Litera
Правильная ссылка на статью:
Неренц Д.В. Особенности фейкового контента в медиапространстве в эпоху развития искусственного интеллекта // Litera. 2024. № 7. С. 107-114. DOI: 10.25136/2409-8698.2024.7.43843 EDN: TETGUG URL: https://nbpublish.com/library_read_article.php?id=43843

Особенности фейкового контента в медиапространстве в эпоху развития искусственного интеллекта

Неренц Дарья Валерьевна

кандидат филологических наук

доцент, кафедра журналистики, Российский государственный гуманитарный университет

125993, Россия, Московская область, г. Москва, Миусская площадь, 6, ауд. 525

ya.newlevel@yandex.ru

Статья из рубрики "Журналистика"

DOI:

10.25136/2409-8698.2024.7.43843

EDN:

TETGUG

Дата направления статьи в редакцию:

20-08-2023

Аннотация: Предметом исследования в данной статье выступает специфика фейкового контента в условиях активного развития искусственного интеллекта (ИИ). Первое полугодие 2023 года продемонстрировало, насколько серьезное влияние может оказать ИИ на жизнь общества в целом, и на информационную повестку в частности. В эпоху беспрерывного потока получаемых данных все сложнее верифицировать контент и критически осмыслять все получаемые сведения. Так называемые дипфейки стали по-настоящему актуальной медиаугрозой, способной вызвать массовые волнения и повлиять на настроения аудитории. В связи с этим актуальным и важным является изучение фейков, созданных ИИ, с целью выявления способов их обнаружения и разоблачения. В статье затронуты такие важные аспекты, как понятие и характерные черты фейкового контента как угрозы информационной безопасности, отличительные особенности дипфейков как продукта ИИ, представлены ключевые шаги по распознаванию фейков. Научная новизна исследования заключается в раскрытии роли дипфейка в современном информационном пространстве, опасности, которые они в себе несут (в том числе и в силу своего мгновенного тиражирования через социальные сети)

и вариантов борьбы с ними в рамках медиапотребления широкой аудитории. Более того, демонстрируется необходимость выработки дальнейших шагов по повышению уровня медиаграмотности и упреждения негативных последствий, которые способны вызвать фейки.

Ключевые слова:

фейк, дипфейк, искусственный интеллект, медиапространство, информация, СМИ, журналистика, медиаграмотность, информационная безопасность, медиаугроза

Работа выполнена в рамках проекта РГГУ «Медиабезопасность в эпоху цифровых трансформаций» (конкурс «Проектные научные коллективы РГГУ»)

Современное медиапространство состоит из непрерывного потока информации, поступающей со всего мира. В 2022 году общий объем данных в мире составил порядка 97 зеттабайт, а, согласно прогнозам, к 2025 году этот объем возрастет до 180 зеттабайт (Share of unique data and replicated data in the global datasphere in 2020 and 2024 // Statista. URL: https://www.statista.com/statistics/1185888/worldwide-global-datasphere-unique-replicated-data). Большое количество из этого объема составлено уже не вручную, а искусственным интеллектом (ИИ). Именно нейросеть на данном этапе развития информационно-коммуникационных технологий создает базы данных, собирает и агрегирует информацию, составляет отчеты и даже пишет посты и тексты для медиаресурсов. Все чаще человек не способен отличить человеческий продукт от продукта, созданного «машиной». Несмотря на очевидные преимущества (экономия времени и снижение энергозатрат, способность переработать гораздо больший объем сведений, ускорение рабочих процессов, возможность уйти от рутинных задач и т.п.), есть целый ряд информационных угроз в медиасреде, которые способны существенно повлиять на поведение и настроения всего общества. В данной статье речь пойдет о роли ИИ в создании и распространении медиаугроз, поскольку их влияние становится по-настоящему заметным.

В первую очередь, это относится к созданию и распространению информации. Нейросети становятся как важным звеном производства цифровых продуктов, так и генератором медиаконтента. Информация в интернете настолько разнообразна и объемна, что даже критически мыслящий человек может быть введен в заблуждение и подвергнуться манипуляции. Данные, которые человек получает, могут быть неактуальными, устаревшими, неполными или специально сфабрикованными. И последний вид представляется наиболее опасный вариантом медиаугрозы. Речь идет о фейках, которые в условиях активного развития новых технологий, также подверглись изменениям и стали практически неотличимы от фактов.

