УДК 004.056.55

Журавлева Марина Олеговна, студентка, Российский технологический университет Московский институт радиотехники, электроники и автоматики, г. Москва

ГЕНЕРАТИВНЫЕ МОДЕЛИ В КОМПЬЮТЕРНОМ ЗРЕНИИ: ПОТЕНЦИАЛЫ И УГРОЗЫ В ЦИФРОВОЙ ЭПОХЕ

Аннотация: В данной работе рассматриваются теоретические и прикладные аспекты использования генеративных моделей в области компьютерного зрения в условиях цифровой эпохи. Особое внимание уделяется анализу архитектур генеративно-состязательных сетей (GAN), вариационных автокодировщиков (VAE) и диффузионных моделей как ключевых инструментов синтеза визуального контента. Проведён обзор актуальных областей применения, включая восстановление изображений, генерацию синтетических данных, суперразрешение, медиавизуализацию и межмодальные преобразования. Отдельно выделены вызовы, связанные с распространением поддельного контента (deepfake), угрозами информационной безопасности, правовыми и этическими дилеммами.

Ключевые слова: генеративные модели; компьютерное зрение; GAN; deepfake; синтетические изображения; информационная безопасность; визуальная дезинформация; диффузионные модели; этика ИИ; цифровое регулирование.

Abstract: This paper explores the theoretical and practical aspects of using generative models in the field of computer vision within the context of the digital age. Special attention is given to the analysis of the architectures of Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models as key tools for visual content synthesis. A review of current application areas is provided, including image restoration, synthetic data generation, super-resolution, media visualization, and cross-modal transformations. The study

highlights challenges associated with the spread of fake content (deepfakes), threats to information security, as well as legal and ethical dilemmas.

Keywords: generative models; computer vision; GAN; deepfake; synthetic images; information security; visual misinformation; diffusion models; AI ethics; digital regulation.

В условиях стремительного развития технологий искусственного интеллекта (ИИ) особое внимание привлекают генеративные модели, представляющие собой методы машинного обучения, способные создавать новые данные, имитирующие исходные распределения. Эти модели, в первую очередь используемые в задачах компьютерного зрения, обеспечивают беспрецедентную гибкость в синтезе изображений, видео и трёхмерных объектов, что открывает широкие горизонты как в научной, так и в прикладной сферах. Генеративные модели уже сегодня трансформируют методы визуального анализа и синтеза, находя применение в медицине, промышленности, криминалистике, а также в индустрии развлечений и искусства. Вместе с тем, их активное внедрение сопряжено с целым рядом угроз: от фальсификации медиа-контента (deepfakes) и атак на системы биометрической идентификации до правовых и этических вызовов, связанных с авторством и ответственностью.

Генеративные модели представляют собой одну из наиболее динамично развивающихся областей в современной архитектуре искусственного интеллекта, где ключевая задача заключается не только в распознавании и интерпретации данных, но и в их воспроизведении — то есть в синтезе новых образов, обладающих статистическими свойствами исходного обучающего множества. В контексте компьютерного зрения, где визуальные данные приобретают всё большее значение в задачах автоматизации, медицины, робототехники, цифровой экономики и информационной безопасности, генеративные модели становятся неотъемлемым элементом

