

УДК 004.421.2:004.77:327.88

## ГЕНЕРАТИВНЫЕ АЛГОРИТМЫ В ИНФОРМАЦИОННЫХ ВОЙНАХ

*А.Д. Вуйкович*

*Национальный исследовательский ядерный университет «МИФИ», Москва, 115409, Россия*

*e-mail: Anastasia.Vuykovich@yandex.ru*

Поступила в редакцию 18.07.2022 г.

После доработки: 15.05.2024

Принята к публикации: 28.05.2024

Развитие информационно-коммуникационных технологий привело к существенному изменению жизни общества в области формирования потребления информации. С распространением сети Интернет начали появляться и различные средства распространения информации такие, как социальные сети, поддерживающие текстовый, визуальный и аудио-форматы. У общества появилась возможность публиковать и создавать контент, что в свою очередь сказывается на информационной среде, которая уже стала ареной информационного противоборства не только государств и организаций, но и отдельных индивидов. Особую роль в информационном противоборстве сегодня занимают методы автоматической генерации контента, основанные на генеративных алгоритмах, которые могут создавать объекты таких форматов, как изображения, видео, аудио и текст. В статье рассматриваются возможности использования генеративных алгоритмов в информационных войнах. Основная трудность обоснования совокупности тематик генеративных алгоритмов и информационных войн – новизна такого инструмента, как генеративно-состязательная сеть, отсутствие нормативно-правовой базы в Российской Федерации, а также подтвержденных случаев использования генеративных алгоритмов в информационной войне.

*Ключевые слова:* информационные войны, генеративный алгоритм, генеративно-состязательные сети, генератор, дискриминатор.

**DOI: 10.26583/vestnik.2024.352**

**EDN STQNGE**

### ВВЕДЕНИЕ

Изучением вопросов информационного противоборства (информационная война – синонимичное понятие) занимались разные ученые как в Российской Федерации, так и за рубежом.

Если рассматривать трактовки понятия ИВ полвека назад и сейчас, то они существенно различаются. В XX веке под ИВ понимали идеологическое и психологическое влияние, сейчас же с развитием технологий это понятие имеет множество толкований и смыслов, которые объединяются только объектной направленностью – информационным оружием.

Введем определения термина «**информационная война**» (далее **ИВ**), данное специалистами из нашей страны.

А.В. Манойло, доктор политических наук, специалист в области кибербезопасности и киберобороны, информационной безопасности, международного терроризма, определяет **ИВ** как процесс противоборства человеческих общностей, направленный на достижение политиче-

ских, экономических, военных или иных целей стратегического уровня, путём воздействия на гражданское население, власти и (или) вооруженные силы противостоящей стороны, посредством распространения специально отобранной и подготовленной информации, информационных материалов, и противодействия таким воздействиям на собственную сторону [1].

Документы, которые были разработаны в МИД РФ, определяют **ИВ** как противоборство между государствами в информационном пространстве с целью нанесения ущерба информационным системам, процессам и ресурсам, критически важным структурам, подрыва политической, экономической и социальной систем, а также массивной психологической обработки населения с целью дестабилизации общества и государства [2].

Термин **ИВ** в Соглашении между правительствами государств – членов Шанхайской организации сотрудничества о сотрудничестве в области обеспечения международной информационной безопасности от 16 июня 2009 года трак-

туется как противоборство между двумя или более государствами в информационном пространстве с целью нанесения ущерба информационным системам, процессам и ресурсам, критически важным и другим структурам, подрыва политической, экономической и социальной систем, массивированной психологической обработки населения для дестабилизации общества и государства, а также принуждения государства к принятию решений в интересах противоборствующей стороны<sup>1</sup>.

Термины **ИВ** и «**информационное противоборство**» (далее **ИП**) зачастую сливаются в единое целое. Введем определение ИП, чтобы сравнить два понятия.

Сулейманова Ш.С., доктор политических наук, и Назарова Е.А., доктор социологических наук, определяют **ИП** как соперничество в информационно-психологической сфере, целью которого является усиление влияния на определенные сферы отношений и установления контроля над стратегическими ресурсами, в результате чего одни участники противоборства получают преимущество, другие – утрачивают [1].

Также под **ИП** понимают борьбу в информационной среде, то есть деструктивное воздействие на информацию, системы, инфраструктуру соперника и защитой собственной информации, систем и инфраструктуры от воздействия подобного рода.

Зарубежные специалисты дают похожие определения. Приведем некоторые из них.