Исследователи определяют фейк по-разному, при этом единогласно сходясь в описании сути этого феномена. Так А.Д. Кривоносов понимает фейк как ложную информацию, которая тиражируется в СМИ в качестве новостного сообщения [1, с. 175]. И.А. Стернин и А.М. Шестерина указывают, что фейк может представлять собой только новость или сообщение, выдающееся за новость и содержащее какое-либо утверждение, поскольку чье-то мнение или оценка не могут быть отнесены к категории фейк-ньюс [2, с. 4]. С.Н. Ильченко определяет фейк как журналистское сообщение, содержащее недостоверную или непроверенную информацию, не соответствующую действительности и опубликованную в СМИ [3, с. 12]. Ю.М. Ершов отмечает, что фейк по-другому можно

назвать дезинформацией, поскольку он представляет собой целенаправленное использование выдуманных новостей с целью подрыва репутации как отдельного человека, так и компании или даже целого института [4, с. 246].

Таким образом, фейк можно обозначить как материал, направленный на введение в заблуждение, иначе говоря ложный или несоответствующий действительности контент. Один из отличительных признаков – сенсационный характер сообщения. Эмоциональная подача первостепенна, а факты вторичны [5, с. 28].

В качестве причин появления фейка выделяются:

1) неверно расслышанная или неверно интерпретированная автором информация;

2) экономические причины (желание получить финансовую выгоду или увеличить количество подписчиков);

3) создание негативного отношения к определенной точке зрения (или конкуренту);

4) намеренное введение в заблуждение с политической целью.

Фейки могут распространяться стихийно и оказывать серьезное влияние на настроения общественности. В качестве факторов, способствующих распространению фейков, можно обозначить недостаточный уровень осведомленности аудитории в теме, «информационный голод» (невозможность проверить данные или получить информацию из официальных источников), восприимчивость к сенсационным заголовкам и шокирующему характеру подачи новости, недоверие или негативное отношение аудитории к эксперту или экспертизе, нежелание пересматривать собственные сформированные взгляды.

Более того, непрерывно растущий поток информации, неспособность отличить главное от второстепенного, активное развитие новых технологий (боты способны бесперебойно тиражировать нужные данные), эмоциональная нестабильность (тролли способны вызвать сильный стресс и ввести в заблуждение с помощью провокационных сообщений) создают благоприятную среду для растущего потока разного рода фейкового контента.

Ложные материалы могут выходить в разных форматах, будь то пост в социальной сети или мультимедийный лонгрид. Однако наиболее разрушительным потенциалом обладают так называемые дипфейки (deepfakes) [6, с. 74]. Дипфейк – фейки, созданные с помощью ИИ. Искусственный интеллект объединяет большое количество фотоизображений и делает из них видеозапись. Программа может с высокой точностью определить, как человек будет реагировать и себя вести в определенной ситуации. Иными словами, суть технологии заключается в том, что часть алгоритма детально изучает изображение объекта и пытается его воссоздать, пока другая часть не перестает различать реальное изображение и созданное нейросетью.

Дипфейки практически невозможно распознать, поскольку видео как правило отличается высокой степенью реалистичности [7, с. 69]. Широкое распространение эта технология получила в 2017 году в США, благодаря разработке технологии «глубокого обучения» (deep learning), в рамках которой ИИ на основе обработки больших данных учится воспроизводить определенные паттерны (модели). На современном этапе дипфейки используются во многих сферах жизнедеятельности. Например, в 2023 году во Флориде фейковый Сальвадор Дали открыл свою выставку (Museum creates deepfake Salvador Dalí to greet visitors // YouTube. URL: https://www.youtube.com/watch?v=64UN-cUmQMs).

Общеизвестны случаи, когда с помощью нейросети рисуют картины, прописывают диалоги в сценариях к сериалам, создают реалистичные фотографии и даже научные тексты. В журналистике дипфейк используют, чтобы скрыть лица источника, пожелавшего остаться анонимным, и при этом не делать изображение размытым. Однако есть и негативные примеры.

Дипфейк стали активно использовать для создания фейк-ньюс и поддельных видео. Одно из них – скандально известный видеоролик 2018 года, на котором бывший президент США Барак Обама прямо оскорбляет действующего на тот момент президента Дональда Трампа (Fagan K. A viral video that appeared to show Obama calling Trump a 'dips---' shows a disturbing new trend called 'deepfakes' // The Insider. URL: https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4). Этим видео автор ролика, Джордан Пил, продемонстрировал реальную информационную угрозу, которая способна серьезно повлиять на общественное мнение и настроения масс. И теперь публичный деятель, чиновник, политик или представитель шоу-бизнеса может обвинить нейросеть и заявить, что его высказывания – результат работы ИИ, и он там никогда не был и никогда такого не говорил. Насколько возможно это доказать – вопрос сложный и пока только начинающий подниматься в правовом, этическом и научном поле. Однако получить поддержку аудитории таким психологическим приемом уже очевидно можно. Например, продюсер Иосиф Пригожин назвал фейком скандальную аудиозапись своего разговора с бизнесменом Фархадом Ахмедовым (Баласян Л. Иосиф Пригожин отрицает подлинность аудиозаписи беседы с миллиардером Ахмедовым с критикой власти // КоммерсантЪ. URL: https://www.kommersant.ru/doc/5899921). Однако доказательств этого никто обнаружить не смог.

