

КРАСОВСКАЯ Наталья Рудольфовна – кандидат психологических наук, доцент кафедры социальной психологии и виктимологии Новосибирского государственного педагогического университета (630126, Россия, г. Новосибирск, Вилюйская ул., 28; goulina@gmail.com)

ГУЛЯЕВ Андрей Анатольевич – кандидат философских наук, доцент Глазовского государственного педагогического института им. В.Г. Короленко (427621, Россия, Республика Удмуртия, г. Глазов, ул. Первомайская, 25; andrey.gulyaev1966@yandex.ru)

ТЕХНОЛОГИИ МАНИПУЛЯЦИИ СОЗНАНИЕМ ПРИ ИСПОЛЬЗОВАНИИ ДИПФЕЙКОВ КАК ИНСТРУМЕНТА ИНФОРМАЦИОННОЙ ВОЙНЫ В ПОЛИТИЧЕСКОЙ СФЕРЕ

Аннотация. Политическое воздействие дипфейков связано с применением определенных психологических инструментов. Основой воздействия дипфейков выступают технологии манипулирования сознанием и поведением людей. Методологическим инструментом данной работы выступил метод системного анализа. Различные манипулятивные техники при использовании дипфейков представляют собой систему, в них изображение и видео играют ключевую роль, а главной их целью является спровоцировать людей на такое поведение, которое является выгодным для манипуляторов. Дипфейк создается с использованием искусственного интеллекта и генеративно-состязательных нейросетей. Ключевым каналом распространения дипфейков являются социальные сети и электронные СМИ.

Ключевые слова: дипфейк, искусственный интеллект, генеративно-состязательные нейросети, психологические инструменты, манипулятивные техники, технология «подталкивания»

Дипфейком называют совокупность технологических трансформаций изображения и видео, созданных с использованием искусственного интеллекта. Это слово появилось в результате синтеза двух понятий (*deep learning* – глубинное изучение; *fake* – подделка). Для создания дипфейка используются генеративно-состязательные нейросети, действующие на основе искусственного интеллекта (*GAN*). Данная технология используется для объединения и наложения изображений и видео.

Дипфейк представляет собой сфабрикованное видео, созданное «с нуля» либо базирующееся на реально существующих материалах, цель которого – воссоздать внешность и голос реального человека, выполняющего действия или говорящего вещи, которые на самом деле им никогда не произносились и не выполнялись. Дипфейки первоначально использовались в порнографии, когда в эротические ленты низкого качества вставлялись лица известных людей либо конкретного объекта. Но время идет, технология развивается, и никто не может дать гарантии, что в ближайшем будущем в дипфейках мы не увидим, например, действующего президента какой-либо страны или другую политическую фигуру, говорящих и делающих то, что захотел автор дипфейка [Ершов 2018].

Создание дипфейка связано прежде всего с работой с изображениями. Части или элементы изображения заменяются на желаемые образы. В этом процесс создания дипфейка подобен редактированию изображений и видео. Но есть главное и существенное отличие – при создании дипфейка для искажения реальности используют искусственный интеллект. Особенностью генеративно-состязательных сетей является использование авторами дипфейка двух алгоритмов. Одна часть обучается на реальных изображениях объекта и создает все более точные копии изображения этого объекта до тех пор, пока вторая часть алгоритма не начнет путать копию с оригиналом. При этом первая часть алго-

ритма вступает в состязание со второй частью и в итоге должна ее победить. Если искусственный интеллект не видит разницы между оригиналом и копией, то, как принято считать, люди тем более не смогут их отличить. Конечно, качество дипфейка во многом зависит от уровня развития искусственного интеллекта. Появление политических дипфейков может иметь самые серьезные последствия для политических деятелей и партий, органов управления и самого государства. В рекламе же, в маркетинге, на телевидении и в киноиндустрии использование дипфейков может, наоборот, принести огромный успех и прибыль данным отраслям [Кудрина 2017].

Большая часть дипфейков представляет собой порноролики (свыше 90%) с использованием лиц известных актеров и т.д. С каждым годом все большее место среди дипфейков занимает политическая продукция. Тревогу по этому поводу начали бить в 2017 г., отмечая, что политические дипфейки могут повлиять на исход выборов, вызвать панические настроения в стране, исказить в позитивном или негативном смысле имидж политика или политической партии и др. [Третьяков и др. 2018].

