

Развитие и распространение технологий «deepfake»: угрозы, риски и вызовы для глобального информационного сообщества

Д.В. Мун

кандидат экономических наук, заместитель директора ФГБУ «Агентство «Эмерком» МЧС России, со-основатель Международного экспертного сообщества «RISK.TODAY»
E-mail: moon@risk.today

В.В. Попета

кандидат технических наук, основатель и президент Международного экспертного сообщества «RISK.TODAY»
E-mail: info@risk.today

Аннотация. В статье исследуется социальный феномен «fake news», анализируются тенденции развития технологий «машинного обучения», позволяющие при помощи искусственного интеллекта создавать визуально «достоверные», не поддающиеся экспертизе «deepfake» фото и видео, даются прогнозы угроз и рисков глобального характера, которые неизбежно последуют в случае отсутствия должного контроля за применением возможностей передовых информационных технологий.

Ключевые слова: фейк-ньюс, deepfakes, Интернет, социальные сети, Facebook, WhatsApp, информационный взрыв, искусственный интеллект, генеративно-состязательные нейросети, «Часы судного дня», информационная культура личности.

Development and Diffusion of «Deepfake» Technologies: Threats, Risks and Challenges for the Global Information Community

D.V. Mun

Ph.D., Deputy Director EMERCOM Agency Ministry of Emergency Situations (Russia), co-founder of the International Expert Society «RISK.TODAY»
E-mail: moon@risk.today

V.V. Popeta

Ph.D., founder and President of the International Expert Society «RISK.TODAY»
E-mail: info@risk.today

Abstract. The article explores the social phenomenon of “fake news”, analyzes the development trends of “machine learning” technologies, which allow using artificial intelligence to create visually “reliable” “deepfakes” photos and videos, that are not amenable to any expertise, and gives the forecasts of global threats and risks which will follow without the proper control over the application of the capabilities of advanced information technologies.

Keywords: fake news, deepfakes, Internet, social networks, Facebook, WhatsApp, information explosion, artificial intelligence, Generative adversarial network (GAN), Doomsday Clock, personality information culture.

DOI: 10.31432/1994-2443-2020-15-1-23-35

Цитирование публикации: Мун Д.В., Попета В.В. Развитие и распространение технологий «deepfake»: угрозы, риски и вызовы для глобального информационного сообщества // Информационная и инновации. 2020. Т. 15, № 1. С. 23-35. DOI: 10.31432/1994-2443-2020-15-1-23-35.

Citation: Mun D.V., Popeta V.V. Development and diffusion of «deepfake» technologies: threats, risks and challenges for the global information community. // Information and Innovations 2020. T. 15, № 1. pp. 23-35. DOI: 10.31432/1994-2443-2020-15-1-23-35.

Эпидемия «фейк ньюс» как социокультурный феномен

Недостоверные, или «фейковые» новости – это информация, которую её авторы и распространители используют с целью извлечения политической и финансовой выгоды, а также для увеличения читательской аудитории и цитируемости. Данный прием начал активно использоваться в политической борьбе ещё с конца XIX века, в эру становления старых медиа – печатных, радио, и телевизионных СМИ, и имел название «газетная утка». Старые медиа представляют собой централизованный и односторонний канал коммуникации. Информация в традиционной медийной среде поступает от создателя контента, зачастую остающегося анонимным, напрямую к потребителю. Роли в традиционной медийной среде как правило фиксированы, существует четкая вертикальная иерархия. Пользователи не имеют возможности напрямую влиять на создание контента [1]. Еще ранее, с незапамятных времен, вброс недостоверных сведений носил название «ложь» или дезинформация. Однако настоящее воздействие на мировые события фейк-ньюс стали производить только с восходом эпохи новых, «горизонтальных медиа», под которыми в первую очередь подразумеваются мессенджеры и социальные сети: средства коммуникаций, которые в состоянии не только моментально и адресно распространять контент, но и оперативно получать обратную связь.

Ведущие эксперты относят сегодня фейк-ньюс к числу основных угроз мирового развития. Правительства всех без исключения стран мира в спешном порядке ищут пути и способы законодательной борьбы с данным явлением. Об угрозе фейковых новостей сделал свое особое заявление даже Папа Римский Франциск, сравнивая их со змеем-искусителем из Библии, «змеиные тактики которого нам необходимо разоблачать».

В ряде стран, например в Малайзии, Испании, Китае, распространителям фейков грозят серьезные тюремные сроки. Жесткие законы собираются принимать в США, Индии и Евросоюзе. В России правительство и законодатели, особенно после «вброшенных» информационных провокаций об «истинных масштабах тра-

гедии в «Зимней вишне» [2] также отреагировали на данный социокультурный феномен, оперативно приняв дополнения к законодательству, которые налагают штрафы за изготовление и распространение общественно опасных фейков. Цитируем: «Распространение в СМИ, а также в информационно-телекоммуникационных сетях заведомо недостоверной общественно значимой информации под видом достоверных сообщений, которая создает угрозу жизни и (или) здоровью граждан, массового нарушения общественного порядка и (или) общественной безопасности, прекращения функционирования объектов жизнеобеспечения, транспортной или социальной инфраструктуры, наступления иных тяжких последствий» нарушителям грозит административная ответственность в виде штрафов от 30 тысяч до полутора миллионов рублей. В свою очередь, интернет-ресурсы, уличенные в распространении фейков и не принявшие срочных мер по удалению недостоверной общественно-опасной информации, подлежат блокировке по запросу Роскомнадзора [3].

Однако на сегодняшний день, с учетом специфики функционирования данного социально-технологического явления, реально остановить или хотя бы сдержать общественную фейк-угрозу не представляется возможным.

