DOI: 10.24412/2500-2872-2021-3-90-101

Злонамеренное использование дипфейков: риски информационно-психологической безопасности Японии

Я.В. Лексютина

Аннотация. Стремительное развитие искусственного интеллекта стимулировало появление целого ряда новых технологий, одной из которых стала технология «дипфейк» (deepfake), позволяющая синтезировать очень реалистичный поддельный видео- или аудиоконтент. Появившаяся совсем недавно, в 2017 г., эта технология буквально за несколько лет стремительно прогрессировала: поддельный контент стал очень правдоподобным, приложения по созданию дипфейков стали более дешёвыми и легкодоступными, позволяя пользователям без специальных компьютерных навыков синтезировать поддельные видео за считанные секунды. Тенденцией последних нескольких лет стал взрывной рост численности и улучшение качества дипфейков в Интернете.

Открывая принципиально новые возможности для целого ряда индустрий (рекламы и медиа, индустрии развлечений и игр, киноиндустрии, медицины и т.д.), эта технология также может быть использована злоумышленниками в преступных целях, для оказания информационно-психологического воздействия на население, а также нанесения вреда межгосударственным отношениям. Риски злонамеренного использования дипфейков являются не менее реальными, чем польза от их применения. В некоторых странах мира технология «дипфейк» даже стала рассматриваться как способная, по мере её совершенствования, бросить вызов национальной безопасности и информационно-психологической безопасности.

В представленной статье даётся общая картина масштабов и направленности использования дипфейков в мире, потенциала и уже имевших место кейсов злонамеренного использования технологии «дипфейк» в политической сфере, а также оцениваются риски злонамеренного использования этой технологии для Японии и восприятие рисков японской стороной.

***Ключевые слова*:** Япония, дипфейки, искусственный интеллект (ИИ), информационно-психологическая безопасность, злонамеренное использование.

***Автор*:** *Лексютина Яна Валерьевна*, доктор политических наук, профессор РАН, профессор кафедры американских исследований, Санкт-Петербургский государственный университет (адрес: 199034, Санкт-Петербург, Университетская наб., д. 7-9). ORCID: 0000-0001-6766-1792; E-mail: lexyana@ya.ru

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Благодарности. Исследование выполнено за счёт гранта РФФИ и ВАОН (№ 21-514-92001).

***Для цитирования*:** Лексютина Я.В. Злонамеренное использование дипфейков: риски информационно-психологической безопасности Японии // Японские исследования. 2021. № 3. С. 90–101. DOI: 10.24412/2500-2872-2021-3-90-101

Malicious use of deepfakes: Risks for Japan's information and psychological security

Ya.V. Leksyutina

Abstract. The rapid development of artificial intelligence has spurred the emergence of a number of new technologies, including the deepfake technology, which allows synthesizing very realistic fake video and audio content. This technology, which appeared quite recently, in 2017, has rapidly advanced in just a few years: fake content has become more realistic, deepfake applications have become cheaper and more accessible, allowing users without special computer qualifications to synthesize fake videos in a matter of seconds. It has resulted in the exponential growth of deepfakes in cyberspace.

Opening up fundamentally new opportunities for a whole range of industries (advertising and media, entertainment and games, film industry, medicine, etc.), this technology can also be used by malevolent actors for criminal purposes, for information and psychological attacks on the population, and also for deliberately harming state-to-state relations. The risks of malicious use of deepfakes are as real as the benefits of using them. In a number of countries, deepfake technology is already seen as presenting a variety of challenges to national security and information and psychological security in the years to come.

This article provides a general overview of the scale and spheres of deepfake usage, the prospects and already registered cases of the malicious use of deepfake technology in the political sphere, and also assesses the risks of malicious use of this technology for Japan and the risk perception by Tokyo.

***Keywords*:** Japan, deepfakes, artificial intelligence (AI), information and psychological security, malicious use.

***Author*:** *Leksyutina Yana V.*, Doctor of Political Sciences, Professor of the Russian Academy of Sciences, Professor of the American Studies Department, Saint-Petersburg State University (7-9, Universitetskaya Emb., Saint-Petersburg, 199034, Russian Federation). ORCID: 0000-0001-6766-1792; E-mail: lexyana@ya.ru

***Conflict of interests*.** The author declares the absence of the conflict of interests.

***Acknowledgements*.** The research is accomplished under the Russian Foundation for Basic Research and the Vietnam Academy of Social Sciences, grant No. 21-514-92001

