. Гольдберг [2], Я. Лекун [7], Ю. Васильев [1]), учебные пособия по грамматике, стилистике, орфографии и пунктуации (Д. Э. Розенталь [10; 11], И. Р. Гальперин [16]).

В качестве эмпирического материала были рассмотрены посты в телеграм-канале Ильи Померанцева, кандидата на пост губернатора Нижегородской области от партии «Гражданская сила» в 2023 г.<sup>1</sup> Данный претендент на участие в выборах лично заявляет, что провёл «первую предвыборную кампанию, придуманную искусственным интеллектом»<sup>2</sup>. В его телеграм-канале «GPT Померанцев» в 38 постах на временном промежутке с 2 августа по 7 сентября 2023 г. встречаются пометки «Пост подготовлен при непосредственном участии искусственного интеллекта» и хештег #ИИ\_пишет, что составляет 30% (практически 1/4) от общего количества опубликованных на канале постов на момент проведения исследования (декабрь 2024) – 127<sup>1</sup>.

Целью данной научной работы является исследование отличительных признаков сгенерированных нейросетью текстов в политических постах в телеграм-каналах. В исследовании авторами представлена классификация для выявления маркеров сгенерированного текста, которая подтверждает гипотезу о тождестве искусственно созданных текстов в медиадискурсе.

Для достижения цели исследования были поставлены следующие задачи:

1. изучить особенности предвыборной кампании в период губернаторских выборов Нижегородской области в 2023 г.;
2. проанализировать современные системы искусственного интеллекта и возможность их применения для генерации политических текстов в медиадискурсе (на примере постов в телеграм-каналах);
3. представить классификацию лингвистических признаков сгенерированного текста;
4. провести эксперимент по выявлению созданных нейросетями текстов в медиадискурсе на основе представленной гипотезы и проанализировать качество сгенерированных текстов.

<sup>1</sup> Телеграм-канал «GPT Померанцев» // Telegram: [сайт]. URL: [https://t.me/s/Pomerantsev\\_Ilya](https://t.me/s/Pomerantsev_Ilya) (дата обращения: 25.11.2024).

<sup>2</sup> Революция в политтехнологии: история первой предвыборной кампании, созданной нейросетью // vc.ru: [сайт]. URL: <https://vc.ru/media/796644-revoluciya-v-politтехнологii-istoriya-pervoi-predvybornoi-kampanii-sozdannoi-neirosetyu> (дата обращения: 25.11.2024).

## ИСПОЛЬЗОВАНИЕ СГЕНЕРИРОВАННОГО КОНТЕНТА В ПОЛИТИЧЕСКОМ МЕДИАДИСКУРСЕ (ПОСТЫ В СОЦИАЛЬНЫХ СЕТЯХ, ПОЛИТИЧЕСКАЯ РЕКЛАМА, НОВОСТИ В СМИ, ДИПФЕЙКИ)

Во время предвыборных кампаний и непосредственно самих выборов влияние на электорат происходит через все каналы СМИ (телевидение, журналы, газеты, социальные сети), однако, если некоторые информационные источники прибегают к искусственному генерированию контента, говорить о вмешательстве в исход выборов не приходится.

Безусловно, нейросети в медиадискурсе имеют широкий спектр применения, поскольку с их помощью можно:

- генерировать изображения и осуществлять дизайн рекламы / постов в социальных сетях;
- генерировать видео;
- оформлять контент в социальных сетях;
- писать репортажи, сценарии и генерировать текст ведущим программ;
- составлять список вопросов для интервью;
- писать статьи для газет, журналов и генерировать к ним заголовки;
- создавать статистическую информацию для журналов.

В современном мире специалисты из различных сфер медиаиндустрии стремятся к применению нейросетевых технологий в профессиональной деятельности, так как это значительно облегчает их труд и сокращает время работы над задачей. Политики могут генерировать тексты публичных выступлений, журналисты – новости и статьи, а блогеры – создавать синтетический контент в социальных сетях. ChatGPT бросил вызов сути коммуникации – созданию истинного знания на благо развития человечества [12, с. 10]. В современных реалиях естественный интеллект не может обойтись без помощи искусственного, что породило революцию в данной сфере. Этот сценарий является результатом постоянных совместных усилий исследователей в области программирования, искусственного интеллекта, инженерии и лингвистики [15, р. 2].