Истоки и причины всемирного фейк-успеха

Почему в настоящее время невозможно остановить производство и распространение фейк-новостей? На это существует как минимум три причины.

Первая причина – глобальное и неконтролируемое распространение информационных технологий среди населения планеты. По состоянию на 2019 год аудитория интернета насчитывает 4,39 миллиарда человек, что на 366 миллионов (9%) больше, чем в январе 2018 года. В социальных сетях зарегистрировано 3,48 миллиарда пользователей. По сравнению с данными на начало прошлого года этот показатель вырос на 288 миллионов (9%). Фактически, весь прошлый год аудитория «всемирной сети» росла со скоростью один миллион новых пользователей в день.



Рис. 1.

На нынешнем этапе технологического развития цивилизации, создание, владение, контроль и распространение информации перестали быть прерогативой социально-правительственных элит и подконтрольных им централизованных/вертикальных СМИ. Информационные технологии подарили человеку никогда прежде не имевшуюся возможность не только «читать, видеть и слышать», но и быть «прочитанным, увиденным и услышанным». Практически каждый человек, имеющий смартфон, интернет, аккаунт в соцсетях, и знающий, как ими пользоваться, стал способен не только потреблять, но также генерировать, изменять, создавать и распространять информацию в неограниченном объеме. И рядовые пользователи сети не замедлили воспользоваться предоставленной им возможностью. Информационно-цифровую

вселенную захлестнул неудержимый вал видео, фото, селфи, постов, блогов, репостов и других проявлений информационной деятельности от индивидуальных пользователей сети.

Компания Cisco [22] отметила 26-кратный рост мобильного трафика только за период 2010-2015 года. Уже сегодня в Интернете хранится 81 триллион изображений и 7 триллионов видеоклипов; в добавок к ним ежедневно в интернет загружается 28 изображений и 2,5 клипа на каждого жителя Земли. Согласно прогнозам, к 2020 году мобильное видео как самый простейший индивидуальный способ создания и потребления информации выйдет на первое место среди всех других ресурсов всемирной сети.

Компания DELL EMC в своих исследованиях прогнозирует, что в 2020 году «цифровая все-

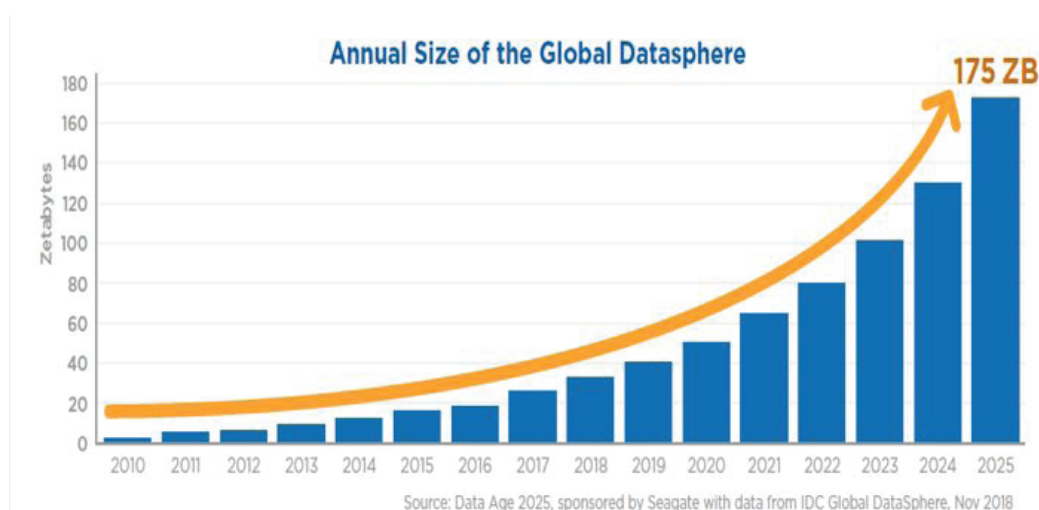


Рис. 2. Прогноз роста объема мировых данных по годам. Источник: seagate.com.

ленная» вырастет до 40 зеттабайт (10^{21}), то есть в 50 раз с уровня того же 2010 года. Если представить, что на всех пляжах планеты около 700 500 000 000 000 000 000 песчинок, то объем информации в 40 зеттабайт в 57 раз больше всего песка мира. А уже к 2025 году объем информации, создаваемой в интернете составит 175 зеттабайт [4].

На фоне данных прогнозов и статистики роста информационно-цифровой вселенной попытки, инвестиции и усилия многих стран мира взять под контроль, а ещё лучше управлять данным процессом, вкладывая значительные средства в формирование программ и технологий «реальной защиты» от фейков, кажутся наивными.

Например, агентство Bloomberg недавно сообщило, что Министерство обороны США заказало программное обеспечение, которое должно выявлять фейковые новости в сообщениях, сетевых изданиях, социальных сетях, аудио- и видеороликах, размещаемых в Интернете. Согласно заявленной стратегии, при помощи нового софта специалисты Пентагона будут отслеживать пропагандистские сообщения в новостях и социальных сетях: специальная программа сможет просматривать до полумиллиона новостных сообщений и постов в социальных сетях на предмет возможного манипулирования общественным мнением и пропаганды, а операторы смогут оперативно реагировать на фейковые вбросы. Предполагается, что программа должна заработать уже в текущем году и предотвратить возможное вмешательство извне в президентские выборы в США, которые должны пройти в 2020 году [5]. Однако мониторинг Пентагоном даже полмиллиона ключевых блогов в целом не позволяет влиять на глобальное информационное пространство.

