

УДК 81'42:070.16:004.8

МЕХАНИЗМЫ ГЕНЕРИРОВАНИЯ ФЕЙКОВОЙ ИНФОРМАЦИИ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ В СОВРЕМЕННОМ МЕДИАДИСКУРСЕ

Оломская Наталья Николаевна

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

Зиньковская Анастасия Владимировна

- e-mail: anastassiat@bk.ru
доктор филологических наук, профессор,
заведующая кафедрой английской филологии,
Кубанский государственный университет
- 350040, г. Краснодар, ул. Ставропольская, д. 149,
Российская Федерация

Для цитирования

Оломская Н. Н., Зиньковская А. В. Механизмы генерирования
фейковой информации искусственным интеллектом в современном
медиадискурсе // Российский социально-гуманитарный журнал. 2024.
№ 2. URL: www.evestnik-mgou.ru

- Статья поступила в редакцию 18.02.2024
- Статья размещена на сайте 29.05.2024

СТРУКТУРА СТАТЬИ

[Аннотация](#)

[Ключевые слова](#)

[Введение](#)

[История вопроса](#)

[Способы создания и распространения фейков: традиционные vs инновационные](#)

[Специфика генерирования фейков нейросетью](#)

[Экспериментальная часть](#)

[Заключение](#)

[Литература](#)

[Article information](#)

АННОТАЦИЯ

Цель состоит в попытке исследовать влияние основных механизмов генерирования фейковой информации на социум в современном медиадискурсе. Гипотеза исследования заключается в том, что с развитием возможностей искусственного интеллекта (ИИ) генерирование фейков происходит с помощью глубинных когнитивно-лингвистических моделей, поскольку «сильный» / «strong» ИИ основан на генеративной архитектуре нейросети, которая способна «учиться», решать сложные комплексные задачи без какого-либо вмешательства со стороны человека, что представляет собой экзистенциальную угрозу существования цивилизации.

Процедура и методы. При анализе эмпирического материала нами использовался ряд методов и приёмов для анализа текстов, сгенерированных с применением нейросетей, таких как контекстуальный анализ, метод обобщения и систематизации, компонентный и сопоставительный анализ.

Результаты. Нами была проанализирована фейковая информация, сгенерированная нейросетями, а также был проведён эксперимент проверки степени правдивости информации, сгенерированной «ChatGPT», которая оказалась на 99% фейком. Кроме того, был проведён эксперимент с использованием опции «аргумент» агрегатора нейросети «GPTunnel», цель которого состояла в том, чтобы выявить качество аргументации на заданную тему. Эксперимент показал, что заявленное «умение логически мыслить и учитывать аргументы-опровержения» не подтвердилось. Однако стремительно развивающийся ИИ и увеличивающиеся возможности нейросетей, возможно, через совсем небольшое время создадут более усовершенствованную нейросеть, способную генерировать логически выстроенную аргументацию.

Теоретическая и / или практическая значимость. Результаты научного исследования вносят вклад в более глубокое понимание механизмов генерирования фейков с помощью генеративных нейросетей в современном медиадискурсе, что в перспективе поможет специалистам в области лингвистики решать проблемы, связанные с идентификацией текстов, созданных ИИ.

КЛЮЧЕВЫЕ СЛОВА

искусственный интеллект (ИИ), лингво-когнитивные модели, манипуляция, медиадискурс, нейросеть, фейк, фейковая информация

ВВЕДЕНИЕ

В современном мире проблема генерирования и распространения фейковой информации становится всё более актуальной, поскольку фейки в значительной степени влияют на формирование реакции социума на те или иные явления общественной и политической жизни.

Стремительное усовершенствование нейросетей привело к тому, что сегодня вполне реально существует экзистенциальная угроза нашей цивилизации. Одной из самых серьёзных проблем является проблема генерирования фейковой (ложной) информации с использованием возможностей искусственного интеллекта. Это тем более опасно, поскольку определить, является ли та или иная информация правдивой или нет, становится всё сложнее и сложнее даже для специалистов, работающих в области искусственного интеллекта (далее – ИИ).

