

Филологические науки. Вопросы теории и практики Philology. Theory & Practice

ISSN 2782-4543 (online) ISSN 1997-2911 (print) 2024. Том 17. Выпуск 7 | 2024. Volume 17. Issue 7 Материалы журнала доступны на сайте (articles and issues available at): philology-journal.ru



Признаки сгенерированного текста в академическом дискурсе: проблема идентификации

Черкасова М. Н., Тактарова А. В.

Аннотация. Современный этап развития академического дискурса характеризуется значительными трансформациями, связанными с внедрением искусственного интеллекта (ИИ). Лицо «академического дискурса» меняется, что требует анализа и идентификации релевантной социально-коммуникативной деятельности в рамках дискурса. Появилась необходимость критического осмысления взаимодействия человека и ИИ в рамках академического взаимодействия, а именно с точки зрения экспертной идентификации сгенерированного текста по языковым признакам. Цель исследования — выявить языковые признаки сгенерированного русскоязычного текста в академическом дискурсе для идентификации устных и письменных академических текстов. Научная новизна исследования заключается в том, что впервые на базе библиометрических показателей представлен комплексный перечень признаков сгенерированного русскоязычного академического текста, полученный в результате анализа отечественных теоретических и экспериментальных исследований (авторский эксперимент с нейросетью с заданным промптом). Результаты исследования позволяют продемонстрировать языковые маркеры русскоязычного сгенерированного текста в рамках академического дискурса на данном синхронном языковом срезе.



Attributes of generated text in academic discourse: the problem of identification

M. N. Cherkasova, A. V. Taktarova

Abstract. The current stage of academic discourse development is characterized by significant transformations associated with the introduction of artificial intelligence (AI). The face of academic discourse is changing, which requires analyzing and identifying relevant socio-communicative activities within the discourse. A need has emerged to critically conceptualize human-AI interaction within academic interaction, namely in terms of expert identification of generated text based on linguistic features. The aim of the study is to identify the linguistic features of the generated Russian-language text in academic discourse for the identification of oral and written academic texts. The scientific novelty of the study lies in the fact that for the first time on the basis of bibliometric indicators a comprehensive list of signs of the generated Russian-language academic text, obtained as a result of analyzing domestic theoretical and experimental studies (author's experiment with a neural network with a given prompt) is presented. The results of the study allow us to demonstrate the linguistic markers of Russian-language generated text within the framework of academic discourse at a given synchronic linguistic slice.

Введение

Актуальность исследования обусловлена тем, что академический дискурс 21 века претерпевает значительные изменения. Являясь институциональным дискурсом, он представляет собой социально-коммуникативную (речевую) деятельность в рамках образования, просвещения и науки, которая объединяет субъекты, объекты и результаты этой деятельности (Стеблецова, 2020, с. 10). Стремительное вхождение Интернет-технологий, поиск модернизаторских решений и изменение когнитивного вектора образования на компетентностный (Афанасьев, Васильева, 2023) становятся причиной формирования нового «лица» академического дискурса и когнитивного академического поведения. Интеграция искусственного интеллекта (ИИ) в академический дискурс становится обязательным элементом учебного процесса. И обучающий, и обучаемый ищут способы и методы, позволяющие оптимально использовать образовательное поле (Кононенко, 2023; Шефиева, Бессарабова, 2023), оптимизировать (Черкасова, 2024) и оценивать (Тактарова, Черкасова 2022) свою деятельность.

Публикационная активность становится маркером компетентности, квалификации и эффективности. Но навыки создания самостоятельного академического текста у современных студентов, магистрантов и аспирантов отсутствуют. В результате появляется скомпилированный текст из разных источников, который не проходит проверку на Антиплагиат-вуз.рф или сгенерированный ИИ текст, который тоже маркируется программой.

Но если письменные работы можно проверить с помощью того же ИИ, который и определяет генерацию в системе Антиплагиат-вуз.рф, то тексты в формате устного академического выступления диагностировать очень сложно, что дает повод использовать нейросети для создания текста. При этом теряется навык самостоятельного создания академического текста, который и носит целеполагающий характер.

Озабоченность по поводу создания текста и его оценки в рамках академического дискурса впервые в России была высказана в январе 2023 г., когда при помощи ChatGPT с ИИ была написана и защищена дипломная работа. Наметилась тенденция к росту сгенерированных работ студентов (Безуглый, Ершова, 2023), что связано с созданием современных мощных нейросетей и их относительной доступностью.