В силу масштабности проблемы, любые попытки государственных и международных структур взять под мониторинг и контроль распространение информационных потоков пока обречены на неудачу.

Вторая причина – сложность установления первичного информационного вброса деструктивного «фейка» в случае использования для передачи информации зашифрованных «мессенджеров». Если вбросы и распространение опасных фейков еще можно пытаться

худобедно отследить через социальные сети (хотя Фейсбук попытался бороться собственными силами с распространением фейк-ньюс и потерпел сокрушительное поражение, о чем будет подробно указано далее), то в случае с чатами типа WhatsApp, где используются алгоритмы двойного шифрования, и у операторов системы нет доступа к переписке пользователей, у властей также отсутствует техническая возможность мониторить все сообщения.

Возьмем в качестве примера Индию, больше других стран пострадавшую от фейков про «банды похитителей детей». В основе информационного вброса, который стремительно передавали друг другу в естественном желании поделиться важной новостью пользователи WhatsApp, лежит «вирусный» ролик, на котором двое мотоциклистов с лицами, закрытыми шлемами, похищают на улице оставленного без присмотра взрослых ребенка. Данное видео, несмотря на низкое качество изображения, стремительно распространилось среди пользователей сразу нескольких индийских штатов и стало основанием для преследования местными жителями любых приезжих, ставших заведомо подозрительными, людей.

Первые инциденты с уличными расправами над незнакомцами начались в 2017 году. Когда же количество жертв стихийных расправ превысило тридцать человек, 4 июля 2018 года правительство Индии обратилось к владеющей WhatsApp компании Facebook, требуя выдать ключи шифрования для выявления зачинщиков и предпринять «немедленные меры» против злоупотреблений платформой. Ведь несмотря на многочисленные разъяснения местного министерства коммуникаций и информационных технологий о «фейковых вбросах», транслировавшихся ежедневно через официальные СМИ, пользователи продолжали пересылать друг другу «секретную, скрываемую властями» информацию о наводнивших страну и орудующих в разных штатах «бандах педофилов-расчленителей-похитителей детей».

WhatsApp отказался представлять «ключи» шифрования индийским властям, но все же, разделяя озабоченность тем, как чудовищно бесплатное приложение используется

By 2020, the number of WhatsApp users will reach 450 million

■ Number of monthly active users (in millions)

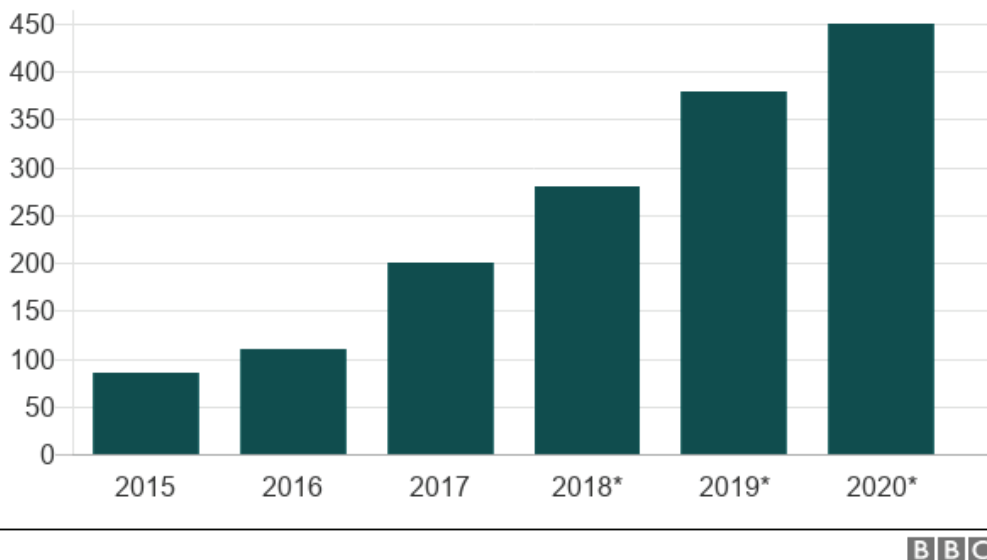


Рис. 3. Количество пользователей WhatsApp в Индии (принадлежит Facebook): <https://www.bbc.co.uk/news/resources/idx-e5043092-f7f0-42e9-9848-5274ac896e6d>

для манипуляции массами, предпринял ряд действий, направленных в первую очередь на ограничение возможностей рассылки сообщений: не более пяти адресатов в Индии и максимум двадцать адресатов по всему миру. Кроме того, в алгоритм была добавлена метка «Переадресовано» (Forwarded) для тех случаев, если сообщение было просто скопировано и переправлено далее.

Самоактуализация через фейк или истинная правда о том, кому на самом деле нужны фейк-ньюс

Третья, и, пожалуй, самая главная причина, по которой нельзя остановить производство и распространение фейк-новостей – это их социальная значимость для производителей, потребителей и распространителей. Истинная правда о том, кому на самом деле нужны псевдо-сенсационные новости, очевидна. Фейк-ньюс нужны практически всем нам. И поэтому ничто – ни штрафы, ни преследования – не в состоянии остановить человека, который решил самоактуализироваться, поднять свою значимость в глазах общественности, проде-

монстрировать собственную элитарность, выражающуюся в доступе к секретному, сокровенному, скрываемому, и, естественно, претендующему на альтернативную истинность знанию.