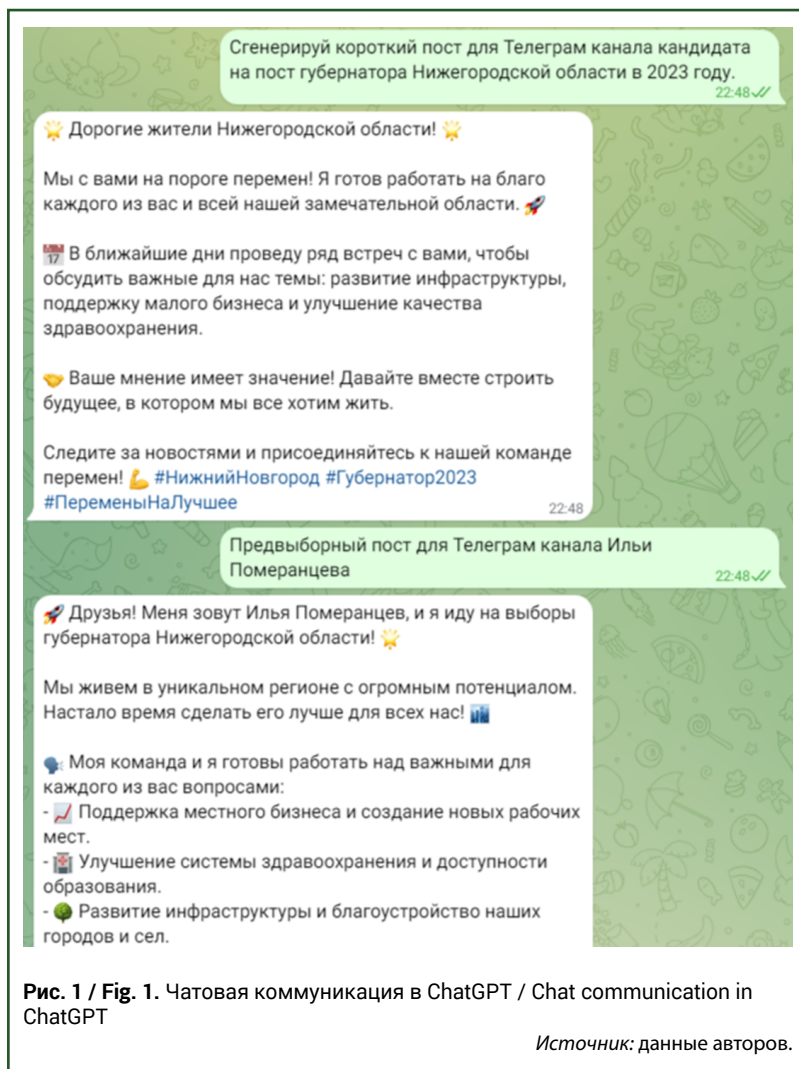
NLP (natural language processing), или область искусственного интеллекта, отвечающая за «понимание» и обработку компьютером естественного языка, даёт возможность создавать синтетический контент для социальных медиа, блогов,

новостей. Нейросети способны анализировать стили авторов и структуру их языка [5]. Например, посты Ильи Померанцева написаны в стиле, соответствующем стилю текстов предвыборной кампании, и, если исключить пометки о генерации текста, распознать искусственно созданный не представляется возможным без дополнительных знаний о лингвистических маркерах сгенерированного текста.

Современный контент, созданный в нейросетевых чат-ботах, выглядит очень реалистично даже с точки зрения наполнения фактической информацией. Программисты научились обучать боты, ChatGPT, в частности, получать информацию из различных открытых источников, например Википедии [1, с. 237]. Несмотря на то, что синтетический текст представляет собой смесь правдивых и вымышленных фактов, отличить его от оригинала не представляется возможным, если заранее не знать правильного варианта [3, с. 27]. Алан Тьюринг разработал тест, сутью которого является коммуникация человека с двумя скрытыми собеседниками (компьютером и другим человеком). Задача человека – «вычислить» компьютер, однако с достижениями в области искусственного интеллекта это становится почти нереально, так как машина может симитировать речь необразованного или аутичного человека, чтобы обосновать свои языковые ошибки [7].

Стоит также отметить, что от запроса и режима работы нейросети зависят результаты, которые мы получим после генерации текста. Нейросети могут работать в двух режимах: «промпт» (также встречается вариант «промт») и «чат». В режиме «чат» нейросеть ведёт коммуникацию в форме диалога и запоминает предыдущие запросы, поэтому уточнять предыдущий можно без необходимости его переписывания снова; в режиме «промпт» каждый запрос нужно формулировать с нуля, так как нейросеть не запоминает предыдущие [8, с. 456]. Например, нейросети ChatGPT, Le Chat, AI Wiz работают в режиме чата, а нейросеть Turbo Text – в режиме промпта. Как мы видим, для ChatGPT достаточно один раз написать промпт, а затем лишь корректировать его (рис. 1).

В процессе исследования был выявлен тот факт, что в открытом доступе отсутствует информация о том, какие СМИ (телеканалы, журналы, газеты и социальные сети) используют искусственный интеллект для создания материала. Бесспорно, применение нейросетей имеет место в медиадискурсе, однако



информации о реальных примерах такого использования в сети Интернет мало, поэтому предположение о том, что тексты могут быть сгенерированы, основывается на гипотезе и определенных признаках, указывающих на генерацию.

Несомненно, электоральная кампания, выстраиваемая кандидатами накануне выборов, вовлекает общество в избира-

тельный процесс. Для того чтобы кампания была эффективной, политикам необходимо правильно подбирать стратегии своих действий и формировать свой имидж политического субъекта. С помощью интернет-технологий и систем искусственного интеллекта увеличиваются скорость и качество проведения предвыборной кампании. Кроме того, интернет и нейросети в XXI в. становятся непосредственными участниками избирательного процесса [4, с. 88] и авторами политических текстов [13, с. 74]. В политическом дискурсе автором текста считается тот, кто берёт ответственность за его содержимое, хотя формально автором может выступать и специально обученный спичрайтер или даже нейросеть [13, с. 74].

В данном исследовании авторы опираются на практический материал, представленный в виде 38 постов в телеграм-канале «GPT Померанцев» Ильи Померанцева, и исследуют аспект сгенерированных текстов в медиадискурсе социальных сетей.