«Объединенная доктрина информационных операций» Министерства обороны США 1998 года определяет **ИВ** как комплексное воздействие или совокупность информационных мероприятий на систему государственного или военного управления противостоящей стороны, на ее военно-политическое руководство, которое уже в мирное время приводило бы к принятию благоприятных для стороны-инициатора информационного воздействия решений, а в ходе конфликта полностью парализовало бы функционирование инфраструктуры управления противника [1].

В документах НАТО **ИВ** трактуется, как операция, проводимая с целью получения информационного преимущества над противни-

ком. Она заключается в контроле собственного информационного пространства, защите доступа к собственной информации, приобретении и использовании информации противника, разрушении его информационных систем и нарушении информационного потока<sup>2</sup>.

Исходя из данных определений термина, можно ввести понятие, которое объединяет в себе все вышесказанное. **ИВ** или **ИП** – форма борьбы сторон с использованием специальных методов, средств и способов, направленная на достижение военных, политических, экономических и других целей, путем воздействия на население, власти и сторону противника, посредством распространения информации, выгодной определенной стороне для защиты собственных интересов.

Понятие генеративного алгоритма впервые появилось в статье А. Гудфеллоу «Генеративно-состязательные сети» в 2014 году [3]. **Генеративно-состязательная сеть**, или **генеративный алгоритм** – алгоритм машинного обучения без учителя, построенный на комбинации нейронных сетей – генеративной и дискриминативной и основанный на теоретико-игровом сценарии с нулевой суммой. Генератор производит образцы. Дискриминатор пытается отличить реальные образцы от сгенерированных.

**Цель генератора** – повысить процент ошибок дискриминатора, а **цель дискриминатора** – усовершенствование точности распознавания результатов.

Когда дискриминатор успешно отличает реальные образцы от сгенерированных, никаких изменений в параметрах его модели не требуется, в отличие от генератора. Когда генератору удастся обмануть дискриминатор, то его параметры не требуют изменения, а параметры дискриминатора обновляются. Когда объекты, созданные генератором, неотличимы от реальных, дискриминатор выдает результат 1/2 и может быть отброшен [4]. Этот принцип можно схематически изобразить (рис. 1). Идеализированную архитектуру генеративно-состязательной сети в книге «Глубокое обучение» описал ученый А. Гудфеллоу [4], его схема представлена на рис. 1. Автор настоящей статьи доработала данную схему [5].

<sup>1</sup> Соглашение между правительствами государств – членов Шанхайской организации сотрудничества о сотрудничестве в области обеспечения международной информационной безопасности // Консорциум КОДЕКС. 2009. [Электронный ресурс]. URL: <https://docs.cntd.ru/document/902289626> (дата обращения: 12.02.2022).

<sup>2</sup> Media – (dis)information – security // NATO. [Электронный ресурс]. URL: [https://www.nato.int/nato\\_static\\_fl2014/assets/pdf/2020/5/pdf/2005-deepportall1-fake-news.pdf](https://www.nato.int/nato_static_fl2014/assets/pdf/2020/5/pdf/2005-deepportall1-fake-news.pdf) (дата обращения: 12.07.2022).