Любознательность – генетически заложенный во всех нас механизм выживания как вида. Новое знание в его нормальном проявлении – вещь чрезвычайно полезная, по сути, двигающая вперед прогресс. Неслучайно во главу угла всех диссертаций ставится в первую очередь научная новизна. Поделиться с собеседником информацией, претендующий в данный момент на сенсацию – не важно, что потом она будет опровергнута и предана забвению – это значит преподнести партнеру по коммуникации свою важность, и соответственно, поднять свой социальный статус. И неважно, что он сиюминутный и виртуальный – для большинства представителей поколения, родившегося со смартфоном и планшетом в руках, разница между миром реальным и виртуальным напрочь отсутствует. Напротив, для «digital natives» виртуальный мир более понятен, они в нем лучше ориентируются, легче преуспевают, следовательно, мир лайков, хайпа и фрагов

для них более важен, чем та среда, в которой и по законам которой существует их «физическая оболочка».

И даже если последовало официальное опровержение, авторитет распространителя «фейков» не сильно упадёт: ведь если ничего плохого не случилось, это уже хорошо. А если, не дай бог, все же случилось где-то что-то подобное, авторитет распространителя резко возрастает: к любому его следующему вбросу начинают относиться как к вещанию пророка. Его могут привлечь в качестве эксперта вертикальные СМИ, он может заработать денег и обрести известность, выступив на ток-шоу и т.д. Таким образом, распространитель заведомых фейков практически ничего не проигрывает, но может и сорвать солидный «куш». Каждый пользователь, вставший на этот путь, понимает, что с фейк-ньюс, как в рулетке, чтобы повезло, нужно чаще ставить на кон. То есть искать, создавать, распространять как можно больше псевдосенсаций.

Как пишет в своей книге «Максимальный репост» Бронислав Козловский [6], «чтение, а главное – репосты новостей — это не столько способ добыть информацию, сколько возможность показать себя другим людям. То есть социальный ритуал, который можно сравнить с посещением воскресной службы в церкви. Или с покупками в дорогом магазине. Нажимая на кнопку «поделиться», вы сообщаете, что такое-то событие важно для вашей картины мира... А если вам важен отклик, то ценность фальшивых и настоящих новостей неодинаковая — и ложные выгоднее истинных...неправдоподобные новости обещают больше нового знания — то есть являются «новостью» в превосходной степени. И привлекают к себе больше внимания».

Согласно исследованиям специалистов Массачусетского технологического института, изучившим более 126 000 фейковых новостей, которые были опубликованы в Twitter за все время его существования с 2006 по 2017 год (опубликовано в журнале Science в 2018 году [7]), заведомо очевидные фейк-новости распространяются в среднем в шесть раз быстрее «реальной» информации из достоверных источников, кроме того, для фейковой новости вероятность перепоста на 70% выше. При этом

максимальной скоростью распространения обладают фейки на темы, связанные с политическими деятелями, «медийными» персонами, катастрофами и стихийными бедствиями.

Глубина каскадов – важнейший параметр, непосредственно характеризующий, как далеко может распространиться новость по цепочкам репостов. И здесь «правда» также оказалась в проигрыше: «фейки» в десять раз быстрее достигали показателя в 19 перепостов, в то время как «правда» ни разу не смогла преодолеть рубеж в 10 переадресаций.

Как утверждают американские когнитивные психологи Крис Белл, Чип Хит и Эмили Стенберг, гипотезу которых поддерживают российские социальные антропологи Александра Архипова и Анна Кирзюк, слухи и фейковые новости (мемы) распространяются в информационной среде намного успешнее объективной правды именно за счет того, что люди попросту «хотят делиться друг с другом отвратительными эмоциями, которые вызывают городские легенды и слухи, и это желание также является следствием эволюционных механизмов», запрограммировавших нас предупреждать соплеменников об опасности [8].

Также исследователи определили, что именно люди, а не роботы (боты), определяют скорость и направления распространения фейковых новостей в информационной среде. При этом следует особо отметить, что активнее всего «ретвитят» фейки не «авторитетные» блогеры и «звёзды инстаграма», имеющие миллионы подписчиков, а непопулярные пользователи мессенджеров и соцсетей, испытывающие острую потребность в повышении своего социально-виртуального статуса.

«Онлайн-платформы как средства мгновенной передачи информации дают канал для выплеска лучшего и худшего, что есть в нас, включая наших страхи и предрассудки...это становится все более очевидным, когда нет эффективной власти, которая может обеспечить нашу безопасность», – утверждает директор Школы коммуникации Ибероамериканского университета (исп. Universidad Iberoamericana Ciudad de México) Мануэль Гутьерро.