интеллектуальных вычислительных систем. Исторически развитие генеративных подходов началось с вариационных автокодировщиков (VAE), в которых реализуется вероятностный принцип латентного кодирования изображений с последующим восстановлением их на основе аппроксимированной функции распределения. Эти модели демонстрируют стабильные результаты, однако в условиях задач, требующих высокого визуального качества, уступают более поздним архитектурам. Качественный прорыв произошёл в 2014 году с появлением генеративно-состязательных сетей (GAN), предложенных Иэном Гудфеллоу и его коллегами. В GAN используется парадигма состязательного взаимодействия двух нейросетевых агентов: генератора, формирующего новые образы, и дискриминатора, оценивающего степень их достоверности по отношению к реальным данным. Благодаря этой архитектуре удалось добиться значительного улучшения фотореалистичности синтезируемых изображений, что позволило внедрять такие системы в области фоторедактирования, генерации лицевых изображений, а также в художественном синтезе. В дальнейшем эволюция генеративных методов продолжилась с разработкой диффузионных моделей (Diffusion Models), основанных на принципе пошагового зашумления данных и их постепенного восстановления при помощи обратного стохастического процесса. Эти модели, в частности DALL·E 2, Imagen и Stable Diffusion, демонстрируют способность к высокоточной генерации изображений по текстовому описанию, что знаменует собой переход от визуально ориентированного генеративного моделирования к мультимодальному взаимодействию, объединяющему визуальные и семантические представления.

Применение генеративных моделей в задачах компьютерного зрения многогранно и охватывает широкий спектр прикладных направлени[1]. Одной из ключевых задач является восстановление утерянной или повреждённой

[1] Гудфеллоу И., Бенжио Й., Курвиль А. Глубокое обучение. — М.: Диалектика, 2018. — 720 с.

информации на изображениях (image inpainting), что особенно актуально в криминалистике и реставрации цифровых архивов. Кроме того, значительное внимание уделяется суперразрешению (super-resolution) — синтезу изображений с повышенным пространственным разрешением, что позволяет использовать такие подходы в диагностической визуализации, спутниковом мониторинге и интеллектуальном видеонаблюдении. Генеративные модели также успешно применяются для синтеза медицинских изображений, включая магнитно-резонансную томографию и компьютерную томографию, где возможность генерации реалистичных и разнообразных данных позволяет существенно повысить обучаемость диагностических систем при дефиците размеченных выборок.

Одним из наиболее перспективных направлений остаётся синтетическая аугментация данных для повышения качества обучения сверточных нейронных сетей. В условиях, когда сбор и аннотирование больших объёмов данных сопряжено с существенными затратами, генеративные модели позволяют формировать высокоразнообразные, но контролируемые выборки изображений, что особенно ценно при обучении моделей в узкоспециализированных прикладных доменах, таких как агрономия, промышленное зрение или патоморфология.

Нельзя не отметить влияние генеративных моделей на общий вектор развития архитектур искусственного интеллекта. Внедрение диффузионных механизмов и трансформерных архитектур открывает путь к моделям нового поколения, обладающим способностью к контекстному управлению генерацией и формированию визуального содержания на основе сложных лингвистических конструкций. Такое объединение семантического и визуального пространств радикально меняет представление о возможностях ИИ в области генерации образов и требует пересмотра не только технических стандартов, но и этико-правовых рамок использования подобных технологий.

Будучи по своей природе обучающимися системами – генеративные модели, ориентированны на синтез новых данных, находят всё более широкое применение в самых различных сферах компьютерного зрения[2]. Их потенциал заключается не только в способности к воссозданию реальности, но и в формировании новых, ранее не существовавших визуальных форм, что открывает кардинально иные парадигмы в обработке изображений, визуальной аналитике и интерфейсных решениях.

Одним из важнейших направлений применения генеративных моделей является восстановление и дополнение изображений. В тех случаях, когда визуальная информация частично утрачена, искажается или фрагментирована, модели типа GAN и diffusion могут достраивать недостающие области, основываясь на латентных закономерностях обучающего множества. Эта способность становится особенно востребованной в задачах криминалистической реконструкции, реставрации архивных документов, а также в медицинской визуализации, где зачастую приходится иметь дело с неполными снимками, вызванными шумами, артефактами или ограниченными условиями съёмки.