***For citation*:** Leksyutina Ya.V. (2021). Zlonamerennoye ispol'zovaniye dipfeykov: riski informatsionno-psikhologicheskoy bezopasnosti Yaponii [Malicious use of deepfakes: Risks for Japan's information and psychological security]. *Yaponskiye issledovaniya* [*Japanese Studies in Russia*], 2021, 3, 90–101. (In Russian). DOI: 10.24412/2500-2872-2021-3-90-101

В стратегиях промышленного и научно-технологического развития современных мировых держав одно из ключевых мест занимает задача развития искусственного интеллекта (ИИ). Широкие перспективы использования ИИ в реальном секторе экономики позволили говорить о начале «четвёртой промышленной революции», а возможности его применения в сфере оборонного строительства − об изменениях в представлениях о военной мощи государств и характере военных действий в будущем. Согласно одному из исследований, ожидается, что в результате использования ИИ мировой ВВП к 2030 г. может увеличиться на 14 % или 15,7 трлн долл. от уровня 2017 г., что делает ИИ крупнейшей коммерческой возможностью в быстро меняющейся современной экономике [Sizing the

prize…]. Завоевание превосходства в сфере ИИ рассматривается в качестве значимого фактора, способствующего повышению конкурентоспособности государства. Некоторые эксперты даже расценивают сферу ИИ как пространство уже протекающего геополитического соперничества держав [Artificial intelligence policies…].

Несколько отставая от США и Китая в сфере ИИ, Япония проводит активную политику, направленную на его развитие и внедрение в разные сферы общества. В 2016 г. японское правительство в 5-м Базовом научно-технологическом плане закрепило разработанную им концепцию «Общество 5.0», под которым подразумевалось «ориентированное на человека общество, которое согласует экономический прогресс с решением социальных проблем с помощью системы, тесно объединяющей киберпространство и физическое пространство»[1]. «Общество 5.0» стало национальной стратегией Японии по построению суперинтеллектуального общества на основе использования новейших технологий и в первую очередь ИИ. В рамках реализации концепции «Общества 5.0» в Японии был создан Стратегический совет по технологии ИИ, специально курирующий исследования, разработки и внедрение ИИ. В 2017 г. Стратегический совет разработал Стратегию развития технологии ИИ, содержащую план развития Японии в области НИОКР и индустриализации с целью развития промышленной экосистемы ИИ к 2030 г.

Возлагая большие надежды на развитие ИИ в деле стимулирования экономического роста и гармонизации общества, государства между тем осознают и серьёзные риски, сопряженные с его развитием. Современный уровень технологического развития позволяет с помощью ИИ создавать технологии, которые потенциально могут быть использованы злоумышленниками для нанесения вреда отдельным индивидам, компаниям, обществу, государствам и т.д. Такие новые технологии, как дипфейк (deepfake), умные боты (smart bots) или фишинговые атаки (phishing), способны создавать поддельный синтезированный видео- и аудиоконтент и ложные новости, влиять на общественно-политический дискурс, подрывать общественное доверие или даже открывать возможности для шантажа чиновников и дипломатов. В международных отношениях эти технологии могут стать эффективным информационным орудием дестабилизации социально-политической ситуации в государствах, а также нанесения вреда межгосударственным отношениям.

Данная статья призвана очертить масштабы и сферы использования дипфейков, их законодательное регулирование в мире и в Японии, раскрыть проблематику злонамеренного использования дипфейков в политических целях, а также оценить риски информационно-психологической безопасности Японии в контексте их злонамеренного использования.

Технология «дипфейк»: масштабы и сферы использования, законодательное регулирование

Достаточно серьёзные опасения в экспертном сообществе связаны с развитием технологии «дипфейк» как открывающей новые возможности для дезинформации, манипулирования медийным пространством и общественным мнением. Технология «дипфейк» состоит в создании с помощью ИИ очень реалистичного поддельного видео- или аудиоконтента. Созданные посредством машинного обучения поддельные видео могут

[1] Society 5.0. https://www8.cao.go.jp/cstp/english/society5_0/index.html (дата обращения: 02.08.2021).

производить реалистичное впечатление, будто человек говорит или делает что-то, чего на самом деле он никогда не говорил и не делал. Используя множество реальных примеров речи и движущихся изображений, обучается так называемая нейронная сеть, которая используется для создания дипфейков и обмана людей.