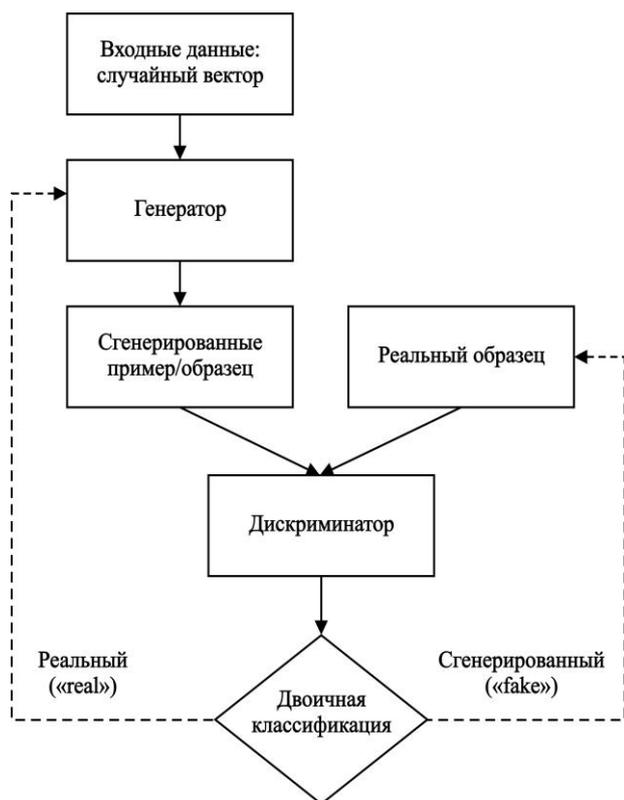


Рис. 1. Архитектура модели генеративно-сопоставительной сети

Изначально генеративно-сопоставительная сеть использовалась для достраивания изображений. А. Гудфеллоу и команда ученых под его руководством провели эксперимент по созданию примеров для набора данных изображений. Они обучили генеративную сеть набору данных MNIST (рис. 2,а), TDF (рис. 2,б), CIFAR-10 (рис. 2,в) и CIFAR-10 (сверточный дискриминатор и «деконволюционный» генератор) (рис. 2,г).

Крайние правые столбцы на рис. 2 (выделены желтой рамкой) показывают ближайшие обучающие примеры из соседних образцов, чтобы продемонстрировать, что модель не запомнила обучающий набор. Образцы являются случайными, а не отобранными.

На рис. 2 [3] изображены образцы, взятые из сети генератора после обучения. Исследователи пришли к выводу, что эти образцы конкурируют с лучшими генеративными моделями и подчеркивают потенциал сопоставительной структуры [3].

В настоящее время генеративный алгоритм используется в информационной среде для освещения событий, в том числе для создания как правдивой, так и ложной информации. Что ставит перед учеными задачу разработать средства программного обеспечения для идентификации сгенерированного контента.



Рис. 2. Визуализация образцов из модели [3]

МЕТОДЫ ИСПОЛЬЗОВАНИЯ

Генеративные алгоритмы могут быть использованы для создания четырех видов объектов информации: изображений, видео, текста и аудио.

**Использование генеративных алгоритмов для создания и преобразования изображений**

Генеративно-состязательным сетям, у которых заданы определенные условия на входе, для преобразования изображения в изображение, таких как преобразование дня в ночь, дискриминатору предоставляются примеры реальных и сгенерированных ночных фотографий, а также реальных дневных фотографий в качестве входных данных. Генератор снабжен случайным вектором из скрытого пространства, исходя из заданных условий, а также реальными дневными фотографиями в качестве входных данных.

Для повышения качества изображений могут быть использованы следующие инструменты: обрезка, переворачивание, масштабирование и другие преобразования существующих изображений в обучающем наборе данных.

В сложных областях или областях с ограниченным объемом данных генеративные алго-

ритмы способны моделировать и обрабатывать недостающие данные.

Приведем несколько примеров использования генеративных алгоритмов в преобразовании изображений.

В научной работе «Преобразование изображения в изображение с помощью условных состязательных сетей» рассматриваются проблемы в обработке изображений: преобразование входного изображения в соответствующее выходное изображение, как показано на рисунке 3 [6]. Решение проблемы было найдено в условных состязательных сетях [6].

Помимо преобразования изображения в изображение исследователи проводили эксперименты с генеративно-состязательными сетями, в которых создавали реалистичные фотографии из текстовых описаний.

Ученые разработали многоуровневые генеративно-состязательные сети (StackGAN) для генерации фотореалистичных изображений размером 256×256, обусловленных текстовыми описаниями. Принцип работы StackGAN состоит из двух этапов:

- первый этап: GAN рисует примитивную форму и цвета объекта на основе заданного текстового описания, получая изображение с низким разрешением (строка (a) на рис. 4) [7];



Рис. 3. Преобразование изображений Google Maps в спутниковый снимок (слева) и преобразование спутникового снимка в изображение Google Maps (справа) [6]



Рис. 4. Результаты работы StackGAN [7]

- второй этап: GAN прорабатывает результаты первого этапа и текстовые описания в качестве входных данных и генерирует изображения с высоким разрешением и фотореалистичными деталями. На втором этапе алгоритм устраняет дефекты в результатах первого этапа и добавляет дополнительные детали в процессе доработки (строка (b) на рис. 4) [7].

Р. Хуан в статье «За пределами поворота лица: глобальное и локальное восприятие GAN для фотореалистичного и сохраняющего идентичность синтеза фронтального вида» демонстрируют использование генеративных алгоритмов для создания фронтальных фотографий человеческих лиц, сделанных под углом. В итоге, сгенерированные фронтальные фотографии могут быть использованы в качестве входных данных для системы проверки или идентификации лиц [8].

Применение генеративно-состязательных сетей в ИП заключается в решении задач, требующих создания новых примеров. На основе приведенных в работе исследований укажем на возможности применения генеративных алгоритмов:

1. Преобразование изображения в изображение: возможность

трансформировать фотографии из разных областей, например, день в ночь, лето в зиму, спутниковые фотографии в Google Maps, черно-белые фотографии в цветные.