Данной проблеме посвящён ряд работ отечественных и зарубежных исследователей: Т. В. Черниговской, В. А. Лекторского, А. А. Смирнова, Д. И. Дубровского, Дж. Фодора и др. Отметим ряд наиболее, с нашей точки зрения, интересных работ по данной проблеме: «Развитие искусственного интеллекта и глобальный кризис земной цивилизации (к анализу социогуманитарных проблем)» в журнале «Философия науки и техники» (2022) (Д. И. Дубровский), «Искусственный интеллект в исследованиях сознания и общественной жизни (к 70-летию статьи А. Тьюринга «Вычислительные машины и разум») (материалы круглого стола)» в журнале «Философия науки и техники» (2022) (В. А. Лекторский), монография Т. В. Черниговской «Чеширская улыбка кота Шрёдингера: мозг, язык и сознание» (2023) и др. Таким образом, актуальность исследования способов создания и распространения фейковой информации не вызывает сомнений, так как понимание и осмысление этих механизмов поможет обществу «отделить зёрна от плевел» и осмыслить способы и методы борьбы с распространением и производством фейковой информации. Представленный в работе эксперимент подтвердил наши опасения о том, что сгенерированный «ChatGPT» текст по заданным критериям оказался 99% фейком (в среде разработчиков ИИ такой феномен носит название «галлюцинация», что, на наш взгляд, вполне оправдано). Современные методы распознавания фейкового сгенерированного текста, к сожалению, далеко несовершенны, и перед

учёными-лингвистами стоит очень трудная задача разработки таких методов распознавания в самой ближайшей перспективе, поскольку, как отметила Т. В. Черниговская, «надежда на то, что когнитивные характеристики искусственных систем приблизятся к уровню человеческих или даже превзойдут их, неоправданно растёт. Вероятнее всего, это связано с тем, что растёт и скорость обработки информации, что, казалось бы, должно обеспечить успех. При этом относительно мало обсуждается вопрос о том, какое именно общение с антропоморфными системами мы будем считать адекватным, чего от него ждём? Это вызывает ряд вопросов не только научных, но и экзистенциальных, и этических» [6, с. 14]. Характер исследования и его основная цель (обоснование гипотезы исследования) объясняют выбор ряда методов исследования. В ходе работы нами применялись контекстуальный анализ, метод обобщения и систематизации, компонентный и сопоставительный анализ. Кроме того, мы провели ряд экспериментов при работе с текстами, сгенерированными «ChatGPT», с целью проверки степени правдивости информации, а также эксперимент с использованием опции «аргумент» агрегатора нейросети «GPTunnel».

ИСТОРИЯ ВОПРОСА

Существование фейков имеет долгую историю: они возникли тогда, когда человечество стало интересоваться степенью правдивости фактов и явлений, происходящих в социально-общественной и политической жизни. Известно, что одним из первых примеров фейковой информации является начертанное на камнях дворцов в Древнем Египте и в Хеттском государстве сообщение о победе в битве, в которой победителей не было. Другим показательным примером распространения фейковой информации является информация в «Нью-Йорк Сан» (The New York Sun) в 1835 г. о том, что астроном в супертелескоп смог рассмотреть жителей Луны. Таких «исторических» примеров фейковой информации можно привести не один десяток.

На протяжении всей своей истории человечество существует в многозначной и противоречивой среде, в которой происходят глобальные «цивилизационные сломы» (термин Т. В. Черниговской) [6], например, падение и исчезновение развитых цивилизаций древности. Однако именно в современную эпоху такие социально-политические изменения происходят стреми-

тельно, что влечёт за собой трансформацию общественных, нравственных, этических ценностей, поскольку исчезает способность верификации правды, так как «проверка» правдивости тех или иных утверждений представляет определённые сложности, в том числе и для экспертов, не говоря уже об обычных людях. Отчасти именно поэтому фейковая информация получила широкое распространение в медиапространстве.

Во время стремительного развития информационных технологий и каналов передачи информации генерирование фейковой информации вышло на значительно более продвинутый уровень по сравнению с прошлыми веками, поскольку в XXI в. создавать и распространять фейковую информацию способен абсолютно каждый человек, имеющий доступ к социальной медиасреде, для которой не существует никаких ограничений, так как адресат информации может взаимодействовать с реципиентами прямо без посредников, не обращая внимания на цензуру, государственные акты и законы. В этом кроется основная трудность отслеживания, контроля и, как следствие, привлечения к ответственности лиц, причастных к генерированию и распространению фейковой информации. Кроме «традиционных» способов создания и распространения фейков, в первой трети XXI в. использование нейросети и искусственного интеллекта (ИИ) при генерировании фейков получило широкое распространение. В этом исследовании мы предлагаем выделить в генерировании фейковой информации «традиционные» и «инновационные» способы и техники создания и распространения.

