

Корепанов Никита Сергеевич

Студент, Федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский государственный энергетический университет», Россия, г. Казань

Сандаков Виталий Дмитриевич

Кандидат технических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский государственный энергетический университет», Россия, г. Казань

ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫЕ СЕТИ (GAN) В СОЗДАНИИ И ОБРАБОТКЕ ЦИФРОВОГО КОНТЕНТА: АРХИТЕКТУРНЫЕ ПРИНЦИПЫ, ПРИКЛАДНЫЕ ЗАДАЧИ И ЭТИЧЕСКИЕ ВЫЗОВЫ

В статье представлен комплексный анализ генеративно-сопоставительных сетей (Generative Adversarial Networks, GAN) – одного из наиболее прорывных направлений в области машинного обучения за последнее десятилетие. Рассмотрены фундаментальные архитектурные принципы GAN, основанные на идее состязательного обучения двух нейронных сетей: генератора и дискриминатора [1]. Детально проанализированы эволюция архитектур (от оригинальной DCGAN до современных StyleGAN и их модификаций) и ключевые технические вызовы, такие как проблема коллапса моды и нестабильность обучения. В работе систематизированы основные прикладные области применения GAN: генерация фотореалистичных изображений и видео, увеличение разрешения (Super-Resolution), трансфер стиля, синтез тренировочных данных для автономных систем, а также применение в науке и медицине. Особое внимание уделено критическому анализу этических и социальных рисков, связанных с технологией, включая создание deepfake-контента, проблемы авторского права и цифровой аутентификации. В

заклучении обсуждаются перспективы развития GAN в контексте формирования этических норм и регуляторных практик, необходимых для безопасного использования технологии.

The article provides a comprehensive analysis of Generative Adversarial Networks (GANs) – one of the most breakthrough areas in machine learning over the past decade. The fundamental architectural principles of GANs, based on the idea of adversarial training of two neural networks (a generator and a discriminator), are considered [1]. The evolution of architectures (from the original DCGAN to modern StyleGAN and its modifications) and key technical challenges, such as the mode collapse problem and training instability, are analyzed in detail. The paper systematizes the main applied areas of GANs: generation of photorealistic images and video, super-resolution, style transfer, synthesis of training data for autonomous systems, as well as applications in science and medicine. Particular attention is paid to a critical analysis of the ethical and social risks associated with the technology, including the creation of deepfake content, copyright issues, and digital authentication. In conclusion, the prospects for the development of GANs in the context of the formation of ethical standards and regulatory practices necessary for the safe use of the technology are discussed.

Ключевые слова: генеративно-сопязательные сети (gan), машинное обучение с учителем, глубокое обучение, синтез данных, deepfake, этика искусственного интеллекта, архитектура нейронных сетей, компьютерное зрение.

Keywords: generative adversarial networks (gan), supervised machine learning, deep learning, data synthesis, deepfake, ai ethics, neural network architecture, computer vision.

Введение

Современный этап цифровой трансформации характеризуется не только увеличением объемов потребляемых данных, но и растущим спросом на технологии их создания и модификации. В этом контексте методы генеративного искусственного интеллекта, способные создавать новые, ранее не существовавшие данные, выдвигаются на передний план исследовательских и коммерческих интересов. Генеративно-сопоставительные сети, предложенные в 2014 году Иэном Гудфеллоу и его коллегами, произвели революцию в этой области, предложив элегантный и мощный фреймворк для обучения генеративных моделей [1]. В отличие от традиционных подходов, GAN позволяют синтезировать данные высокой степени реалистичности, часто неотличимые от реальных для человеческого восприятия. Актуальность исследования GAN обусловлена их стремительным проникновением в самые разные сферы – от цифрового искусства и медиа до фундаментальной науки и медицины. Однако бурное развитие технологии порождает комплекс серьезных этических, правовых и социальных вызовов, делая критический анализ ее возможностей и ограничений необходимым условием ответственного внедрения.

1. Архитектурные принципы и эволюция генеративно-сопоставительных сетей

Фундаментальная идея GAN заключается в организации сопоставительного процесса между двумя нейронными сетями: генератором (G) и дискриминатором (D). Генератор получает на вход шумовой вектор (латентное пространство) и пытается сгенерировать синтетические данные (например, изображение). Дискриминатор, в свою очередь, получает на вход как реальные данные из тренировочного набора, так и сгенерированные образцы, и обучается отличать одни от других [2]. Цель генератора – обмануть дискриминатор, создавая максимально реалистичные данные, в то время как цель дискриминатора – стать идеальным классификатором. Этот процесс

описывается как минимаксная игра, в ходе которой обе сети одновременно улучшают свои характеристики.

