

УДК 32.019.51
DOI: 10.21209/2227-9245-2021-27-6-101-106

ДИПФЕЙК КАК ФЕНОМЕН ПОЛИТИЧЕСКОЙ КОММУНИКАЦИИ

DEEPFAKE AS A PHENOMENON OF POLITICAL COMMUNICATION



М. А. Фалалеев,
Дальневосточный федеральный
университет, г. Владивосток
falaleev.mike@gmail.com



Н. А. Ситдикова,
Дальневосточный федеральный
университет, г. Владивосток
nata-sitdikova@mail.ru



Е. Е. Нечай,
Дальневосточный федеральный
университет, г. Владивосток
nechay.ee@dvfu.ru

M. Falaleev,
Far Eastern Federal University, Vladivostok

N. Sitdikova,
Far Eastern Federal University, Vladivostok

E. Nechay,
Far Eastern Federal University, Vladivostok

Развитие цифровых технологий, вкупе с прогрессом в разработке самообучающихся программ на базе нейросетей искусственного интеллекта (Artificial Intelligence – AI), несет в себе очевидные преимущества в области повышения эффективности информационного воздействия на людей по всему миру. В течение 2010-х гг. исследователи фиксировали тенденцию использования искусственного интеллекта для конструирования и распространения медиаконтента в целях манипулирования политическим дискурсом на национальном и глобальном уровнях. Исследательский интерес в контексте данной проблематики представляет то, как стремительное развитие AI-технологий влияет на политическую коммуникацию. Предмет исследования – дипфейк как явление современной политической коммуникации. Соответственно, цель исследования – описать и спрогнозировать влияние технологии дипфейков на политическую коммуникацию на глобальном и национальном уровнях. В работе представлена дефиниция дипфейка, дана оценка его характеристик в зависимости от способов и целей распространения, а также проанализированы перспективы использования этого инструмента для воздействия на политический дискурс в современной России. Для изучения предметного поля исследования применены методы систематизации теоретических данных, классификации, анализа совокупности факторов и прогнозирования. Практическая значимость работы обусловлена тем, что в ней представлено авторское определение и типологизация феномена дипфейка, а также описано его значение как фактора политической коммуникации на примере конкретной страны. Результаты работы будут полезны для исследователей, изучающих проблематику цифровизации медиапространства и современных средств дезинформации в политике как на локальном, так и на глобальном уровнях.

Ключевые слова: дипфейк; политическая коммуникация; AI системы; машинное обучение; искусственный интеллект; цифровизация; цифровые угрозы; деформация медиасреды; трансформация политических институтов; политическая стабильность

The development of digital technologies, coupled with progress in the development of self-learning programs based on AI (Artificial Intelligence), has obvious advantages in improving the effectiveness of information impact on people around the world. During the 2010s, researchers have documented trends in the use of artificial intelligence for the construction and distribution of media content to indirectly manipulate political discourse at the national and global levels. Special interest in the context of this issue is how the rapid development of AI technologies affects political communication. The object of consideration within the framework of this article is

the deepfake technology. Based on this, as a subject, the authors define deepfake as a phenomenon of modern political communication. Accordingly, the purpose of the study is to describe and predict the impact of deepfake technology on political communication at the global and national levels. The paper presents the definition of deepfake, assesses its characteristics depending on the methods and purposes of its distribution, and analyzes the prospects for using this tool to influence political discourse in modern Russia. To study the subject field of the research, methods of systematizing theoretical data, classification, analysis of a set of factors and forecasting have been applied. The practical significance of the work is presented by the authors' definition and typology of the phenomenon of deepfake and describes its significance as a factor of political communication on the example of a particular country. The results of the work will be useful for researchers studying the problems of digitalization of the media space and modern means of disinformation in politics, both at the local and global levels

Key words: *deepfake; political communication; AI systems; machine learning; artificial intelligence; digitalization; digital threats; media environment deformation; transformation of political institutions; political stability*

Введение. За последние 10 лет мир публичной политической коммуникации пережил настоящую революцию. В нулевые годы визионерские прогнозы о глобальном доступе к цифровым технологиям воспринимались со значительной долей скепсиса. Теперь мы видим, что даже в государствах, имеющих традиционный уклад экономики и квазифеодальную систему распределения власти, у большей части населения есть доступ к международной инфокоммуникационной сети Интернет. Это в значительной степени повышает значимость контроля не только над ресурсами для распространения информации, сколько над возможностями моделирования содержания политического дискурса [3].