ИИ способен создавать настолько качественные фейки, что они неотличимы от оригинала. В будущем такие технологии из-за создания ложного видеоролика или фотоизображения могут не только испортить жизнь одному человеку, но привести к массовым беспорядкам, митингам и даже военным столкновениям.

Еще одним нашумевшим примером стала фейковая фотография Папы Римского, одетого в модный пуховик и гулящего по улицам Нью-Йорка, которая была сделана с помощью нейросети Midjorney (Гостева А. Вирусное фото папы римского в стильном пуховике оказалось фейком // Lenta.ru. URL: https://lenta.ru/news/2023/03/27/thepope1). Или фотографии с арестом Д. Трампа, которого якобы насильно сажают в полицейскую машину (Макарычев М. В Сети появились фейковые снимки «силового ареста» Трампа полицейскими // Российская газета. URL: https://rg.ru/2023/03/21/v-seti-poiavilis-fejkovye-snimki-silovogo-aresta-trampa-policejskimi.html). Нередкими становятся случаи, когда ИИ используется для модуляции голоса и создания поддельного номера телефона, что позволяет обмануть доверчивых граждан и выманить у них большие суммы денег.

Подобные примеры демонстрируют серьезные угрозы для неподготовленной аудитории со стороны ИИ. Этим «оружием» могут умело пользоваться манипуляторы и мошенники, преследующие свои цели. Особенно остро этот вопрос встал с созданием нейросети ChatGPT-4, тексты, фотографии, видео которой порой невозможно распознать и выявить генерацию. В Российской Федерации к развитию ИИ относятся осторожно, не стремясь внедрить его во всех сферы производства резко и оперативно.

При этом в качестве серьезных рисковдля аудитории с позиции использования ИИ в медиа является создание фейкового контента, преднамеренно или по неосторожности. Социальные сети и блогерский контент является источником информации не только для молодежной аудитории, но и для многих журналистов, которые в погоне за трафиком и

популярностью стремятся не столько проверить новость, сколько стать первым и опубликовать «эксклюзив». Благодаря ИИ дипфейки могут обмануть даже опытных репортеров.

Другой вариант – когда сама нейросеть ошибается, неверно прочитав обозначения (например, вместо 1–7% указывая 17% или вместо 1925 г. говоря о 2025 г.). Такие фактические ошибки в рамках новостей о котировках акций или финансовых сделках могут привести к глобальным последствиям. В текстах, созданных нейросетью, может отсутствовать контекст происходящего, что также приведет читателя к неверному выводу.

С этической точки зрения важным аспектом является отсутствие указания, когда материал опубликован ИИ, а когда – реальным журналистом. Например, новостная лента во многих интернет-изданиях построена по принципу отсутствия авторства, что не позволяет читателю увидеть, кто и каким способом написал этот текст. Например, спортивное издание Sports.ru применяет ИИ при ведении спортивной хроники и генерации различных заголовков, которые иногда имеют ошибки и неточности. И тут уже встает следующий вопрос: кто несет ответственность за неточности, фактические ошибки, неверную интерпретацию данных или комментариев? Согласна ли аудитория читать тексты, сгенерированные нейросетями или смотреть новости с виртуальными ведущими?

Недоверие аудитории к нейросетям и компьютерным алгоритмам может стать существенной преградой на пути к развитию ИИ в медиасфере в целом. Согласно исследованию ВЦИОМ, в 2022 году более трети опрошенных (32%) россиян не доверяют технологиям ИИ. Самыми сильными страхами названы утечка собираемых им данных, использование в корыстных целях и риск принятия решений, за которые никто не несет ответственности (Россияне назвали свои главные страхи перед искусственным интеллектом // РБК. URL: https://www.rbc.ru/society/28/12/2022/63ab45de9a7947664c3ef893). Помимо названного, респонденты считают, что ИИ приведет к деградации населения (что, в целом, может быть близко к истине, учитывая последние тенденции в сфере образования, когда студенты считают приемлемым сдавать сгенерированный нейросетями текст в качестве реферата, курсового проекта и даже ВКР), а также слишком велики риски систематических сбоев и ошибок в работе.