СПОСОБЫ СОЗДАНИЯ И РАСПРОСТРАНЕНИЯ ФЕЙКОВ: ТРАДИЦИОННЫЕ VS ИННОВАЦИОННЫЕ

В современных исследованиях широко представлены «сценарии» создания и распространения фейковой информации. Самыми распространёнными «традиционными» фальсификациями являются манипуляции с фотографиями, видеотропки, подтасовка фактов, выступления различных псевдоэкспертов, манипуляции с данными, ссылка на «уважаемые» СМИ и т. д.

Манипуляции с фотографиями – пожалуй, самый «популярный» приём среди распространителей фейковой информации, когда используется редактирование фотографий с помощью специальных программ или фотография не редактируется, но

выдаётся за реальную фотографию события с соответствующим комментарием. Например, в телеграм-канале была опубликована фотография с фестиваля «Burning Man» с утверждением о том, что на ней изображена «пробка» на границе России и Казахстана после объявления о частичной мобилизации в Российской Федерации. Подобная информация оказалась 100% фейком и была удалена адресатом, но получила широкое распространение в сети.

К «инновационным» способам и техникам создания и распространения фейковой информации относится так называемый дипфейк (deep fake) [4]. Дипфейки представляют собой манипулирующий визуальный контент, который создаётся искусственным интеллектом (ИИ) с помощью генерирования ложного звукового ряда и изображений (визуальные картинки), транслирующих «реальную» информацию. Как правило, реальные события и люди в данном визуальном контенте подменяются подобным ложным контентом, который достаточно сложно идентифицировать как фейковый.

Самым знаменитым примером дипфейка является видео, выложенное актёром Дж. Пилом, в котором он записал речь голосом Б. Обамы, совместив его с видеовыступлением экс-президента США. Во второй части актёр продемонстрировал, как выглядят обе части этого видеофейка, тем самым показав реальную опасность технологий ИИ¹.

СПЕЦИФИКА ГЕНЕРИРОВАНИЯ ФЕЙКОВ НЕЙРОСЕТЬЮ

Тем не менее технологии усовершенствования ИИ развиваются настолько стремительно, что распознавать «фальшивки» становится всё труднее. Так, к примеру, учёные из Вашингтонского университета и Института Аллена, работающие в области исследования ИИ, смогли создать нейросеть, автоматически генерирующую фейковую информацию, которую трудно отличить от правдивой. Суть эксперимента заключалась в том, что в нейросеть загрузили паттерны, содержащие тексты газеты «Уолл-стрит джорнэл» (The Wall Street Journal), в которых были учтены лексико-стилистические особенности стиля современной журналистики, научив тем самым ИИ «писать» новостной материал. ИИ сгенерировал материал о добыче полезных иско-

¹ Obama tricks Trump – fake video // YouTube: [сайт]. URL: https://www.youtube.com/watch?v=_gVTvITMjV4 (дата обращения: 05.01.2024).

паемых на Луне, который невозможно отличить от оригинальных статей в «Уолл-стрит джорнэл»².

Другой пример. Разработчики «OpenAI» создали усовершенствованную языковую модель, известную как «ChatGPT», поскольку эта модель способна генерировать текст, используя огромный массив данных, накопленный человечеством на протяжении всей своей истории. Конструкция «ChatGPT» позволяет выполнять широкий спектр задач с использованием возможностей естественного языка, например, перевод языка, уплотнение текста, ответы на вопросы и создание кода. В основу «ChatGPT» была заложена обширная информация из интернета, что позволяет ему выполнять огромное количество самых разнообразных задач. «ChatGPT» представляет собой подкласс глубоких нейронных сетей. Кроме того, учёные из института «OpenAI» в Сан-Франциско разработали проект, в котором с помощью ИИ сгенерировали «новость» о том, что из-за политики Д. Трампа Россия объявила войну США:

*Russia has declared war on the United States after Donald Trump accidentally fired a missile in the air. Russia said it had "identified the missile's trajectory and will take necessary measures to ensure the security of the Russian population and the country's strategic nuclear forces." The White House said it was "extremely concerned by the Russian violation" of a treaty banning intermediate-range ballistic missiles. The US and Russia have had an uneasy relationship since 2014, when Moscow annexed Ukraine's Crimea region and backed separatists in eastern Ukraine*³.