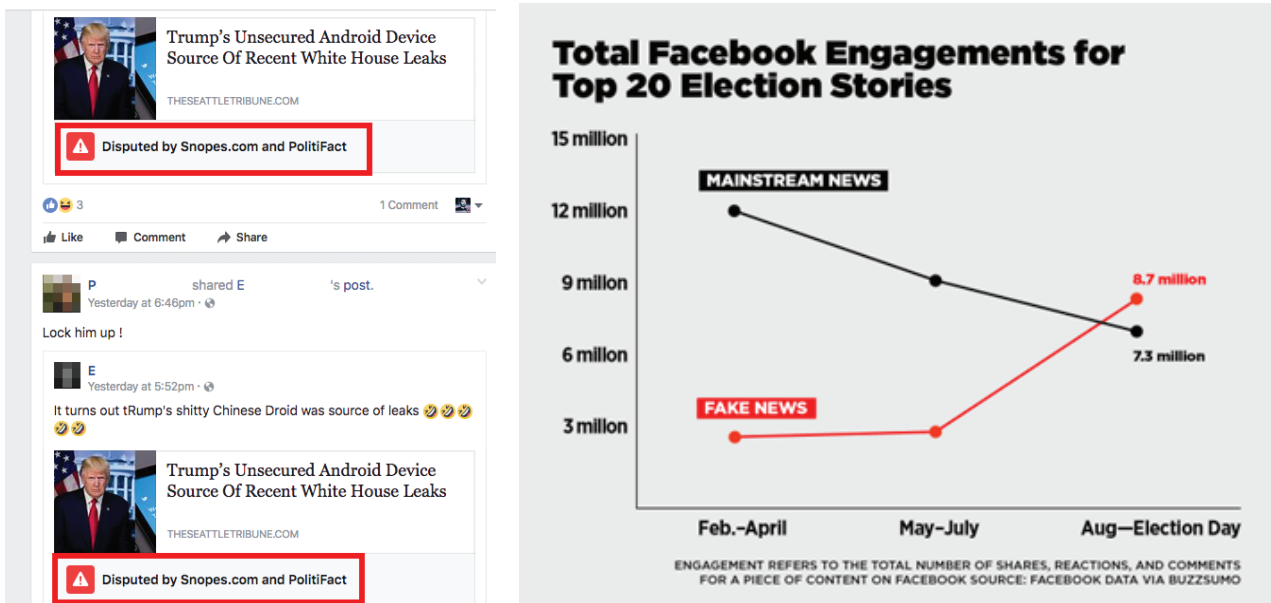


Рис. 4. Популярность «фейковых» новостей в Facebook резко выросла в течение трёх месяцев перед президентскими выборами в США 2016 года. Инфографика: BuzzFeed.

Источник: <http://pr-cy.ru/news/p/6119>

Если сравнивать между «лучшим» и «худшим», что есть в нас, по статистике Facebook худшее пока с большим отрывом побеждает. И подтверждает этот печальный факт неудавшаяся попытка бороться с распространением фейк-ньюс в крупнейшей в мире социальной сети её отца-основателя Марка Цукерберга. Долгое время администрация Фейсбук пыталась уйти от ответственности за контент, утверждая, что является «всего лишь технологической компанией», однако, после ряда официальных обвинений со стороны правоохранительных органов США во вмешательстве в президентские выборы, Цукерберг и его команда вытащили наконец «голову из песка» и публично пообещали с 2017 года проверять подлинность выкладываемых на страницах пользователей новостей и пресекать её стихийное распространение.

При этом разработчики Facebook решили не заниматься наймом платных экспертов, а опираться на ресурсы добровольцев-волонтеров, найденных в этой же сети, которые были готовы на безвозмездной основе определять фактическую достоверность новостей и сообщать админам, что те или иные посты содержат недостоверную информацию. Такие посты не удалялись, а помечались специальной отметкой «спорные» (англ. disputed). Как было написано

на двери клуба «четырёх коней» из бессмертного произведения И.Ильфа и Е.Петрова, «спасение утопающих – дело в первую очередь рук самих утопающих».

И вот чем это «благое начинание» закончилось. В марте 2017 года соцсеть начала помечать отдельные новостные истории на английском, квалифицированные волонтерами как «фейк», специальной предупреждающей плашкой — красный восклицательный знак в треугольнике и рядом метка «disputed». И с момента, как только сообщение помечалось как «disputed», количество его просмотров и репостов мгновенно и многократно возрастало. Любые другие попытки на административном уровне бороться с фейками в сети порождали обратный эффект у её пользователей. Спустя девять месяцев с начала всемирно объявленного «крестового похода» против фейк-ньюс, Цукерберг и К° махнули на все это рукой, и расписавшись в собственном бессилии, прекратили эту неравную борьбу.

Согласно независимым исследованиям, число репостов фальшивых новостей не только в Facebook, но и в других глобальных социальных сетях, по-прежнему многократно превышает число репостов добросовестных сообщений [9].

«Deepfake» — технологии: начало

В декабре 2017 года сотрудники исследовательского центра NVIDIA – одного из пионеров освоения возможностей искусственного интеллекта – опубликовали на страницах своего сайта [10] статью, в которой продемонстрировали, как можно при помощи GAN – генеративно-сопоставительных нейросетей – полностью видоизменять исходное цифровое фото, превратив исходные зимние пейзажи в летние. В отличие от «банального» фотешопа, подделку которого человек может распознать невооруженным взглядом, разработанный в NVIDIA алгоритм «глубокого» нейросетевого машинного обучения фактически позволял создавать новую цифровую реальность. На продемонстрированных картинках, созданных при помощи технологии «глубинного обучения» (англ. Deep learning), даже эксперты не смогли распознать подделку.

Неделю спустя на популярном социальном новостном сайте Reddit появились ссылки на порнографическое видео, на котором одна из самых известных и высокооплачиваемых голливудских актрис Галь Гадот занимается сексом со своим «сводным братом» [11]. Автор оказавшегося поразительно достоверным фейкового

видео, скрывающийся под ником Deepfakes, моментально стал чрезвычайно популярен. В последовавших за скандальной публикацией многочисленных интервью Deepfakes не только рассказал о себе, но и раскрыл общественности, как он создал свое «творение». Человек, который скрывается под данным ником – рядовой, ничем не выделяющийся программист, который признался, что использовал идеи и алгоритмы NVIDIA для того, чтобы просто воплотить свои эротические фантазии. Понятное дело, что Галь Гадот, будучи одной из самых красивых и успешных женщин мира, не только никогда не снималась в «домашнем порно» у Deepfakes, но даже не была уведомлена последним о том, что стала «главной героиней» его «кино для взрослых». Программист-эротоман для создания своего «шедевра» воспользовался обычным порнофильмом, в котором на тело неизвестной порноактрисы наложил изображения лица Гадот, позаимствованные из открытых источников: многочисленных фото и видеоматериалов из Google и YouTube. Deepfakes пояснил, что созданием таких роликов может заниматься каждый, кто умеет пользоваться интернетом и понимает принципы работы генеративно – сопоставительных нейросетей. Он также назвал



Рис. 5. Из зимы в лето: как переделывалось зимнее видео в летнее.
Скриншот с сайта <https://blogs.nvidia.com/blog/2017/12/03/nvidia-research-nips/>

программное обеспечение (которое бесплатно и находится в открытом доступе), использовавшееся для создания нового жанра в порноиндустрии, названного в последствие в его честь «Дипфейк» (англ. deer – глубокий, и fake – подделка).