Нельзя не согласиться, что современная информационная насыщенность – путь к многочисленным информационным ресурсам, с одной стороны, но, с другой, появляется необходимость критического осмысления этой информации (Короткина, 2017, с. 9). От пассивного потребления мы перешли к интерактивному потреблению информации, которое и является релевантным в академическом поле. Но использование чатбота при написании академического текста возвращает нас к пассивному, часто неадекватному академическому поведению в рамках дискурса, нарушается один из основных признаков академической грамотности – критический (Короткина, 2017; Green, Beavis 2012). По мнению И. Б. Короткиной, именно это и удерживает «социально-критический фокус научной мысли» (Короткина 2017, с. 10), формирующий академический дискурс с точки зрения динамики техно-семиотического ландшафта.

Несмотря на регулярные публикации в этом направлении проблема идентификации искусственных текстов в академическом дискурсе на основе универсальных признаков и акцентуаторов сгенерированного текста находится в стадии разработки и является дискуссионной как в России, так и за рубежом.

Для достижения поставленной цели необходимо решить следующие научно-исследовательские задачи:

- проанализировать работы, посвященные идентификации сгенерированного текста в академической среде с целью выявления универсальных признаков сгенерированного текста (далее СТ);
- провести эксперимент с CT по заданному промпту (команде) для иллюстрации универсальных признаков и выявления новых маркеров искусственного текста.

При решении задач, определенных в работе, использовались следующие методы исследования: методы библиометрического и теоретического анализа для выяснения степени разработанности проблемы (задача 1) согласно персоналиям; экспериментальный метод для фиксации средств и приемов (задача 2), позволяющих получить корректные данные для дальнейшей иллюстрации и интерпретации проблемы. Комплекс общенаучных (анализ, синтез, обобщение и систематизация) и частных методов (метод стратификации для изучения 2-х типов текста, метод субституции для иллюстрации релевантной замены единицы в сгенерированном тексте, метод лингвостилистического анализа, ориентированный на выявление характерных/нехарактерных признаков употребления единиц в тексте) для презентации универсального списка маркеров сгенерированного текста в данный период.

Теоретической базой исследования послужили работы в области академического текста (Афанасьев, Васильева, 2023; Короткина, 2017; Стеблецова, 2020; Green, Beavis 2012), критического мышления (Шефиева, Бессарабова, 2023), интеграции ИИ в академический дискурс (Безуглый, Ершова, 2023; Капустина, Ермакова, Калюжная, 2023; Кононенко, 2023; Сысоев, Филатов, 2023; Черкасова, 2024) и работы, затрагивающие проблемы идентификации сгенерированного текста (Айдагулова, 2023; Прохоров, Асадчая, 2023; Тельпов, Ларцина, 2023; Туркулец, 2023).

Практическая ценность работы заключается в создании перечня идентифицирующих лингвистических и экстралингвистических признаков сгенерированного академического текста, которые помогут распознать такой текст при оценке устных и письменных текстов без инструментальных способов идентификации. Результаты проведенного исследования могут быть использованы в университетском курсе по теории языка.

Обсуждение и результаты

Благодаря исследованиям, по большей части эмпирического характера за 2023 и 2024 гг., начал формироваться перечень признаков, идентифицирующих текст как сгенерированный нейросетью. В силу новизны самой проблемы, работы в этом направлении только начаты и не носят массовый характер. Вектор исследования намечен лишь пунктирно. В этом разделе мы ограничились только работами отечественных исследователей (Айдагулова, 2023; Безуглый, Ершова, 2023; Капустина, Ермакова, Калюжная, 2023; Сысоев, Филатов, 2023; Тельпов, Ларцина, 2023; Туркулец, 2023) по объективным причинам: исследуется именно признаки СТ на русском языке, работы по признакам СТ на английском языке есть, но они не могут выступать теоретико-методологической базой для нашего исследования. Благодаря заданным параметрам были выявлены признаки СТ, которые представлены в таблице с точки зрения качественной характеристики (признак), демонстрации (пример/описание) и персоналии (см. Таблица 1).

