

УДК 004.8

Царева Виктория Олеговна
магистрант кафедры перевода
ФГБОУ ВО РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Viktoria Tsareva
Master's degree student
The Herzen State Pedagogical
University of Russia
Saint Petersburg, Russia
v2o2@yandex.ru

О ВОЗМОЖНОСТЯХ ГЕНЕРАЦИИ ЮМОРА СИСТЕМАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Генерация юмора с помощью искусственного интеллекта – это активно развивающаяся область на стыке лингвистики и машинного обучения. В данной статье представлен обзор современных методов создания юмористических текстов с использованием больших языковых моделей, таких как GPT. Рассматриваются ключевые преимущества ИИ, такие как скорость генерации текстов, грамматическая точность и способность воспроизводить стилистические особенности, а также основные ограничения: зависимость от качества обучающих данных, отсутствие контекстуального понимания и трудности передачи невербальных аспектов юмора. Особое внимание уделено этическим проблемам, включая предвзятость данных и риски генерации оскорбительного контента. Анализируются методы оценки юмора, такие как экспертные оценки и «коэффициент комического», а также перспективы дальнейшего развития технологий, включая мультимодальные решения. Статья подчеркивает потенциал ИИ как инструмента поддержки творчества, но отмечает, что создание по-настоящему оригинального юмора, учитывающего нюансы языкового, культурного, ситуативного и контекстуального порядка, остается сложной задачей.

Ключевые слова: искусственный интеллект; генерация юмора; большие языковые модели; обработка естественного языка; оценка юмора.

ON POSSIBILITIES OF HUMOUR GENERATION BY ARTIFICIAL INTELLIGENCE SYSTEMS

Artificial intelligence humour generation is an actively developing field at the intersection of linguistics and machine learning. This article presents an overview of current methods of humour text generation using large language models such as GPT. Key advantages of artificial intelligence, such as speed of text generation, grammatical accuracy and the ability to reproduce stylistic features, are discussed, as well as major limitations, such as dependence on the quality of training data, lack of contextual understanding and difficulties in conveying non-verbal aspects of humour. Particular attention is paid to ethical issues, including data bias and the risks of generating offensive content. Methods for assessing humour, such as expert judgement and the “humour quotient”, are analysed, as well as prospects for further technological developments, including multimodal solutions. The article highlights the potential of artificial intelligence as a tool to support creativity, but notes that creating truly original humour that is linguistically, culturally, situationally and contextually nuanced remains a challenge.

Key words: artificial intelligence; humour generation; large language models; natural language processing; humour evaluation.

Генерация юмора с помощью искусственного интеллекта (ИИ) – это активно развивающаяся область лингвистики и машинного обучения. Современные языковые модели, такие как GPT-3.5 и GPT-4, могут применяться для генерации шуток, языковой игры, стендап выступлений и других юмористических текстов. Однако способности ИИ остаются предметом дискуссий: насколько ИИ понимает юмор, может ли он создавать по-настоящему оригинальные и смешные тексты и каковы этические и технические ограничения таких систем?

Существует несколько основных исследовательских направлений в области цифровой лингвистики, изучающей юмор, – распознавание юмора, генерация юмора, а также различные модели оценки юмора.

Генерация юмора направлена на создание юмористических текстов – шуток, анекдотов, текстов стендап-выступлений, каламбуров. Современные методы генерации юмора можно разделить на три основные категории: шаблонные, нейросетевые и гибридные методы. Шаблонные методы основаны на заранее определённых структурах или шаблонах, которые заполняются словами или фразами, отобранными по определённым критериям. Методы, основанные на шаблонах, «используют лексическую замену, такую как использование синонимии, меронимии, гипонимии» [1, с. 17831], а также подбирают рифмующиеся слова [2, с. 6]. Заполнение шаблона позволяет создавать шутки, соответствующие определённому стилю или жанру, но ограниченность предопределёнными структурами снижает их оригинальность и не позволяет выйти за рамки определённого жанра [3, с. 4]. Нейросетевые методы используют современные языковые модели, такие как GPT, и глубокое обучение для генерации юмора без каких-либо структурных ограничений. Нейронные сети обучаются на больших объемах текстовых данных, содержащих юмористические тексты, и затем используют полученные знания для генерации новых шуток. Сгенерированные юмористические тексты оригинальнее шаблонных, однако некоторые исследователи отмечают, что процесс генерации юмора нейронными сетями непрозрачен для пользователей-людей, и пользователи «не могут понять, как модель придумала шутку, или дать ей указание генерировать шутки определённым образом, что может привести к созданию неуместного или оскорбительного юмора» [3, с. 1]. Некоторые исследователи создают собственные гибридные методы, объединяющие в себе преимущества шаблонных и нейронных подходов [1, с. 17826].