Сама по себе технология «дипфейк», с одной стороны, открывает обширные перспективы для целого ряда индустрий, в первую очередь работающих с изображениями: рекламы и медиа, индустрии развлечений и игр, виртуальных ассистентов, киноиндустрии[2], медицины (например, создание изображений МРТ для обучения медицинского персонала) и пр. С другой стороны, дипфейки как новая форма дезинформации становятся настоящей проблемой в коммуникативной среде из-за лёгкости распространения поддельного контента через социальные сети и онлайн-новости. Появление первого настоящего дипфейка в Интернете принято датировать сентябрем 2017 г., когда на сайте Reddit.com пользователь под ником DeepFakes опубликовал серию компьютерных видеороликов порнографического содержания с людьми, чьи лица были заменены на лица известных актрис. Это стало началом появления различного контента для взрослых с «участием» известных актрис. Более того, с момента своего создания в 2017 г. инструменты и алгоритмы, позволяющие изменять лица и звуки в аудиовизуальном контенте, развились до такой степени, что появились мобильные приложения и Интернет-сервисы, позволяющие обычным пользователям создавать всё более убедительный поддельный контент. Приложения по созданию дипфейков стали легкодоступными и простыми в использовании, и пользователи без специального компьютерного инженерного образования получили возможность генерировать поддельное видео за считанные секунды. На сегодняшний день существует большое количество подобного рода приложений, включающих, но не ограниченных приложениями ZAO, REFACE, FakeApp, FaceSwap, DFaker (развёрнутый список существующих приложений и их классификацию см. [Masood, Nawaz 2021, p. 5]).

Количество дипфейков в Интернете стремительно растёт. Согласно докладу компании Deeptrace, лишь за первые девять месяцев 2019 г. число обнаруженных в Интернете дипфейк-видео увеличилось почти вдвое с 7964 до 14678 [The state of deepfakes: landscape…], уже к декабрю 2020 г. достигнув 85 тысяч [The state of deepfakes: updates…]. При этом, согласно экспертным оценкам, подавляющее большинство обнаруженных дипфейков нацелены на США и Великобританию (50,1 и 10,9 % от всех обнаруженных дипфейков на середину 2020 г. соответственно), далее в порядке убывания следуют Республика Корея (9,6 %), Индия (5 %) и Япония (4 %) [Hofesmann 2020].

Технология «дипфейк» нашла широкое применение для производства порнографического контента. Так, на сентябрь 2019 г. 96 % размещенного в Интернете дипфейк-видео приходилось именно на подобный контент [The state of deepfakes: landscape…]. Согласно экспертным оценкам, чаще всего жертвами дипфейков порнографического характера становятся британские и американские актрисы, а также южнокорейские K-pop исполнители [The state of deepfakes: landscape…, p. 7]. В Японии подобный контент также получил широкое распространение. Так, осенью 2020 г. Столичная полиция Токио сообщила национальному телеканалу NHK, что с ними связались около

[2] Почувствуйте разницу // Информационные технологии. 31 марта 2021. № 55. С. 16. https://www.kommersant.ru/doc/4740464 (дата обращения: 14.08.2021).

200 знаменитостей женского пола, чтобы пожаловаться на дипфейки порнографического содержания[3].

В сентябре 2020 г. в Японии были произведены первые аресты граждан, подозреваемых в создании и распространении дипфейков порнографического содержания (с целью получения материальной выгоды) − Столичное управление полиции Токио и полиция префектуры Тиба арестовали двух мужчин по обвинению в клевете и в распространении в Интернете подделанных ими порнографических видео, где лица актрис на исходных видео были заменены лицами знаменитостей[4]. Уже в ноябре того же года ещё трое японцев были задержаны по схожим обвинениям. Примечательно, что в Японии, как и в большинстве стран мира, пока отсутствуют законодательные нормы, регулирующие технологию «дипфейк». Редким новатором в сфере разработки правового регулирования дипфейков является Китай, где в ноябре 2019 г. были обнародованы правила, регулирующие видео- и аудиоконтент в Интернете. В них, в частности, запрещается поставщикам и пользователям сетевых аудио- и видеоинформационных услуг публиковать и распространять ложную информацию или дипфейки в Интернете без чёткого обозначения того, что соответствующий контент был создан с использованием ИИ или виртуальной реальности[5]. В США, которые чрезвычайно обеспокоены исходящими от дипфейков вызовами (в частности, дипфейки там рассматриваются как способные представлять вызовы демократии и национальной безопасности), на федеральном уровне пока не было принято законодательных норм, регулирующих дипфейки, но в отдельных американских штатах подобная законодательная практика уже представлена. Так, в штате Вирджиния был принят закон, запрещающий дипфейки порнографического содержания, а в Техасе − запрещающий дипфейки, используемые для влияния на выборы [Nelson 2019]. В штате Калифорния в 2019 г. также были приняты две законотворческие нормы, запрещающие распространение политических дипфейков в ходе избирательных кампаний (AB730) и дипфейков порнографического содержания (AB602)[6].