2. Создание реалистичных фотографий из текстовых описаний.

3. Использование фотографий для системы проверки или идентификации лиц.

По мнению Гудфеллоу генеративные алгоритмы смогли создавать фотографии настолько реалистичные, что даже эксперты не могли отличить сгенерированные алгоритмом изображения от действительности [4].

#### **Использование генеративных алгоритмов для создания видеоматериалов**

Исследователи из Facebook изобрели генераторы, которые имеют специально созданные структуры взаимосвязанных слоев. Каждый из уровней генератора состоит из определенного алгоритма, применяемого к входным данным, которые получает сеть. Первый уровень запускает алгоритм, который извлекает необработанные пиксели и простые мотивы из набора данных, представляющего изображение. Следующий уровень объединяет эти мотивы в более сложные композиции. Другие уровни обнаруживают части объектов, собирают их в единое целое и создают отдельные сцены, пока не будет получено все изображение.

Генеративно-состязательная сеть смогла научиться рисовать определенные объекты по мере прохождения обучения. Учеными были протестированы прогностические возможности генератора – использование необработанных

данных для создания видеоклипов. Генератору передали 4 кадра видео и заставили его создать другие 2 кадра на основе этих данных. Полученные кадры, сгенерированные искусственным интеллектом, выглядели как реалистичное продолжение действия [9].

**Использование генеративного алгоритма для создания текста**

М. Хоссам и его команда ученых разработали возможность составления текста генеративным алгоритмом, используя набор данных IMDB movie reviews. Набор данных состоял из 100 000 отзывов, извлеченных с веб-сайта IMDB и был разделен на 3 группы: 12 500 положительных, 12 500 отрицательных и 50 000 немаркированных отзывов, а тестовый набор состоял из 12 500 положительных и отрица-

тельных отзывов в каждой. Во всех экспериментах использовались предложения длиной 20 слов и объемом словарного запаса 10 000 [10].

Чтобы оценить качество полученного результата, ученые показали сгенерированные выборки с низкими (0 – 33,33 %) или средними (38,33 – 50 %) показателями BLEU (рис. 5 [10]). Показатель BLEU – это алгоритм, который оценивает качество автоматически переведенного текста: чем ближе машинный перевод к профессиональному человеческому, тем ниже показатель BLEU и, соответственно, лучше качество перевода. Предложения с низким уровнем BLEU имеют грамматическую структуру.

Ученые доказали, что генеративный алгоритм может генерировать грамматически и семантически верные предложения с высоким разнообразием.

**Образцы с низким/средним показателем BLEU**

BLEU-3	Образцы предложений
00.00%	a creative , well-made comedy that focuses into a scary film by peter jackson , are heading for <UNK>
00.00%	this direction struck me , that's ! this dreadful zombie movie , generally and curtis are terrific but doesn't .
00.00%	but effort , rural accent must be developed from joke a video <UNK> this new version starts ! some say
11.11%	okay exceptionally documentary about steve hayes ( who doesn't even destroy to the <UNK> ! that doesn't act better than
11.11%	sorry you're a fan , might even like your a spoiler as a master of shakespeare ; a little girl
11.11%	the wish i can relate to seeing many kind of kills the other well-made film adaptations than the godfather documentary
22.22%	wow hitchcock's prince meets thriller of the modern relationships in <UNK> terror and <UNK> above <UNK> the war film and
27.77%	the fine ending is , not even close to " princess <UNK> , " that it'd be extremely cool .
27.77%	this is an excellent , unbelievable , rather colorful show . brian <UNK> did never become a straight to part
33.33%	brilliant carpenter's halloween is actually my favorite films of my life . it made thriller a few laughs of new
33.33%	seriously imdb many previous reviews posted here about this movie due to a pair of art killings & <UNK> combined
33.33%	i haven't been such a clever flick , which is neither so <UNK> a mystery . the acting is somewhere
33.33%	so is a lovely film by any <UNK> . it makes a dumb decision to play fairly typical thing .
38.33%	wow surprised how wanted to see this . i thought about two of steven johnson are superb . however ,
44.44%	the tense 1931 melodrama is one of the finest entries to all ideas . while this is a great anime
44.44%	i believe i know things i can say , you always come out of my time and the way max
44.44%	the movie is a piece of crap . a group of guy lives off the shelf who meets shrek ,
50.00%	this izzard is a one-in-a-million comic genius ! many stars in this film ? <br / ><br / <UNK> the

Рис. 5. Образцы с низким/средним показателем BLEU [10]

**Использование генеративных алгоритмов для создания аудиозаписей**

Для обучения генеративного алгоритма Р. Аленкар использовал базу данных Freesound-Audio-Tagging, содержащую более 10 000 аудиосэмплов с разрядностью 16 бит и частотой дискретизации 44,1 кГц, монофоническими аудиофайлами, каждый из которых имеет разную длительность. Кроме того, все примеры разделены на 41 различные категории, такие как «Труба», «Аплодисменты», «Скрипка» и другие.