Первые архитектуры, такие как Deep Convolutional GAN (DCGAN), заложили базовые принципы устойчивого обучения, используя сверточные слои и батч-нормализацию [3]. Однако они страдали от известных проблем: нестабильности обучения (поиск равновесия Нэша в высокоразмерном пространстве параметров является сложной задачей) и коллапса моды, когда генератор начинает производить одно или несколько ограниченных разнообразием изображений, «обманывая» дискриминатора. Преодоление этих ограничений стало драйвером эволюции GAN.

Последующие разработки, такие как Wasserstein GAN (WGAN), заменили функцию потерь оригинального GAN (потерю кросс-энтропии) на расстояние Вассерштейна, что позволило получить более стабильные градиенты и осмысленную метрику качества в процессе обучения [4]. Архитектура StyleGAN совершила качественный скачок в генерации лиц, введя концепцию стилевого пространства, которое позволяет независимо контролировать различные аспекты генерируемого изображения (прическу, позу, мимику) на разных масштабах, и используя отображение латентного пространства для более равномерного его покрытия [5]. Это позволило не только повысить реалистичность, но и обеспечить беспрецедентный уровень контроля над процессом синтеза.

2. Прикладные области и практическая реализация технологии GAN

Мощность GAN-архитектур нашла применение в широком спектре задач:

Генерация фотореалистичного контента: Создание изображений людей, животных, ландшафтов, интерьеров и предметов, не существующих в реальности. Это активно используется в дизайне, концепт-арте, рекламе и игровой индустрии для быстрого прототипирования визуальных концепций.

Увеличение разрешения и качества изображений (Super-Resolution): Технология SRGAN позволяет детально восстанавливать высокочастотные текстуры и детали на сильно увеличенных изображениях, что превосходит по

качеству традиционные интерполяционные методы [6]. Это применяется в медицине (улучшение качества снимков), реставрации архивного видео, работе с фотографиями.

Трансфер стиля и модификация изображений: GAN позволяют изменять стиль изображения (например, превращать фотографию в картину в стиле Ван Гога), менять время года на снимке, добавлять или удалять объекты (в том числе и в видео), редактировать атрибуты лиц (возраст, эмоции, аксессуары).

Синтез данных для обучения других моделей ИИ: В задачах, где сбор реальных данных дорог, сложен или сопряжен с этическими проблемами (например, медицинские изображения редких патологий или данные с датчиков автономных автомобилей для аварийных ситуаций), GAN используются для генерации дополнительных синтетических тренировочных наборов, что повышает точность и робастность моделей [7].

Научные и медицинские приложения: Генерация молекулярных структур с заданными свойствами для фармакологии, синтез астрономических изображений, улучшение качества медицинской визуализации (МРТ, КТ) для более точной диагностики.

3. Этические, социальные и регуляторные вызовы технологии Deepfake

Наиболее острые дискуссии вокруг GAN связаны с их использованием для создания deepfake-контента – гиперреалистичных поддельных изображений, аудио- и видеозаписей, в которых люди совершают действия или произносят слова, которых не было в реальности.

Технически deepfake чаще всего реализуется на основе архитектур типа Autoencoder или GAN, которые обучаются подменять лицо и/или голос целевого человека на исходный контент [8]. Доступность opensource-библиотек и сервисов резко снизила порог входа для создания такого контента, породив ряд серьезных угроз:

Угроза личной репутации и приватности: Создание компрометирующих материалов с лицом конкретного человека.

Дестабилизация общественно-политической жизни: Генерация фейковых заявлений публичных лиц или новостных репортажей, что может манипулировать общественным мнением, влиять на ход выборов или провоцировать конфликты.

Криминальное использование: Мошенничество с использованием синтезированного голоса и видео (например, для авторизации в системах или вымогательства).

Эрозия доверия к цифровым медиа: Размывание понятия объективной истины, формирование среды тотального скептицизма, где любое доказательство может быть объявлено подделкой («эффект Липпмана»).

Эти вызовы делают актуальным разработку комплексных мер противодействия:

Технические меры (детекция): Развитие методов обнаружения deepfake, основанных на анализе артефактов генерации (несовершенства в моргании глаз, неестественные движения губ, артефакты в области ушей, несоответствия в освещении), с использованием, в том числе, тех же GAN и других моделей глубокого обучения [9].

Правовые и регуляторные меры: Формирование законодательной базы, криминализирующей создание и распространение malicious deepfake, разработка стандартов цифрового водяного знака и аутентификации контента (например, на базе блокчейна).

Медиаграмотность и образование: Просветительская работа с населением по распознаванию потенциально фейкового контента.