Важным является и недостаток квалифицированных кадров для настройки и управления таким программным обеспечением. Это не позволяет редакции углубляться и изучать все возможности нейросети, а, что называется, идти «по верхам», используя только доступные и понятные всем функции. Подобный подход приводит к низкому качеству создаваемого контента и негативному отношению аудитории к подобным опытам. А редакция в результате теряет большие деньги и не может продолжать свои эксперименты.

Значительной сложностью является отсутствие полной картины мира (алгоритмы создают «информационный пузырь»), когда именно ИИ решает за пользователя, что главное, а что второстепенно и может быть проигнорировано, тем самым, не позволяя человеку увидеть все происходящее вокруг, а сосредоточиться только на маленькой части. Это, в свою очередь, способствует поляризации мнений, ведь современная аудитория все реже прибегает к рефлексии, отказываясь доверять или даже воспринимать сведения, которые не отражают ее убеждения.

В целом, дипфейки представляют собой актуальную медиаугрозу, поскольку в отличие от фотографий, коррекция которых не является чем-то уникальным, видеозаписям пользователи склонны по-прежнему доверять. Поэтому дипфейки в видеоформате с большой долей вероятности будут восприняты как подлинные, а благодаря социальным сетям могут быть растиражированы за считанные минуты. И здесь главным оружием борьбы выступает только критическое мышление. Чем в большей опасности себя чувствуют люди, тем более они подвержены они влиянию. Они менее критичны, меньше сил тратят на проверку информации. Даже рациональные люди могут быть введены в заблуждение. Этим умело пользуются манипуляторы. В этом случае определенный алгоритм позволит убедиться в подлинности получаемых данных:

1) прочесть полностью материал (заголовок не всегда отражает содержание или суть текста);

2) определить автора и дату публикации (это крупное СМИ или чей-то авторский блог, насколько автор текста авторитетен и знает тему);

3) проверить адресную строку (это могут быть «клоны», которые копируют страницы настоящих медиа, изменяя букву или добавляя какой-то знак пунктуации, или фейковые аккаунты, которые также отличаются буквально одним символом в названии);

4) найти и изучить первоисточник информации;

5) в фейковых материалах большое количество эмоциональных высказываний, субъективных суждений, оценочных выражений (упор на сенсацию или эксклюзив). Как правило, в журналистских материалах на первое место выходит фактическая информация, а эмоциональные высказывания играют второстепенную роль.

Однако если речь идет о дипфеках, этих правил может быть недостаточно. При сгенерированном видео стоит обращать внимание на качество изображения, естественность телодвижений, блики в глазах или на очках, естественность освещения и фоновых цветов и пр. Иными словами, увидеть искусственно созданные кадры может только крайняя внимательность и целенаправленный поиск.

По большей части отличить фейк от правды способен только сам пользователь. Конечно, наиболее крупные и резонансные фейки будут разоблачены публично, однако это лишь незначительная часть той неправды, с которой сталкивается ежедневно человек. На первый план выходит способность критически осмыслять любую важную информацию, обращаться только к проверенным источникам, перепроверять данные. Иными словами, как никогда важно продолжать повышать медиаграмотность и соблюдать основные правила информационной безопасности.

Библиография

1. Кривоносов А.Д. Эволюция фейков в эпоху диджитализации // Известия Санкт-Петербургского государственного экономического университета. 2022. № 6 (138). С. 174-177.
2. Стернин И.А., Шестернина А.М. Маркеры фейка в медиатекстах. Рабочие материалы. Воронеж: ООО «РИТМ», 2020.
3. Ильченко С.Н. Фейк-контроль, или Новости, которым не надо верить: как нас дурачат СМИ. Ростов н/Д: Феникс, 2021.
4. Ершов Ю.М. Феномен фейка в контексте коммуникационных практик // Вестник Томского государственного университета. Филология. 2018. № 52. С. 245-256.
5. Галяшина Е.И. и др. Фейковизация как средство информационной войны в интернет-