Как видим из приведённого текста «горячей новости», стилистика высказывания абсолютно совпадает со стилистическими особенностями современной журналистики, например, такими как наличие маркированных структур, имеющих негативную коннотацию, «*an uneasy relationship*», «*annexed*», «*separatists*», при этом подача двух точек зрения «*Russia said*», «*The White House said*», достаточно часто используемая манипулятивная технология текстов современной журналистики, придаёт тот политический контекст, который прежде всего выгоден манипулятору.

² Хвостик Е. Искусственный интеллект научился делать натуральные фейковые новости // Коммерсантъ: [сайт]. [16.10.2019]. URL: <https://www.kommersant.ru/doc/4126694> (дата обращения: 18.12.2023).

³ Knight W. An AI that writes convincing prose risks mass-producing fake news // MIT Technology Review: [сайт]. [14.02.2019]. URL: <https://www.technologyreview.com/2019/02/14/137426/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish> (дата обращения: 18.12.2023).

Сами исследователи-разработчики данного проекта признаются, что они намеревались разработать языковой алгоритм общего назначения, который *обучается* (курсив наш – Н. О., А. З.) на огромном количестве текстов из интернета. Кроме того, данный языковой алгоритм может быть способен переводить текст, отвечать на вопросы и выполнять другие полезные задачи. Однако разработчики допустили возможность использования языкового алгоритма для генерирования фейковой информации. Директор по политике «OpenAI» Д. Кларк говорит: «Мы начали его тестировать и быстро обнаружили, что можно легко генерировать вредоносный эскиз-контент». Д. Кларк отмечает, что такая программа, как ИИ, может использоваться для автоматизации генерации убедительных фейковых новостей, публикаций в социальных сетях или другого текстового контента. Этот инструмент может быть использован для создания фейков, которые, например, способны фальсифицировать скандальные разоблачения во время выборов. Д. Кларк уверен, что осталось совсем недолго, прежде чем ИИ сможет надёжно производить фальшивые истории, фальшивые твиты или двуличные комментарии, которые являются ещё более убедительными⁴.

В этой связи, на наш взгляд, возникает прямая угроза информационно-безопасности социуму в целом и индивидууму в частности, поскольку практически полностью отсутствует так называемый фактчекинг (проверка информации) адреса информации. Как справедливо отметил Д. И. Дубровский: «...система такого рода ... способна к «самодетальности»; она сможет легко выходить из-под контроля человека, более того – генерировать собственные цели и даже конкурировать с ним» [1, с. 105].

Более того, в последнее время появляются учебники, содержащие подробные инструкции, как использовать «OpenAI» для написания текстов различной направленности – от научных до художественных. Так, к примеру, книга Шона А. Уильямса «ChatGPT for Writers Unlocking the Power of Artificial Intelligence to Enhance Your Nonfiction Writing Process (Learn how to use ChatGPT for brainstorming, research, and productivity)» содержит подробную инструкцию к созданию текстов, книг, научных иссле-

⁴ Knight W. An AI that writes convincing prose risks mass-producing fake news // MIT Technology Review: [сайт]. [14.02.2019]. URL: <https://www.technologyreview.com/2019/02/14/137426/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish> (дата обращения: 18.12.2023).

дований и проч. Автор пишет: «ChatGPT можно использовать для генерации идей для вашей книги, предоставив предложения для названий глав, подтем, исследовательских вопросов и многого другого. Вы можете подсказать ChatGPT общую тему вашей книги, и она создаст список идей, связанных с этой темой. Например, если вы пишете книгу об истории конкретной страны, вы можете дать ChatGPT подсказки – такие как «идеи для глав об истории страны X», и она создаст список названий глав – таких как «Ранняя цивилизация Икс», «Колониальная эра в Икс», «Современная эра в Икс» и так далее» [7]. Таким образом, возможности ИИ подводят человечество к мысли о том, что нет необходимости думать, испытывать творческие порывы, получать образование, совершенствовать свои навыки и умения, развивать интеллект, поскольку есть «ChatGPT» и он всё сделает за вас. Умение ИИ генерировать «свои» идеи представляется опасным, так как подобные идеи, как указывалось выше, не только могут быть фейковыми, но и «помогают» усовершенствовать возможности ИИ в целом и приближают то время, когда ИИ сможет обладать потенциалом, опасным для цивилизации.

