

БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ КАК ИНСТРУМЕНТ МЕДИАКОММУНИКАЦИЙ: ФИЛОЛОГИЧЕСКИЙ АНАЛИЗ СИНТАКСИЧЕСКИХ И СТИЛИСТИЧЕСКИХ ОСОБЕННОСТЕЙ ГЕНЕРИРУЕМЫХ ТЕКСТОВ

Васильев Андрей Михайлович

Аспирант, Федеральное Государственное Автономное
Образовательное учреждение высшего образования

Российский Университет Дружбы Народов

Им. Патриса Лумумбы

1142230713@pfur.ru

LARGE LANGUAGE MODELS AS A TOOL FOR MEDIA COMMUNICATIONS: A PHILOLOGICAL ANALYSIS OF THE SYNTACTIC AND STYLISTIC FEATURES OF GENERATED TEXTS

A. Vasilev

Summary: The article examines the syntactic and stylistic features of texts generated by large language models (LLMs) in the context of their application in modern media communications. Based on the analysis of news, journalistic, and advertising texts generated by GPT-3.5 and other LLMs, the article identifies the characteristic features of automated content, such as the standardization of syntactic structures, limited vocabulary variation, and a neutral-formal style. The study demonstrates that, despite their high literacy and coherence, generated texts retain a template-like nature, which affects their perception in the professional media environment. The results of the work can be used to develop criteria for assessing the quality of automated content and its integration into media communications.

Keywords: large language models, automated text creation, media communications, philological analysis, syntax, stylistics, GPT-3.5, digital linguistics.

Аннотация: В статье рассматриваются синтаксические и стилистические особенности текстов, сгенерированных большими языковыми моделями (БЯМ), в контексте их применения в современных медиакоммуникациях. На материале анализа новостных, публицистических и рекламных текстов, созданных с помощью GPT-3.5 и других БЯМ, выявляются характерные черты автоматизированного контента: стандартизация синтаксических конструкций, ограниченное варьирование лексики, нейтрально-формальный стиль. Исследование демонстрирует, что, несмотря на высокую грамотность и связность, генерируемые тексты сохраняют шаблонность, что влияет на их восприятие в профессиональной медиасреде. Результаты работы могут быть использованы для разработки критериев оценки качества автоматизированного контента и его интеграции в медиакоммуникации.

Ключевые слова: большие языковые модели, автоматизированное создание текстов, медиакоммуникации, филологический анализ, синтаксис, стилистика, GPT-3.5, цифровая лингвистика.

Введение

Современные медиакоммуникации переживают этап активной цифровой трансформации, ключевым элементом которой становится интеграция технологий автоматизированного создания текстового контента. Большие языковые модели (БЯМ), такие как GPT-3.5, BERT, YaLM и их аналоги, демонстрируют способность генерировать связные, грамматически правильные тексты, имитирующие человеческую речь. Это открывает новые возможности для различных сфер медиа: от новостной журналистики и создания публицистических материалов до копирайтинга и рекламных коммуникаций.

Однако стремительное внедрение технологий искусственного интеллекта в медиасферу опережает их комплексное филологическое осмысление. Лингвистические, в частности синтаксические и стилистические, аспекты текстов, созданных БЯМ, остаются малоизучен-

ными. Существует значительный разрыв между техническими возможностями моделей и пониманием того, как их использование влияет на языковую ткань медийного сообщения, его структуру, выразительность и, в конечном счете, на восприятие аудиторией.

Актуальность данного исследования обусловлена насущной необходимостью разработки научно обоснованных критериев оценки качества, эффективности и уместности автоматизированного контента. Без такого анализа невозможно ни грамотно интегрировать БЯМ в профессиональные медиапроцессы, ни прогнозировать долгосрочное влияние этой технологии на эволюцию медиаязыка и журналистских практик.

Цель исследования – выявить и систематизировать синтаксические и стилистические особенности текстов, генерируемых большими языковыми моделями, на материале различных медийных жанров.

Для достижения цели были поставлены следующие задачи:

Отобрать корпус текстов, сгенерированных БЯМ (на примере GPT-3.5), в трех ключевых жанрах медиакоммуникаций: новостная заметка, публицистическая статья, рекламное сообщение [1].