Основными категориями, использованными для эксперимента, были «Саксофон» и «Скрипка». Все аудиофайлы были повторно обработаны до 16 кГц, с продолжительностью 4 секунды, а затем нормализованы.

Эксперимент Р. Аленкара показал, что сгенерированные аудиодорожки более шумные по сравнению с оригинальными, а части тишины в оригинальных образцах не могут быть воспроизведены генератором, как изображено на рис. 6, 7 [11].

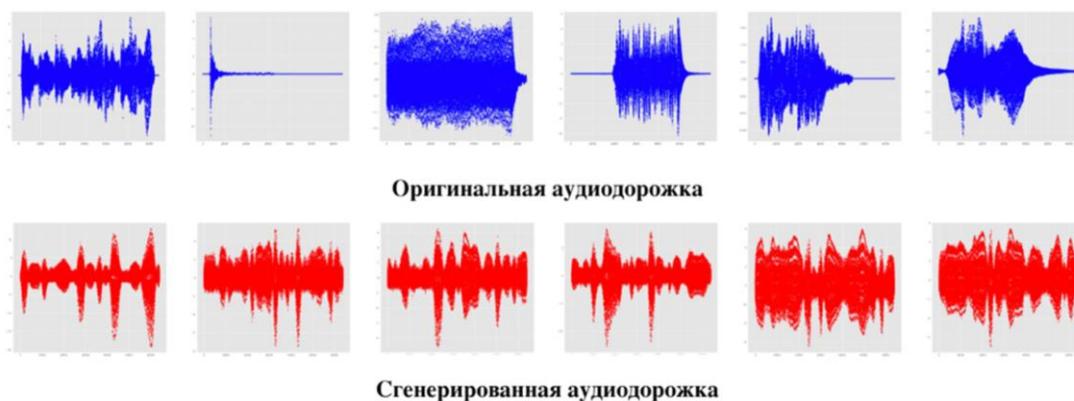


Рис. 6. Сравнение оригинальной аудиодорожки и сгенерированной [11]

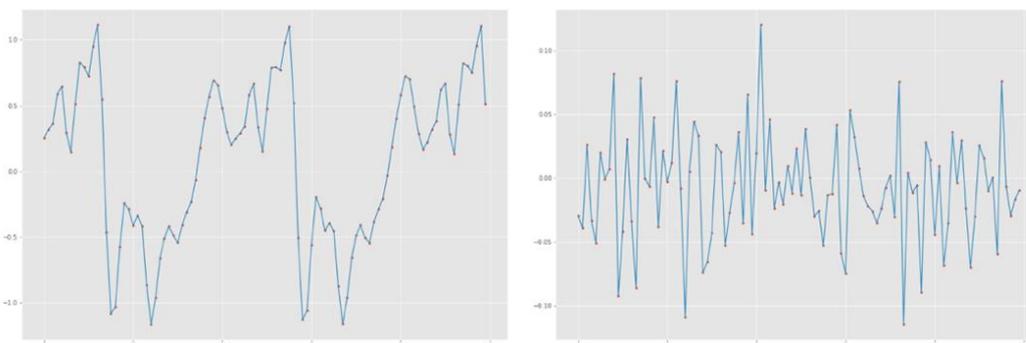


Рис. 7. Оригинальная аудиодорожка слева и сгенерированная аудиодорожка справа (в увеличении) [11]

Оригинальная волна выглядит лучше, чем сгенерированная. Ее форма определяет тембр, который отличает различные типы производства звука. Сгенерированная волна выглядит как случайный шум. Поэтому, чтобы улучшить эту модель, нужно обратить внимание на тембр этого звука и попытаться улучшить качество его генерации [11].

По сравнению с исследованиями 2019 года сгенерированные генеративными алгоритмами аудиозаписи сложно отличить от реальных, не используя специальные технологии. Нейросети научились создавать аудиоматериалы, которые подражают человеческой манере речи.