## ПРАКТИЧЕСКИЙ АСПЕКТ ИССЛЕДОВАНИЯ: КЛАССИФИКАЦИЯ ЛИНГВОСТИЛИСТИЧЕСКИХ МАРКЕРОВ

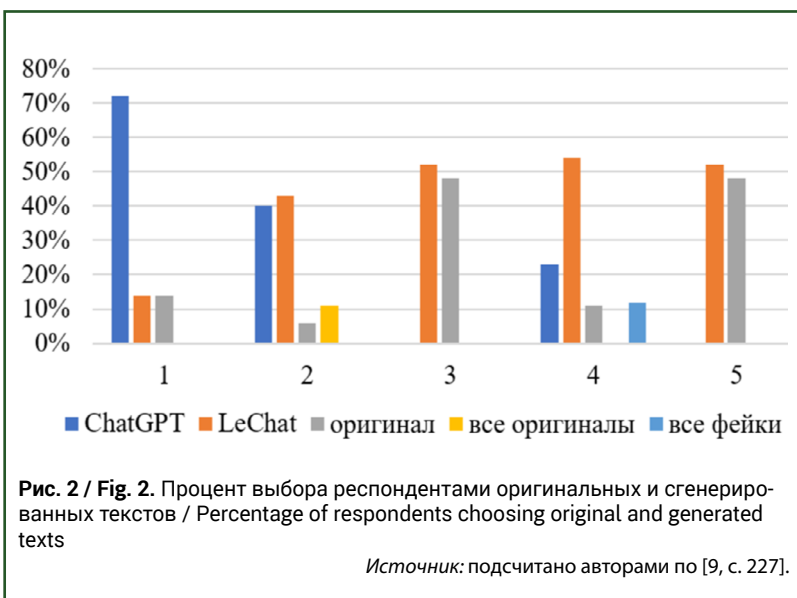
*Этап 1. Выявление маркеров сгенерированного текста.*

Данное исследование было разделено на два этапа, в ходе первого из которых был продолжен эксперимент по исследованию способности респондентов определять сгенерированные и оригинальные тексты среди общего массива. Респондентам (филологам, лингвистам и журналистам) были предложены 15 текстов с общей политической тематикой, созданные в период различных предвыборных кампаний. В каждом пункте опроса были представлены 3, из которых 1 был оригинальной, реально существующей новостью и 2 сгенерированы по её подобию нейросетями ChatGPT, Le Chat на основе ключевых слов. Оригинальные тексты были взяты из следующих новостных изданий: Politico, The New York Times, The Guardian, CNN, World Socialist Web Site (первые 2–3 абзаца текста новостной статьи).

Опрос состоял из 5 пунктов: в сумме респонденты получили 15 текстов (5 оригинальных и 15 сгенерированных; 3 текста в одном пункте). Участникам опроса необходимо было прочитать предложенные тексты и высказать предположение о том, какие из них являются реальными новостями, а какие – сге-

нерированными, а также выделить лингвистические маркеры искусственно созданных текстов и отметить, что отличает эти «новости» друг от друга (рис. 2) [9].

Как видно из диаграммы, процент выбора сгенерированных текстов больше или равен проценту выбора оригинальных. Кроме того, респонденты отмечали, что все новости являются либо сгенерированными фейками, либо оригиналами. Эти данные дают возможность предположить, что распознать, был ли текст сгенерирован или нет, – сложная задача, в том числе и для будущих профессионалов-филологов. Н. Р. Куликова полагает, что написанные нейросетями новости невозможно отличить от авторских [6, с. 67].



*Этап 2. Распознавание и классификация маркеров сгенерированного текста.*

Задачей второго этапа эксперимента являлись распознавание и классификация маркеров сгенерированного текста (на данном этапе учитывались маркеры, выделенные респондентами во время первого этапа исследования).

Согласно гипотезе к признакам сгенерированного текста относятся:

- неестественное и необоснованное повторение лексических структур и однотипных идей в предложениях;
- шаблонная структура;
- клише;
- избыточная терминология;
- фактические ошибки;
- логические ошибки;
- ошибочные цитаты;
- нарушение хронотопа (описания времени и места события; также относим к фактическим ошибкам);
- лексические ошибки;
- лексический повтор;
- грамматические ошибки (морфологические и синтаксические);
- орфографические ошибки;
- пунктуационные ошибки.

Поскольку эмпирическая часть данного исследования состоит в анализе предвыборных постов в телеграм-канале «GPT Померанцев», для подтверждения гипотезы и поиска доказательств того, что большая часть сгенерированных текстов имеет одинаковый набор лингвистических признаков, авторы обратились к данному «предвыборному феномену».

Приведём лингвистический анализ некоторых постов, представленных в данном телеграм-канале и содержащих описанные выше наиболее типичные признаки генерации.

Тексты были предварительно очищены от смайлов и эмодзи. Выделим лингвистические маркеры, которыми насыщен пост-приветствие с потенциальными избирателями от 2 августа 2023 г.<sup>3</sup>.

1. Лексическая ошибка («край» не может быть «передним» или «задним»; данное словосочетание имеет место как военный термин в значении оборонительной линии): *«Наша команда объединяет тех, кто находится на переднем крае созидательного труда в различных сферах нашего общества».*

2. Клише (коллокация «крайне важны» является типичной для нейросетей): *«Сегодня я бы хотел поделиться с вами частью из тех направлений, которые крайне важны для меня и нашей партии».*

<sup>3</sup> Пост от 2 августа 2023 г. в телеграм-канале «GPT Померанцев» // Telegram: [сайт]. URL: [https://t.me/Pomerantsev\\_Ilya/43](https://t.me/Pomerantsev_Ilya/43) (дата обращения: 25.11.2024).