Представленный набор признаков – это обобщенный результат работы ИИ-детектора и экспертаисследователя. Таким образом, смешанный метод позволяет с математической точностью рассчитать количественное соотношение словоупотреблений с СТ и ЕТ, используя цифровые способы идентификации при участии исследователя-эксперта, который корректирует, направляет, мониторит и контролирует процесс работы с ИИ. 2228 Теория языка

Таблица 1. Признаки сгенерированного текста

Признак	Пример / описание	Персоналии
Лексическое	Лексика естественного текста значительно богаче, разнооб-	Р. Е. Тельпов, С. В. Ларцина
однообразие:	разнее (Р. Е. Тельпов, С. В. Ларцина, 2023, с. 60-61)	
Семантическая неточ-	Подробнее* в работе Т. А. Безуглого, М. Е. Ершовой (2023, с. 210)	Т. А. Безуглый, М. Е. Ершова
Ность отдельных слов	VIDALIJIJAGIVAA LIGINIJIJAJIJA DVAGITA VIDAVVIJAGIVAA DAĞA VADAVVIJA	Т. А. Безуглый, М. Е. Ершова
Нарушение лексической сочетаемости в устойчи-	- хроническое нарушение вместо хроническое заболевание; - иметь большую роль вместо (значение) (Т. А. Безуглый,	1. А. везуглый, м. Е. Ершова
вых словосочетаниях	м. Е. Ершова, 2023, с. 209)	
Лексические повторы	Подробнее в работах А. Р. Айдагуловой (с. 155), Р. Е. Тельпо-	А. Р. Айдагулова; Р. Е. Тельпов,
(слов и словосочетаний)	ва, С. В. Ларциной (2023, с. 57-59)	С. В. Ларцина
во всем тексте		-
Частотный повтор слов	Частота употребления ключевых слов (из заголовка) в СТ выше	Р. Е. Тельпов, С. В. Ларцина
из заголовка	более чем в 2 раза, чем в естественном тексте, что нецеле-	
	сообразно (Р. Е. Тельпов, С. В. Ларцина, 2023, с. 57-59)	m + n
Тавтология	уровень глюкозы в крови, облегчая поглощение глюкозы (Т. А. Безуглый, М. Е. Ершова, 2023, с. 209)	Т. А. Безуглый, М. Е. Ершова
Частотность полнознач-	в сгенерированных текстах выше приблизительно в 2 раза	Р. Е. Тельпов, С. В.Ларцина
ных слов	(Р. Е. Тельпов, С. В. Ларцина, 2023, с. 62)	
Повтор одного и того же предлога в предложении	в крови в результате (Т. А. Безуглый, М. Е. Ершова, 2023, с. 209)	Т. А. Безуглый, М. Е. Ершова
Однотипные	1. Подлежащее и сказуемое выражены существительными:	Т. А. Безуглый, М. Е. Ершова
синтаксические	Диабет – это	
конструкции	2. Сложноподчиненное предложение с придаточным опре-	
	делительным и изъяснительным (с союзами/союзными сло-	
	вами «который», «что»)	
	3. Последовательная связь в предложении при помощи ука-	
	зательного местоимения «это»	
	4. Причастные обороты в постпозиции 5. Прямой порядок слов	
	6. Частотное выражение предиката с помощью глагола «мочь»	
	7. Длинные предложения с обилием причастных и деепри-	
	частных оборотов, придаточных предложений	
	(Т. А. Безуглый, М. Е. Ершова, 2023, с. 209-210)	
Нарушение логической	Подробнее в работе А. Р. Айдагуловой (2023, с. 155)	А. Р. Айдагулова
связи между сегментом		
СТ и естественного текста (далее ЕТ)		
Расплывчатость форму-	Подробнее в работах А. Р. Айдагуловой (2023); Т. А. Безугло-	Все авторы (единое мнение
лировок, отсутствие	го и М. Е. Ершовой (2023); Л. В. Капустиной, Ю. Д. Ермаковой	отмечается как у отече-
конкретики	и Т. В. Калюжной (2023); П. В. Сысоев и Е. М. Филатов (2023);	ственных, так и у зарубеж-
	Р. Е. Тельпов и С. В. Ларцина (2023); И. А. Туркулец (2023)	ных исследователей, биб-
		лиометрические показатели
		которых не были приведены в таблице)
Пунктуационные ошибки	Подробнее в работе Т. А. Безуглого и М. Е. Ершовой (2023, с. 209)	Т. А. Безуглый, М. Е. Ершова
(схожи с ошибками в ЕТ)		, ,
Нерелевантные примеры,	Необходима проверка и корректировка. По мнению Л. В.	А. Р. Айдагулова;
искажение фактов,	Капустиной, Ю. Д. Ермаковой, Т. В. Калюжной (2023, с. 126),	Л. В. Капустина, Ю. Д. Ерма-
ИИ-галлюцинации	это может быть преимуществом в образовательной среде	кова и Т. В. Калюжная;
	и мотивацией к обучению. Подробнее в работах А. Р. Айдагуловой (2023, с. 155); Л. В. Капустиной, Ю. Д. Ермаковой,	П. В. Сысоев, Е. М. Филатов
	Т. В. Калюжной (2023, с. 126); П. В. Сысоева и Е. М. Филатова	
	(2023, c. 288)	
Отсутствие цитат	Подробнее в работе А. Р. Айдагуловой (2023, с. 155)	А. Р. Айдагулова
Нумерованный список,	Подробнее в работе И. А. Туркулец (2023)	И. А. Туркулец
начинающийся с ввод-		
ной информации и за-		
канчивающийся корот-		
ким резюме (часто спи- сок нелогичен)		
Отсутствие ссылок	ИИ объясняет этот факт ограниченной базой данных, хотя	П. В. Сысоев, Е. М. Филатов.
на русском языке	запрос формулировался на русском языке (Сысоев, Филатов,	Об условных ограничениях
	2023, c. 287)	базы данных говорят и зару-
		бежные ученые