В стремлении добиться максимального влияния на граждан, политические силы и медиа при помощи современных технологий обработки и анализа больших объемов данных (*big-data*) создают компьютерные алгоритмы различного назначения. Одним из таких назначений является целенаправленная дезинформация объектов воздействия с целью манипуляции их поведением и восприятием в интересах субъекта [13]. В современной политической науке единицу влияния с помощью недостоверных данных о реальности обозначают термином фейк (от англ. «fake» – ложное, недостоверное) [1].

Развитие машинного обучения на основании *big-data* позволило создать самообучаемые нейросети искусственного интеллекта, или AI. Эта технология дает возможность автоматически создавать уникальный медиаконтент на основании заданных ограничений в короткие сроки [8]. Постепенное распространение AI, помимо очевидного

повышения эффективности работы СМИ, способствует стремительной эволюции традиционных фейков. Важнейшей ступенью в данном процессе стала адаптация аффективных вычислений для анализа и моделирования человеческой речи, мимики, голоса, а также письма [15]. Благодаря этой технологии, создатели недостоверной информации получили возможность частично симулировать коммуникационные практики людей, что позволило сделать традиционные фейки значительно более аутентичными и трудно верифицируемыми без специальных знаний или программного обеспечения.

Первыми исследователями, которые зафиксировали и развили в своих работах тезис о том, что термин «фейк» не в полной степени подходит для описания дезинформации нового типа, стали Р. Чесни и Д. Цитрон (Chesney R., Citron D.), а также М. Вестрелунд (Westerlund M.) [6]; [14]. Поэтому в данной работе для характеристики объекта исследования используется понятие «дипфейк» (от англ. «deepfake» – глубоко (в значении усиления) недостоверный, ложный), которое обозначает подвид фейка, то есть единицу информации, сгенерированную самообучающимся AI для дезинформации людей, в целях искажения их восприятия реальности.

Целью исследования является описание и прогнозирование влияния технологии дипфейков на политическую коммуникацию на глобальном и национальном уровне.

Методология работы включает систематизацию теоретических данных, классификацию, анализ совокупности факторов и прогнозирование.

Результаты исследования. В рамках описания феномена и его значения в полити-

ческой коммуникации необходимо раскрыть его характеристики. Основными критериями для анализа проявлений дипфейков в политике для нас станут формат генерации, цель их применения, субъект-ретранслятор.

Важной особенностью самообучающихся AI является то, что они могут на основе ключевых черт текстов, речей или изображений реальных объектов создавать их уникальные аналоги, сохраняя при этом особенности прототипов [11; 5]. Это значит, что дипфейки способны иметь аудиальный, визуальный и аудиовизуальный форматы. Таким образом, например, они могут быть применены при создании крайне аутентичных текстов, аудиофайлов и видеороликов, где определённый субъект якобы выражает какую-то позицию или совершает некое действие. Разнообразие форм генерации дипфейков делает их универсальным и эффективным инструментом воздействия на человека. Это является одной из важнейших предпосылок для широкого применения технологии и постепенного отхода от применения классических фейков.

В наши дни созданную при помощи AI дезинформацию используют для достижения различных результатов, что дает нам возможность типологизировать её через критерий цели применения. Первоочередной целью использования дипфейков в современной политической коммуникации является дискредитация того или иного субъекта посредством создания контента, где он будто выражает провокационные мысли или совершает порицаемые в обществе действия [10]. Данная технология применима и для продвижения определённых политических акторов или сообществ. При генерации дипфейков также могут достигаться цели по созданию провокаций, которые предназначены для стимулирования политического субъекта на совершение активных действий, заданных логикой распространяемой дезинформации [3]. Более того, дипфейки способны эффективно привлекать массовую аудиторию для просмотра монетизированного контента, поскольку они позволяют помещать различных политических деятелей или представителей сообществ в нереалистичные условия, а неординарность контента порождает интерес у людей. Из этого мы делаем вывод, что технология может быть использована в политической коммуникации в коммерческих

целях. Таким образом, по цели применения дезинформации, сгенерированной AI, мы выделяем дискредитирующие, продвигающие, провокативные и коммерческие дипфейки. Целевая многовекторность технологии подтверждает тезис о её универсальности и практической применимости для реализации проектов в сфере политической коммуникации.