Основой действия политических дипфейков являются техники манипулирования людьми. В этих техниках в качестве субъекта могут использоваться копии изображений политических деятелей (например, президент США Д. Трамп или спикер палаты представителей США Н. Пелоси), а в качестве объекта – масса зрителей по всему миру. В продвижении дипфейков важнейшую роль играют электронные СМИ, социальные сети (например, Фейсбук), видеохостинги (например, *You Tube*) [Ершов 2018].

Копии изображения политического деятеля или видео оказывают психологическое воздействие на зрителя. Это воздействие будет тем больше, чем меньше зритель сомневается в подлинности изображения или видео. Развитие информационных технологий и Интернета, программных средств обработки и использования «больших данных», увеличение быстродействия и мощности суперкомпьютеров, опережающие разработки «искусственного интеллекта» ставят на повестку дня проблему политического воздействия дипфейков на массовую аудиторию, посредством которых обходится языковой барьер и при этом используются различные психологические инструменты.

В психологическом инструментарии воздействия на индивидуальное и массовое сознание одну из главных ролей играют техники манипулирования, лежащие в основе концепции «подталкивания» человека к нужному для манипулятора поведению. Эти техники создают у человека ложную иллюзию свободы и самостоятельности в принятии решений. Принятие решений и поведение манипулируемого или толпы манипулируемых будет выгодно только манипулятору [Ковалев, Матвиенко 2016]. К таким техникам манипулирования бихевиористского характера можно отнести использование социальных норм, раскрытие информации, заранее согласованную стратегию, напоминания, заранее высказанное намерение, информирование людей о последствиях их предыдущих выборов, упрощение, стереотипизацию, сенсационность, эффект «плохих» изображений и видео [Санстайн 2016].

Такие техники манипулирования, как информирование людей о последствиях их предыдущих выборов, упрощение, повторение, утверждение, стереотипизация, напоминания, сенсационность, эффект «плохих» изображений и видео, представляют собой систему, состоящую из элементов, каждый из которых является структурной единицей, соединенной с другими единицами многочисленными связями в определенной последовательности.

Упрощение является фундаментальной техникой манипулирования человеком. Массовое сознание активно реагирует на достаточно простые изобра-

жения и видео. В социальных сетях и видеохостингах (например, в *Facebook* или *YouTube*) больше всего лайков собирают достаточно примитивные, но эмоционально насыщенные (чаще всего с негативной коннотацией) изображения и видео с обильным применением просторечий и жаргонизмов [Кугушева 2016].

В целях манипулирования для лучшего понимания и запоминания месседжа содержание ролика значительно упрощают, примитивизируют. Затем происходит его семантизация, т.е. находят несколько слов (клише, характерные понятия), которыми описывается этот упрощенный по содержанию месседж. «Сообщение должно иметь уровень понятности, соответствующий коэффициенту интеллектуальности примерно на десять пунктов ниже среднего коэффициента того социального слоя, на который рассчитано сообщение» [Шейнов 2010]. С точки зрения психологии это правило объясняется склонностью человека на подсознательном уровне к простым объяснениям сложных проблем. При процедуре упрощения сообщения коммуникатор должен иметь представление о том, какова его целевая аудитория и насколько высок ее интеллектуальный уровень.

Для целей усиления манипулятивного эффекта упрощение сочетается с другими техниками манипуляции сознанием. Упрощение должно быть в форме напоминания. Сообщение в изображении и видео должно внушать некую мысль аудитории в краткой, энергичной и впечатляющей форме – в форме утверждения (как приказ – без возможности возражения). Утверждение в месседже означает отказ от обсуждения, поскольку власть человека или идеи, которая может подвергаться обсуждению, теряет всякое правдоподобие. Это означает также просьбу к аудитории, к толпе принять идею без обсуждения такой, какая она есть, – без взвешивания «за» и «против», и отвечать «да», не раздумывая [Московичи 2018].

Вслед за утверждением следует повторение месседжа. Повторение изображения и видео постепенно внедряется в глубины подсознания, где зарождаются мотивы наших действий. Повторение дает импульс утверждению и превращает его в навязчивую идею. Воспринимая упрощенные утверждения раз за разом, в различных версиях и по самому разному поводу, в конце концов объекты воздействия начинают проникаться ими. Повторение придает сообщению осязаемость и очевидность, которые заставляют принять их целиком. Повторение становится барьером против отличающихся, а тем более противоположных мнений. Таким образом, оно сводит к минимуму рассуждения и быстро превращает мысль в действие, на которое у массы уже сформировался условный рефлекс. С помощью повторения мысль отделяется от своего автора. Она превращается в очевидность, не зависящую от времени, места, личности. Она не является более выражением человека, который говорит, но становится выражением предмета, о котором он говорит. Повторение является одним из главных инструментов рекламы и пропаганды.