3. В вышеуказанном предложении была выделена морфологическая ошибка (существительное «часть» сочетается с дополнением в родительном падеже без предлога). После исправления данной ошибки было получено следующее предложение: *«Сегодня я бы хотел поделиться с вами частью тех направлений, которые крайне важны для меня и нашей партии».*

4. Необоснованный лексический повтор «власти» в предложении: *«Ответственность власти перед гражданским обществом власти».*

5. Нарушение лексической сочетаемости в предложении *«Влиятельный и профессиональный средний класс».* В данном фрагменте автор поста имеет в виду то, что представители среднего класса имеют высшее образование и занимаются высококвалифицированной работой, таким образом, имеет место словосочетание *«социально-профессиональный статус среднего класса».*

6. Лексический повтор компонента «инициатива» в конце обоих предложений: *«Важно защищать интересы всех граждан, поддерживая развитие частной инициативы. Я обещаю не возводить административные барьеры перед бизнесом и общественной инициативой».*

7. Клише (типичные для нейросетевого текста коллокации со словом «инициатива» – «частная инициатива»): *«Важно защищать интересы всех граждан, поддерживая развитие частной инициативы».*

8. Ещё одно клише со словом «инициатива» («общественная инициатива»): *«Я обещаю не возводить административные барьеры перед бизнесом и общественной инициативой».*

9. В вышеуказанном предложении встречается использование клише со словами «барьеры» и «административный»: *«административные барьеры».*

10. Клише со словом «ключ» и его однокоренными словами: *«Для нас демократия – это ключ к общему благосостоянию».*

11. Шаблонные структуры:

*Мы выбираем следующие принципы:*

*Демократическое государство и развитое гражданское общество.*

*Обеспечение гражданских прав и свобод.*

*Ответственность власти перед гражданским обществом власти.*

*Независимая судебная власть.*

*Адекватный уровень налогов и социальных расходов.*

*Незыблемость и гарантия прав частной собственности.*

*Влиятельный и профессиональный средний класс.*

*Вместе мы сможем сделать Нижегородскую область ещё лучше и обеспечить процветание для всех нас!*

12. Избыточная терминология (*гражданское общество, демократия, общественная инициатива, административные барьеры*).

Следующий пост, опубликованный в тот же день (2 августа 2023 г.), был также проанализирован на наличие признаков сгенерированности<sup>4</sup>. Обнаружены следующие маркеры:

1. Пунктуационная ошибка (необоснованное использование тире в придаточном предложении после производного предлога «несмотря на»): «Статистика говорит нам о том, что *несмотря на все обстоятельства – малый бизнес продолжает расти и развиваться*».оборот с производным предлогом «несмотря на» всегда обособляется запятыми с обеих сторон: «Статистика говорит нам о том, что, *несмотря на все обстоятельства, малый бизнес продолжает расти и развиваться*».

2. Пунктуационная ошибка (отсутствие точки в конце предложения): «*Давайте разбираться*».

3. Морфологическая ошибка (существительное «поддержка» управляет родительным падежом без предлога, т. е. «поддержка кого-то / чего-то»): «Необходимо увеличить объёмы финансирования по имеющимся госпрограммам поддержки для малого бизнеса, создать новые программы налоговых льгот». «Необходимо увеличить объёмы финансирования по имеющимся госпрограммам поддержки малого бизнеса, создать новые программы налоговых льгот».

4. Пунктуационная ошибка (вводное слово «например» выделяется запятыми, а не двоеточием): «Например: с 2022 года IT-компании имеют право выделять свои подразделения в отдельные юридические лица, что не расценивается как “дробление бизнеса”, а реально помогает снизить налогообложение».

5. Лексическая ошибка в вышеуказанном предложении (лексический повтор однокоренных слов «выделять ... в отдельные»). Чтобы исправить ошибку, необходимо заменить одно из однокоренных слов на синонимичное, например, «выделять» на «обособлять»: «Например, с 2022 года IT-компании имеют право обособлять свои подразделения в отдельные юридические лица, что не расценивается как “дробление бизнеса”, а реально помогает снизить налогообложение».

<sup>4</sup>Пост от 2 августа 2023 г. в телеграм-канале «GPT Померанцев» // Telegram: [сайт]. URL: [https://t.me/Pomerantsev\\_Ilya/45](https://t.me/Pomerantsev_Ilya/45) (дата обращения: 25.11.2024).

6. Синтаксическая ошибка в некорректном использовании противительного союза «не как ..., а». Противопоставляемые члены предложения также являются разными частями речи: «... не как "дробление бизнеса", а реально помогает». Таким образом, в предложении нарушается логика. Учитывая все ошибки, исправляем предложение: «Например, с 2022 года IT-компании имеют право обособлять свои подразделения в отдельные юридические лица, что расценивается не как "дробление бизнеса", а реальная помощь по снижению налогообложения».

7. Клише (коллокации со словом «процессы») является типичной для нейросетей): «Медлительность бюрократических процессов мешает регистрации, лицензированию и прочим аспектам развития бизнеса».

8. К клише можно отнести также коллокации со словом «бюрократический»: «Медлительность бюрократических процессов мешает регистрации, лицензированию и прочим аспектам развития бизнеса».

9. Пунктуационная ошибка в предложении: «Новым предприятиям тяжело конкурировать с бизнес-гигантами и, зачастую, даже просто выйти на рынок не представляется возможным». Слово «зачастую» не является вводным и не выделяется запятыми.

10. Клише (коллокации со словом «инициатива») являются также типичными для сгенерированных текстов – «региональные инициативы»): «Нужна всесторонняя поддержка региональных инициатив».

11. Избыточная терминология (дробление бизнеса, цифровая платформа МСП, бизнес-гиганты, бизнес-инкубаторы, технопарки, мировые практики, НКО, электронный документооборот).