Менее чем через месяц создание дипфейковых порновидеороликов со «звездами» превратилось в мощную индустрию. Ролики с любыми исходными «лицами» стали делать быстро, дешево, на заказ. И хотя юристы призывают «оскорблённых звезд» подавать на новоявленных «порнорежиссеров» в суд за незаконную обработку персональных данных, «звезды», памятуя об известном «Эффекте Стрейзанд» (социальный феномен, выражающийся в том, что попытка изъять определённую информацию из публичного доступа приводит лишь к её более широкому распространению в интернете) [12], не торопятся выступать с официальными опровержениями, дабы не привлекать в данном контексте лишнего внимания к своим персонам со стороны пользователей всемирной сети.

Эволюция дипфейка

И было бы, наверное, все хорошо, если бы дело ограничилось только дипфейковым порно. Однако, в скором времени у мира появятся куда более серьезные проблемы, чем несанкционированное использование домашними эротоманами находящихся на всеобщем обозрении «персональных данных» звезд гламура и кино в сочетании с бесплатными сетевыми возможностями искусственного интеллекта.

Как мы уже ранее упоминали, с нынешнего года мобильное видео выходит на первое место как самый простой способ потребления информации. И главным «хабом» – распространителем этого видео является YouTube – самый популярный видеохостинг в мире.

Летом 2018 года рекордсменом по просмотрам на YouTube стало видеообращение, на котором бывший президент США Барак Обама публично оскорбляет действующего президента США Дональда Трампа. Мир буквально взорвался от такой горячей новости и сразу же разделился на два активно противоборствующих лагеря: на тех, кто поддерживал Обаму, считая, что Трамп «получил по заслугам», и тех, кто был уверен, что ни Обама, ни кто-либо другой не должен позволять себе подобных высказываний.

Видео было настолько достоверным, что поначалу никто не поверил, что оно является подделкой. Дипфейк-подделкой. Обама за свои два президентских срока много чего наговорил, и в открытом доступе находится большое количество цифрового фото и видео с его непосредственным участием. Однако Трампа он никогда не оскорблял. И в данном конкретном случае он вообще оказался ни при чем.

Авторами провокационного ролика выступили американский комик и режиссер Джордан Пил и новостная интернет-медиа компания BuzzFeed. Дипфейк-видео, едва не ставшее поводом для «гражданской войны» между сторонниками республиканцев и демократов, было создано Пилом при помощи графического редактора Adobe After Effects и простой бесплатной программы Fakeapp, выложенной в открытом доступе на вышеупомянутом сайте Reddit для всех желающих самостоятельно поупражняться в создании «порно со звёздами».

По словам Пила, главная цель, которую он, как режиссер, преследовал при создании своего «фильма», благородна: предупредить людей о том, что «настали опасные времена. В будущем мы должны быть более бдительными в вопросе, чему верить в сети. Это время, когда мы должны полагаться на авторитетные новостные источники, которым стоит доверять», — заклю-

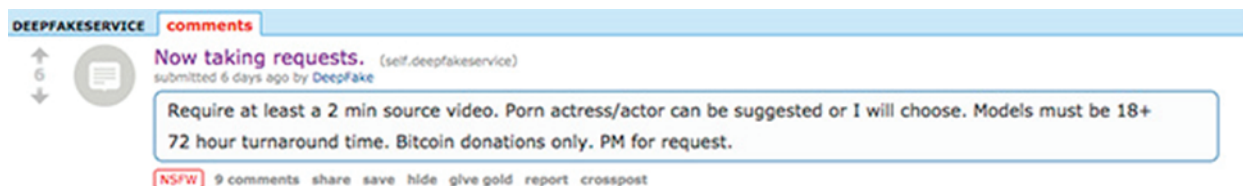


Рис. 6. Скриншот рекламы с сайта Reddit.

Источник: <https://lenta.ru/articles/2018/02/09/deepfake/>

чил Пил в конце своего скандального видеоролика устами Обамы [13].

В непременно следующих за подобными эпизодами интервью журналистам Джордан Пил также поведал следующее: «Я знал, что это возможно, но не был уверен, действительно ли это сработает. Когда оно сработало, я подумал: «Это безумие! Посмотрите на это! Знаете, пока это еще достаточно сложно, но это гораздо легче, чем должно быть» [14].

Сегодня, с учетом количества и качества исходного цифрового материала, хранящегося во всемирной сети и доступного каждому, а также не только политики и селебритиз, но и любой рядовой пользователь соцсетей, имевший ранее неосторожность выложить свои селфи и/или видео на всеобщее обозрение, рискует стать жертвой не просто розыгрыша, но и серьезной провокации или мошенничества.

Преступники не дремлют: они активно осваивают современные информационные технологии. И вот уже всемирные новостные ленты начали наполняться описаниями случаев подделки голосов директоров и CEO, по устному указанию которых сотрудники переводили деньги мошенникам; «дипфейковыми видео» со звездами экрана, призывающими перейти по ссылке и получить приз, и так далее.

