УДК 327; 339.9
DOI:10.30546/3006-0346.2024.1.79.66

ДИПФЕЙКИ, КАК РАЗНОВИДНОСТЬ МАССОВОЙ ЦИФРОВОЙ ДЕЗИНФОРМАЦИИ

ЛЯМАН СУЛЕЙМАН-ЗАДЕ
Бакинский Государственный Университет, докторант
E-mail: lyama13@gmail.com

На торжественном приеме по случаю Дня Республики в 2019 году Президент Азербайджанской Республики Ильхам Алиев заявил: «Азербайджан - страна, на протяжении многих лет являющаяся мишенью информационной войны. Искаженная, ложная информация, вымышленные факты, фиктивные новости о стране получили широкий размах. Мы должны адекватно отвечать на это, доводить до мира свои реалии».

12 октября 2023 г. Государственная служба специальной связи и информационной безопасности Азербайджанской Республики и Агентство развития медиа Азербайджанской Республики сделали совместное заявление, о распространенном в ряде сегментов социальных сетей видеообращении, якобы адресованном азербайджанскому народу от имени Президента Азербайджанской Республики господина Ильхама Алиева. Служба заявила, что упомянутое видеообращение является ложной не отражающей реальности информацией, созданной с использованием технологии «deepfake». Случаи создания и распространения фальшивых видеороликов с использованием технологии «deepfake», были выявлены и предотвращены и во время крупной анти-азербайджанской информационной интернет-кампании, определенно носившей характер использования технологий искусственного интеллекта, в сентябре-ноябре 2020 года в период 44-х дневной Отечественной войны.

Всего три десятилетия назад практические возможности отдельных лиц и организаций передавать изображения, аудио и видео (как подлинные, так и нет) были ограничены. Распространение информации на национальном и глобальном уровнях контролировали крупнейшие теле- и радиосети, газеты, журналы и книжные издательства. Хотя правительства, рекламодатели и видные политические деятели могли влиять на зарубежные средства массовой информации, большинству из них приходилось заниматься передачей контента лишь на локальном уровне. Информационная революция разрушила эту модель.

Обмен информацией в Интернете стал сложной проблемой, особенно в контексте различных глобальных взглядов на право слова, свободу выражения мнений, роль средств массовой информации и основные ценности Интернета. Хотя дискуссии о дезинформации существовали на протяжении тысячелетий, 2016 год ознаменовал переход в то, что некоторые называют «эрой постправды», когда информация, как правдивая, так и ложная, стала оружием политической борьбы. В конце 2016 г. Оксфордский словарь английского языка (OED) объявил «постправду» словом года. Оно определяется как «обстоятельства, при которых объективные факты оказывают меньшее влияние на формирование общественного мнения, чем обращение к эмоциям и личным убеждениям». По данным OED, использование этого термина увеличилось более чем на 2000% с 2015 г. - скорее всего, в связи с референдумом Соединенного Королевства о выходе из Европейского Союза и президентскими выборами в США, которые были окутаны кампаниями по дезинформации [10].

Проблемы дезинформации глубоко переплетены с развитием цифровых медиа. Они движимы государственными или негосударственными политическими субъектами, коммерческими субъектами, средствами массовой информации, гражданами, индивидуально или в группах - а также манипулятивным применением коммуникационных инфраструктур, которые используются для производства, распространения и усиления дезинформации в более крупных масштабах, чем раньше, часто новыми способами, которые до сих пор плохо обозначены и поняты. Так в отчете от 2022 г. неправительственной организации Freedom House, специализирующейся на исследованиях состояния политических и гражданских свобод, зарегистрировано две тенденции: снижение свободы Интернета и рост дезинформации в информационном поле. Среди 65 опрошенных стран более 30-ти занимались де-

зинформацией и влиянием на деятельность информационных агентств внутри и за пределами своих границ [7, с.2].