В другой своей книге «ChatGPT for Coders Unlock the Power of AI with ChatGPT: A Comprehensive Guide to Efficient and Effective Coding» Шон А. Уильямс утверждает, говоря о будущем ChatGPT: «Очень важно знать возможности ... ChatGPT, чтобы максимально использовать преимущества, которые он предлагает ... Кроме того, важно знать о любых возможных опасностях и ограничениях, которые могут существовать» [8]. Таким образом, сами разработчики допускают возможные опасности, но при этом активно совершенствуют возможности ИИ к самообучению. Как справедливо заметил В. А. Лекторский: «Человек – это несовершенная интеллектуальная система, а ИИ может быть более совершенной» [2, с. 8].

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Другая проблема, связанная с использованием нейросетей, заключается в том, что задача, которую ставит пользователь перед нейросетью, не всегда имеет адекватное решение. Так, к примеру, специалистам по ИИ известен термин «галлюцинация», который подразумевает, что, если ИИ не способен решить задачу, не знает ответа на вопрос и проч., он «сочиняет» прав-

доподобный ответ и тем самым вводит в заблуждение пользователя, при этом такой «ответ» может содержать элементы правдивой информации, чтобы придать информации максимально правдоподобный вид. Таким образом, масштаб возможностей генерирования и распространения ложной информации поистине не имеет границ, что, бесспорно, может иметь весьма существенные последствия.

Так, к примеру, ИИ был нами введён запрос: «Подбери литературу по теме: «Проблема времени в произведениях У. Фолкнера». Книги и статьи, вышедшие в печать после 2015 года». Мы получили список из 10 наименований:

1. The Temporal Poetics of William Faulkner by Hans J. Staats (2015);
2. Faulkner and Time: Faulkner and Yoknapatawpha, 2015-2016 edited by Jay Watson and Ann J. Abadie (2015);
3. Time in Faulkner's fiction by Phillip M. Weinstein (2015);
4. Time and Timelessness in William Faulkner edited by Emir Ulgener (2016);
5. William Faulkner: Time and Myth edited by Clark Blaise and A. A. Devy (2017);
6. Faulkner's Temporal Landscapes: Personal and Political Time in His American Fiction by Dominique Pitavy-Souques (2017);
7. Time and Modernism in Faulkner, Woolf, and Moore by Cristanne Miller (2017);
8. The Sounds of Slowness: Faulkner's Temporal Imaginary by Aric Putnam (2018);
9. Faulkner and Formalism: Returns of the Text edited by James Grimshaw, W. L. Howatson, and Joanna P. McAfee (2019);
10. Memory in Motion: Reckoning with Time in Faulkner, Woolf, and Nabokov by Hanna Roman (2019).

Как видим из примера, список, на первый взгляд, кажется вполне реальным. Однако при проверке на «подлинность» оказалось, что данный список является примером «галлюцинации» ИИ, и, таким образом, реципиенту предлагается 90% фейк, из 10 наименований – 9 оказались фейками.

Не так давно в телеграм-канале *1337: IT, ChatGPT, Midjourney* (14–15 ноября 2023 г.) была анонсирована достаточно интересная информация о том, что появился новый агрегатор нейросетей «GPTunnel». Разработчики утверждают, что за «три клика» пользователь имеет возможность выбрать «ChatGPT» 4-й версии, «Midjourney», «Faceswap» с дипфейками или «Clipchamp»

для синтеза речи и поставить ИИ любую задачу. Впечатляет также то, что «GPT-4 128K» Vision, которая была заявлена за неделю до анонса, уже доступна пользователям. Скорость создания ИИ действительно впечатляет. В меню агрегатора входит целый перечень возможностей, например, «лицоzaменитель» позволяет легко менять лица на фотографиях, «диктор» озвучивает текст с использованием различных голосов, а также «создать бота» и проч. Больше всего впечатляет опция «аргумент», которая помогает создавать доводы на основе доказательств по различным темам.

Нами был проведён эксперимент с использованием опции «аргумент» агрегатора нейросети «GPTunnel». Цель эксперимента состояла в том, чтобы выявить качество аргументации на заданную тему. Ниже приведём часть «диалога» между нейросетью и пользователем (Н. Н. Оломская), состоящую из вопросов (наших) и ответов (нейросети):

Н. О.: Объясни суть конфликта России и Украины.