Важной областью применения остаётся генерация синтетических данных. В условиях дефицита размеченных изображений, например, при обучении моделей для редких заболеваний, в геоинформационных системах или в задачах промышленного контроля, генеративные модели позволяют формировать большие массивы достоверных, но синтетических изображений. Эти данные существенно увеличивают объём обучающих выборок, снижая риск переобучения и повышая обобщающую способность систем. В частности, доказано, что использование синтетических изображений,

---

[2] Сергеев А.Ю., Вишняков В.А. Глубокое обучение в задачах компьютерного зрения. — СПб.: БХВ-Петербург, 2020. — 384 с.

сгенерированных при помощи GAN, может повышать точность классификаторов на десятки процентов при ограниченной реальной выборке.

Существенное внимание в последние годы уделяется применению генеративных моделей в медицинских задачах. Визуальные данные, полученные из таких источников, как МРТ, КТ или ультразвуковое исследование, нуждаются в расширенной интерпретации. Генеративные архитектуры, обученные на медицинских выборках, позволяют не только синтезировать новые изображения, но и выполнять межмодальные преобразования, например, трансформировать МР-снимок в КТ-аналог, что особенно ценно при ограниченном доступе к многомодальной диагностике[3].

Не менее значимым направлением является применение генеративных моделей в области цифрового дизайна, искусства и медиа. Современные архитектуры, такие как DALL·E или MidJourney, способны по текстовому описанию создавать высококачественные иллюстрации, анимации и графику. Это открывает возможности для автоматизированного проектирования интерфейсов, архитектурных макетов, генерации рекламных материалов и даже создания цифровых двойников людей, что уже находит коммерческое применение в игровой индустрии и в сфере виртуальных помощников.

Генеративные модели также играют важную роль в задачах повышения качества изображений. Речь идёт не только о классическом super-resolution, когда изображение восстанавливается в более высоком разрешении, но и о таких задачах, как устранение шума, коррекция экспозиции, цветовой перенос и реконструкция визуальной информации на основе ограниченного набора пикселей. Эти методы оказываются особенно полезными в условиях плохого

---

[3] Петухов В.Е., Балыкин А.В. Генеративные состязательные сети в задачах синтеза изображений // *Вестник ВГУ. Серия: Системный анализ и информационные технологии*. — 2021. — №2. — С. 42–50.

освещения, съёмки с мобильных устройств или камер видеонаблюдения, а также в космической съёмке, где технические ограничения диктуют необходимость последующей генеративной обработки.

Значительный потенциал генеративных моделей прослеживается в развитии систем человеко-машинного взаимодействия. Объединяя визуальный синтез с языковыми моделями, современные генеративные архитектуры становятся основой для интерфейсов нового поколения: голосовых ассистентов, способных интерпретировать и визуализировать запросы пользователя, умных систем проектирования, генераторов видеоконтента на основе сценариев. Такие системы постепенно формируют так называемое креативное ИИ-пространство, где границы между технической реализацией и художественным воображением оказываются стираемыми.

Несмотря на значительный инновационный потенциал генеративных моделей, их широкомасштабное внедрение в цифровую среду сопровождается целым спектром угроз, приобретающих всё более актуальный и системный характер. Парадоксально, но технологии, изначально разработанные для расширения возможностей анализа и синтеза визуальных данных, сегодня становятся источником дестабилизации информационного пространства, подрывая доверие к визуальным доказательствам, способствуя распространению дезинформации и провоцируя этические, юридические и социальные конфликты[4].

Одной из ключевых угроз, сопряжённых с использованием генеративных моделей, является распространение поддельного визуального контента, получившего наименование deepfake. Технологии глубокого подражания, основанные на GAN и аналогичных архитектурах, позволяют с высокой степенью фотореализма моделировать лица, голоса и движения

[4] Смирнов К.Л., Мешков А.Ю. Риски и вызовы использования генеративных моделей в цифровом обществе // *Информационное общество*. — 2022. — №3. — С. 57–63.

людей, формируя видеозаписи, неотличимые от реальных. Подобные материалы уже использовались в политических манипуляциях, шантаже, фейковых новостях и даже в подмене доказательств в судебных разбирательствах. В этом контексте особую озабоченность вызывает то, что возможности создания дезинформации становятся доступны не только государственным или корпоративным структурам, но и отдельным пользователям, не обладающим высокой технической квалификацией.