«Фейковые новости» (новая риторика сегодняшнего дня - общий термин, охватывающий все: от целевой рекламы до пропагандистских кампаний и новостных статей, созданных с целью обмана общественности) - стали серьезной угрозой публичному дискурсу и демократии. В сочетании с охватом и скоростью социальных сетей ложная информация способна моментально завладеть вниманием миллионов людей, оказывая негативное воздействие на общество в целом. Дезинформацию не так-то просто выявить и раскрыть, кампании по манипулированию информацией часто сочетают ее с элементами мизинформации (вводящие в заблуждение утверждения) и правдивой информации [12,с.1]. Трудно отследить источники дезинформации, проблематично установить действующие лица и тем более вопросы атрибуции, стоящие за фейковым контентом. Глобальное общество «постправды» изобилует ложными сообщениями, пропагандистскими машинами, предвзятыми алгоритмами, вредоносными приложениями искусственного интеллекта, ботами в социальных сетях и вводящими в заблуждение информационными кампаниями.

Одним из последних событий, усугубляющих проблему распространения дезинформации, с учетом того, что каждый пятый пользователь Интернета на сегодняшний день получает новости через видеохостинг YouTube, уступающий только социальной сети Facebook, стало появление дипфейков (комбинация двух терминов: глубокое обучение и фейки) - гиперреалистичных видеороликов, в которых применяется искусственный интеллект (ИИ).

Дипфейк - это название синтетических медиа, в которых человек на существующем изображении или видео заменяется своим подобием и одновременно общее название технологии «использования методов глубокого обучения для тренировки алгоритмов визуальных манипуляций» [1,с.7]. Фактором, меняющим правила игры в дипфейках, является масштаб и сложность задействованных технологий, поскольку практически любой, у кого есть компьютер или смартфон, может создавать фейковые видеоролики, практически неотличимые от подлинных медиа. В настоящее время программное обеспечение для дипфейков легко доступно для бесплатной загрузки (например, DeepFaceLab, DeepNude, FaceSwap, FakeApp, Zao), а сотни видеороликов на YouTube предлагают обучающие материалы для новичков. И хотя ранние дипфейковые материалы явно демонстрировали признаки манипуляции, благодаря технологическим новшествам с каждым днем их содержание становится все более убедительным. Дипфейки нацелены на платформы социальных сетей, где легко распространяются заговоры, слухи и дезинформация, поскольку непрекращающийся «информационный апокалипсис» подталкивает людей доверять информации, исходящей исключительно из них [14,с.2]. Причинами стремительного вирусного распространения дипфейков считаются три явления - динамика «информационного каскада», человеческое влечение к негативной и новой информации и «пузыри фильтров».

Чтобы бороться с дипфейками, необходимо понимать технологии, лежащие в их основе и причины их существования. Инновационная технология дипфейков использует мощные методы искусственного интеллекта и машинного обучения для создания видео- и аудио-контента с высоким потенциалом обмана и манипулирования. Методы машинного обучения, используемые для создания дипфейков, основаны на глубоком обучении нейронных сетей: генеративно-состязательных (GAN) и вариационных автокодировщиков (VAE) [2, с.2]. Основная идея заключается в одновременном обучении двух алгоритмов нейронной сети («дискриминатора» и «генератора») посредством состязательного и итеративного процесса на одном и том же наборе данных: изображений, видео или звуков. Первая пытается создать новые образцы, чтобы ввести в заблуждение вторую сеть, определяющую, является ли новый носитель подлинным. Со временем генератор учится выдавать все более реалистичные выходные данные, в результате чего дискриминатору становится все труднее отличать данные, на которых он обучался, от данных, создаваемых генератором. Этот двухуровневый производственный процесс позволяет сетям учиться на своих ошибках, корректируя алгоритмы и тем самым совершенствуя дипфей-

ки. GAN может просмотреть тысячи фотографий человека, создавая новый портрет, который на выходе будет соответствовать исходным фотографиям, не являясь точной копией ни одной из их. Нейронные сети анализируют большие наборы данных для имитации мимики человека, его манер, голоса и интонации, после чего передают видеозапись в алгоритм глубокого обучения для их преобразования [2, с.7].

Кто создает дипфейки? Существует как минимум четыре основных типа их производителей:

✓ иностранные правительства, политические игроки, политические агитаторы, хактивисты, террористы;

✓ злонамеренные субъекты и мошенники;

✓ законные субъекты, телевидение и средства массовой информации;

✓ сообщества любителей.[11, с.305]

Изначально дипфейки фокусировались на знаменитостях и корпоративных лидерах, ввиду большого количества исходного материала в Интернете: фотографий и видео необходимых для обучения систем искусственного интеллекта. Однако, в 2018 году дипфейки стремительно ворвались в политический мейнстрим. Самыми запоминающимися из них стали:

✓ Созданный Голливудским режиссером Джорданом Пилом видеофрагмент, представивший экс-президента США Б.Обаму, обсуждающего опасность фейковых новостей и высмеивающего на тот момент действующего президента Дональда Трампа.