12. Шаблонная структура:

Какие проблемы возникают в этом сегменте и как их решить? Давайте разбираться

*Высокие налоги и сложности с кредитованием.*

*Административные барьеры.*

*Цифровизация!*

*Недостаточное вовлечение государства.*

*Создание бизнес-инкубаторов и технопарков, проведение форумов, конференций, выставок ...*

*Существуют мировые практики по помощи в переориентации бизнеса ...*

*... бесплатный доступ к сервисам электронного документооборота.*

*Нехватка кадров.*

*Доступность образования, увеличение числа бюджетных мест, а также специальные программы мотивации и выплаты молодым специалистам.*

*Чаще всего молодые специалисты стремятся пойти на практику в крупные корпорации.*

*Уверен, что те, кто сталкивался с запуском своего бизнеса, могут поделиться множеством подводных камней. Был бы рад обсудить и их. Пишите в комментариях!*

13. Логическая ошибка. Ввиду шаблонной структуры текста становится очевидным его посыл – перечисление проблем в сегменте малого бизнеса. Однако по мере повествования в список проблем вклиниваются их возможные решения, при этом список продолжается без каких-либо пометок о том, что речь уже идёт об их устранении.

В посте, опубликованном 4 августа 2023 г.<sup>5</sup>, были идентифицированы следующие признаки нейросетевого «творчества»:

1. Клише со словом «ключевой», которое является типичным для сгенерированного текста: *«Образование и научный резерв: каждый из ключевых вузов имеет профильный факультет и планируется, что к 2024 году ежегодно будет выпускаться до 3 тысяч высококвалифицированных кадров».*

2. Ещё одно типичное для нейросетей клише со словом «ключевой»: *«ЦИПР – ключевая конференция для современных технологий уже четвёртый год подряд размещается на Нижегородской площадке».*

3. Пунктуационная ошибка («ключевая конференция для современных технологий» является приложением и в позиции после имени собственного должно обособляться либо двумя запятыми, либо двумя тире): *«ЦИПР – ключевая конференция для современных технологий уже четвёртый год подряд размещается на Нижегородской площадке».* Получаем следующий корректный вариант предложения: *«ЦИПР, ключевая конференция для современных технологий, уже четвёртый год подряд размещается на Нижегородской площадке».* Иначе данное предложение могло бы выглядеть так: *«ЦИПР – ключевая конференция для современных технологий – уже четвёртый год подряд размещается на Нижегородской площадке».*

<sup>5</sup>Пост от 4 августа 2023 г. в телеграм-канале «GPT Померанцев» // Telegram: [сайт]. URL: [https://t.me/Pomerantsev\\_Ilya/46](https://t.me/Pomerantsev_Ilya/46) (дата обращения: 25.11.2024).

4. Синтаксическая ошибка в нарушении согласования между подлежащим и сказуемым: «*Органы гос.власти и бизнес-сообщество активно обменивается опытом и демонстрирует лучшие достижения, что положительно влияет на развитие сферы в регионе*». Так как в предложении находятся однородные подлежащие, однородные сказуемые должны согласовываться с ними во множественном числе: «*Органы гос.власти и бизнес-сообщество активно обмениваются опытом и демонстрируют лучшие достижения, что положительно влияет на развитие сферы в регионе*».

5. Логическая ошибка в вышеназванном предложении (из контекста не ясно, о какой «сфере» идёт речь).

6. Пунктуационная ошибка (отсутствие точки в конце предложения): «*GP-days – масштабная конференция для владельцев и руководителей digital-агентств, продакшенов и заказчиков цифровых услуг*».

7. Нарушение лексической сочетаемости в предложении: «*Активная работа и поддержка сообщества со стороны властей помогут сделать Нижний Новгород местом, где процветают современные технологии и рождаются идеи будущего*». Коллокация «*идеи будущего*» не существует, её можно заменить сочетанием «*идеи для будущего*».

8. Логическая ошибка (непонятно, о каком «профильном факультете» идёт речь): «*Образование и научный резерв: каждый из ключевых вузов имеет профильный факультет и планируется, что к 2024 году ежегодно будет выпускаться до 3 тысяч высококвалифицированных кадров*».

9. В данном предложении видим также третье нейросетевое клише со словом «ключевой»: «*Образование и научный резерв: каждый из ключевых вузов имеет профильный факультет и планируется, что к 2024 году ежегодно будет выпускаться до 3 тысяч высококвалифицированных кадров*».

10. Логическая ошибка («регион» не может стать «столицей»): «*Чего же не хватает региону, чтобы по праву стать IT-столицей?*»

11. Пунктуационная ошибка (оборот с союзом «как» в данном предложении можно заменить синонимичным с «в качестве», и в таком случае выделяться запятыми он не будет): «*Я, как кандидат в Губернаторы, считаю развитие IT-сферы в регионе одной из приоритетных задач!*» «*Я как кандидат в Губернаторы считаю развитие IT-сферы в регионе одной из приоритетных задач!*»

12. Излишняя терминология (инкубаторы, технопарки, стартапы, хакатон, digital-агентства, продакшены).

### 13. Шаблонная структура:

*У нашего региона есть преимущества:*

*Образование и научный резерв ...*

*Привлекательность для инвестиций ...*

*Технопарки и инкубаторы ...*

*А ещё Нижний Новгород может похвастаться событиями все-  
российского масштаба:*

*ЦИПР*

*IT-summit*

*GP-days*

*Digital Оттепель*

*Всероссийский Хакатон*

*Startup Weekend NN*

*Чего же не хватает региону, чтобы по праву стать IT-столицей?*

*Инфраструктура ...*

*Образование и поддержка стартапов ...*

*Взаимодействие с властями ...*

Активная работа и поддержка сообщества со стороны властей помогут сделать Нижний Новгород местом, где процветают современные технологии и рождаются идеи будущего.