К числу фундаментальных техник манипулирования сознанием относится и стереотипизация. Стереотипизация как техника манипулирования основана на процессе человеческого восприятия как цепочки рефлексов. Восприятие человеком нового является подгонкой этого нового явления под стереотип – заранее сформированную человеком мыслительную оценку чего-либо, которая может отражаться в соответствующем поведении. Человек должен воспринимать изображение и видео без усилий и безоговорочно, без внутренней борьбы и критического анализа.

Другими средствами манипуляций сознанием является срочность, сенсационность и эффект «плохих» изображений и видео. Эти три, по сути, отдельные

манипулятивные техники объединяет одно общее воздействие на человека – помимо общего внушения, он оказывается под мощным воздействием ситуативного внушения. Ситуативное внушение обостряется в момент, когда для человека рушится привычная социальная ситуация, даже если это разрушение является чисто виртуальным или сильно преувеличенным реальным. В этот момент, когда срочные (человек не успевает их осмыслить и проверить), сенсационные (эмоционально оглушают, вызывают либо паралич, либо импульсивные действия), «плохие» изображения и видео (человек под воздействием ситуативного внушения способен на неадекватные действия, легко подчиняется паническому настроению, подпадает под власть толпы и стихийного коллективного безответственного поведения) атакуют общество, немалое число людей становятся орудиями манипуляторов. В эти моменты критическое восприятие у человека является пониженным, внушаемость – повышенной [Кара-Мурза 2015]. Для формирования такого эффекта необходимо создать такую ситуацию, при которой за относительно короткий период времени на общество выплескивается поток «плохих» изображений и видео о реальных или вымышленных событиях. Здесь также стоит учитывать феномен лавины, когда ком «плохих» изображений и видео имеет тенденцию к моментальному субъективному «нарастанию» в силу особенностей человеческой психики и восприятия (повышенный интерес к происшествиям и подсознательное ожидание их).

После техники манипулирования с использованием эффекта «плохих» изображений и видео коммуникаторы реализуют еще одну технику манипуляции индивидуальным и массовым сознанием, которая связана с информированием людей о последствиях их предыдущих выборов. Чтобы направить волю людей в нужном для манипулятора направлении, необходимо информировать и напоминать им о том, что в недавнем прошлом или даже накануне они сделали «неправильный», «плохой» выбор (проголосовали не за того кандидата, не вышли на митинг протеста и т.д.).

В результате последовательная система техник манипулирования сознанием составляет ядро технологии «подталкивания» человека к нужному для манипулятора поведению.

На основании изложенного выше можно сделать следующие выводы.

Масштабы лжи каждый год практически удваиваются. В 2019 г. только с апреля по сентябрь Фейсбук удалил 3,2 млрд фальшивых аккаунтов. В 2018 г. за аналогичный период было удалено 1,55 млрд аккаунтов. Также интенсивно растет сегмент клонирования голоса. По оценкам *Markets&Markets*, в ближайшие 2 года этот рынок достигнет размеров в 1,8 млрд долл., демонстрируя в настоящее время рост более, чем на 30% в год¹. Правительство Китая с 1 января 2020 г. ввело уголовное наказание за изготовление и распространение дипфейков [Третьяков и др. 2018]. В России также существует озабоченность проблемой защиты от дипфейков. Компании *Screenlife Technologies* и «Стафори» разрабатывают цифровую платформу *Vera Voice*, которая в т.ч. будет применяться для защиты прав обладателей голоса от *deepfake* и отслеживать появление голосовых клонов, определяя факт синтеза голоса и его оригинальность².

Как отмечают эксперты США, на законодательном уровне можно с бороться с авторами и распространителями дипфейков внутри государства, но практи-

¹ Рем Е. 2019. Vera Voice позволит «говорить» голосом любой знаменитости. Доступ: <https://techfusion.ru/vera-voice-pozvolit-govorit-golosom-lyuboj-znamenitosti/> (проверено 10.07.2020).

² Там же.

чески неэффективно бороться с иностранными изготовителями данной продукции [Westerlund 2019]. И эта задача требует дальнейшего поиска решений.

В борьбе с дипфейками необходимо объединить усилия государства, общества и производителей технологий. Для этого надо заключить с этой триадой нечто вроде общественного договора. Те, кто продумывает технологии и затем их внедряет, являются серьезной силой, и с ними нужно договариваться. Необходима такая корпоративная политика, которая препятствовала бы созданию и распространению дипфейков [Westerlund 2019].