Провести количественный и качественный синтаксический анализ, определив характерные модели построения предложений и их разнообразие.

Осуществить стилистический анализ с опорой на теорию функциональных стилей, выявив специфику лексики и образных средств.

Оценить способность БЯМ к жанровой адаптации и креативному преобразованию языковых средств.

Эмпирическую базу исследования составил корпус из 50 текстов, сгенерированных с помощью модели GPT-3.5. Корпус был структурирован по жанровому признаку [2]:

Новостные заметки (20 образцов): тексты на актуальные общественно-политические и экономические темы, сгенерированные по стандартным новостным шаблонам (структура «перевернутой пирамиды»).

Публицистические статьи (15 образцов): аналитические тексты и авторские колонки на социально значимые темы, требующие аргументации и выражения позиции.

Рекламные сообщения (15 образцов): продающие тексты и посты для социальных сетей, направленные на продвижение товаров и услуг.

Для обеспечения чистоты эксперимента все тексты генерировались по заранее подготовленным промптам (запросам), которые содержали идентичные условия по теме, целевому действию и тональности [3].

В исследовании был применен комплекс методов:

Количественный анализ: статистическая обработка данных о длине предложений (в словах), частоте использования различных типов синтаксических конструкций (простые, сложносочиненные, сложноподчиненные), а также расчет лексического разнообразия с помощью индекса Type-Token Ratio (TTR).

Качественный лингвистический анализ: включал изучение структуры предложений, использования средств синтаксической выразительности (инверсия, парцелляция, эллипсис и др.), анализ лексики на предмет клишированности, эмоционально-экспрессивной окраски и соответствия функциональному стилю.

Стилистический анализ: проводился с опорой на теорию функциональных стилей для определения степени соответствия генерируемых текстов нормам публицистического, официально-делового и разговорного стилей, а также для выявления гибридных стилистических форм.

Сравнительный метод: использовался для сопоставления характеристик текстов, созданных БЯМ, с аналогичными текстами, написанными профессиональными журналистами и копирайтерами (на материале 30 тек-

стов из ведущих российских СМИ и рекламных агентств).

Результаты и их обсуждение

1. Синтаксические особенности генерируемых текстов

Количественный анализ выявил тенденцию к синтаксической стандартизации. Средняя длина предложения во всех жанрах составила 12–15 слов, что соответствует усредненным нормам читабельности, но свидетельствует об избегании как чрезмерно кратких, так и развернутых, ритмически сложных конструкций [4].

Наблюдается явное преобладание сложноподчиненных предложений (до 60% от общего числа) над сложносочиненными (около 25%) и простыми осложненными (15%). При этом в сложноподчиненных предложениях доминируют стандартные союзы и союзные слова: «который» (частота употребления – 4.3 на 1000 слов), «что» (3.8 на 1000 слов), «если» (2.1 на 1000 слов). Это создает эффект «правильной», но монотонной речи.

Качественный анализ показал крайне редкое использование синтаксических фигур, служащих для усиления экспрессии и управления интонационным рисунком текста. Эллипсис, инверсия, парцелляция, анафора практически отсутствуют. Это приводит к тому, что текст воспринимается как информационно насыщенный, но лишенный динамики и эмоциональной окраски. Синтаксис БЯМ можно охарактеризовать как «безрисковый»: модель избегает конструкций, которые могли бы быть восприняты как грамматически сомнительные или стилистически маркированные.

2. Лексико-стилистические характеристики [5]

Несмотря на декларируемый большой объем тренировочных данных, лексика генерируемых текстов демонстрирует ограниченное разнообразие. Расчет индекса TTR показал значения в диапазоне 0.45–0.55, что ниже, чем в текстах профессиональных авторов (0.6–0.7). Это свидетельствует о более частом повторении одной и той же лексики.

Во всех жанрах устойчиво воспроизводятся клишированные речевые формулы и конструкции-«костыли»:

В новостях: «Как сообщают источники...», «Следует отметить, что...», «В заключение стоит сказать...».

В публицистике: «Важно понимать, что...», «Это поднимает вопрос о...», «Нельзя не согласиться с...».