Большинство современных исследований сфокусировано на генерацию юмора с помощью больших языковых моделей (БЯМ). Генерация юмора искусственным интеллектом – одна из сложнейших задач в области обработки естественного языка. Сгенерированные юмористические тексты имеют ряд недостатков. Качество сгенерированного юмора напрямую зависит от качества данных, на которых обучается нейросеть. Нейросетям необходимы большие объёмы информации, которая в силу своего объёма может включать в себя неподходящие данные. Некачественно составленный корпус юмо-

ристических текстов, имеющих, например, повторяющиеся шуточки или неюмористические тексты, влияет на полученный результат [4, с. 5]. Также при создании корпуса для обучения исследователи сталкиваются с проблемой предвзятости данных, собираемых с онлайн-платформ, использующих, например, системы сбора материала на основе количества лайков как показателя успешности юмора. Алгоритмы продвижения контента создают ситуацию, когда определенные шутки становятся гораздо более заметными, чем другие, и собирают больше положительных оценок и комментариев. Не все онлайн-платформы обеспечивают свободный доступ к количеству просмотров какой-либо записи, из-за чего возникает трудность вычета процентного соотношения просмотров к лайкам или комментариям. Кроме того, на количество положительных оценок влияют факторы, не имеющие прямого отношения к качеству шутки, такие как популярность автора или использование определенных ключевых слов [5, с. 6]. Эти факторы приводят к тому, что корпус может оказаться нерепрезентативным и предвзятым. Кроме того, корпус данных для обучения может содержать оскорбительный юмор, который затем будет воспроизведен в сгенерированном тексте [4, с. 5]. Однако, системы модерации сгенерированных текстов могут значительно ограничивать тематику юмористических текстов. При генерации текстов стендап-выступлений, которые затрагивают табуированные в том или ином обществе темы и критику политических и социальных институтов, исследователи столкнулись с тем, что модели зачастую отказываются генерировать подобный контент, рассматривая запросы пользователей как «потенциально опасные» или «неинклюзивные». Недостаточно разнообразный материал для обучения нейросети также может привести к культурной и социальной предвзятости и воспроизведению стереотипов [7, с. 3]. Некоторые исследования также обнаружили, что модель GPT повторяет шуточки из обучающего корпуса, не внося никаких изменений [4, с. 5; 6, с. 325].

Другим недостатком сгенерированного юмора является отсутствие контекстуального понимания. Исследователи отмечают, что сгенерированный юмор «находится везде и нигде одновременно» [7, с. 10]. Юмор, предназначенный для узкой аудитории и основанный на личном опыте, «будет трудно создавать без источника информации о говорящих и их прошлом» [8, с. 16]. Успешный юмор также крайне зависим от культурного и географического контекста. ИИ, ограниченный только тем контекстом, который предоставлен в промпте, не способен полностью адаптировать материал под конкретную аудиторию или ситуацию, особенно при генерации стендап-выступлений. Использование различных методов промптинга и уточнение деталей не всегда улучшает полученные результаты [7, с. 7].

БЯМ создают письменные юмористические тексты, и одним из основных недостатков сгенерированного юмора является ограниченность в подаче. Шутки, представленные только в текстовом формате, часто оцениваются как менее смешные, чем те же шутки, но представленные в таких форматах, как

аудиовизуальные ролики, юмористические изображения или стендап-выступления [9, с. 10]. Подача (delivery) играет ключевую роль в создании комического эффекта. Подача включает в себя как вербальные, так и невербальные элементы – мимику, жесты и позы, интонацию, скорость и громкость речи, использование различных акцентов и изменение произношения, а также тайминг – координацию времени произнесения реплик, пауз и других элементов выступления, чтобы достичь оптимального комического эффекта [10, с. 379]. БЯМ могут сгенерировать качественные юмористические тексты, но комический эффект не будет таким же сильным. Исследователи D. Gorenz и N. Schwarz отмечают, что активное развитие технологий искусственного интеллекта может устранить этот недостаток, так как «уже разрабатываются и совершенствуются инструменты создания голоса, имитирующие человеческую речь, и инструменты генерации видео, что в будущем позволит интегрировать создание и подачу юмора в единый процесс» [9, с. 10].