В приведённых выше случаях задержания японских граждан при их арестах применялись законодательные нормы, связанные с клеветой: согласно японскому законодательству, обвинение в клевете может быть предъявлено тому, кто оскорбляет честь или наносит ущерб репутации другого человека. Его можно применять в случаях, когда жертва известна, поскольку это может повлиять на его или её общественное положение

[3] Ryall J. Celebrity deepfake porn cases in Japan to rise in sex-related cybercrime // South China Morning Post. November 20, 2020. https://www.scmp.com/news/china/diplomacy/article/3145113/chinese-actor-zhang-zhehan-faces-domestic-boycott-over-2018?module=perpetual_scroll&pgtype=article&campaign=3145113 (дата обращения: 14.08.2021).

[4] Two men arrested over deepfake pornography videos // The Japan Times. October 2, 2020. https://www.japantimes.co.jp/news/2020/10/02/national/crime-legal/two-men-arrested-deepfake-pornography-videos/ (дата обращения: 12.08.2021).

[5] Ванло инь шипинь синьси фуу гуаньли гуйдин [Положения об администрировании сетевых аудио- и видеоинформационных служб]. Ст. 10–12. http://www.cac.gov.cn/2019-11/29/c_1576561820967678.htm (дата обращения: 18.03.2021). (На кит.).

[6] Halm K.C., Kumar A., et al. Two new California laws tackle deepfake videos in politics and porn // DWT LLP. November 10, 2019. https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2019/10/california-deepfakes-law (дата обращения: 11.08.2021).

и карьеру, но это также применимо к обычным гражданам на тех же основаниях, что и нанесение вреда репутации[7].

Вместе с тем, несмотря на широкое использование дипфейков в порнографическом контенте, большинство дипфейков, представленных в настоящее время на открытых социальных платформах, таких как YouTube, Facebook или Twitter, носят всё же безобидный, развлекательный характер [Masood, Nawaz 2021, p. 4], что, однако, не исключает риски злонамеренного использования дипфейков злоумышленниками. Созданные при помощи технологии «дипфейк» ложно-негативные аудио и видео представляют собой новую форму кибератак, способных нанести серьёзный ущерб отдельным индивидам, компаниям, обществу, государству.

Злонамеренное использование дипфейков в политических целях

В зависимости от преследуемых целей злонамеренное использование дипфейков можно условно подразделить на три уровня. К первому уровню относится хулиганство, при котором у злоумышленников отсутствует намерение получения выгоды, а целью является либо просто развлечение, либо демонстрация целевой аудитории своих возможностей (например, нанесение вреда репутации звёздам киноиндустрии). Второй уровень уже подразумевает преследование злоумышленниками цели извлечения материальной или иной выгоды (например, финансовое мошенничество, социальный инжиниринг, шантаж, распространение платного порнографического контента, нанесение репутационного вреда конкурентам, получение доступа к личной информации). Широкую известность получил случай использования преступниками голосового дипфейка в 2019 г., когда управляющий директор британской энергетической компании, получив звонок от якобы главы материнской компании в Германии (была синтезирована речь главы компании), перевёл преступникам 243 тыс. долл. [The state of deepfakes: landscape…, p. 14].

И, наконец, третий уровень злонамеренного использования дипфейков − это их применение в политической сфере с целью нанесения вреда политическим деятелям и партиям (например, во время избирательных компаний), государству (например, посредством дестабилизации политической или социально-экономической ситуации) и межгосударственным отношениям. В руках злоумышленников дипфейки могут превращаться в эффективное орудие информационно-психологического воздействия на население. Так, дезинформация, переданная с помощью дипфейков, может стать проблемой во время выборов, поскольку любой политический деятель может попытаться дискредитировать оппонента или спровоцировать политический скандал с целью продвижения своих собственных интересов. Как следствие, после воздействия дипфейка граждане могут изменить своё отношение как к соответствующему политику, подвергнувшемуся информационной атаке посредством дипфейка, так и к партии политика в целом.