✓ В аналогичном дипфейке того же года, Дональд Трамп называет изменение климата фальшивкой и озвучивает решение о выходе США из Парижского соглашения по климату. На сей раз видео было создано бельгийской социалистической политической партией SP.A с целью сбора подписей под онлайн-петицией, призывающей бельгийское правительство принять срочные меры по борьбе с изменением климата [8, с.633].

✓ В конце мая 2019 г. в Интернете распространилось видео с Нэнси Пелоси, на котором спикер Палаты представителей США выглядела пьяной и невнятно произносила слова. Видео на Facebook за считанные дни было просмотрено более 2,5 млн. раз, им поделились видные политические лидеры, что стало примером преднамеренно измененного аудио-визуального контента, усиленного эффектом социальных сетей [1, с.3].

✓ Чтобы подчеркнуть потенциальную угрозу дипфейков выборам 2020 года, демократическая партия США в том же году подделала собственного председателя Тома Переса.

✓ 2 марта 2022 г., вскоре после российского вторжения в Украину, на новостном сайте «Украина 24.1» появилось видеообращение президента страны Владимира Зеленского, умолявшего украинцев сложить оружие и сдаться. Неудивительно, что видео быстро распространилось в ВКонтакте, Telegram и других социальных сетях, где его подхватили и сообщили о нем многие мировые СМИ.

Обладая беспрецедентным уровнем реализма, скоростью и способностью персонализировать дезинформацию, дипфейки представляют собой серьезную угрозу обществу, политической системе и бизнесу поскольку:

✓ угрожают национальной безопасности путем осуществления идеологической пропаганды и внешнего вмешательства в политику суверенной страны;

✓ подрывают доверие граждан к информации, исходящей от легитимной власти;

✓ наносят вред кибербезопасности;

✓ оказывают давление на СМИ, затрудняя распознавание фейковых новостей.

Дезинформация все чаще рассматривается как вопрос национальной безопасности. Субъектами дезинформации могут быть как внутренние, так и внешние силы: внутриполитические игроки могут подвергнуть общественность страны политическим манипуляциям, способствуя вирусной цифровой войне, провоцирующей внутренние беспорядки, иностранные игроки также могут вмешиваться во внутренние процессы суверенного государства, привнося ложь и конфликты в его публичную сферу. Различные политические акторы, в том числе политические агитаторы, хактивисты, террористы и иностранные государства могут использовать дипфейки в кампаниях по дезинформации с целью манипулирования общественным мнением и подрыва доверия к институтам страны [13, с.7]. Дезинформация, передаваемая с помощью дипфейков, может стать проблемой во время выборов, любой политический деятель может попытаться дискредитировать оппонента или спровоциро-

вать политический скандал с целью продвижения собственной кандидатуры. Дипфейк способен изменить отношение граждан к политику или к партии политика, и в результате повлиять на голосование в соответствии с целями политического игрока, стоящего за дипфейком.

Технология дипфейков может использоваться и более тонкими способами манипулирования, например, посредством «подманивания толпы» - метода, используемого в области кампаний маркетинга, для создания видимости (или отсутствия) народной поддержки [5, с.11]. Вскоре станет возможным настраивать прямую трансляцию, ретушируя выражения лиц людей, например, слушающих речь политического деятеля, так чтобы казалось, что они ухмыляются, хмурятся или выглядят скучающими во время выступления. Дипфейки могут использоваться для создания материалов шантажа, ложно инкриминирующих жертву, для шантажа избранных должностных лиц или тех, кто имеет доступ к секретной информации, в целях шпионажа или оказания давления. Дипфейки также становятся все более популярными среди мошенников с целью рыночных и фондовых манипуляций. Эксперты прогнозируют, что злоумышленники будут стремиться монетизировать использование дипфейков, начав предлагать их как услугу, предоставляя менее опытным или знающим хакерам инструменты для использования атак посредством лишь нажатия кнопки за небольшую оплату [4, с.3].