медиа: научно-практическое пособие. М.: Блок-Принт, 2023.
6. Колесникова Е.В. Технология deepfake в журналистике // Мир современных медиа: новые возможности и перспективы: сборник научных трудов / под общ. ред. Д.В. Неренц. М.: Знание-М, 2022. С. 74-77.
7. Смирнов А.А. «Глубокие фейки». Сущность и оценка потенциального влияния на национальную безопасность // Свободная мысль. 2019. № 5 (1677). С. 63-84.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Оценка и анализ фейкового контента в медиапространстве все чаще встречается в научной сфере. Действительно, тема вполне актуальна, интересна, причем она рассматривается с разных точек зрения, например, появления данного сегмента, или же его распространение, или определение роли / функции указанного блока. Следовательно, выбор автора вполне оправдан, а работа является неким дополнением уже существующей достаточно большой массы источников (РИНЦ, сеть Internet, ВАК-издания и т.д.). Причем фейк-вариант и его распространение связывается с галопирующим развитием искусственного интеллекта. Работа, на мой взгляд, имеет реферативно-обобщенный характер, в ней нет продуктивно-концептуальной идеи, которая была бы отлична, оригинальна, явно нова. Однако общая системная оценка имеющихся данных тоже необходимо, ибо она позволяет прогнозировать развития научной мысли далее. Цель и задачи, которые ставит пред собой автор рецензируемого сочинения, воплощены; методология анализа ориентирована на эмпирико-системный тип, не помешала бы работе ярко выраженная гипотеза природы фейкового контента, м.б. полярная раскладка этой достаточно непростой проблемы. Стиль тяготеет к научному типу. Например, это проявляется в следующих фрагментах: «фейки могут распространяться стихийно и оказывать серьезное влияние на настроения общественности. В качестве факторов, способствующих распространению фейков, можно обозначить недостаточный уровень осведомленности аудитории в теме, «информационный голод» (невозможность проверить данные или получить информацию из официальных источников), восприимчивость к сенсационным заголовкам и шокирующему характеру подачи новости, недоверие или негативное отношение аудитории к эксперту или экспертизе, нежелание пересматривать собственные сформированные взгляды», или «ложные материалы могут выходить в разных форматах, будь то пост в социальной сети или мультимедийный лонгрид. Однако наиболее разрушительным потенциалом обладают так называемые дипфейки (deepfakes). Дипфейк – фейки, созданные с помощью ИИ. Искусственный интеллект объединяет большое количество фотоизображений и делает из них видеозапись. Программа может с высокой точностью определить, как человек будет реагировать и себя вести в определенной ситуации. Иными словами, суть технологии заключается в том, что часть алгоритма детально изучает изображение объекта и пытается его воссоздать, пока другая часть не перестает различать реальное изображение и созданное нейросетью», или «важным является и недостаток квалифицированных кадров для настройки и управления таким программным обеспечением. Это не позволяет редакции углубляться и изучать все возможности нейросети, а, что называется, идти «по верхам», используя только доступные и понятные всем функции. Подобный подход приводит к низкому качеству создаваемого контента и негативному отношению аудитории к подобным опытам. А редакция в результате теряет большие деньги и не может продолжать свои

эксперименты» и т.д. Не лишена работа практической составляющей, хотя большая часть высказанного автором, «находится на поверхности». Нарочито автор обозначает в тексте т.н. предупредительный характер «фейков» и «дипфейков», думается, что это вполне рационально. Например, «в целом, дипфейки представляют собой актуальную медиаугрозу, поскольку в отличие от фотографий, коррекция которых не является чем-то уникальным, видеозаписям пользователи склонны по-прежнему доверять. Поэтому дипфейки в видеоформате с большой долей вероятности будут восприняты как подлинные, а благодаря социальным сетям могут быть растиражированы за считанные минуты. И здесь главным оружием борьбы выступает только критическое мышление. Чем в большей опасности себя чувствуют люди, тем более они подвержены они влиянию. Они менее критичны, меньше сил тратят на проверку информации. Даже рациональные люди могут быть введены в заблуждение. Этим умело пользуются манипуляторы. В этом случае определенный алгоритм позволит убедиться в подлинности получаемых данных…» и .т.д. Материал, на мой взгляд, затрагивает «свежую» тему и может быть интересен / полезен как опытным читателем, так и тем, кто только приближается к проблеме анализа «фейковых материалов». Основные требования издания учтены, структура выдержана, наличного объема достаточно для раскрытия темы, но примеров / иллюстраций могло быть введено более. Итоги по тексту объективны, они соотносятся с основным блоком данных: «по большей части отличить фейк от правды способен только сам пользователь. Конечно, наиболее крупные и резонансные фейки будут разоблачены публично, однако это лишь незначительная часть той неправды, с которой сталкивается ежедневно человек. На первый план выходит способность критически осмыслять любую важную информацию, обращаться только к проверенным источникам, перепроверять данные. Иными словами, как никогда важно продолжать повышать медиаграмотность и соблюдать основные правила информационной безопасности». Библиографические источники актуальны, тип изданий вариативен. Рекомендую статью «Особенности фейкового контента в медиапространстве в эпоху развития искусственного интеллекта» к публикации в журнале «Litera».