[7] Ryall J. Celebrity deepfake porn cases in Japan to rise in sex-related cybercrime // South China Morning Post. November 20, 2020. https://www.scmp.com/news/china/diplomacy/article/3145113/chinese-actor-zhang-zhehan-faces-domestic-boycott-over-2018?module=perpetual_scroll&pgtype=article&campaign=3145113 (дата обращения: 14.08.2021).

Технология «дипфейк» может также потенциально быть использована в политических целях для дискредитации национальных лидеров, распространения дезинформации и манипулирования общественным мнением, подстрекательств к актам насилия в отношении нацменьшинств, распространения идеологий экстремистских или террористических групп, разжигания социальных волнений и политической поляризации общества и пр. Созданные с использованием технологии «дипфейк» фальшивые видеоролики о жёстоком обращении правительства с гражданами могут вызвать возмущение общественности, угрожая стабильности политической системы. Эксперты обращают внимание на то, что, например, использование технологии «дипфейк» для имитации голоса высокопоставленных правительственных чиновников или военных может иметь серьёзные последствия для национальной безопасности [Masood, Nawaz 2021, p. 5−6]. Передавая фальшивую информацию, дипфейки могут использоваться для обмана военных аналитиков, например, показывая фальшивый мост через реку, вводя в заблуждение при развертывании войск. Дипфейки также могут быть использованы для шантажа высокопоставленных должностных лиц или лиц, имеющих доступ к секретной информации [Sayler 2021]. Уже есть свидетельства того, что оперативники внешней разведки использовали дипфейк-фотографии для создания поддельных учётных записей в социальных сетях, где они предпринимали попытки вербовки [Sayler 2021].

Известные к настоящему моменту случаи дипфейков с политическими деятелями преимущественно носили характер политической сатиры. Например, это появившийся в 2018 г. дипфейк с 44-м президентом США Б. Обамой, в котором он якобы оскорбляет 45-го президента Д. Трампа, или два дипфейк-видео 2020 г., на которых северокорейский лидер Ким Чен Ын и президент РФ В. Путин якобы высказываются по поводу выборов в США и уязвимости демократии[8].

Однако имели место и случаи намеренного использования дипфейков или манипулирования проблематикой дипфейков в политических целях и оказания информационно-психологического воздействия на население. Например, в 2018 г. одна из политических партий Бельгии создала фальшивое видео выступления Д. Трампа, в котором он призывал Бельгию последовать примеру США и выйти из Парижского климатического соглашения. По замыслу его создателей, дипфейк был призван обратить внимание населения на необходимость принятия мер в связи с изменением климата и убедить людей подписать петицию, призывающую правительство принять меры против изменения климата[9].

Возможности манипулирования проблематикой дипфейков были продемонстрированы в конце 2018 − начале 2019 гг. в ходе политических событий в Габоне. Тогда, на фоне слухов о состоянии здоровья президента Габона Али Бонго, который продолжительное время не появлялся на публике, правительство выпустило видео, на котором Бонго произносил традиционное новогоднее обращение. Необычное поведение Бонго на видео побудило многих в социальных сетях, в том числе его политических оппонентов, заявить, что видео

[8] *Hao K.* Deepfake Putin is here to warn Americans about their self-inflicted doom // MIT Technology Review. September 29, 2020. https://www.technologyreview.com/2020/09/29/1009098/ai-deepfake-putin-kim-jong-un-us-election/ (дата обращения: 12.08.2021).

[9] Следует отметить, однако, что в конце видео было обозначено, что это дипфейк. См.: Yans von der Burchard. Belgian socialist party circulates ‘deep fake’ Donald Trump video. May 21, 2018. https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/ (дата обращения: 10.08.2021).

было дипфейком и правительство скрывает нездоровье или смерть президента. Через неделю после публикации видео была предпринята неудачная попытка государственного переворота. Последующая экспертиза не обнаружила признаков подтасовки видеоролика, и Бонго после этого неоднократно появлялся на публике.

В июне 2019 г. в Малайзии разразился настоящий политический скандал в связи с видео, на котором тогдашний министр экономики (2018–2020 гг.), а ныне министр международной торговли и промышленности (с 2020 г.) этой страны Мохамед Азмин Али занимался сексом с однополым партнёром, что само по себе является нарушением закона в Малайзии. По утверждению политика, видео являлось дипфейком, но специалисты не смогли однозначно определить, было ли видео подделкой.