У большинства людей нет осознания ответственности за распространение изображений и видео. Поэтому ключевой вопрос при этом, в т.ч. в сфере законодательства, – воспитание информационной культуры. Причем воспитание строгое и массовое: с позиций того, что любые распространяемые слова и изображения должны соотноситься с ответственностью за то содержание, которое пытаешься распространить. Чем меньше наши знания о чем-либо, тем больше вероятность купиться на дипфейк. Поэтому к самой большой группе риска относятся дети [Ершов 2018]. Необходимо принять активные меры в сфере образования, касающиеся распространения знаний среди всех учащихся по поводу существования дипфейков [Westerlund 2019].

Использование дипфейков основано на техниках манипулирования поведением человека. Сами техники манипулирования образуют систему, состоящую из элементов, которые по отдельности и в определенном сочетании традиционно применяются в пропаганде, рекламе и т.д. Использование дипфейков имеет принципиальную новизну в том, что эти элементы используются комплексно с системой техник манипулирования на основе применения искусственного интеллекта, генеративно-состязательных нейросетей (*GAN*) и их воздействие направлено на большие массы людей. Каналами распространения подобных изображений и видео являются социальные сети и интернет-медиаресурсы. Главной и наиболее сложной задачей является создание технологий распознавания политических дипфейков. Здесь требуется активное участие всех инновационных компаний, работающих в деле повышения эффективности применения искусственного интеллекта, которое станет препятствием для изготовления и распространения политических дипфейков тех сил, которые находятся вне договорных отношений или нарушают их [Westerlund 2019].

Список литературы

Ершов Ю.М. 2018. Феномен фейка в контексте коммуникационных практик. – *Вестник Томского государственного университета. Филология*. № 52. С. 62-74.

Кара-Мурза С.Г. 2015. *Манипуляция сознанием. Век 21*. М.: Алгоритм. 525 с.

Ковалев В.И., Матвиенко Ю.А. 2016. «Поведенческая» война как одна из перспективных невоенных угроз безопасности России. – *Информационные войны*. Т. 40. № 4. С. 20-24.

Кугушева А. 2016. От информационных войн к поведенческим. – *Информационные войны*. Т. 37. № 1. С. 11-23.

Кудрина Л.В. 2017. Актуальные заимствования в языке СМИ на примере лексем «фейк». – *ART LOGOS*. № 1. С. 215-224.

Московичи С. 2018. *Век толл*. Минск: Харвест. 432 с.

Санстайн К. 2016. *Иллюзия выбора. Кто принимает решения за нас, и почему это не всегда плохо*. М.: Альпина. 416 с.

Третьяков А.О., Филатова О.Г., Жук Д.В., Горпушина Н.Н., Пучковская А.А. 2018. Метод определения русскоязычных фейковых новостей с использованием

искусственного интеллекта. — *International Journal of Open Information Technologies*. Т. 6. № 12. С. 99-105.

Шейнов В.П. 2010. *Манипулирование сознанием*. Минск: Харвест. 718 с.

Westerlund M. 2019. The Emergence of Deepfake Technology: A Review. — *Technology Innovation Management Review*. Vol. 9. Is. 11. November. P. 39-52.

KRASOVSKAIA Nataliya Rudol'fovna, *Cand.Sci. (Psych.)*, Associate Professor of the Chair of Social Psychology and Victimology, Novosibirsk State Pedagogical University (28 Vilyuiskaya St, Novosibirsk, Russia, 630126; goulina@gmail.com)

GULYAEV Andrey Anatol'evich, *Cand.Sci. (Philos.)*, Associate Professor of V.G. Korolenko Glazov State Pedagogical Institute (25 Pervomaiskaia St, Glazov, Udmurt Republic, Russia, 427621; andrey.gulyaev1966@yandex.ru)

TECHNOLOGIES OF MIND CONTROL BY USING DEEPPAKES AS AN INSTRUMENT OF INFORMATION WAR IN THE POLITICAL SPHERE

Abstract. *The political impact of deepfakes is associated with the use of certain psychological tools. Technology of manipulating the consciousness and behavior of people forms the basis of deepfake influence. A methodological tool for this paper is the system analysis method. Various manipulative techniques used for deepfakes form a system in which image and video play a key role, and their main goal is to ignite people to such behavior that is beneficial for manipulators. Deepfake is created by using artificial intelligence and generative-competitive neural networks. Social networks and electronic media are the key channels for the distribution of deepfakes.*

Keywords: *deepfake, artificial intelligence, generative-competitive neural networks, psychological tools, manipulative techniques, «nudging» technology*