Второй существенной проблемой становится подрыв доверия к визуальному восприятию как таковому. В условиях, когда граница между реальным и синтетическим изображением стирается, возникает феномен информационного релятивизма, в котором любое изображение может быть поставлено под сомнение. Это, в свою очередь, ставит под угрозу функционирование журналистики, научной верификации, судебной экспертизы и других институтов, опирающихся на достоверность визуальных свидетельств. В обществе формируется так называемая «кризисная оптика», когда даже правдивый визуальный материал начинает восприниматься с недоверием, а это открывает пространство для манипуляции массовым сознанием.

Не менее значимыми являются и этические дилеммы, возникающие в связи с автономной продукцией визуального контента. Вопрос об авторстве и интеллектуальной собственности приобретает новую форму: кто является владельцем изображения, созданного генеративной моделью — разработчик модели, пользователь, сгенерировавший запрос, или сама модель как функциональная система? Прецеденты, связанные с отказом ведущих арт-платформ принимать ИИ-созданные изображения для участия в конкурсах, демонстрируют, что общество пока не готово к однозначному решению этой проблемы. Более того, в случаях, когда генеративные модели обучаются на массиве изображений, содержащем произведения, защищённые авторским

правом, возникает риск скрытого плагиата, против которого действующее законодательство оказывается бессильным.

Отдельного внимания заслуживают риски, связанные с использованием генеративных моделей в кибербезопасности. Возможность генерации поддельных изображений документов, лиц для обхода биометрической верификации, а также фальшивых логотипов и интерфейсов создаёт угрозу для защищённых систем и может использоваться в рамках фишинговых атак или инженерного социального воздействия. Подобные атаки уже зафиксированы в банковской сфере, в системе онлайн-голосования и в области идентификации персонала.

Кроме того, генеративные модели поднимают вопрос о цифровом неравенстве и доступе к технологии. Расширение возможностей синтеза визуального контента, сопряжённое с высокой вычислительной стоимостью (особенно в случае диффузионных моделей), создаёт неравномерный доступ к технологиям: крупные корпорации и государственные структуры получают в своё распоряжение средства влияния, недоступные малым организациям и частным лицам. Это усугубляет дисбаланс в сфере цифровой власти и может способствовать технологической монополизации визуального пространства

Активное распространение генеративных моделей в цифровой среде, сопровождаемое ростом числа злоупотреблений, провоцирует необходимость выработки эффективных регуляторных механизмов. Технологическое сообщество, юристы, государственные институты и этические комитеты сегодня сталкиваются с задачей установления баланса между инновационным развитием и защитой общественных интересов, частной жизни, информационной безопасности и прав личности. Многоуровневая природа рисков, описанных в предыдущей главе, требует комплексного подхода, включающего как технические средства противодействия, так и нормативные инициативы на национальном и международном уровнях.

Одним из наиболее активно развивающихся направлений становится детектирование синтетического контента, целью которого является выявление изображений и видеозаписей, созданных с использованием генеративных моделей. Здесь применяются как классические методы цифровой криминалистики, основанные на выявлении аномалий сжатия, артефактов обработки и неестественных особенностей текстуры, так и методы машинного обучения, способные классифицировать изображения на основе латентных признаков, ускользающих от человеческого глаза. В частности, нейросети, обученные на задачах бинарной классификации “реальное/синтетическое”, демонстрируют высокий уровень точности при идентификации deepfake-контента, однако эффективность этих методов существенно снижается при появлении новых, более совершенных генераторов[5].