ИИ: Конфликт между Россией и Украиной начался после того, как украинский президент Виктор Янукович отказался подписать соглашение о сотрудничестве с Европейским союзом в 2013 году, что привело к массовым протестам в Киеве. В результате этого конфликта Янукович был свергнут с поста президента, а Россия аннексировала Крым.

Н. О.: Россия присоединила Крым, а не аннексировала.

ИИ: Крым был присоединён к России в 2014 году после референдума, в результате которого большинство жителей этого региона выразили желание вступить в состав России. Таким образом, российские власти считают присоединение Крыма законным.

Н. О.: Приведи аргументы и факты в пользу присоединения Крыма.

ИИ: Аргументы в пользу присоединения Крыма:

Историческая принадлежность: Крым был частью Российской империи до 1954 года, когда был передан Украине без согласия народа.

Воля народа: По результатам референдума, проведённого в 2014 году, подавляющее большинство крымчан выразили желание присоединиться к России.

Н. О.: Приведи аргументы и факты в пользу аннексии Крыма.

ИИ: Аргументы за аннексию Крыма:

Историческая принадлежность: Крым входил в состав Российской империи с 1783 года до 1954 года, когда был передан в состав Украины без участия населения полуострова. Волеизъявление жителей Крыма: в 2014 году жители Крыма проголосовали на референдуме".

Таким образом, в ходе проведённого анализа отметим два момента. Во-первых, один и тот же поставленный вопрос получил по сути два разных ответа с точки зрения эмоционально-оценочного компонента: «Крым был частью Российской империи до 1954 года, когда был передан Украине *без согласия народа*. Воля народа: По результатам референдума, проведённого в 2014 году, *подавляющее большинство крымчан выразили желание присоединиться к России*» и «Крым входил в состав Российской империи с 1783 года до 1954 года, когда был передан в состав Украины *без участия населения полуострова*. Волеизъявление жителей Крыма: в 2014 году *жители Крыма проголосовали на референдуме*». Коннотация существительных «народ» и «население» достаточно идеологически различна. В Конституции нашей страны декларируется: «Мы, многонациональный народ Российской Федерации...». Кроме того, «народ» является носителем суверенитета страны, и наш президент клянётся верно служить народу. «Население» – это некая узколокальная общность, ограниченная рамками географических границ. Поэты и писатели призывают к народу, а не к населению, и с этой точки зрения «народ» имеет идеологически яркий коннотативный компонент. Во-вторых, маркированные паттерны «*подавляющее большинство крымчан выразили желание присоединиться к России*» и «*жители Крыма проголосовали на референдуме*» стилистически также неоднородны. Следует отметить слабость аргументации, поскольку изначально разработчиками было заявлено, что «ассистент поможет создать убедительные доводы на основе доказательств по различным темам. Часто нам может понадобиться умение логически мыслить и учитывать аргументы-опровержения, чтобы наша точка зрения была максимально сильной и убедительной. С помощью «Аргумент» ты сможешь разбить свою аргументацию на логические блоки, предоставить доказательства и факты, а также привести релевантные примеры для подкрепления своих утверждений. Благодаря этому твои доводы будут крепкими и удерживающимися подробностями». В ходе проведённого эксперимента «умение логически мыслить и учитывать аргументы-опровержения,

чтобы наша точка зрения была максимально сильной и убедительной», не проявилось. Согласимся с мнением Р. И. Тельпова и С. В. Ларциной о том, что «заметно повышенная частотность ключевых слов в текстах, сгенерированных нейронной сетью, позволяет предположить, что «ChatGPT» на текущем этапе разработки не может производить тексты, которые содержат отступления в сторону от первоначального запроса пользователя» [5, с. 59]. Однако не стоит не учитывать стремительно развивающегося ИИ и увеличивающихся возможностей нейросетей. Возможно, через совсем небольшое время мы получим совершенную нейросеть, способную генерировать логически выстроенную аргументацию (и не только!).