14. Благодаря выделению шаблонной структуры сразу становится очевидной логическая ошибка: сначала автор заявляет, что «образование» – это преимущество региона, затем – что его не хватает.

15. Повтор одной и той же идеи в последующих трёх абзацах:

*«... важно обеспечить высокоскоростной интернет, развить технопарки и инкубаторы, а также привлечь инвестиции для создания современной IT-инфраструктуры.*

*... специалисты IT-сферы нуждаются в постоянном обучении и развитии. Кроме того, стартапам необходима поддержка и финансирование для успешного старта и роста.*

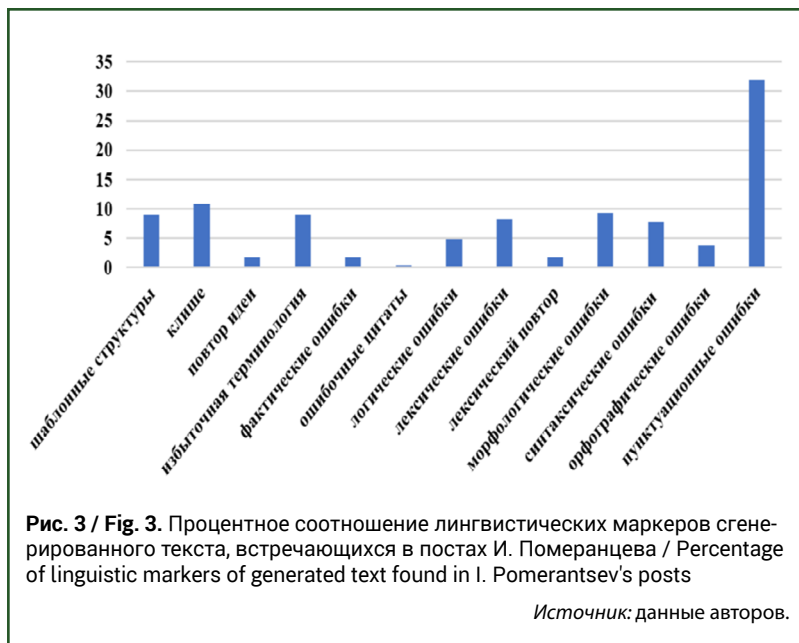
*... активная поддержка со стороны государства и местных властей для формирования благоприятного для развития IT-индустрии законодательства и налоговых условий».*

## ЗАКЛЮЧЕНИЕ

Следует принять во внимание, что формат статьи не позволяет привести в качестве примера все сгенерированные посты, однако в исследовании был проанализирован массив данных

объёмом 38 постов с пометками «Пост подготовлен при непосредственном участии искусственного интеллекта» и хештегом #ИИ\_пишет в телеграм-канале Ильи Померанцева, кандидата на пост губернатора Нижегородской области в 2023 г. По результатам исследования в 38 постах Ильи Померанцева было выявлено 423 лингвистических маркера сгенерированного текста (рис. 3). Анализ количественного и процентного содержания в каждом из видов выявил: 38 шаблонных структур (в каждом посте) – 8,9%; 46 клише (10,9%); 7 повторов одной и той же идеи в стоящих рядом предложениях (1,7%); 38 примеров избыточной терминологии (в каждом посте) – 8,9%; 7 фактических ошибок (1,7%); 1 ошибочная цитата (0,2%); 21 логическая ошибка (4,9%); 35 лексических ошибок (8,3%); 7 лексических повторов (1,7%); 39 морфологических ошибок (9,2%); 33 синтаксические ошибки (7,8%); 16 орфографических ошибок (3,8%); 136 пунктуационных ошибок (32%).

Следует отметить, что шаблонные структуры и избыточная терминология присутствовали в каждом из постов, что даёт возможность позиционировать их как лингвистические маркеры сгенерированного текста со стопроцентной степенью уверенности.



Как видно из представленной выше классификации, наиболее типичными оказались для нейросети пунктуационные (32%) и грамматические (морфологические и синтаксические) ошибки (17%). Нейросеть многократно неверно ставила запятые, тире, использовала неправильные паронимы, писала аббревиатуры строчными буквами, употребляла некорректную падежную форму у существительных и прилагательных, ошибалась в управлении глаголов, а также писала отрывистые предложения, которые не были связаны с остальным текстом логически. Кроме того, нейросеть использовала шаблонную структуру практически в каждом из постов (введение с упоминанием проблемы, список подпунктов с краткими заголовками, заключение с примерными фразами «Мы обсудили основные моменты, и это важно»)<sup>6</sup> – 8,9%. Стоит отметить, что при перечислении подпунктов нейросеть добавляла смайлы и эмодзи для их разграничения. Отметим, что 10,9% признаков занимают клише. По нашим наблюдениям, типичными словами для сгенерированных текстов оказались «инициатива», «процессы», «решения», «бюрократический», «гражданский», «общественный», «ключевой»; также наблюдались клишированные сочетания слов «не просто ...», «не то что», «не только ...» и составной союз «не только ..., но и ...»<sup>7</sup>. Данные слова использовались в различных коллокациях 46 раз.

Пунктуационные ошибки связаны с невозможностью распознавания нейросетью пауз и интонаций устной речи на письме.

Кроме того, в исследовании было выявлено, что в каждом из постов присутствует избыточная терминология (8,9%), что не должно быть свойственно текстам электоральной направленности. Особенностью политической лингвистики, согласно Е. И. Шейгал, является её доступность для понимания гражданами и потенциальными избирателями и деспециализация политических терминов [14].