Субъектная сторона феномена дипфейка представляет высокий исследовательский интерес для изучения и прогнозирования. В связи с повышением глобальной транспарентности, технологии создания продвинутых форм дезинформации оказались доступны и применимы как коллективными, так и индивидуальными политическими акторами. Практика показывает, что характеристики и ресурсы субъекта воздействия в значительной степени определяют, однако не детерминируют эффективность применения AI-систем. На основании критерия ретранслятора продвинутых форм дезинформации мы выделяем государственные, корпоративные, гражданские дипфейки. Конечно, сгенерированная искусственным интеллектом недостоверная информация при распространении государственными структурами или корпоративными медиа имеет априори больший потенциальный охват нежели аналогичный контент, созданный активистом или гражданским объединением. Однако в ряде случаев последние оказываются более успешными, поскольку в зависимости от характеристики социальной системы и политической культуры конкретного общества может варьироваться уровень доверия людей государству и крупным медиахолдингам. Особенно четко это видно на примере развитых обществ, в частности США, где маргинальные гражданские организации выступают в качестве главных игроков рынка создания и распространения дипфейков, а государство и цифровые гиганты разрабатывают продвинутые системы верификации информации, тиражируемой в медиапространстве [7; 12].

Типологизация дипфейков позволяет выделить важнейшие характеристики этого феномена, что дает возможность спроектировать их на реальность при помощи инструментария прогнозирования. Для формулирования валидных выводов мы рассмотрели влияние технологии на уровне глобального и локального, в нашем случае российского фрейма.

Глобальное распространение дипфейков создает большие риски для стабильности международного порядка.

Во-первых, регионы земного шара неравномерно технологически развиты, поэтому ряд государств или корпораций могут получить чрезвычайный уровень инициативы в вопросах информационного давления на систему политических коммуникаций в странах, компаниях или гражданских общностях условного противника [2].

Во-вторых, доступность применения дипфейков чревата значительными искажениями глобального информационного поля, что может стать причиной негативных сценариев развития, вызванных агрессивными действиями тех или иных государств, решения и мотивы которых определены эффектами деформированной медиасреды.

Аудиовизуальные дипфейки провокативного и дискредитирующего типов, созданные транснациональными корпоративными или гражданскими субъектами, представляют наибольшую угрозу, поскольку воздействуют на наибольшее количество каналов восприятия информации по форме, способствуют разбалансировке политической стабильности по цели, а их распространение не может быть полностью заблокировано на законодательном уровне из-за наднациональной специфики актора-примирителя.

Оценивая потенциал применения AI-систем по созданию дезинформации в России, мы пришли к заключению, что они также представляют значительную угрозу для устойчивости политической системы страны. Данное предположение основано на анализе совокупности факторов, характеризующих предпосылки успешной разработки и применения дипфейков в РФ:

1) высокая цифровизация публичного пространства и каналов трансляции политического дискурса, о чем свидетельствуют данные статистики [9];

2) обширная теоретическая и материально-техническая база, а также наличие специалистов, имеющих навыки создания и настройки AI-систем, что позволяет создавать дипфейки всех описанных форм;

3) слабые регуляционные механизмы как на законодательном, так и на исполнительном уровнях власти [4];

4) конкурентная цифровая медиасреда в области распространения политического

контента, что позволяет говорить о высоком уровне полисубъектности источников дипфейков.