ЗАКЛЮЧЕНИЕ

К сожалению, использование возможностей ИИ для генерирования фейков и, соответственно, расширение вариативности языкового манипулирования социумом остановить в данный момент не представляется возможным. Да и можно ли вообще остановить этот процесс? Распознавание ложной информации становится всё более сложным, что, как мы полагаем, будет иметь негативные последствия. Как справедливо отметил А. Ведяхин, говоря о возможных сценариях развития ситуации в мире при использовании возможностей ИИ: «Куда вероятнее сценарий обострения конкурентной борьбы за ресурсы, зоны контроля и умы людей. При этом общества, претендующие на политический суверенитет и долю мирового рынка, должны будут выигрывать в гонке за адаптацию, разрабатывать жизнеспособную управленческую модель, транслировать позитивный образ будущего. А для этого им неизбежно понадобятся AGI-технологии. Соответственно, технологическое и научное лидерство в области AGI или, как минимум, технологический паритет с другими государствами / сообществами стоит рассматривать как насущную необходимость, ключевой компонент борьбы за выживание, безопасность и процветание. При этом важно соблюдать этические принципы при разработке общего искусственного интеллекта, чтобы избежать возможных катастроф и сделать технологии ИИ максимально безопасными, полезными и доступными для всех людей на планете» [3, с. 230].

Однако всегда есть опасность того, что, выпустив сегодня «джина из бутылки», мы не сможем контролировать его завтра.

ЛИТЕРАТУРА

1. Дубровский Д. И. Развитие искусственного интеллекта и глобальный кризис земной цивилизации (к анализу социогуманитарных проблем) // Философия науки и техники. 2022. Т. 27. № 2. С. 100–107.
2. Лекторский В. А. Искусственный интеллект в исследованиях сознания и общественной жизни (к 70-летию статьи А. Тьюринга «Вычислительные машины и разум») (материалы круглого стола) // Философия науки и техники 2022. Т. 27. № 1. С. 5–33.
3. Сильный искусственный интеллект: на подступах к сверхразуму / А. А. Ведяхин и др. М.: Интеллектуальная Литература, 2021. 232 с.
4. Смирнов А. А. «Глубокие фейки». Сущность и оценка потенциального влияния на национальную безопасность // Свободная мысль. 2019. № 5. С. 63–84.
5. Тельпов Р. Е., Ларцина С. В. Типовые различия естественных и сгенерированных нейронной сетью текстов в квантитативном аспекте // Научный диалог. 2023. Т. 12. № 7. С. 47–65.
6. Черниговская Т. В. Чеширская улыбка кота Шредингера: мозг, язык и сознание. М.: АСТ, 2023. 496 с.
7. Williams S. A. ChatGPT for Writers Unlocking the Power of Artificial Intelligence to Enhance Your Nonfiction Writing Process (Learn how to use ChatGPT for brainstorming, research, and productivity) [Электронный ресурс]. URL: https://psv4.userapi.com/c240331/u399904795/docs/d55/8d3d8502d136/ChatGPT_for_Writers.pdf?extra=XF-Gdv4N5Uzve5MWBQyCFV_oYtq-vQX683IDiXB2qHUGfv8FoLqEhSwvKEsUg1x2kUJ_EhLEwQbPNGkgLbdyfwUEC0Si4isHUYimM7oDYFRhPddeqk1pQE63bG80SFJGFJCH8H7hu-jJENWJ9dectok (дата обращения: 07.01.2024).
8. Williams S. A. ChatGPT for Coders Unlock the Power of AI with ChatGPT: A Comprehensive Guide to Efficient and Effective Coding [Электронный ресурс]. URL: https://psv4.userapi.com/c240331/u399904795/docs/d11/17806604b649/ChatGPT_for_Coders_1.pdf?extra=7D8biSkrUEHBhbzblyV5Odvj3S0YGVqZA8YT2hbCd1hErhBF2UhhvUbsR5ulnAOrd4qtsHfrMGQ_baWmMYTQxJdSFQ5cAr7z2fHzLLrpwbhyu1VixtNZ3HwynRLOJf6KQYsWLSzFvGvCDfnNT6E9_xEO (дата обращения: 07.01.2024).

ARTICLE INFORMATION

Authors

Natalya N. Olomsкая

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

Anastasia V. Zinkovskaya

- e-mail: anastassiat@bk.ru;
Dr. Sci. (Philology), Prof., Department Head,
Department of English Philology,
Faculty of Romano-Germanic Philology;
Kuban State University
ul. Stavropolskaya 149, Krasnodar 350040, the Russian Federation

For citation

Olomsкая N. N., Zinkovskaya A. V. Mechanisms of generation of fake information by artificial intelligence in modern media discourse. In: *Russian Social and Humanitarian Journal*, 2024, no. 2.