^{*}Подробное ознакомление с примерами возможно по ссылкам на работы, указанные в библиографическом списке.

Для дальнейшей иллюстрации выявленных признаков СТ нами был проведен эксперимент с участием чат-бота. Мы предложили ИИ на базе https://gpt-chatbot.ru/chat-gpt-ot-openai-dlya-generacii-teksta написать научный текст на 500 слов на тему «Проблема идентификации сгенерированного текста в академическом дискурсе» с целью иллюстрации выделенных в таблице признаков. Мы намеренно в запросе повторили название своей статьи. Для удобства анализа полученный текст (335 слов) мы разбили на 5 сегментовпримеров (рисунки 1-5), которые повторяют рубрики сгенерированного текста.

Проблема идентификации сгенерированного текста в академическом дискурсе

Введение

С развитием искусственного интеллекта и технологий обработки естественного языка (Natural Language Processing, NLP) возникают новые вызовы в области академического исследования. Одним из таких вызовов является проблема идентификации сгенерированных текстов в академическом дискурсе. Сгенерированные тексты, порожденные автоматически, могут быть трудно отличить от текстов, созданных реальными авторами. Это создает потенциальные проблемы с точки зрения академической честности, а также может привести к недоверию к исследованиям и публикациям.

Рисунок 1. Пример 1

Статус исследований

С № С О КОПТЕКСТИВНИИ ПОДСКАЗКАМИ

Исследования в этой области находятся на ранней стадии, но уже сделали значительный прогресс. Многие работы включают в себя использование методов машинного обучения и алгоритмов анализа текста для выявления особенностей, характерных для сгенерированных текстов. Однако, существует необходимость в разработке более сложных и надежных моделей, способных эффективно различать сгенерированные тексты от реальных.

Рисунок 2. Пример 2

Характеристики сгенерированных текстов

Сгенерированные тексты могут иметь ряд характеристик, которые отличают их от текстов, созданных реальными авторами. Некоторые из них включают:

- Отсутствие семантической связности: Сгенерированные тексты могут содержать логические ошибки, несвязанные идеи или противоречивые утверждения из-за ограничений алгоритмов генерации.
- Статистические аномалии: Сгенерированные тексты могут иметь неправильное распределение слов, частоту повторяющихся фраз или использование необычных слов или конструкций.
- Стилевые отличия: Автоматически сгенерированные тексты могут не иметь определенного стиля, характерного для реальных авторов, либо иметь стиль, характерный для определенных алгоритмов генерации.

Рисунок 3. Пример 3

- Машинное обучение: Модели машинного обучения могут быть обучены на больших корпусах текстов для выявления статистических особенностей, характерных для сгенерированных текстов.
- Анализ семантики и контекста: Методы NLP могут быть использованы для анализа семантики и контекста текста, чтобы выявить несвязанные или противоречивые утверждения.
- Статистические тесты: Различные статистические тесты могут быть применены для определения аномалий в распределении слов или фраз.