Сочетание этих факторов, периодические вспышки протестной активности и сложная международная конъюнктура дают нам основания прогнозировать, что дипфейки могут стать эффективным инструментом дестабилизации политической стабильности в России в среднесрочной перспективе, особенно в контексте выборов в Государственную Думу РФ VIII созыва. Тем не менее, последние пять лет в стране ведется разработка национальных систем распознавания AI-дезинформации, а также наблюдается активизация законотворческой деятельности, направленной на выработку институционального ответа на угрозу. Мы констатируем, что политическое руководство РФ осознает и посильно реагирует на проблему применения дипфейков, как инструментов политической коммуникации, а значит риски дестабилизации политической ситуации в России на данный момент имеют контролируемый характер.

Заключение. Таким образом, следует заключить, что развитие AI-технологии играет большую роль в трансформации политической коммуникации и это требует от субъектов различных уровней адаптироваться к новым условиям медиасреды. В работе дана характеристика AI-дезинформации при помощи метода классификации. Рассматривая этот феномен по критерию формата его генерации, выделяются аудиальный, визуальный и аудиовизуальный типы. По цели создания дипфейки подразделяются на дискредитирующие, продвигающие, провокативные и коммерческие виды. На основании критерия ретранслятора продвинутых форм дезинформации авторы выделяют государственные, корпоративные, гражданские типы. Анализ проблемы детекции и нейтрализации AI-дезинформации показывает, что она представляет большую угрозу для глобальных и локальных акторов. Рассмотрение российского опыта в контексте исследования дополнительно актуализирует вопросы государственного противодействия распространению дипфейков, а также задает фрейм для оценки и прогнозирования влияния описываемого феномена на национальные политические системы.

Список литературы

1. Ершов Ю. М. Феномен фейка в контексте коммуникационных практик // Вестник Томского государственного университета. Филология. 2018. № 52. С. 62–74.
2. Ковалев В. И., Матвиенко Ю.А. «Поведенческая» война как одна из перспективных невоенных угроз безопасности России // Информационные войны. 2016. Т. 40, № 4. С. 20–25.
3. Красовская Н. Р. Технологии манипуляции сознанием при использовании дипфейков как инструмента информационной войны в политической сфере // Власть. 2020. Т. 28, № 4. С. 93–98.
4. Минбалаев А. В. Проблемы регулирования искусственного интеллекта // Вестник Южно-Уральского государственного университета. Право. 2018. Т. 18, № 4. С. 82–87.
5. Bendel O. The synthetization of human voices // AI & Society. 2019. P. 83–89.
6. Chesney R., Citron D. Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security // California Law Review. 2019. P. 1753–1820.
7. Creating a dataset and a challenge for deepfakes. URL: <https://ai.facebook.com/blog/deepfake-detection-challenge> (дата обращения: 15.06.2021). Текст: электронный.
8. Diakopoulos N. Automating the News: How Algorithms are Rewriting the Media. Cambridge, MA: Harvard University Press. 2019. 336 р.
9. Digital 2020: Global Digital Overview. URL: <https://datareportal.com/reports/digital-2020-global-digital-overview> (дата обращения: 16.06.2021). Текст: электронный.
10. Fletcher J. Deepfakes, artificial intelligence, and some kind of dystopia: the new faces of online post fact performance // Theatr J. 2018. P 455–471.
11. Kim H., Garrido P., Tewari, A. Deep video portraits // ACM Transactions On Graphics. 2018. P. 1–14.
12. Malicious Deep Fake Prohibition Act of 2018 // Congress.gov. URL: <https://www.congress.gov/bill/115th-congress/senate-bill/3805/text> (дата обращения: 16.06.2021). Текст: электронный.
13. Tandoc E. C., Zheng W. L., Ling R. Defining ‘Fake News’: a typology of scholarly definitions // Digital Journalism. 2019. P. 137–153.
14. Westerlund M. The Emergence of Deepfake Technology: A Review // Technology Innovation Management Review. 2019. Vol. 9. P. 39–52.
15. Yonck R. Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence // New York: Arcade Publishing. 2017. P. 150–153.