Схожий инцидент имел место в Японии, но с менее счастливым концом для его фигуранта. Весной 2018 г. Генеральный секретарь Министерства финансов Японии Фукуда Дзюнъити был обвинен в сексуальных домогательствах к женщине-репортёру во время интервью. В качестве доказательства вины высокопоставленного чиновника была предоставлена аудиозапись, подлинность которой Фукуда отрицал [Rini 2020]. В конечном счёте, обвинённый в домогательствах японский политик был вынужден оставить свой пост.

Инциденты с Фукуда Дзюнъити и Мохамедом Азмин Али отражают проблематику так называемого эффекта дивидендов лжеца, введённого в научный оборот Д.К. Ситроном и Р. Чесни и означающего ситуацию, когда люди утверждают, что контент является дипфейком и на этом основании отрицают подлинность соответствующего аутентичного контента, изображающего неподобающее или преступное поведение [Sayler 2021].

Злонамеренное использование дипфейков в международных отношениях: риски для Японии

Тот факт, что вплоть до текущего момента технология «дипфейк» крайне редко злонамеренно использовалась с преступными или политическими целями [Ciancaglini, Gibson 2020, p. 59], а применялась преимущественно в развлекательных целях, не исключает потенциала её злонамеренного использования по мере совершенствования этой технологии. Для антисоциальных негосударственных акторов, таких как террористические организации, преступные группы, оппозиционные политические силы, корпоративные группы интересов или секты, технология «дипфейк» может оказаться мощным, недорогим и легкодоступным инструментом оказания информационно-психологического давления на целевую аудиторию.

Злонамеренное использование дипфейков может проявиться и в международных отношениях. Например, заинтересованные страны способны при помощи ИИ создавать ложные информационные поводы, фабриковать «доказательства», оправдывающие их вмешательство во внутренние дела других государств (по аналогии с фабрикацией «данных», сделавших возможной операцию в Ираке) [Лун Кунь, Ма Юэ 2019, с. 26]. Недружественные государства могут также, создавая ложно-негативный контент, разжигать внутренние противоречия в государствах, дискредитировать национальных лидеров, правящие политические партии, инспирировать «цветные революции» и пр.

Не исключена ситуация, когда недружественное государство или антисоциальные акторы вознамерятся имитировать информационные поводы для нанесения вреда межгосударственным отношениям третьих стран, способствовать разрушению взаимного

доверия между ними или даже провоцировать межгосударственные конфликты. Например, злонамеренные информационные «вбросы» по «чувствительным вопросам» японско-китайских отношений (по вопросам принадлежности островов Сэнкаку или касающиеся исторических событий 1931–1945 гг.), японско-южнокорейских (по вопросам исторического наследия периода колониального правления Японии на Корейском полуострове в 1910–1945 гг.), японско-северокорейских (по вопросам похищения Пхеньяном японских граждан в конце 1970-х − начале 1980-х гг.) или японско-российских (по вопросам принадлежности Курильских островов) отношений способны оказывать дестабилизирующий эффект на социально-политическую обстановку в соответствующих странах и на их межгосударственные отношения.

Исключительно гипотетически, для Японии подобного рода риски, связанные с использованием дипфейков в целях оказания информационно-психологического воздействия, могут проистекать как от недружественных государств (сейчас в Японии в этом контексте главным образом рассматривают Китай и Северную Корею), так и со стороны, например, религиозных сект, террористических и экстремистских сил, преступных групп и пр.

В контексте обеспечения национальной безопасности в Японии сейчас огромное значение придаётся рискам, вызовам и угрозам, связанным с технологическим прогрессом и инновациями и, в особенности, с появлением и развитием таких технологий, способных кардинально изменить характер боевых действий в будущем, как искусственный интеллект, гиперзвуковые технологии и пр.[10]. В связи с этим, наибольшую обеспокоенность вызывает Китай, занимающий ныне первое место в списке угроз безопасности Японии [Лексютина, Михалевич 2021, с. 78] и, по японским оценкам, придающий большое значение укреплению своих операционных возможностей, направленных на обретение информационного превосходства и расширение возможностей в космосе, киберпространстве и электромагнитном спектре[11].

Второе место в списке угроз безопасности Японии − после Китая − ныне занимает Северная Корея, чьё военное развитие и ракетно-ядерная программа оцениваются как представляющие «серьёзную и неминуемую угрозу безопасности Японии»[12]. Помимо развития ракетно-ядерного потенциала и общего повышения операционных возможностей северокорейских вооруженных сил, особую озабоченность Токио выражает в связи с тем, что Северная Корея целенаправленно развивает киберсилы в русле укрепления своего асимметричного военного потенциала, занимается кражей военных секретов и развивает возможности для атак на критическую инфраструктуру зарубежных стран[13].