Но самой большой опасностью, по мнению профессионалов информационной безопасности, являются дипфейки, которые будут использованы при ведении грядущих информационных войн. Так, профессор Калифорнийского университета в Лос-Анджелесе Джон Вильясенор утверждает, что «технология deepfake станет мощнейшим инструментом в руках тех, кто захочет дезинформировать массы и напрямую



Рис. 7. Вы не поверите, что Обама говорит на этом видео! Дипфейк Видео: BuzzFeedVideo. Источник <https://youtu.be/cQ54GDm1eL0>

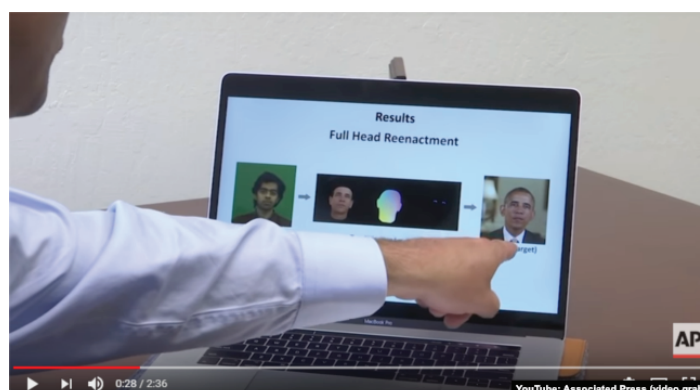


Рис. 8. Скриншот с обучающего видео в Ютубе, на котором объясняется, как работать с программами по созданию «дипфейков»

влиять на политический процесс». В свою очередь, американские военные считают, что самой перспективной разработкой являются редактирование и вбросы противникам поддельных спутниковых фото и видео, которые смогут их быстро дезориентировать и повергнуть мир в полный хаос [15].

В интернете уже несколько раз появлялись «пробные» видеоролики анонимных «энтузиастов», на которых руководители двух ядерных супердержав: президент США Дональд Трамп и президент России Владимир Путин сообщают о планах глобальных вооруженных нападений [16]. Многим зрителям хватило ума отнестись данным видео как к провокационным розыгрышам и не распространять панику в сети. Этому способствовало и низкое качество представленных на «суд общественности» доказательств.

Однако, нейросетевые технологии машинного обучения стремительно совершенствуются, на основании чего можно сделать вывод, что создание абсолютно достоверных «дипфейк-видео», которые будет невозможно, подвергнув экспертизе, отличить от реально снятых кадров, не «за горами».

Сегодня неуправляемое распространение недостоверной и непроверяемой информации начинает подрывать сами фундаментальные основы существования общества: взаимное доверие, доверие к государственным институтам, репутацию и авторитеты, презумпцию невиновности и так далее.

Если верить прогнозу одного из пионеров всемирной индустрии дипфейк-видео, профессора Хао Ли, «счет идет на месяцы, а не на годы. А это значит, что правдоподобные дипфейки, неотличимые от реальности, появятся во время предвыборной кампании в США — и, вполне возможно, изменят ее ход» [17]. Также в недавнем интервью CNBC Хао Ли отметил, что «такие дипфейки смогут создавать обычные пользователи без особой технической подготовки или уникального оборудования, и у нас есть лишь от шести до 12 месяцев, чтобы научиться распознавать подделки» [18].

Заключение

Было бы неверно утверждать, что проблеме создания специальных программ – «противоядия против дипфейков» – не уделяется должного внимания со стороны правительств, корпораций и частных лиц. Некоторые страны оперативно реагируют на информационные угрозы. Например, в Китае уже приняли закон, запрещающий с 1 января 2020 года публиковать фейковые новости и вводящие в заблуждение «дипфейковые» видео, созданные с помощью искусственного интеллекта. За публикацию вводящих в заблуждение роликов предусмотрена уголовная ответственность [19]. Однако успеют ли последовать данному примеру другие страны?

Кроме того, в настоящее время скорость распространения информации на порядок выше, чем оперативность реагирования на нее в виде проведения проверки, экспертизы и при необходимости создания соответствующего опровержения (с учетом того, что алгоритмы экспертизы дипфейков ещё предстоит разработать и внедрить). Поэтому кого, как и за что будут судить – вопрос весьма дискуссионный.

К сожалению, как это постоянно происходит в истории, сначала люди изобретают технологию, и лишь потом задумываются, как себя обезопасить от нее. Средства безопасности и защиты человека от достижений технического прогресса всегда сильно запаздывают в погоне за технологическим прогрессом. Сначала люди освоили огонь: и лишь тысячелетия спустя додумались изобрести огнетушитель и стали организовывать профессиональные пожарные команды; люди начали производить автомобили, разгоняющиеся до ста и более километров в час, и лишь десятилетия спустя изобрели трехточечный ремень безопасности [20]; люди научились извлекать энергию из деления ядер урана, но до сих пор так и не научились эту энергию контролировать. А к чему могут привести «шалости и игры» дилетантов с неизведанными пока возможностями искусственного интеллекта – даже страшно себе представить...

«Наша близость к катастрофе сейчас выражается в секундах — не в часах и даже не в мину-

тах...Мы сейчас сталкиваемся с настоящей чрезвычайной ситуацией — абсолютно неприемлемым состоянием дел в мире, которое не оставляет возможности для ошибок или задержек [с исправлением положения]» [21], утверждает исполнительный директор издания *Bulletin of the Atomic Scientists* Рэйчел Бронсон.

Сегодня не обязательно быть мировым лидером, чтобы запустить одним нажатием кнопки «репост» необратимый процесс распространения очередного «забавного фейка», который может начать на планете ядерную войну. Поэтому на каждом пользователе всемирной сети лежит персональная ответственность за то, что он/она читает, пишет, и распространяет.