Однако, современные исследования выделяют несколько преимуществ генерации юмора [7, с. 7; 9, с. 9]. В первую очередь, они отмечают скорость генерации текстов. P. Mirowski et al в своей работе наблюдает эффективность ИИ в качестве инструмента для быстрой генерации первоначального контента и структуры стендап-выступлений. Респонденты их исследования позитивно отзывались о скорости генерации и помощи в построении структуры, предоставляющий основу, которую затем можно самостоятельно доработать [7, с. 7]. Сгенерированные тексты грамматически верны и не содержат ошибок [6, с. 328]. Галлюцинации, то есть склонность больших языковых моделей выдавать ложную информацию за достоверную, не являются столь значительным недостатком при генерации юмористических текстов, поскольку, в отличие от журналистских текстов или научных статей, юмористические тексты не выполняют функции передачи фактической информации и используют многочисленные средства нарушения логики, поэтому они не обязаны соответствовать строгим стандартам точности [9, с. 1].

БЯМ также хорошо справляются с генерацией шаблонного юмора или юмористических текстов со строгой структурой [8, с. 15]. Исследование S. Kim и L. B. Chilton показало, что обученный ИИ успешно генерирует юмористические подписи к фотографиям, которые затем получили такую же высокую оценку, как и подписи, написанные человеком [8, с. 17]. Специально обученная языковая модель генерировала более качественные тексты, по сравнению с GPT. Модель успешно подбирала «аналогичные и соотносимые социальные ситуации», подстраивалась под аудиторию (использовала молодежный слэнг) и следовала структуре шуток [8, с. 15]. Работа Y. Chen et al также показывает эффективность специально обученных моделей, по сравнению с GPT-3 и GPT-3.5 [1, с. 17830]. Кроме того, ИИ способен успешно подражать стилю шуток, указанных в промпте. Работа D. Gorenz и N. Schwarz показала, что сатирические заголовки, созданные ИИ на основе корпуса заголовков сатирической газеты “The Onion”, были оценены как столь же

смешные, как и написанные профессиональными авторами. Это свидетельствует о способности ИИ анализировать и воспроизводить стилистические особенности, характерные для конкретного источника [9, с. 6]

Исследователи предлагают различные методы оценки качества юмора. В основном, исследователи смешивают сгенерированные и написанные человеком шутки, а затем предлагают оценить их, используя, например, шкалу Ликерта [5, с. 6]. В своей работе A. Mittalatal использует «коэффициент комического», который рассчитывается на основе интенсивности и продолжительности смеха, а также последующую оценку смеха человеком [11, с. 1]. В исследовании, проведенном N. Joshi, участники оценивали юмор, сгенерированный ИИ и созданный человеком. Результаты исследования показывают, что сгенерированные шутки получали одинаковую оценку с человеческими, а в случаях генерации игры слов шутки ИИ оценивались выше [12, с. 4]. Примечательно, что, когда создатель (ИИ или человек) шутки не был указан, оценки сгенерированного юмора значительно улучшались, что свидетельствует о наличии предвзятости, связанной с осведомленностью о происхождении текстов. Участники также с трудом отличали сгенерированные шутки от человеческих, несмотря на высокую уверенность в своих ответах, что подчеркивает прогресс ИИ в имитации человеческого юмора [12, с. 4]. Однако, исследования также показывают, что БЯМ генерируют однотипный юмор, который респонденты находят неоригинальным и несмешным [6, с. 332; 7, с. 11].

Генерация юмора с помощью искусственного интеллекта – это динамично развивающаяся область. Современные языковые модели способны создавать грамматически правильные и структурно сложные юмористические тексты, демонстрируя потенциал в качестве инструментов для творчества. Однако их способность генерировать оригинальный юмор, который бы полностью учитывал контекст, остается ограниченной. Ключевые проблемы включают зависимость от качества обучающих данных, отсутствие глубокого понимания культурных и социальных нюансов, а также трудности в передаче невербальных аспектов юмора. Несмотря на эти ограничения, ИИ может служить полезным инструментом для быстрой генерации идей и структурирования материала, особенно в шаблонных жанрах. Дальнейшее развитие технологий, включая улучшение моделей и интеграцию мультимодальных возможностей, может помочь преодолеть существующие барьеры. В то же время важно учитывать этические аспекты, такие как предотвращение предвзятости и контроль над генерацией потенциально оскорбительного контента, чтобы обеспечить ответственное использование этих технологий.