В докладе «Оборона Японии в 2020 г.» особо выделяются Китай и Россия как страны, которые усиливают наступательные кибервозможности своих вооруженных сил с целью

[10] 2020 Defense of Japan. Pamphlet. P. 1. https://www.mod.go.jp/en/publ/w_paper/wp2020/DOJ2020_Digest_EN.pdf (дата обращения: 14.08.2021).

[11] 2020 Defense of Japan. Pamphlet. P. 3. https://www.mod.go.jp/en/publ/w_paper/wp2020/DOJ2020_Digest_EN.pdf (дата обращения: 14.08.2021).

[12] National defense program guidelines for FY 2019 and beyond. December 18, 2018. https://warp.da.ndl.go.jp/info:ndljp/pid/11591426/www.mod.go.jp/j/approach/agenda/guideline/2019/pdf/20181218_e.pdf (дата обращения: 14.08.2021).

[13] National defense program guidelines for FY 2019 and beyond. December 18, 2018. https://warp.da.ndl.go.jp/info:ndljp/pid/11591426/www.mod.go.jp/j/approach/agenda/guideline/2019/pdf/20181218_e.pdf (дата обращения: 14.08.2021).

нанесения вреда силам противников и разрушения их инфраструктуры[14]. В основном озабоченность Японии связывается с кибератаками против информационно-коммуникационных сетей правительственных организаций, военных ведомств, корпораций и академических институтов, а также с атаками, преследующими цель кражи критических технологий, секретов или личной информации[15].

В целях усиления способностей противостоять вызовам и угрозам в киберсфере в Японии придают приоритетное значение усилению специальной Группы кибербезопасности в составе Сил самообороны (в т.ч. посредством увеличения персонала и финансирования), улучшению компьютерных сетей и систем, развитию человеческих талантов в киберсфере, использованию передовых кибертехнологий[16]. Также Япония уделяет внимание участию в многосторонних формах сотрудничества по вопросам обеспечения кибербезопасности. В частности, в сентябре 2020 г. Япония присоединилась к инициированному Вашингтоном многостороннему форуму «Партнёрство в области ИИ для обороны», включающему также Австралию, Канаду, Данию, Эстонию, Финляндию, Францию, Израиль, Норвегию, Южную Корею, Швецию и Великобританию. Этот форум направлен на «обеспечение основанного на ценностях глобального лидерства» в вопросах внедрения ИИ в поддержание безопасности и на согласование действий «в целях содействия ответственному использованию ИИ, продвижения общих интересов и лучших практик по внедрению этики ИИ, установления рамок для облегчения сотрудничества, координации стратегических выступлений о политике в области ИИ»[17].

Заключение

С момента появления в Интернете первого дипфейка в 2017 г., эксперты фиксируют стремительный рост применения технологии «дипфейк» не только в развлекательных целях, но и в менее безобидных формах. Злоумышленники стали широко использовать эту технологию в преступных целях для извлечения материальной выгоды (прежде всего, для создания порнографического контента и атак с использованием социальной инженерии) и в редких случаях − даже для оказания информационно-психологического воздействия на население. В отличие от США и Великобритании, на которые приходится бóльшая часть дипфейк-атак, для Японии проблема злонамеренного использования дипфейков пока не носит характер массового явления, хотя случаи ареста японских граждан, подозреваемых в создании и распространении дипфейков порнографического содержания, уже имели место. В политической сфере манипулирование проблематикой дипфейков в Японии имеет пока единичный характер, что вероятно объясняет недостаточное внимание японских властей к данной проблематике. В Японии отсутствуют законодательные нормы, регулирующие использование технологии «дипфейк», а проблематика дипфейков слабо представлена в японском экспертном и общественно-политическом дискурсе. Несмотря на наличие рисков

[14] 2020 Defense of Japan. Pamphlet. P. 8. https://www.mod.go.jp/en/publ/w_paper/wp2020/DOJ2020_Digest_EN.pdf (дата обращения: 14.08.2021).

[15] 2020 Defense of Japan. Pamphlet. P. 8. https://www.mod.go.jp/en/publ/w_paper/wp2020/DOJ2020_Digest_EN.pdf (дата обращения: 14.08.2021).

[16] 2020 Defense of Japan. Pamphlet. P. 12, 14. https://www.mod.go.jp/en/publ/w_paper/wp2020/DOJ2020_Digest_EN.pdf (дата обращения: 14.08.2021).

