ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Ушакова Егора Николаевича «Метод условной генерации изображений в диффузионных моделях со структурированным текстовым условием», представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Работа Ушакова Е. Н. посвящена актуальной задаче повышения управляемости условной генерации изображений в диффузионных моделях. В современных text-to-image моделях качество и воспроизводимость результата в значительной мере определяются формой представления текстового условия. Свободное описание содержит повторы, отрицания, сравнительные и анамнестические конструкции и фрагменты без визуального соответствия, а ограниченная длина входа текстового энкодера (77 токенов в Stable Diffusion 1.5) приводит к потере части значимой информации. Эти трудности особенно выражены для радиологических отчетов. Целью исследования является разработка метода формирования и использования структурированного текстового условия, позволяющего выделить визуально обоснованные признаки без изменения архитектуры диффузионной модели, его программная реализация и экспериментальная валидация. Разработка такого метода имеет научную и практическую значимость.

Содержание диссертации и автореферата. Диссертация изложена на 119 страницах, состоит из введения, трех глав, заключения и списка литературы из 73 наименований, содержит 9 рисунков и 22 таблицы.

Во введении обоснована актуальность темы, сформулированы цель и задачи исследования, научная новизна, теоретическая и практическая значимость, три положения, выносимые на защиту; приведены сведения об апробации, публикациях и личном вкладе автора.

Первая глава содержит обзор генеративных моделей и механизмов задания условия в латентных диффузионных моделях, анализ радиологических отчетов как источника условия и недостатков неструктурированного текста (синонимия, отрицания, неопределенность, сравнительный контекст, ограничение длины), сравнение форм структурированного представления (набор меток, список находок, граф сущностей) и методов оценки синтетических медицинских изображений. Сформулированы требования к разрабатываемому методу.

Во второй главе предложен метод преобразования текстового описания в структурированный список клинических находок: введена формальная модель отчета, находки как кортежа (тип, статус, локализация, выраженность) и списка находок; описаны выделение релевантной секции, нормализация, извлечение сущностей, обработка отрицаний и неопределенности, фильтрация, объединение дубликатов и сериализация в формат finding | location | presence | severity. Доказаны семь лемм и три теоремы: при фиксированных параметрах схемы сложность формирования списка линейно ограничена длиной отчета, объем хранения — числом находок, а замена условия не изменяет асимптотическую сложность обучения и генерации Stable Diffusion. Представлена архитектура программного комплекса.

В третьей главе проведено сравнение исходного и структурированного условий на полном наборе MIMIC-CXR при единой архитектуре Stable Diffusion 1.5 и одинаковых гиперпараметрах. Исследованы компактность условия, качество генерации (MSE, FID, KID), полезность синтетических данных для обучения классификаторов с проверкой на независимых данных CheXpert и три аспекта конфиденциальности (близость к обучающим изображениям,

membership inference, риск повторной идентификации пациента). Дополнительный эксперимент на VinDr-Mammo посвящен переносу принципа структурированного условия на маммографию.

В заключении сформулированы основные результаты исследования. Автореферат в целом отражает содержание диссертации и ее основные положения.

Новизна и значимость результатов.

Научная новизна работы состоит, во-первых, в методе разложения неструктурированного текстового условия, в котором клиническое утверждение представляется кортежем из типа признака, статуса, локализации и выраженности, а политика обработки отрицаний, неопределенных и сравнительных формулировок и нерелевантных фрагментов зафиксирована явно и единообразно применяется при обучении и генерации. В отличие от известных работ по генерации рентгенограмм (RoentGen и др.), где условием служит секция отчета в свободной форме, и от подходов с фиксированным набором меток (CheXpert, CheXbert), представление сохраняет переменное число находок с локализацией и выраженностью и передается через штатный текстовый интерфейс без изменения архитектуры, функции потерь и процедуры сэмплирования. Во-вторых, метод доведен до формально описанных алгоритмов с оценками сложности, что соответствует профилю специальности 2.3.5. В-третьих, экспериментально установлено, что структурирование сокращает долю условий, превышающих ограничение энкодера, с 38,2 % до 8,0 %, улучшает FID (64,16 → 61,25) и KID (0,0185 → 0,0159) и downstream-показатели, а также дополнительно снижает показатели повторной идентификации пациента по синтетическим изображениям (Top-1 0,0135 → 0,0091; устойчивость подтверждена парным бутстрэпом), то есть выступает и как фактор защиты конфиденциальности.

Теоретическая значимость определяется формализацией преобразования текстового условия и доказательством инвариантности асимптотической сложности обучения и генерации при ограниченной длине текстового входа. Вместе с тем полученные оценки носят преимущественно счетный характер и опираются на допущение об ограниченном числе операций на токен; их глубина соответствует скорее строгому обоснованию применимости метода, чем самостоятельному математическому результату (см. замечание 5). Практическая значимость заключается в создании воспроизводимого программного комплекса, объединяющего подготовку условий, обучение, генерацию и многоуровневую оценку синтетических изображений, и в показанной возможности использовать синтетические данные для расширения обучающих выборок без роста рисков конфиденциальности в рассмотренных постановках.

Достоверность основных положений обеспечивается контролируемым сравнением (единые архитектура, данные, разбиение, гиперпараметры и начальные состояния генератора случайных чисел), использованием общепринятых метрик и независимой тестовой выборки CheXpert, положительным контролем и бутстрэп-анализом в privacy-экспериментах, а также корректно сформулированными автором ограничениями. Основные результаты представлены в четырех публикациях в изданиях, рекомендованных ВАК, и апробированы на профильных конференциях.