Однако, стоит заметить, что технология deepfake имеет также и положительное применение во многих отраслях и сферах человеческой деятельности, включая кино, образовательные средства массовой информации, цифровые коммуникации, игры, развлечения, социальные сети, музейную деятельность, здравоохранение, различные сферы бизнеса, такие как мода и электронная коммерция [3, с.1]. Потенциальные преимущества целевых дипфейков настолько значительны для различных государственных целей, что можно с уверенностью сказать, государственные субъекты различных стран проводят секретные исследования в этой области. Однако неясно, отстают ли секретные исследования от коммерческих и академических усилий или опережают их. По крайней мере, можно с уверенностью сказать, что промышленность, научные круги и правительства имеют мотив, средства и возможность продвигать эту технологию быстрыми темпами [13, с.2].

Проблемы обнаружения глубоких фейков значительны, среди них:

✓ отсутствие юридических работников, специально обученных методам визуальной проверки;

✓ растущая сложность и одновременное снижение стоимости технологий глубокого обучения, что делает их доступными для широкого круга участников;

✓ информационная экосистема, с потерей доверия к фактам и их источникам.

Методами борьбы с дипфейками в этой связи видятся:

✓ законодательные меры и регулирование борьбы с цифровой дезинформацией;

✓ технология борьбы с дипфейками, включающая обнаружение, аутентификацию контента и предотвращение дезинформации;

✓ лишение социальных сетей юридического иммунитета от контента, который публикуют их пользователи;

✓ усовершенствованная корпоративная политика и обучение специалистов.

Дезинформация в контексте международных отношений предстает в виде преднамеренного распространения ложной или несбалансированной информации иностранными государствами (или соответствующими негосударственными субъектами) с целью увеличить собственное международное влияние. Основная техника международной дезинформации хорошо известна. Вычислительные системы стимулируют и автоматизируют манипулятивный медиа-контент, который широко и целенаправленно распространяется хакерами, троллями, ботами, фейковыми группами пользователей (astroturf) и другими акторами цифровой сферы, вовлеченными в распространение предвзятых фейковых сообщений и контента [9, с.221].

Дезинформация как инструмент внешней политики может быть частью гораздо более масштабного и опасного комплекса международной государственной деятельности в киберсфере, включая кибератаки, хакерские атаки и другие подрывные действия, объединяемые под по-

нятием «гибридной войны». Дипфейки могут оказать глубокое влияние на международные отношения: создать напряженность в дипломатических отношениях, спровоцировать вооруженный конфликт, вызвать общее недоверие к надежности отснятого материала, что может усилить нежелание международного сообщества участвовать в гуманитарной интервенции на основе видео- или аудиоматериалов, предположительно доказывающих совершение геноцида или других массовых нарушений прав человека, подлинность которых может быть трудно оценить. В целевых военных и разведывательных операциях дипфейки могут быть использованы для фальсификации приказов военачальников, с целью посеять хаос в вооруженных силах и среди гражданского населения, придать легитимность войнам и восстаниям. Осознание опасности манипулирования информацией в политических целях резко возросло после неоднократного иностранного вмешательства во внутриполитический процесс стран, особенно во время выборов. Феномен дипфейков потенциально влияет на цифровой суверенитет государств, поскольку предлагает технологические средства для манипулирования оцифрованными или цифровыми средствами идентификации, включая официальные средства идентификации [6, с.3].

Природа манипулирования информацией становится все более заметным аспектом глобальной политики. Одним из важных элементов контекстуального фона является текущая глобальная конкуренция за лидерство, происходящая в области технологий искусственного интеллекта и машинного обучения. В первую очередь это касается двух ведущих мировых технологических полюсов - США и Китая, а также Европейского Союза и России. Способность создавать глубокие фейки наверняка быстро распространится, независимо от того, какие усилия будут предприняты для защиты от них. Потенциал зависит не от дефицита материальных ресурсов, а скорее от доступа к технологиям, таким как нейронные сети и подходы к машинному обучению.