#### **Примеры использования генеративных алгоритмов в информационных войнах**

Наиболее часто в контексте информационных войн используются фото, видео и аудиоматериалы. В период предвыборной кампании использование «грязных» методов или черного пиара возрастает в разы, чтобы дискредитировать оппонента или очистить доброе имя. Ниже приведены некоторые примеры.

Один из них связан с фальшивым голосом Джо Байдена, который использовался в автоматически сгенерированных робозвонках, чтобы отговорить демократов от участия в первичных выборах. Байдена даже не было в избиратель-

ном бюллетене, хотя от его имени была запущена кампания по внесению изменений<sup>1</sup>. Стоит отметить, что аудиозаписи, созданные с помощью генеративных алгоритмов, не только копируют голос человека, но и манеру речи, а также фирменные слова и фразы, поэтому достаточно сложно сходу понять, что это фальшивка.

Так в предвыборной гонке в Индии сами кандидаты просили специалистов, создающих контент с помощью генеративных алгоритмов, наложить лица соперников на порнографические изображения или же создать некачественные поддельные видеоролики с кандидатом, которые впоследствии могут быть опубликованы, чтобы поставить под сомнение любые изображения реальные видеоролики, появившиеся во время выборов<sup>2</sup>.

<sup>1</sup> The deepfake era of US politics is upon us // CNN, 2024 . [Электронный ресурс] URL: <https://edition.cnn.com/2024/01/24/politics/deepfake-politician-biden-what-matters/index.html> (дата обращения: 13.05.2024).

<sup>2</sup> AI deepfakes threaten to upend global elections. No one can stop them. // The Washington Post, 2024. [Электронный ресурс] URL: <https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/> (дата обращения: 13.05.2024).



**Рис. 8.** Схема использования генеративных алгоритмов в информационной войне

Легкий доступ к инструментам искусственного интеллекта, таким как ChatGPT, приводит к тому, что в сети распространяются фейковые видео, записи разговоров, картинки, которые влияют на все сферы жизни общества, а не только на политическую, как приведено в примерах выше. Для того, чтобы сфабриковать видео, аудио или фото требуется всего несколько минут и умения составлять промпты (запросы-подсказки) для нейросетей. Как правило, используется техника few-shot prompting (обучение с примерами): пользователь приводит примеры и детально описывает нейросети, что хо-

чет получить на выходе, нейросеть в свою очередь отправляет пользователю файл в соответствии с его запросом. Далее информация распространяется с помощью социальных сетей, где набирает просмотры, лайки и репосты, а потом на волне популярности появляется в различных ток-шоу и новостях, где будет разобрана подробно и, возможно, разоблачена. Проще этот процесс представить схематично, как показано на рис. 8.

## РЕЗУЛЬТАТЫ И ВЫВОДЫ

Изображения, видео, тексты и аудиозаписи, созданные с помощью генеративных алгоритмов, могут быть использованы как объекты информационного противостояния. Некоторые изображения, прошедшие через генеративно-состязательные сети, не обладают высоким качеством, но могут быть использованы в информационных войнах. Камеры видеонаблюдения, фотографии, сделанные в экстренных ситуациях, также не отличаются высоким разрешением, поэтому могут быть использованы в корыстных целях. Броский заголовок, опора на авторитеты и непрерывное раскручивание темы приводят к желаемому результату – общественность поверит и примет происходящее на видео или изображении за реальность.

Короткий текст, созданный генеративным алгоритмом, имеет право на существование в условиях информационной войны, так как он не лишен смысла и несет в себе верную грамматическую структуру, то есть для продвижения на платформах с ограниченным количеством знаков в сообщении, как никогда подойдет и впишется органично, донося до читателей нужную повестку. Аудиозаписи, сгенерированные с помощью генеративно-состязательных сетей, ранее не передавали тембров, а лишь сплошной шум. Спустя некоторое время, аудиоматериалы, созданные нейросетями, почти что неотличимы от оригинала и могут подстраиваться под речь человека, используя его фирменные слов и выражения.

Технологии, разработанные в лаборатории Facebook, отличаются лучшим качеством видео и возможностью генерировать продолжение кадров. Так как кадры выглядят реалистично, что даже эксперт с трудом сможет отличить сгенерированное видео от реального, мировая общественность безоговорочно поверит в происходящее. Видео будет «кочевать» из одного ресурса в другой, «обрастать» комментариями, покрупными разборами и каждый участник

информационной войны будет трактовать увиденное в выгодном для него свете.