К сильным сторонам работы можно отнести:

1. Корректную постановку контролируемого эксперимента, в котором единственным варьируемым фактором является форма текстового условия, что позволяет интерпретировать различия именно как эффект структурирования.

2. Архитектурную нейтральность метода: структурирование выполняется как этап предобработки и переносимо на любую text-to-image модель со стандартным текстовым интерфейсом.
3. Клинически продуманный дизайн представления: раздельная обработка отрицаний, неопределенных формулировок (статус probable), разрешившихся и хронических изменений, технических и сравнительных фрагментов.
4. Масштаб и комплексность оценки: полный MIMIC-CXR для обучения, внешний CheXpert для downstream-оценки, три взаимодополняющих privacy-эксперимента с положительным контролем и бутстрэпом; аккуратная интерпретация без преувеличения гарантий конфиденциальности.
5. Воспроизводимость и научную добросовестность: задокументированы версии, гиперпараметры и начальные состояния; автор сам указывает на малость различий в downstream-метриках и непрямой характер метрик по отношению к клинической корректности.

Замечания

К содержанию и оформлению работы имеются следующие замечания и пожелания:

1. **Качество преобразования FINDINGS → список находок не оценено.** Требования Grounded, Format и NoDuplicate сформулированы как обязательные, однако точность и полнота извлечения находок, корректность полярности и локализации не измерены ни на размеченной вручную подвыборке, ни в сопоставлении с имеющимися для MIMIC-CXR аннотациями CheXbert и RadGraph. Ошибки извлечения напрямую переносятся в условие генерации.
2. **Не измерено соответствие сгенерированных изображений заданному условию.** Целью работы заявлено повышение управляемости генерации, однако оценка ограничена распределительными метриками и downstream-полезностью, характеризующими это свойство косвенно. Применение к синтетическим изображениям классификатора патологий, обученного на реальных данных, и сравнение его предсказаний с находками в условии (как в RoentGen) было бы наиболее прямым подтверждением основного тезиса.
3. **Статистическая обоснованность и полнота сравнения.** Различия по FID, KID и особенно по downstream-метрикам (Macro ROC AUC 0,6205 против 0,6252) получены в единственном запуске без доверительных интервалов; абсолютный уровень downstream-показателей невысок (Macro F1 около 0,16, обучение на 2348 синтетических изображениях), а референсный классификатор на реальных данных того же объема отсутствует. Предложенное представление сопоставлено только с секцией FINDINGS: не рассмотрены секция IMPRESSION, сериализованные метки CheXpert, простое усечение текста до той же длины, а обсуждаемые в главе 1 решения (порядок полей, разделители, включение отрицательных находок) не проверены абляцией. В результате вклад разложения на находки не отделен от эффекта простого сокращения длины условия.

   Рекомендуется учесть данное замечание в будущей работе в случае ее продолжения.
4. **К теоретической части.** Леммы 1–7 и теоремы 1–3 корректны, но во многом непосредственно следуют из допущений об ограниченности d, C, a и S; для нейросетевого экстрактора допущение о постоянном числе операций на токен требует отдельной оговорки об ограниченной длине контекста. Содержательно более интересные утверждения — о выполнении алгоритмом требований Grounded, Format, NoDuplicate —

не сформулированы. Список находок вводится как множество $A = \{a_1, \ldots, a_m\}$, тогда как при сериализации трактуется как упорядоченная последовательность, причем порядок, по словам автора, значим.

5. **Эксперимент на маммографических данных (раздел 3.7).** Условие строится из готовых атрибутов исследования, алгоритм разложения текста не применяется, поэтому эксперимент подтверждает переносимость принципа, но не метода; формулировки результата 6 следовало бы уточнить. Не указаны конфигурации генеративной модели и классификатора и число положительных примеров в тесте (по значениям Recall — около двух десятков), так что различия ROC AUC 0,9406 → 0,9494 находятся в пределах погрешности. Раздел желательно дополнить рисунком с примерами синтезированных маммограмм и структурированных условий. Также не представлена оценка изображений врачами-рентгенологами; даже ограниченная слепая оценка усилила бы выводы о клинической правдоподобности.
6. **По оформлению.** Обучение выполнено при разрешении 256 × 256, что ниже штатного для Stable Diffusion 1.5 и следовало бы оговорить в ограничениях. В тексте встречаются отдельные опечатки (например, «над элементом вход», с. 76).

Указанные замечания в значительной части носят характер пожеланий к дальнейшему развитию исследования, не затрагивают основных научных результатов и не снижают общей положительной оценки диссертации.

Диссертация Ушакова Егора Николаевича «Метод условной генерации изображений в диффузионных моделях со структурированным текстовым условием» является завершенной научно-квалификационной работой, в которой решена научная задача формирования и использования структурированного текстового условия в диффузионной генерации изображений. Работа соответствует требованиям пункта 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, предъявляемым к кандидатским диссертациям, а также паспорту специальности 2.3.5. Ушаков Егор Николаевич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент
кандидат технических наук,
директор Института искусственного интеллекта
Автономной некоммерческой организации
высшего образования «Университет Иннополис»

Рамиль Фуатович Кулеев

«28» октября 2026 г.