[17] AI partnership for defense joint statement. September 16, 2020. https://www.ai.mil/docs/AI_PfD_Joint_Statement_09_16_20.pdf (дата обращения: 10.08.2021).

злонамеренного использования дипфейков в политических целях, в Японии в русле обеспечения национальной безопасности, информационно-психологической безопасности и защиты демократии эта проблематика не является приоритетной. Придавая большое значение рискам, вызовам и угрозам национальной безопасности, связанным с развитием искусственного интеллекта, Япония сфокусирована на противодействии кибератакам и кибершпионажу, а также на мониторинге кибервозможностей своих потенциальных оппонентов, мало внимания уделяя рискам, сопряжённым с технологией «дипфейк». Противодействие злонамеренному использованию дипфейков в Японии преимущественно идет по линии борьбы с порнографическим контентом, распространяемым в Интернете.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Лексютина Я.В., Михалевич Е.А., Нго Хыонг Лан. Характер межгосударственных отношений, межгосударственные противоречия и вызовы информационно-психологической безопасности в Северо-Восточной Азии. *Россия и Китай: история и перспективы сотрудничества: материалы XI международной научно-практической конференции (Благовещенск, 11–12 мая 2021 г.)* / отв. ред. Д.В. Кузнецов. Вып. 11. Часть 4. 2021. Благовещенск: Изд-во БГПУ. С. 76−83.

REFERENCES

Leksyutina, Ya., Mikhalevich, E., Ngo Houng Lan. (2021). Kharakter mezhgosudarstvennykh otnoshenii, mezhgosudarstvennye protivorechiya i vyzovy informatsionno-psikhologicheskoi bezopasnosti v Severo-Vostochnoi Azii [Inter-state relations, major contradictions, and challenges to the information and psychological security in Northeast Asia]. In D. Kuznetsov (ed.), *Rossiya i Kitaya: istoriya i perspektivy sotrudnichestva: materialy XI mezhdunarodnoi nauchno-prakticheskoi konferentsii (Blagoveshchensk, 11-12 maya 2021 g*.) [Russia and China: history and prospects of cooperation: proceedings of the XI scientific and practical conference (Blagoveshchensk, May 11-12, 2021)] (p. 76−83). Blagoveshchensk: Izd-vo BGPU. (In Russian).

* * *

Long Kun, Ma Yue. (2019). Shēndù wèizào duì guójiā ānquán de tiǎozhàn jí yìngduì [The challenges and responses of deepfake to national security], *Xìnxī ānquán yǔ tōngxìn bǎomì* [Information Security and Communication Confidentiality], 10, 21−34. (In Chinese).

Artificial intelligence policies in East Asia: an overview from the Canadian perspective. 2019. Retrieved August 10, 2021, from https://www.asiapacific.ca/sites/default/files/filefield/ai_report_2019.pdf

Ciancaglini, V., Gibson, Cr., Sancho, D., et. al. (2020). *Malicious uses and abuses of Artificial intelligence: Europol public information*. Retrieved May 10, 2021, from https://documents.trendmicro.com/assets/white_papers/wp-malicious-uses-and-abuses-of-artificial-intelligence.pdf

Hofesmann, E. (November 19, 2020). *The state of deepfakes in 2020*. Retrieved August 11, 2021, from https://www.skynettoday.com/overviews/state-of-deepfakes-2020

Masood, M., Nawaz, M., et al. (2021). *Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward*. Retrieved August 15, 2021, from https://arxiv.org/ftp/arxiv/papers/2103/2103.00484.pdf

Nelson, A. (October 23, 2019). *Trust your eyes? Deepfakes policy brief*. Retrieved August 13, 2021, from https://www.csis.org/analysis/trust-your-eyes-deepfakes-policy-brief

Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosopher's Imprint*, 20(24), 1−16. Retrieved August 15, 2021, from https://philpapers.org/archive/RINDAT.pdf

Sayler, K. (June 8, 2021). *Deep fakes and national security*. CRS report. Retrieved August 15, 2021, from https://crsreports.congress.gov/product/pdf/IF/IF11333

Sizing the prize. (2017). What's the real value of AI for your business and how can you capitalise? *PwC report*. Retrieved August 10, 2021, from https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

The state of deepfakes: landscape, threats, and impact. (2019). Retrieved August 12, 2021, from https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

The state of deepfakes: updates on statistics and trends. (2021). Retrieved August 15, 2021, from https://sensity.ai/reports/

Поступила в редакцию 16.08.2021 Received 16 August 2021