Особую роль в борьбе с недобросовестным использованием генеративных моделей играет политика цифровой прозрачности, предполагающая открытость алгоритмов, аудит выборок данных и обеспечение воспроизводимости генеративных систем. Некоторые технологические компании начали внедрение watermarking-систем — невидимых цифровых меток, внедряемых в изображения и видео с целью последующего отслеживания их происхождения. Однако на текущем этапе данные методы сталкиваются с рядом технических и юридических ограничений: во-первых, маркеры могут быть легко удалены при минимальной постобработке; во-вторых, единых стандартов маркировки пока не существует, а их внедрение требует согласованных действий на международном уровне.

---

[5] Калачев С.И. Deepfake как феномен цифровой эпохи: правовые и социальные аспекты // *Юридическая информатика*. — 2021. — №1. — С. 31–38.

Параллельно с мерами контроля важное значение приобретает формирование культуры ответственного ИИ, предполагающей осознанное проектирование генеративных систем с учётом потенциальных социальных последствий. Ведущие исследовательские институты и корпорации вводят этические кодексы, согласно которым разработка ИИ должна сопровождаться оценкой рисков, анализом возможных злоупотреблений и механизмами отказа от внедрения систем, способных нанести вред. В некоторых случаях предлагается концепция “контролируемой генерации”, когда пользователь, получая доступ к модели, ограничен в типах запросов или получает визуальный фидбэк о границах допустимого

Развитие генеративных моделей в контексте компьютерного зрения представляет собой одно из наиболее значительных достижений современной науки о данных, демонстрируя потенциал к радикальному преобразованию процессов визуального анализа, синтеза и взаимодействия человека с цифровой средой. Архитектуры, основанные на генеративно-состязательных сетях, вариационных автокодировщиках и диффузионных процессах, наделяют искусственный интеллект способностью к продуктивному воспроизводству информации, выходящему за рамки традиционных алгоритмических парадигм.

Задача будущего состоит не в отказе от генеративных моделей, а в создании инфраструктуры прозрачности, доверия и устойчивости, при которой интеллектуальные технологии смогут служить общественному благу, не становясь источником деструктивных трансформаций. Только в условиях такого баланса можно будет говорить о гармоничном сосуществовании человека и генеративного ИИ в сложной и противоречивой цифровой реальности XXI века.

Литература

1. Гудфеллоу И., Бенжио Й., Курвиль А. Глубокое обучение. — М.: Диалектика, 2018. — 720 с.
2. Сергеев А.Ю., Вишняков В.А. Глубокое обучение в задачах компьютерного зрения. — СПб.: БХВ-Петербург, 2020. — 384 с.
3. Петухов В.Е., Балыкин А.В. Генеративные состязательные сети в задачах синтеза изображений // *Вестник ВГУ. Серия: Системный анализ и информационные технологии*. — 2021. — №2. — С. 42–50.
4. Смирнов К.Л., Мешков А.Ю. Риски и вызовы использования генеративных моделей в цифровом обществе // *Информационное общество*. — 2022. — №3. — С. 57–63.
5. Калачев С.И. Deepfake как феномен цифровой эпохи: правовые и социальные аспекты // *Юридическая информатика*. — 2021. — №1. — С. 31–38.

Literature

1. Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. — Moscow: Dialektika, 2018. — 720 p.
2. Sergeev, A.Yu., & Vishnyakov, V.A. *Deep Learning in Computer Vision Tasks*. — St. Petersburg: BHV-Petersburg, 2020. — 384 p.
3. Petukhov, V.E., & Balykin, A.V. Generative Adversarial Networks in Image Synthesis Tasks // *Vestnik of Voronezh State University. Series: Systems Analysis and Information Technologies*. — 2021. — No. 2. — Pp. 42–50.
4. Smirnov, K.L., & Meshkov, A.Yu. Risks and Challenges of Using Generative Models in the Digital Society // *Information Society*. — 2022. — No. 3. — Pp. 57–63.
5. Kalachev, S.I. Deepfake as a Phenomenon of the Digital Age: Legal and Social Aspects // *Legal Informatics*. — 2021. — No. 1. — Pp. 31–38.