Заключение

Генеративно-состязательные сети представляют собой мощный и универсальный инструмент, открывший новую эру в синтезе и обработке цифровых данных. Их архитектурное развитие от базовых DCGAN до контролируемых стилевых моделей демонстрирует впечатляющий прогресс в достижении фотореалистичности и управляемости. Широта прикладных

областей – от креативных индустрий до фундаментальной науки – подтверждает их трансформационный потенциал.

Однако сила технологии одновременно является источником ее главных рисков. Проблема deepfake стала ярким примером того, как инновации могут быть обращены против общества. Таким образом, дальнейшее развитие и внедрение GAN должно носить сбалансированный характер. Приоритетами должны стать не только достижение новых технических рубежей, но и активное формирование этических рамок, развитие технологий детекции, совершенствование законодательства и повышение цифровой грамотности. Будущее GAN лежит не в сфере запретов, а в области ответственного и осознанного использования, где потенциал для созидания будет максимально защищен от возможностей для разрушения. Исследования в этом направлении требуют междисциплинарного подхода, объединяющего специалистов в области ИИ, права, социологии и политологии.

Список литературы

1. Goodfellow I., Pouget-Abadie J., Mirza M. et al. Generative Adversarial Networks // Communications of the ACM. 2020. Vol. 63, No. 11. P. 139–144.
2. Генеративно-сопоставительные сети (GAN) для начинающих // Хабр. 2020. URL: <https://habr.com/ru/companies/wunderfund/articles/529436/> (дата обращения: 17.01.2026).
3. Radford A., Metz L., Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks // arXiv preprint. 2015. arXiv:1511.06434. URL: <https://arxiv.org/abs/1511.06434> (дата обращения: 17.01.2026).
4. Arjovsky M., Chintala S., Bottou L. Wasserstein GAN // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. Vol. 70. P. 214–223.

5. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks // arXiv preprint. 2018. arXiv:1812.04948. URL: <https://arxiv.org/abs/1812.04948> (дата обращения: 17.01.2026). (Исправлено: указан год первоначального препринта 2018 вместо журнальной версии 2021 года, так как это более корректно для цитирования базовой архитектуры).
6. Ledig C., Theis L., Huszár F. et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 4681–4690.
7. Frid-Adar M., Klang E., Amitai M. et al. Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification // IEEE 15th International Symposium on Biomedical Imaging (ISBI). 2018. P. 289–293.
8. Что такое Deepfake: технология, риски и методы борьбы // SecurityLab. 2023. URL: <https://www.securitylab.ru/analytics/525199.php> (дата обращения: 17.01.2026).
9. Rossler A., Cozzolino D., Verdoliva L. et al. FaceForensics++: Learning to Detect Manipulated Facial Images // IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 1–11.
10. Этика искусственного интеллекта: основные принципы и проблемы // ТАСС. Наука. 2022. URL: <https://tass.ru/nauka/15713825> (дата обращения: 17.01.2026).

References

1. Goodfellow I., Pouget-Abadie J., Mirza M. et al. Generative Adversarial Networks // Communications of the ACM. 2020. Vol. 63, No. 11. P. 139–144.
2. Generative Adversarial Networks (GAN) for Beginners // Habr. 2020. URL: <https://habr.com/ru/companies/wunderfund/articles/529436/> (accessed: 17.01.2026).
3. Radford A., Metz L., Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks // arXiv

- preprint. 2015. arXiv:1511.06434.
URL: <https://arxiv.org/abs/1511.06434> (accessed: 17.01.2026).
4. Arjovsky M., Chintala S., Bottou L. Wasserstein GAN // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. Vol. 70. P. 214–223.
 5. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks // arXiv preprint. 2018. arXiv:1812.04948. URL: <https://arxiv.org/abs/1812.04948> (accessed: 17.01.2026).
 6. Ledig C., Theis L., Huszár F. et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 4681–4690.
 7. Frid-Adar M., Klang E., Amitai M. et al. Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification // IEEE 15th International Symposium on Biomedical Imaging (ISBI). 2018. P. 289–293.
 8. What is Deepfake: Technology, Risks and Countermeasures // SecurityLab. 2023. URL: <https://www.securitylab.ru/analytics/525199.php> (accessed: 17.01.2026).
 9. Rossler A., Cozzolino D., Verdoliva L. et al. FaceForensics++: Learning to Detect Manipulated Facial Images // IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 1–11.
 10. AI Ethics: Basic Principles and Problems // TASS. Science. 2022. URL: <https://tass.ru/nauka/15713825> (accessed: 17.01.2026).