Сегодня Азербайджан – это уверенный и независимый региональный игрок, сохранивший и развивающий партнерские отношения с различными геополитическими центрами. Именно поэтому против Азербайджана и по сей день ведется война в киберпространстве, распространяются провокационные видеоролики. Несмотря на это, Азербайджан уверенно победил во второй Карабахской войне не только на боевом, но и на информационном поле. Лидером информационной борьбы Азербайджана был его Президент - Ильхам Алиев. Главным фактором, обеспечившим нашу победу в этой борьбе, стали интервью главы государства ряду влиятельных зарубежным СМИ, среди которых российские телеканалы «Россия-1», «Первый», РБК, агентства «РИА Новости», ТАСС и «Интерфакс», турецкие TRT Haber, CNN-Türk, Haber Global, Haber Türk, HTB, A Haber, французские телеканалы Sky News, France 24, газета La Figaro, немецкое телевидение ARD, японская газета Nikkei, американское телевидение Fox News, итальянские Rai-1 TV, газета La Repubblica, испанское информационное агентство EFE, а также Al Jazeera, Al Arabia, Euronews, CNN International и BBC, аудитории которых охватывают сотни миллионов человек по всему миру.

Дипфейки - это зарождающийся технологический феномен, который, как ожидается, будет иметь далеко идущие философские, юридические и моральные последствия, с точки зрения того, как мы воспринимаем истину и реальность. Дипфейки сами по себе не хороши и не плохи. Это всего лишь технология, которая выражает конструктивные или деструктивные (сознательные или бессознательные) намерения людей. Их не следует демонизировать, но и нельзя игнорировать реальную угрозу их использования в антисоциальных целях. История показывает, что те, кто пытается манипулировать средствами массовой информации, часто оказываются на шаг впереди тех, кто пытается защититься от таких манипуляций.

Потенциал дипфейков выглядит особенно пугающим в рамках продолжающейся информационной войны, направленной против Азербайджана. Уже сейчас необходима широкая подготовка специалистов по искусственному интеллекту, фрагментарно востребованных в нашей стране в сфере информационных технологий, финансовых услуг, телекома и ритейла. Большой потенциал цифровых технологий проявляет себя в дипломатии и внешнеполитической деятельности не в полную меру, а ведь

искусственный интеллект способен оперативно хранить в памяти и производить анализ больших объемов информации, выявлять общее и различия, причинно-следственные связи, поддерживать системы автоматического анализа и поиска подтверждений относительно фейковых видео, аудио и текстовых материалов для своевременного дипломатического реагирования на провокации в социальных сетях, Интернете и т.п. Чтобы подготовить конкурентно-способных специалистов в области искусственного интеллекта, нужны совместные усилия высших учебных заведений и государства по созданию специальных междисциплинарных программ, поскольку в рамках отдельной специальности данную задачу решить практически невозможно. Дефицит кадров для цифровых экспертных систем, имитирующих мышление человека – это в первую очередь технологическое отставание страны, что для Азербайджанской Республики недопустимо.

СПИСОК ЛИТЕРАТУРЫ:

1. Appel M., Prietzel F. The detection of political deep fakes. Journal of Computer-Mediated Communication 27.4.2022, p.1-13

2. Artificial Intelligence and National Security 2020/https://crsreports. congress.gov/product / pdf/R/R45178

3. Dack S. Deep fakes, fake news, and what comes next. The Henry M. Jackson School of International Studies, 2019

4. Danielle K.C., Chesney R. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. 107 California Law Review 1753, 2019

//https://scholarship.law.bu.edu/faculty_ scholarship/640

5. Jacobsen B.N., Simpson J. The tensions of deep fakes. Information, Communication & Society, 2023, p.1-15

6. Horne C.L. Internet governance in the post-truth era: Analyzing key topics in "fake news" discussions at IGF. Telecommunications Policy 45.6.,2021, p.102-150

7. Freedom House, Freedom on the Net 2022. Manipulating Social media to Undermine Democ¬racy, N.p., n.d., https://freedomhouse.org/ report/freedom-net/freedom-net-2022

8. Kocsis E. Deepfakes, shallowfakes, and the need for a private right of action //Dickinson law review. – Carlisle, 2022, Vol.126, N 2, p.621-650

9. Mullen M. A new reality: deepfake technology and the world around us // Mitchell Hamline law review. – Saint Paul, 2022, Vol.48, N1, p.210-234