ЛИТЕРАТУРА

1. Chen Y. et al. Talk Funny! A Large-Scale Humor Response Dataset with Chain of-Humor Interpretation [Electronic resource] // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. Vol. 38. № 16. P. 17826–17834. DOI: <https://doi.org/10.1609/aaai.v38i16.29736>.

2. Winters T. Computers Learning Humor Is No Joke. [Electronic resource] // Harvard Data Science Review. 2021. Vol. 3. № 2. P. 1–18. DOI: <https://doi.org/10.1162/99608f92.f13a2337>.
3. Chen Y., Shi B., Si M. Prompt to GPT-3: Step-by-Step Thinking Instructions for Humor Generation [Electronic resource] // arXiv preprint arXiv:2306.13195. 2023. P. 1–5. DOI: <https://doi.org/10.48550/arXiv.2306.13195>.
4. Faulkner R., Rex R., Ryan M. Knock Knock: Neural Joke Generation and Classification Final Report [Electronic resource] // Georgia Institute of Technology Technical Report. 2023. URL: <https://michryan.com/project/humor-bot/> (date of access: 18.06.2025).
5. Winters T., Nys V., De Schreye D. Automatic joke generation: Learning humor from examples [Electronic Resource] // Distributed, Ambient and Pervasive Interactions: Technologies and Contexts. DAPI 2018. Lecture Notes in Computer Science. 2018. Vol. 10922. P. 360–377. DOI: https://doi.org/10.1007/978-3-319-91131-1_28.
6. Jentzsch S., Kersting K. ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models [Electronic resource] // Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. 2023. P. 325–340. DOI: <https://doi.org/10.48550/arXiv.2306.04563>.
7. Mirowski P. W., Love J., Mathewson K., Mohamed S. A Robot Walks Into a Bar: Can language models serve as creativity support tools for comedy? An evaluation of LLMs' humour alignment with comedians [Electronic resource] // Proceedings of the ACM Conference on Fairness, Accountability, and Transparency. 2024. P. 1622–1636. DOI: <https://doi.org/10.1145/3630106.365899>.
8. Kim S., Chilton L. B. AI Humor Generation: Cognitive, Social and Creative Skills for Effective Humor [Electronic resource] // arXiv preprint arXiv:2502.07981. 2025. P. 1–18. DOI: <https://doi.org/10.48550/arXiv.2502.07981>
9. Gorenz D., Schwarz N. How funny is ChatGPT? A comparison of human- and A.I.-produced jokes [Electronic Resource] // PLoS ONE. 2024. Vol. 19. № 7. DOI: <https://doi.org/10.1371/journal.pone.0305364>.
10. Double O. Getting the Joke: The Inner Workings of Stand-up Comedy. London : Bloomsbury Publishing, 2014. 536 p.
11. Mittal A. et al. “So You Think You're Funny?": Rating the Humour Quotient in Standup Comedy [Electronic resource] // arXiv preprint arXiv:2110.12765. 2021. P. 1–7. DOI: <https://doi.org/10.48550/arXiv.2110.12765>.
12. Joshi N. N. Evaluating human perception and bias in AI-generated humor [Electronic resource] // Proceedings of the 1st Workshop on Computational Humor (CHum). 2025. P. 79–87. URL: <https://aclanthology.org/2025.chum-1.9/> (date of access: 18.06.2025).
13. Borji A. A categorical archive of chatgpt failures [Electronic resource] // arXiv preprint arXiv:2302.03494. 2023. P. 1–41. DOI: <https://doi.org/10.48550/arXiv.2302.03494>.

14. Avetisyan H., Safikhani P., Broneske D. Laughing Out Loud – Exploring AI-Generated and Human-Generated Humor [Electronic resource] // 4th International Conference on NLP & Artificial Intelligence Techniques (NLAI 2023). 2023. P. 59–76. DOI: <https://doi.org/10.5121/csit.2023.132406>.
15. Cowie R. Computational research and the case for taking humor seriously [Electronic resource] // HUMOR. 2023. Vol. 36. № 2. P. 207–223. DOI: <https://doi.org/10.1515/humor-2023-0021>.
16. Kalloniatis A., Adamidis P. Computational humor recognition: a systematic literature review [Electronic resource] // Artificial Intelligence Review. 2024. Vol. 58. P. 1–52. DOI: <https://doi.org/10.1007/s10462-024-11043-3>.
17. Barattieri di San Pietro C., Frau F., Mangiaterra V., Bambini V. The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent [Electronic resource] // SistemiIntelligenti. 2023. Vol. 35. № 2. P. 379–399. DOI: <https://doi.org/10.1422/108136>.