Данное исследование показало, что несовершенство обработки естественного языка нейросетью связано с:

– полисемией слов;

<sup>6</sup> 6 симптомов, что текст написала нейросеть. Как поймать копирайтеров, авторов и экспертов на генерации контента // Дзен: [сайт]. URL: <https://dzen.ru/a/ZfkeX9icdjKpZa7d> (дата обращения: 27.11.2024).

<sup>7</sup> Штмп на штампе: 10 любимых клише ChatGPT, которые выдают вашу статью с головой (один неожиданный) // vc.ru: [сайт]. URL: <https://vc.ru/chatgpt/1558247-shtamp-na-shtampe-10-lyubimyh-klishe-chatgpt-kotorye-vydyayut-vashu-statyu-s-golovoi-odin-neozhidannyi> (дата обращения: 28.11.2024).

- быстрым пополнением языка неологизмами;
- изменением семантики предложений в связи с эмоциональным компонентом в устной речи, которая не может отразиться в письменном тексте;
- использованием невербальных средств общения, в связи с которыми могут быть изменены значения слов в устной речи. Например, крик «Давай!» при пасае мяча в футболе [2, с. 167].

На современном этапе развития нейросетей понимание и генерация ответов на естественном языке является для них трудной задачей [2, с. 19], которую можно лишь «облегчить» с помощью правильно написанного промпта в запросе. Естественный язык добавляет двусмысленность в интерпретацию команд нейросетью [18, р. 177], поэтому текст для задания должен быть похож на набор данных, на котором изначально была обучена нейросеть [17, р. 174]. Мы полагаем, что, несмотря на то, что Илья Померанцев, вероятно, использовал качественно написанные промпты для нейросетей, полученные тексты необходимо подвергнуть корректировке. Хотя NLP даёт возможность автоматически создавать текстовый контент, контроль и авторская редакция все ещё являются необходимыми действиями [5].

## ЛИТЕРАТУРА

1. Васильев Ю. Обработка естественного языка. Python и spaCy на практике. СПб.: Питер, 2021. 256 с.
2. Гольдберг Й. Нейросетевые методы в обработке естественного языка / пер. с англ. М.: ДМК Пресс, 2019. 282 с.
3. Зашихина И. М. Подготовка научной статьи: справится ли ChatGPT? // Высшее образование в России. 2023. Т. 32. №8–9. С. 24–47. DOI: 10.311992/0869-3617-2023-32-8-9-24-47.
4. Каминченко Д. И. Политический имидж в электоральный период: по материалам избирательной кампании в Нижегородской области в 2023 году // Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2024. Т. 24. №1. С. 80–89. DOI: 10.18500/1818-9601-2024-24-1-80-89.
5. Картер Д. Нейросети. Обработка естественного языка / пер. с англ. М.: Автор, 2023. 231 с.
6. Куликова Н. Р. Автоматическая генерация текста (на примере новости фигурного катания) // Цифровая наука. 2020. №6. С. 64–74.
7. Лекун Я. Как учится машина: Революция в области нейронных

- сетей и глубокого обучения / пер. с фр. М.: Альпина ПРО, 2021. 423 с.
8. Мальцев Н. Д. Структурные и филологические особенности текстовых генеративных нейронных сетей // *Неофилология*. 2024. №10. С. 452–464. DOI: 10.20310/2587-6953-2024-10-2-452-464.
  9. Оломская Н. Н., Юрова Е. А. Лингвистические маркеры сгенерированных новостей (на материале президентских выборов 2024 в США) // *Междисциплинарные аспекты лингвистических исследований: сборник научных трудов* / под ред. Н. Н. Оломской. Краснодар: Кубанский государственный университет, 2025. С. 220–229.
  10. Розенталь Д. Э. Русский язык. Орфография и пунктуация. М.: Эксмо, 2011. 288 с.
  11. Розенталь Д. Э. Справочник по русскому языку. Практическая стилистика. М.: ОНИКС 21 век: Мир и образование, 2001. 381 с.
  12. Тихонова Е. В. ChatGPT в контексте научной коммуникации // *Хранение и переработка сельхозсырья*. 2023. №3. С. 8–12. DOI: 10.36107/spfp.2023.518.
  13. Чудинов А. П. Политическая лингвистика. М.: Флинта: Наука, 2006. 256 с.
  14. Шейгал Е. И. Семиотика политического дискурса: дис. ... докт. филол. наук. Волгоград, 2000. 440 с.
  15. Becker J., Wahle J. P., Gipp B., et al. Text Generation: A Systematic Literature Review of Tasks, Evaluation, and Challenges // *Cornell University. ArXiv.org* (электронный научный журнал). 2024. URL: <https://arxiv.org/abs/2405.15604> (дата обращения: 05.12.2024). DOI: 10.48550/arXiv.2405.15604.
  16. Galperin I. R. *English Stylistics*. М.: URSS, 2014. 332 p.
  17. Grinin I. L. The generation of coherent texts, analysis of neural network mechanics. Mechanics two – learning model for working with a neural network // *Innovation & Investment*. 2020. No. 11. P. 173–176.
  18. Sboev A. G., Gryaznov A. V., Rybka R. B., et al. A deep learning method based on language models for processing natural language Russian commands in human robot interaction // *Research result. Theoretical and applied linguistics*. 2023. Vol. 9. №1. P. 174–191. DOI: 10.18413/2313-8912-2023-9-1-1-1.