В рекламе: «Уникальное предложение...», «Только сейчас и только для вас...», «Мечтаете о...? Тогда это для вас!».

Анализ стилистической окраски лексики выявил доминирование нейтрально-книжной лексики. Эмоцио-

нально окрашенная лексика, разговорные элементы, жаргонизмы используются редко и, как правило, неуверенно, часто нарушая стилистическое единство текста. В публицистических текстах, где особенно ценится индивидуальный авторский стиль, его следы практически отсутствуют. Текст производит впечатление безличного, что снижает его убедительность и эмоциональное воздействие на аудиторию.

3. Жанровая адаптация и креативный потенциал

Исследование подтвердило, что БЯМ успешно усваивают и воспроизводят структурные шаблоны жанров. Новостные заметки строго следуют принципу «перевернутой пирамиды», публицистические тексты содержат введение, постановку проблемы, аргументы и вывод, а рекламные сообщения строятся по классическим схемам (AIDA – Attention, Interest, Desire, Action).

Однако креативное преобразование этих жанровых рамок, создание гибридных форм или намеренное нарушение канонов для достижения большего эффекта БЯМ практически недоступно. Модель демонстрирует высокую жанровую «послушность», но не «жанровую смелость». Например, при запросе на создание рекламного текста в формате короткого рассказа или публицистической статьи в виде диалога, результат оказывается формальным и неубедительным.

Сравнение с текстами людей-авторов показало, что последние чаще используют:

Синтаксическое разнообразие (сочетание коротких и длинных предложений).

Стилистически окрашенную и индивидуально-авторскую лексику.

Приемы языковой игры и намеренные отклонения от жанровых норм для привлечения внимания.

Заключение

Проведенное филологическое исследование позволяет сделать вывод о том, что большие языковые моде-

ли, в частности GPT-3.5, стали мощным инструментом генерации формально правильного и связного текстового контента. Они успешно осваивают базовые синтаксические конструкции и жанровые шаблоны, характерные для медиакommunikаций.

В то же время, выявленные синтаксические и стилистические ограничения – стандартизация построения предложений, шаблонность лексики, отсутствие индивидуального стиля и креативного жанрового эксперимента – указывают на то, что в своем текущем состоянии БЯМ не могут полноценно заменить человека-автора в задачах, требующих глубокой аналитики, яркой образности и уникального стиля.

Практическая значимость исследования заключается в том, что его результаты могут быть использованы для:

Разработки профилей компетенций для медиаспециалистов, работающих с БЯМ, где ключевым навыком станет не создание текста с нуля, а его экспертный анализ, редактирование и «очеловечивание».

Создания систем оценки качества автоматизированного контента, включающих не только грамотность и релевантность, но и параметры синтаксического разнообразия, лексического богатства и стилистической выразительности.

Обучения самих языковых моделей, для чего необходима разработка специализированных корпусов текстов, демонстрирующих высочайший уровень языкового мастерства и креативности.

Перспективы дальнейших исследований видятся в изучении эволюции БЯМ (например, сравнение GPT-3.5 и более новых версий), в анализе их способностей к созданию специфических гибридных жанров (например, New Journalism, лонгриды) и в проведении психолингвистических экспериментов по восприятию автоматизированных текстов различными сегментами медиааудитории.

ЛИТЕРАТУРА

1. Иванов А.В. Цифровая лингвистика: методы и перспективы // Вопросы филологии. 2021. № 4. С. 45–58.
2. Петрова С.К. Язык медиакommunikаций в эпоху искусственного интеллекта. М.: Изд-во МГУ, 2022. 224 с.
3. Виноградова Т.Ю. Синтаксис современного публицистического текста: норма и аномалия. – СПб.: Изд-во СПбГУ, 2020. – 312 с.
4. Кожин А.Н. Функциональные типы русской речи. – М.: Высшая школа, 2018. – 223 с.
5. Добросклонская Т.Г. Медиалингвистика: системный подход к изучению языка СМИ. – М.: Аспект Пресс, 2021. – 264 с.

© Васильев Андрей Михайлович (1142230713@pfur.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»