Рисунок 4. Пример 4

Заключение

Проблема идентификации сгенерированных текстов в академическом дискурсе представляет собой важное направление исследований. Развитие более точных и эффективных методов идентификации является ключевым для поддержания академической честности и доверия к результатам исследований. Дальнейшие исследования в этой области могут привести к разработке инструментов и методов, способных эффективно различать сгенерированные тексты от реальных, и тем самым содействовать развитию академической среды.

2230 Теория языка

Основываясь на выявленных признаках СТ и данных эксперимента, мы смогли получить дополнительные маркеры (акцентные признаки) СТ.

Логико-композиционная корректность и релевантность смысла

Композиция СТ отвечает требованиям для композиции научного текста. Предложена релевантная рубрикация. Но при этом само изложение материала дано довольно размыто, особенно введение и заключение. Если в Примерах 1 (введение) и 5 (заключение) заменить анализируемое понятие «сгенерированный текст», то текст заключения с незначительными корректировками будет универсальным для всех исследований. Пример 2 содержит уже терминологию (машинное обучение, сгенерированный текст, реальный текст), относящуюся непосредственно к исследованию, но опять-таки чрезмерная обобщенность позволяет использовать такую модель презентации рубрики «Статус исследования» для широкого круга проблем. Пример 3 представлен адекватным набором признаков СТ, но без демонстрации примеров, иллюстрирующих теоретические положения. Пример 4 (Методы идентификации), на наш взгляд, корректен. Такая триада методов идентификации совершенно оправдана, так как затрагивает предельные величины (только ИИ – метод 1, только человек – метод 2 и промежуточную точку (человек+ ИИ – метод 3)). Таким образом, при конечном вынесении вердикта о СТ мнение человека-эксперта должно учитываться. Иначе происходит крен только в сторону инструментального метода, который мы не можем игнорировать при современном высоком уровне развития технологий. Но отметим, что в СТ нет отсылки к экспертному анализу именно человеком, а это подразумевает семантико-контекстуальный анализ. Смешанный метод, на наш взгляд, коррелирует со «статистическими тестами», когда эксперт, используя статистические данные, полученные при помощи машины, и определяет по частотности тех или иных единиц их релевантное/нерелевантное словоупотребление.

Языковые признаки СТ

- 1) Для СТ характерен такой тип речи, как описание;
- 2) повторы ключевых слов, а именно единиц из заявленной темы исследования (заголовка). Частотный повтор слов из заголовка является на наш взгляд, одним из самых ярких признаков СТ, что во многом приводит к речевой избыточности (автоматически сгенерированный текст) и излишней акцентуации темы исследования;
- 3) частотность предиката «мочь» со значением «предположительности», «вероятности» составляет 10 раз на СГ (335 слов), что свидетельствует об однотипности синтаксической структуры предложений (См. Пример 4: могут использоваться, могут быть обучены, могут быть использованы, могут быть применены. В четырех предложениях Примера 4 одна и та же синтаксическая модель с предикатом «мочь».);
 - 4) стилистические ошибки, в результате которых нарушается лексико-грамматическая сочетаемость:
- Сгенерированные тексты, порожденные автоматически, могут быть трудно отличить от текстов, созданных реальными авторами. На наш взгляд, ошибка такого типа возникла в результате перевода с английского на русский язык, так как сеть, в которой проводился эксперимент, создана не в России. Налицо просто «калька» с английского «it can be difficult to differ», но неудачно интегрированная в русский текст: соблюдена категория числа тексты ... могут (множественное число), но инфинитив из английского языка не преобразован в страдательное причастие (трудно отличимы). Если же убрать предикат «могут», то исчезнет его модальность.
- Исследования в этой области находятся на ранней стадии, но уже сделали значительный прогресс. В предложении нарушена лексическая сочетаемость. Очевидно, что сами исследования прогресс сделать не могут, ученые могут сделать прогресс. Поправка с учетом ошибки: отмечается прогресс в исследованиях. Ошибки такого типа свидетельствуют о несовершенстве языковой модели ИИ;
- 5) частотность цепной связи во Введении и Заключении, что иллюстрирует шаблонность и схематизм СТ. Пример 1: Новые вызовы... Одним из таких вызовов <... > идентификация сгенерированных текстов... Сгенерированные тексты... Пример 5: Проблема идентификации сгенерированных текстов... Развитие <... > методов идентификации является <... > к результатам исследований. Дальнейшие исследования в этой области...
- 6) морфологическая характеристика. В тексте преобладают знаменательные части речи, отсутствуют частицы, позволяющие внести дополнительный смысл в изложение материала, вводные слова, усиливающие эмоциональную нагрузку текста и маркирующие последовательность презентации материала, демонстрирующие уверенность/неуверенность или служащие аттракторами, что является маркером рассуждения, повествования и аргументации. Отсутствие таких частей речи делает текст штампованным и клишированным, сухим и безобразным, что характерно для официально-делового стиля. Трафаретность сгенерированности отчетливо видна в Примерах 1 и 5.
 - 7) пунктуационные и графические ошибки.
- Ошибочное обособление «однако» в значении «но» в начале предложения. Отметим, что такая ошибка отмечалась также во всех сгенерированных текстах, которые были проанализированы и за рамками данного исследования.
- Графическая ошибка (6 раз) (оформление части предложения после двоеточия строчной буквой (*Машинное обучение*: *Модели машинного обучения*), носящая системный характер в тексте. На наш взгляд, эти системные ошибки следствие несовершенства алгоритма генерации, что еще раз подчеркивает автоматизм и механическое составление текста.