## ARTICLE INFORMATION

### Author

Natalya N. Olomsкая

- e-mail: olomnat@mail.ru,  
Dr. Sci. (Philology), Assoc. Prof., Prof., Department of English Philology,  
Kuban State University  
ul. Stavropolskaya 149, Krasnodar 350040, Russian Federation

Elizaveta A. Yurova

- e-mail: elizabeth1999fedorova@gmail.com,  
Assistant, Department of English Philology,  
Kuban State University  
ul. Stavropolskaya 149, Krasnodar 350040, Russian Federation

### For citation

Olomsкая N. N., Yurova E. A. 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). In: *Russian Social and Humanitarian Journal*, 2025, no. 2.

**Available at:** [www.evestnik-mgou.ru](http://www.evestnik-mgou.ru)

### Abstract

**Aim.** To study the distinctive features of neural AI-generated texts on example of political posts on Telegram channels, and to make their classification to both identify the markers of the generated text and prove the presence of such artificially created texts in media discourse.

**Methodology.** Descriptive methods (collection of linguistic material, classification of found linguistic units, description of their features in relation to media discourse), and the methods of content analysis (qualitative and quantitative analysis of markers of the generated texts) were used in the research, which implied AI-generated text analysis.

**Results.** According to the results of the study, 423 linguistic features of AI-generated text have been detected in Ilya Pomerantsev's 38 posts. The quantitative and percental content of each type of marker are presented.

**Research implications.** The results of this study are valuable since they offer insight to the principles of neural networks, the process of text generation made by AI and its natural language processing. The experiment

clearly demonstrates the linguistic level of generative neural networks. The practical significance of this study is the proposed classification of the linguistic peculiarities (markers), which can help specialists analyze the text and then identify the elements of AI-generation.

### Keywords

*text generation, gubernatorial elections 2023, artificial intelligence (AI), features of a generated text, neural network, political media discourse*

### References

1. Vasiliev, Yu. (2021). *Natural Language Processing. Python and spaCy in Practice*. St. Petersburg: Piter publ. (in Russ.).
2. Goldberg, Y. (2019). *Neural Network Methods for Natural Language Processing*. Moscow: DMK Press publ. (in Russ.).
3. Zashikhina, I. M. (2023). Scientific Article Writing: Will ChatGPT Help? In: *Higher education in Russia*, 32 (8-9), 24–47. DOI: 10.31992/0869-3617-2023-32-8-9-24-47 (in Russ.).
4. Kaminchenko, D. I. (2024). Political Image during the Electoral Period: Based on Materials of the Pre-election Campaign in the Nizhny Novgorod Region in 2023. In: *Izvestiya of Saratov University. Sociology. Politology*, 24 (1), 80–89. DOI: 10.18500/1818-9601-2024-24-1-80-89 (in Russ.).
5. Carter, J. (2023). *Neural Networks Beginnings*. Moscow: Avtor publ. (in Russ.).
6. Kulikova, N. R. (2020). Automated Text Generation (On the Figure Skating News Example). In: *Digital Science*, 6, 64–74 (in Russ.).
7. Le Cun, Y. (2021). *Quand la machine apprend: La révolution des neurones artificiels et de l'apprentissage profond*. Moscow: Alpina Pro publ. (in Russ.).
8. Maltsev, N. D. (2024). Structural and Philological Features of Text Generative Neural Networks. In: *Neophilology*, 10, 452–464. DOI: 10.20310/2587-6953-2024-10-2-452-464 (in Russ.).
9. Olomskaya, N. N. & Yurova, E. A. (2025). Linguistic Markers of Generated News (Based on the Presidential Elections in the USA, 2024). In: *Interdisciplinary Aspects of Linguistic Research*. Krasnodar: Kuban State University publ., pp. 220–229. (in Russ.).
10. Rosenthal, D. E. (2011). *Russian Language. Spelling and Punctuation*. Moscow: Eksmo publ. (in Russ.).
11. Rosenthal, D. E. (2001). *Handbook of the Russian language. Practical Stylistics*. Moscow: ONIX 21<sup>st</sup> century publ., World and Education publ. (in Russ.).

12. Tikhonova, E. V. (2023). ChatGPT in the Context of Scientific Communication. In: *Storage and processing of Farm Products*, 3, 8–12. DOI: 10.36107/spfp.2023.518 (in Russ.).
13. Chudinov, A. P. (2006). *Political Linguistics*. Moscow: Flinta publ., Science publ. (in Russ.).
14. Sheigal, E. I. (2000). *Semiotics of Political Discourse* [dissertation]. Volgograd (in Russ.).
15. Becker, J., Wahle, J. P., Gipp, B., et al. (2024). Text Generation: A Systematic Literature Review of Tasks, Evaluation, and Challenges. In: *Cornell University. ArXiv.org*. URL: <https://arxiv.org/abs/2405.15604> (accessed: 05.12.2024). DOI: 10.48550/arXiv.2405.15604.
16. Galperin, I. R. (2014). *English Stylistics*. Moscow: URSS.
17. Grinin, I. L. (2020). The Generation of Coherent Texts, Analysis of Neural Network Mechanics. Mechanics Two – Learning Model for Working with a Neural Network. In: *Innovation & Investment*, 11, 173–176.
18. Sboev, A. G., Gryaznov, A. V., Rybka, R. B., et al. (2023). A Deep Learning Method based on Language Models for Processing Natural Language Russian Commands in Human Robot Interaction. In: *Research Result. Theoretical and applied linguistics*, 9 (1), 174–191. DOI: 10.18413/2313-8912-2023-9-1-1-1.