Помимо создания фейков и вбросов, генеративный алгоритм можно использовать для преобразования изображений в изображение, улучшения качества изображений, создания фотографий из текстовых описаний, а также достраивания профиля лица на фотографиях, чтобы впоследствии использовать их для проверки и идентификации лиц. Эти возможности также важны в информационных войнах для создания качественного контента и доказательной базы.

Основная трудность обоснования использования генеративных алгоритмов в информационных войнах – новизна такого инструмента, как генеративно-состязательная сеть, отсутствие нормативно-правовой базы в большинстве стран мира. Так в Соединенных Штатах Америки<sup>1</sup> и Китайской Народной Республике<sup>2</sup> за распространение фейков с помощью генеративных алгоритмов и отсутствие маркировки на фото- и видеоконтенте, созданных с помощью нейросетей, предусмотрены штрафы и заключение под стражу. Даже при жестком правовом регулировании остаются трудности в поиске создателей фейкового контента, а также его маркировке, так как материалы распространяются лавинообразно. Поэтому целесообразно проработать нормативно-правовую базу, например, в Российской Федерации не регулируется распространение материалов, созданных генеративными алгоритмами<sup>3</sup>, а пользователям критично относиться к поступающей информации и перепроверять ее.

<sup>1</sup> H.R.3230 – DEEP FAKES Accountability Act [Электронный ресурс] // CONGRESS.GOV, 2019. [Электронный ресурс]. URL: <https://www.congress.gov/bill/116th-congress/house-bill/3230/text> (дата обращения: 14.05.2024).

<sup>2</sup> 国家互联网信息办公室关于《互联网信息服务深度合成管理规定（征求意见稿）》公开征求意见的通知 [Электронный ресурс] // Office of the Central Cyberspace Affairs Commission, 2022. [Электронный ресурс]. URL: [http://www.cac.gov.cn/2022-01/28/c\\_1644970458520968.htm](http://www.cac.gov.cn/2022-01/28/c_1644970458520968.htm) (дата обращения: 14.05.2024).

<sup>3</sup> ГК РФ Статья 152.1. Охрана изображения гражданина [Электронный ресурс] // КонсультантПлюс: [сайт]. [2022]. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_5142/14c6c3902cffa17ab26d330b2fd4fae28e5cd059/](http://www.consultant.ru/document/cons_doc_LAW_5142/14c6c3902cffa17ab26d330b2fd4fae28e5cd059/) (дата обращения: 14.05.2024).

## СПИСОК ЛИТЕРАТУРЫ

1. Сулейманова Ш.С., Назарова Е.А. Информационные войны: история и современность: учеб. пособие. М.: Международный издательский центр «Этносоциум», 2017. 126 с.
2. Ромашикина Н.П., Ковалев В.И., Топычканов П.В., Евсеев В.В., Идаятов А.К. Угрозы информационной безопасности в кризисах и конфликтах XXI века. М.: ИМЭМО РАН, 2015. 151 с.
3. Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets, 2014. [Электронный ресурс]. URL: <https://arxiv.org/pdf/1406.2661.pdf> (дата обращения: 01.04.2022).
4. Goodfellow I., Bengio Y., Courville A. Deep learning. Massachusetts Institute of Technology, 2016. 700 p.
5. Вуйкович А.Д. Генеративные алгоритмы в информационных войнах // Сборник тезисов докладов IX Международной молодежной научной школы-конференции «Современные проблемы физики и технологий», 26–28 апреля 2022 г. М.: НИЯУ МИФИ, 2022. С. 247–250.
6. Isola P., Zhu J.Y., Zhou T., Efros A.A. Image-to-Image Translation with Conditional Adversarial Networks, 2018. [Электронный ресурс]. URL: <https://arxiv.org/pdf/1611.07004.pdf> (дата обращения: 02.04.2022).
7. Zhang H., Xu T., Li H., Zhang S., Wang X., Huang X., Metaxas D. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks, 2017. [Электронный ресурс]. URL: <https://arxiv.org/pdf/1612.03242.pdf> (дата обращения: 02.04.2022).
8. Brownlee J. 18 Impressive Applications of Generative Adversarial Networks (GANs) // Machine Learning Mastery, 2019. [Электронный ресурс]. URL: <https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/> (дата обращения: 03.04.2022).
9. Greenemeier L. When Will Computers Have Common Sense? Ask Facebook // SCIENTIFIC AMERICAN, 2016. [Электронный ресурс]. URL: <https://www.scientificamerican.com/article/when-will-computers-have-common-sense-ask-facebook/> (дата обращения: 19.03.2022).
10. Hossain M., Le T., Papasimeon M., Huynh V., Phung D. Text Generation with Deep Variational GAN, 2021. [Электронный ресурс]. URL: <https://arxiv.org/pdf/2104.13488.pdf> (дата обращения: 04.04.2022).
11. Alencar R. Audio Generation with GANs // Medium, 2019. [Электронный ресурс]. URL: <https://medium.com/neuronio/audio-generation-with-gans-428bc2de5a89> (дата обращения: 04.04.2022).