Экстралингвистические признаки

- 1) В тексте отсутствуют примеры, позволяющие проиллюстрировать теоретические положения, нет ссылок на авторитетные исследования и персоналии, что недопустимо в научном тексте.
 - 2) Разностилевое оформление списка признаков (Пример 2, 3).
- 3) Ссылки не были приведены в тексте, хотя задача была «написать научный текст». Отметим, что проблема некорректных ссылок, приравненных к ИИ-галлюцинациям, отмечается нами при анализе СТ и за рамками данного исследования, при конкретной просьбе дать ссылки на работы.

Выявленные на основе анализа теоретических и экспериментальных исследований универсальные признаки и маркеры сгенерированного текста являются ориентиром для идентификации искусственного (нерелевантного с точки зрения академической этики) текста в академическом дискурсе.

В результате предложен перечень дополнительных маркеров сгенерированного текста с точки зрения логико-композиционной корректности и релевантности смысла, языковых и экстралингвистических признаков. Приведенный перечень (на данном этапе развития интеллектуальных систем) идентификаторов и акцентуаторов сгенерированного текста не претендует на статус завершенности, так как исследования СТ находятся в начальной стадии разработки.

Заключение

Впервые сформулирован комплексный список признаков СТ (Таблица 1): лексическое однообразие, семантическая неточность отдельных слов, нарушение лексической сочетаемости в устойчивых словосочетаниях, лексические повторы (отдельных слов и словосочетаний) во всем тексте, частотный повтор слов из заголовка, тавтология, частотность полнозначных слов, повтор предлога в предложении, однотипные синтаксические конструкции, нарушение логической связи между сегментами текста (в случае с интеграцией СТ в ЕТ), системные пунктуационные ошибки, нерелевантные примеры и искажение фактов, известные как ИИгаллюцинации, отсутствие цитат и ссылок на авторитетные работы, нерабочие или некорректные ссылки, обязательный нумерованный список (часто некорректный и нелогичный).

В результате эксперимента выявлены дополнительные маркеры СТ, затрагивающие лексико-композиционный и смысловой уровень (шаблонность, неконкретность, размытость, трафаретность содержания) и акцентуаторы СТ на уровне лексики (многообразные повторы ключевых слов, часто приводящие к речевой избыточности), грамматики (доминирование предиката «мочь», цепной связи в тексте, знаменательных слов и однотипных синтаксических конструкций при отсутствии вводных конструкций); экстралингвистические признаки (отсутствие примеров, ссылок, персоналий, но с обязательным присутствием нумерованного списка).

Перечень особенностей СТ носит условно универсальный характер для определенного языка, так как сам ИИ находится в процессе самообучения, что позволяет ему учиться на своих же ошибочных текстах, которые он сам и генерирует.

В качестве перспективы для дальнейших исследований возможен компаративный анализ признаков СТ, так как для каждой нейросети и каждого языка характерны свои маркеры.