ЛИТЕРАТУРА

1. Barbara Becker, Josef Wehner. *Electronic Networks and Civil Society*. p. 81 — SUNY Press, 2001. — ISBN 978-0-79145-016-1.

2. Ульяна Скойбеда «400 фейков кемеровского пожара: Почему люди так легко поверили в большую ложь после трагедии в ТРЦ «Зимняя вишня», «Комсомольская правда» от 02.04.18 г.

3. URL: <http://publication.pravo.gov.ru/Document/View/0001201903180031>

4. URL: <https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#b5fe21054597>

5. URL: <https://www.rbc.ru/politics/01/09/2019/5d6ae9a79a79475cc886b53c>

6. Борислав Козловский. «Максимальный репост: как соцсети заставляют нас верить фейковым новостям», М, Альпина Паблишер, 2018.

7. The spread of true and false news online «Soroush Vosoughi1, Deb Roy1, Sinan Aral2,*Science 09 Mar 2018.

8. URL: <https://www.vedomosti.ru/opinion/articles/2019/03/10/796034-mozhno-pobedit-zakonom>

9. URL: <http://6abc.com/news/fake-news-is-dominating-facebook/1621221/>

10. URL: <https://blogs.nvidia.com/blog/2017/12/03/nvidia-research-nips/>

11. URL: https://www.vice.com/en_us/article/gdydym/gal-gadot-fake-ai-porn

12. The Streisand Effect and Censorship Backfire, *International Journal of Communication* 9 (2015): pp. 656—671.

13. URL: <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peelee-buzzfeed>

14. URL: <https://ru.krymr.com/a/29347608.html>

15. URL: <https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/>

16. URL: <https://lenta.ru/articles/2020/01/01/cybercrime2020>

17. URL: <https://hightech.plus/2019/09/24/dipfeiki-stanut-neotlichimi-ot-originalov-v-techenie-6-12-mesyacev>

18. URL: <https://www.cnn.com/2019/09/20/hao-li-perfectly-real-deepfakes-will-arrive-in-6-months-to-a-year.html>

19. URL: <https://www.scmp.com/tech/apps-social/article/3039978/china-issues-new-rules-clamp-down-deepfake-technologies-used>

20. Мун, Д.В. От «Титаника» до «Зимней вишни». Алгоритмы грядущих катастроф : 3-е изд., доп. и перераб. – Москва ; Берлин : Директ-Медиа, 2018. – 352 с. ISBN 978-5-4475-9669-9

21. URL: <https://www.npr.org/2020/01/23/799047659/the-end-may-be-nearer-doomsday-clock-moves-within-100-seconds-of-midnight>

22. Cisco результаты исследования Интернета «Наглядный индекс развития сетевых технологий декабрь 2018 г.», URL: <https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html>

REFERENCES

1. Barbara Becker, Josef Wehner. *Electronic Networks and Civil Society*. p. 81 — SUNY Press, 2001. — ISBN 978-0-79145-016-1

2. Ul'yana Skojbeda «400 fejkov kemerovskogo pozhara: Pochemu lyudi tak legko poverili v bol'shuyu lozh' posle tragedii v TRC «Zimnyaya vishnya», «Komsomol'skaya pravda» ot 02.04.18 g.

3. URL: <http://publication.pravo.gov.ru/Document/View/0001201903180031>
4. URL: <https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#b5fe21054597>
5. URL: <https://www.rbc.ru/politics/01/09/2019/5d6ae9a79a79475cc886b53c>
6. Borislav Kozlovskij. «Maksimal'nyj repost: kak socseti zastavlyayut nas verit' fejkovym novostyam», M, Al'pina Pablisher, 2018.
7. The spread of true and false news online «Soroush Vosoughi1, Deb Roy1, Sinan Aral2,*Science 09 Mar 2018.
8. URL: <https://www.vedomosti.ru/opinion/articles/2019/03/10/796034-mozhno-pobedit-zakonom>
9. URL: <http://6abc.com/news/fake-news-is-dominating-facebook/1621221/>
10. URL: <https://blogs.nvidia.com/blog/2017/12/03/nvidia-research-nips/>
11. URL: https://www.vice.com/en_us/article/gdydym/gadot-fake-ai-porn
12. The Streisand Effect and Censorship Backfire, International Journal of Communication 9 (2015): pp. 656-671.
13. URL: <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peelee-buzzfeed>
14. URL: <https://ru.krymr.com/a/29347608.html>
15. URL: <https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/>
16. URL: <https://lenta.ru/articles/2020/01/01/cybercrime2020>
17. URL: <https://hightech.plus/2019/09/24/dipfeiki-stanut-neotlichimi-ot-originalov-v-techenie-6-12-mesyacev>
18. URL: <https://www.cnn.com/2019/09/20/hao-li-perfectly-real-deepfakes-will-arrive-in-6-months-to-a-year.html>
19. URL: <https://www.scmp.com/tech/apps-social/article/3039978/china-issues-new-rules-clamp-down-deepfake-technologies-used>
20. Mun, D.V., Popeta V.V. Ot «Titanika» do «Zimnej vishni». Algoritmy` gryadushhix katastrof: 3-e izd., dop. i pererab. – Moskva; Berlin: Direkt-Media, 2018. – 352 s. ISBN 978-5-4475-9669-9
21. URL: <https://www.npr.org/2020/01/23/799047659/the-end-may-be-nearer-doomsday-clock-moves-within-100-seconds-of-midnight>
22. Cisco rezul'taty` issledovaniya Interneta «Naglyadny`j indeks razvitiya setevy`x texnologij dekabr` 2018 g.», URL: <https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html>