Available at: www.evestnik-mgou.ru

Abstract

Aim. To try to investigate the impact of the main mechanisms of fake information generation on society in the modern media discourse. *The hypothesis* of the study is that with the development of artificial intelligence (AI) capabilities, the generation of fake information occurs with the help of deep cognitive-linguistic models, because "strong" AI is based on the generative architecture of neural networks, which is able to "learn", to solve complex tasks without any human intervention, which is an existential threat to civilization.

Methodology. In analyzing the empirical material, we used a number of methods and techniques to analyze the texts generated using neural networks, such as contextual analysis, generalization and systematization method, component and comparative analysis.

Results. We have analyzed fake information generated by neural networks, and also conducted an experiment to check the degree of truthfulness of

information generated by ChatGPT which turned out to be 99% fake. In addition, we conducted an experiment using the “argument” option of the GPTunnel neural network aggregator, the purpose of which was to reveal the quality of argumentation on a given topic. The experiment showed that the claimed “ability to think logically and consider arguments” was not confirmed. However, rapidly developing AI and increasing capabilities of neural networks, might in a very short time create a more advanced neural network capable of generating logically constructed argumentation.

Research implications. The results of the research contribute to a deeper understanding of the mechanisms of fake generation using generative neural networks in modern media space, which in the future will help linguists solve problems related to the identification of texts generated by AI.

Keywords

Artificial intelligence (AI), linguistic-cognitive models, manipulation, media discourse, neural network, fake, fake information

References

1. Dubrovsky D. I. [The development of artificial intelligence and the global crisis of earthly civilization (to the analysis of socio-humanitarian problems)]. In: *Filosofiya nauki i tekhniki* [Philosophy of Science and Technology], 2022, vol. 27, no. 2, pp. 100–107.
2. Lektorsky V. A. [Artificial intelligence in the research of consciousness and in social life (in honor of 70-years anniversary of A. Turing's paper “Computing Machinery and Intelligence”) (papers of the “round table”)]. In: *Filosofiya nauki i tekhniki* [Philosophy of Science and Technology], 2022, vol. 27, no. 1, pp. 5–33.
3. Vedyakhin A. A., et al. Strong artificial intelligence: On the approaches to superintelligence. Moscow, Intellektualnaya Literatura Publ., 2021. 232 p.
4. Smirnov A. A. [“Deepfakes”. Essence and assessment of the potential impact on national security]. In: *Svobodnaya mysl'* [Free Thought], 2019, no. 5, pp. 63–84.
5. Telpov R. E., Lartsina S. V. [Typological differences of natural and neural network-generated texts in a quantitative aspect]. In: *Nauchnyi dialog* [Scientific dialogue], 2023, vol. 12, no. 7, pp. 47–65.
6. Chernigovskaya T. V. *Cheshirskaya ulybka kota Shredingera: mozg, yazyk i soznanie* [Cheshire smile of Schrodinger's cat: brain, language and consciousness]. Moscow, AST Publ., 2023. 496 p.
7. Williams S. A. ChatGPT for Writers Unlocking the Power of Artificial Intelligence to Enhance Your Nonfiction Writing Process (Learn how to use ChatGPT for brainstorming, research, and productivity).

Available at: https://psv4.userapi.com/c240331/u399904795/docs/d55/8d3d8502d136/ChatGPT_for_Writers.pdf?extra=XF-Gdv4N5Uz-ve5MWBQyCFV_oYtq-vQX683IDiXB2qHUGfV8FoLqEhSwvKEsUg-1x2kUJ_EhLEwQbPNGkgLbdyfvwUEC0Si4isHUYimM7oDYFRhPd-deqk1pQE63bG8OSFJGFJCH8H7hu_jJENWJ9dectok (accessed: 07.01.2024).

8. Williams Sean A. ChatGPT for Coders Unlock the Power of AI with ChatGPT: A Comprehensive Guide to Efficient and Effective Coding. Available at: https://psv4.userapi.com/c240331/u399904795/docs/d11/17806604b649/ChatGPT_for_Coders_1.pdf?extra=7D8biSkr-UEHBhbzblyV5Odvj3S0YGVqrZA8YT2hbCd1hErhBF2UhvUbRs5u-InAOrd4qtshfrMGQ_baWmMYTQxJdSFQ5cAr7z2fHzlLrpwbhyu-1VixtNZ3HwynRLOJf6KQYsWLsZFvGvCDfnNT6E9_xEo (accessed: 07.01.2024).