## GENERATIVE ALGORITHMS IN INFORMATION WARFARE

A.D. Vuykovich

National Nuclear Research University MEPhI, Moscow, 115409 Russia

e-mail: [Anastasia.Vuykovich@yandex.ru](mailto:Anastasia.Vuykovich@yandex.ru)

Received July 18, 2022; revised May 15, 2024; accepted May 28, 2024

The development of information and communication technologies has led to a significant change in the life of society in the field of information consumption formation. With the spread of the Internet, various means of information dissemination began to appear, such as social networks that support text, visual and audio formats. The society has the opportunity to publish and create content, which in turn affects the information environment, which has already become the arena of information confrontation not only between states and organizations, but also individuals. Methods of automatic content generation based on generative algorithms that can create objects of such formats as images, video, audio and text play a special role in the information confrontation today. The article discusses the possibilities of using generative algorithms in information warfare. The main difficulty in substantiating the totality of topics of generative algorithms and information warfare is the novelty of such a tool as a generative adversarial network, the lack of a regulatory framework in the Russian Federation, as well as confirmed cases of the use of generative algorithms in information warfare.

*Key words:* information warfare, generative algorithm, generative adversarial network, GAN, generator, discriminator.

### REFERENCES

1. Suleymanova Sh.S., Nazarova E.A. *Informatsionniye voyni: istoria i sovremennost'* [Information wars: history and modernity]. Moscow, International Publishing Center «Ethnosocium» Publ., 2017. Pp. 4–46.
2. Romashkina N.P., Kovalev V.I., Topychkanov P.V., Evseev V.V., Idayatov A.K. *Ugrozy informatsionnoi bezopasnosti v krizisah i ugrozah XXI veka* [Threats to information security in crises and conflicts of the XXI century]. Moscow, IMEMO RAS, 2015, Pp. 21–30.
3. Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. *Generative Adversarial Nets* 2014. Available at: <https://arxiv.org/pdf/1406.2661.pdf> (accessed: 01.04.2022).
4. Goodfellow I., Bengio Y., Courville A. *Deep learning*. Massachusetts Institute of Technology, 2016, Pp. 700.
5. Vuykovich A.D. *Generativnyye algoritmy v informatsionnykh voynah* [Generative algorithms in information wars]. *Sbornik tezisev dokladov IX Mezhdunarodnoj molodezhnoj nauchnoj shkoly-konferencii «Sovremennye problemy fiziki i tekhnologii»*, 26–28 aprelya 2022g. [Collection of abstracts of the IX International Youth Scientific School-Conference «Modern Problems of Physics and Technologies», April 26–28, 2022]. Moscow, NIYAU MIFI Publ., 2022. Pp. 247–250 (in Russian).
6. Isola P., Zhu J.Y., Zhou T., Efros A.A. *Image-to-Image Translation with Conditional Adversarial Networks* 2018. Available at: <https://arxiv.org/pdf/1611.07004.pdf> (accessed: 02.04.2022).
7. Zhang H., Xu T., Li H., Zhang S., Wang X., Huang X., Metaxas D. *StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks* 2017. Available at: <https://arxiv.org/pdf/1612.03242.pdf> (accessed: 02.04.2022).
8. Brownlee J. *18 Impressive Applications of Generative Adversarial Networks (GANs)*. *Machine Learning Mastery*, 2019. Available at: <https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/> (accessed 03.04.2022).
9. Greenemeier L. *When Will Computers Have Common Sense? Ask Facebook*. *Scientific American*, 2016. Available at: <https://www.scientificamerican.com/article/when-will-computers-have-common-sense-ask-facebook/> (accessed: 19.03.2022).
10. Hossam M., Le T., Papasimeon M., Huynh V., Phung D. *Text Generation with Deep Variational GAN*, 2021. Available at: <https://arxiv.org/pdf/2104.13488.pdf> (accessed: 04.04.2022).
11. Alencar R. *Audio Generation with GANs*. *Medium*, 2019. Available at: <https://medium.com/neuronio/audio-generation-with-gans-428bc2de5a89> (accessed: 04.04.2022).