Источники | References

- 1. Айдагулова А. Р. Особенности текстов, сгенерированных искусственным интеллектом // Вестник Башкирского государственного педагогического университета им. М. Акмуллы. 2023. № 4 (72).
- 2. Афанасьев Н. Д., Васильева А. А. Академический дискурс как основа формирования академической грамотности студента-международника // Przeglad Rusycystyczny. 2023. № 1 (181).
- 3. Безуглый Т. А., Ершова М. Е. Использование текстовых нейросетей и искусственного интеллекта в учебных работах студентов // Проблемы современного образования. 2023. № 5.
- 4. Капустина Л. В., Ермакова Ю. Д., Калюжная Т. В. ChatGPT и образование: вечное противостояние или возможное сотрудничество? // Научно-методический электронный журнал «Концепт». 2023. № 10. https://e-koncept.ru/2023/231099.htm
- 5. Кононенко А. П. Популяризация чат-ботов в лингвистическом и методическом аспектах // Преподаватель высшей школы в XXI веке. Труды 20-й Юбилейной международной научно-практической конференции. Ростов-на-Дону, 2023.
- **6.** Короткина И. Б. Академическая грамотность и методы глобальной научной коммуникации // Научный редактор и издатель. 2017. Т. 2. № 1. https://doi.org/10.24069/2542-0267-2017-1-8-13
- 7. Стеблецова А. О. Академический дискурс в западных исследованиях на рубеже XX–XXI вв.: эволюция направлений и концепций // Вестник Волгоградского государственного университета. Серия 2. Языкознание. 2020. Т. 19. № 5. https://doi.org/10.15688/jvolsu2.2020.5.1
- 8. Сысоев П. В., Филатов Е. М. ChatGPT в исследовательской работе студентов: запрещать или обучать? // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2023. Т. 28. № 2. https://doi.org/10.20310/1810-0201-2023-28-2-276-301

2232 Теория языка

9. Тактарова А. В. Амбивалентный характер оценивания образовательного процесса высшей школы периода COVID-19 в динамике (на примере прагмалингвистического опроса студентов) // Тактарова А. В., Черкасова М. Н. // Современные исследования социальных проблем. 2022. Т. 14. № 3.

- **10.** Тельпов Р. Е., Ларцина С. В. Типовые различия естественных и сгенерированных нейронной сетью текстов в квантитативном аспекте // Научный диалог. 2023. Т. 12. № 7. https://doi.org/10.24224/2227-1295-2023-12-7-47-65
- 11. Туркулец И. А. Композиционные особенности текстов, сгенерированных chatgpt, как маркер несамостоятельности выполнения работ студентами // Правовая реальность в условиях цифровизации общества: материалы Всероссийской научно-практической конференции (г. Хабаровск, 9-11 ноября 2023 г.). Хабаровск: Дальневосточный государственный университет путей сообщения, 2023.
- 12. Черкасова Е. А. Комплекс команд по выполнению упражнений и заданий в рамках обучения грамматике английского языка студентов технических вузов посредством чат-бота ChatGPT // Филологические науки. Вопросы теории и практики. 2024. Т. 17. № 5. https://doi.org/10.30853/phil20240236
- **13.** Шефиева Э. Ш., Бессарабова О. Н. Обучение иностранному языку студентов неязыкового вуза как технология развития критического мышления (возможности и ограничения в условиях цифровизации) // Вестник Южно-Уральского государственного гуманитарно-педагогического университета. 2023. № 2 (174).
- **14.** Literacy in 3D: An Integrated Perspective in Theory and Practice / Green B., Beavis C. (Eds.). Camberwell, Victoria: Australian Council for Educational Research (ACER), 2012.

Информация об авторах | Author information



Черкасова Марина Николаевна¹, к. филол. н. **Тактарова Анна Валерьевна**², к. филол. н.

1,2 Ростовский государственный университет путей сообщения



Marina Nikolaevna Cherkasova¹, PhD Anna Valer'evna Taktarova², PhD

1, 2 Rostov State Transport University

Информация о статье | About this article

Дата поступления рукописи (received): 06.06.2024; опубликовано online (published online): 22.07.2024.

Ключевые слова (keywords): сгенерированный текст; академический текст; идентификация текста; нейросеть; ChatGPT; академический дискурс; generated text; academic text; identification of the text; neural networks; ChatGPT; academic discourse.

¹ chercasovamn-rostov@rambler.ru, ² annataktar@yandex.ru