10. Oxford English Dictionary (OED) https:// www.oed.com/?tl=true

11. Pérez D., Jesús A. and other. Deep fakes on Twitter: which actors control their spread?, Journal Media and Communication, 2021, 9, p. 301-312

12. Somogyi A. Are deep fakes a threat? Redefining deep fake - AI through popular culture & the everyday. Diss. Central European University, 2023

13. Tahraoui M., Krätzer C., Dittmann J. Defending Informational Sovereignty by Detecting Deep fakes: Risks and Opportunities of an AI-Based Detector for Deep fake - Based Disinformation and Illegal Activities. Weizenbaum Conference" Practicing Sovereignty: Interventions for Open Digital Futures". DEU, 2023

14. Vaccari C., Chadwick A. Deep fakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Figshare, 2020//https://hdl.handle. net/2134/11548140.v1

XÜLASƏ

DİPFEYK KÜTLƏVİ RƏQƏMSAL DEZİNFORMASİYA NÖVÜ KİMİ

L.S. Süleyman-zadə

Bakı Dövlət Universiteti, doktorant

Beynəlxalq münasibətlərdə informasiya ilə manipulyasiya, təbliğat və dezinformasiya hər hansı bir ölkənin, siyasi rejimin səlahiyyəti, müasir və ya yeni bir fenomen deyil - bunlar bir çox dövlətlərin xarici siyasətini həyata keçirmək üçün klassik alətlərdir. Süni intellekt texnologiyalarının meydana çıxması ilə dezinformasiyanın ötürülməsinin yeni və çox cəlbedici üsulu – dipfeyk yaranıb. Zərərli aktorlar tərəfindən yaradılan dipfeyklərin kütləvi istehsalı və yayılması perspektivi onlayn ictimai-siyasi diskursun həqiqiliyinə ciddi problem yarada bilər. Məqalədə dipfeyk nədir, onları kim istehsal edir, dipfeyk texnologiyasının üstünlükləri və təhdidləri nələrdir kimi suallara cavab tapmağa,

dipfeyklərdən nümunələr verməyə və onlarla mübarizə üsullarını nəzərdən keçirməyə çalışılır.

Açar sözlər: *Dezinformasiya, dipfeyk, süni intellekt, neyron şəbəkələr, informasiya müharibəsi, "post-həqiqət"*

РЕЗЮМЕ

ДИПФЕЙКИ, КАК РАЗНОВИДНОСТЬ МАССОВОЙ ЦИФРОВОЙ ДЕЗИНФОРМАЦИИ

Л.С.Сулейман-заде

Манипулирование информацией, пропаганда и дезинформация в международных отношениях не являются прерогативой какой-либо одной страны, политического режима, типично современным или новым явлением - это классические инструменты реализации внешней политики многих государств. С появлением технологий искусственного интеллекта возник новый и весьма привлекательный метод передачи дезинформации - дипфейки. Перспектива массового производства и распространения дипфейков, созданных злоумышленниками, может представлять собой серьезный вызов аутентичности общественно-политического онлайн-дискурса. Статья пытается найти ответы на вопросы, что такое дипфейки, кто их производит, каковы преимущества и угрозы технологии дипфейков, привести примеры дипфейков и рассмотреть методы борьбы с ними.

Ключевые слова: *Дезинформация, дипфейк, искусственный интеллект, нейронные сети, информационная война, «постправда»*

SUMMARY

DEEPFAKES AS A TYPE OF MASS DIGITAL DISINFORMATION

L.S.Suleyman-zada
Baku State University, doctoral student

Manipulation of information, propaganda and disinformation in international relations are not the prerogative of any country, political regime, a typically modern or new phenomenon - these are classic tools for implementing the foreign policies of many states. With the advent of artificial intelligence technologies, a new and very attractive method of transmitting disinformation has emerged - deepfakes. The prospect of mass production and distribution of deepfakes created by malicious actors may pose a serious challenge to the authenticity of online socio-political discourse. The article tries to find answers to the questions of what deepfakes are, who produces them, what are the advantages and threats of deepfake technology, give examples of deepfakes and consider methods of combating them.

Key words: *Disinformation, deepfake, artificial intelligence, neural networks, information war, "post-truth".*