References

1. Ershov Yu. M. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologija* (Bulletin of Tomsk State University. Philology), 2018, no. 52, pp. 62–74.
2. Kovalev V. I., Matvienko Ju. A. *Informatsionnye voyny* (Information wars), 2016, vol. 40, no. 4, pp. 20–25.
3. Krasovskaya N. R. *Vlast (Power)*, 2020, vol. 28, no. 4, pp. 93–98.
4. Minbalaev A. V. *Vestnik yuzhno-uralskogo universiteta* (Bulletin of the South Ural State University), 2018, vol. 18, no. 4, pp. 82–87.
5. Bendel O. *AI & Society* (AI & Society), 2019, pp. 83–89.
6. Chesney R., Citron D. *California Law Review* (California Law Review), 2019, pp. 1753–1820.
7. *Creating a dataset and a challenge for deepfakes* (Creating a dataset and a challenge for deepfakes). Available at: <https://ai.facebook.com/blog/deepfake-detection-challenge> (date of access: 15.06.2021). Text: electronic.
8. Diakopoulos N. *Automating the News: How Algorithms are Rewriting the Media* (Automating the News: How Algorithms are Rewriting the Media). Cambridge: MA: Harvard University Press, 2019, 336 p.
9. *Digital 2020: Global Digital Overview* (Digital 2020: Global Digital Overview). Available at: <https://data-reportal.com/reports/digital-2020-global-digital-overview> (date of access: 16.06.2021). Text: electronic.
10. Fletcher J. *Theatr J.* (Theatr J.), 2018, no. 11, pp. 455–471.
11. Kim H., Garrido P., Tewari A. *ACM Transactions On Graphics* (ACM Transactions On Graphics), 2018, no. 4, pp. 1–14.
12. *Congress.gov* (Congress.gov). Available at: <https://www.congress.gov/bill/115th-congress/senate-bill/3805/text> (date of access: 16.06.2021). Text: electronic.
13. Tandoc E., Zheng W., Ling R. *Digital Journalism* (Digital Journalism), 2019, no. 6, pp. 137–153.
14. Westerlund M. *Technology Innovation Management Review* (Technology Innovation Management Review), 2019, vol. 9, pp. 39–52.
15. Yonck R. *Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence* (Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence). New York: Arcade Publishing, 2017, pp. 150–153.

Информация об авторе

Фалалеев Михаил Алексеевич, лаборант, кафедра политологии, Дальневосточный федеральный университет, г. Владивосток, Россия. Область научных интересов: политические науки, влияние цифровых технологий на развитие политических систем, цифровизация политических институтов
falaleev.mike@gmail.com

Ситдикова Наталья Андреевна, лаборант, кафедра политологии, Дальневосточный федеральный университет, г. Владивосток, Россия. Область научных интересов: политические науки, цифровизация политической коммуникации, процессы трансформации политических институтов
nata-sitdikova@mail.ru

Нечай Екатерина Евгеньевна, канд. полит. наук, доцент кафедры политологии, Восточный институт – Школа региональных и международных исследований, Дальневосточный федеральный университет, г. Владивосток, Россия. Область научных интересов: политические науки, политические институты и процессы, цифровизация политического дискурса
nechay.ee@dvfu.ru

Information about the author

Mikhail Falaleev, laboratory assistant, Political Science department, Vladivostok, Russia. Scientific interests: political sciences, impact of digital technologies on the development of political systems, digitalization of political institutions

Natalia Sitdikova, laboratory assistant, Political Science department, Vladivostok, Russia. Scientific interests: political sciences, digitalization of political communication, processes of transformation of political institutions
nata-sitdikova@mail.ru

Ekaterina Nечay, candidate of political sciences, associate professor, Political Science department, Eastern Institute – School of Regional and International Studies, Far Eastern Federal University, Vladivostok, Russia. Scientific interests: political science, political institutions and processes, digitalization of political discourse.

Для цитирования

Фалалеев М. А., Ситдикова Н. А., Нечай Е. Е. Дипфейк как феномен политической коммуникации // Вестник Забайкальского государственного университета. 2021. Т. 27, № 6. С. 101–106. DOI: 10.21209/2227-9245-2021-27-6-101-106.

Falaleev M., Sitdikova N., Nечay E. Deepfake as a phenomenon of political communication // Transbaikal State University Journal, 2021, vol. 27, no. 6, pp. 101–106. DOI: 10.21209/2227-9245-2021-27-6-101-106.

Статья поступила в редакцию: 25.06.2021 г.
Статья принята к публикации: 05.07.2021 г.