

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Гуманитарно-педагогический институт
(наименование института полностью)

Кафедра Журналистика и социология
(наименование)

42.04.02 Журналистика
(код и наименование направления подготовки)

Журналистика данных
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему Использование нейросетей для генерации дата-контента: современный опыт,
проблемы и перспективы

Обучающийся Луговая Е.Д.
(Инициалы Фамилия) (личная подпись)

Научный кандидат филологических наук, доцент, Л. В. Иванова
руководитель (ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Оглавление

Введение.....	3
Глава 1 Теоретические основы применения нейронных сетей в генерации дата-контента.....	10
1.1 Место нейронных сетей в понятийном аппарате исследуемой области: ключевые термины и взаимосвязи между ними.....	10
1.2 Применение нейросетей для генерации данных в журналистике: возможности, проблемы и вызовы.....	21
Глава 2 Анализ опыта, качественная оценка и методические рекомендации по использованию генеративного контента в журналистике	52
2.1 Анализ современного опыта использования нейронных сетей в сфере журналистики	52
2.2 Результаты качественного анализа генеративного контента в актуальном медиадискурсе	67
Заключение	132
Список используемой литературы и используемых источников.....	136

Введение

В последние годы нейронные сети активно применяются в журналистской практике, особенно в журналистике данных. Являясь важнейшим компонентом современного искусственного интеллекта (ИИ), нейросети охватывают разные подходы к созданию интеллектуальных систем: благодаря способности обучаться на больших наборах данных и решать сложные задачи, они доминируют в области ИИ, достигают значительных результатов в компьютерном зрении и обработке естественного языка. Эти технологии применяются в журналистике данных, позволяют специалистам быстро анализировать и визуализировать большие массивы информации, создавать контент на их основе, выявлять скрытые закономерности в данных.

По данным Института изучения журналистики агентства Рейтер [44], в 2023 году 67 % СМИ использовали искусственный интеллект для создания контента, 28 % из них применяли эту технологию постоянно. Исследование 2024 года [45] показало развитие этой тенденции: приоритетом 56 % медиакомпаний стала автоматизация базовых процессов – почти вдвое больше показателя 2022 года (29 %). Некоторые редакции использовали в работе виртуальных журналистов, которые генерировали до 5 % публикаций. Прогнозы на 2025 год [43] свидетельствуют о еще более глубокой интеграции искусственного интеллекта и нейросетей в журналистскую практику: 96 % респондентов планируют автоматизировать редакционные задачи и сделать это приоритетным направлением своего развития, 87 % опрошенных признают значительное влияние искусственного интеллекта на работу редакций, что указывает на переход от экспериментального применения искусственного интеллекта в журналистике к системному.

Согласно материалам Yandex Cloud [25], применение нейросетей и ИИ в российской медиаиндустрии становится популярным направлением. Нейросети способны создавать не только короткие новости, но и емкие, более сложные материалы. Компания уже предлагает решения [24] для анализа

контента, предсказания предпочтений аудитории и персонализации. Исследования сервиса журналистских запросов Pressfeed за январь [15] и октябрь [6] 2024 года показывают динамику внедрения нейросетей в медиасферу: если в начале года только 19,2 % респондентов отмечали существенное влияние технологии на их работу, то к октябрю уже 79 % выразили положительное отношение, а 65 % активно применяли на практике. Большинство пользователей (72 %) используют нейросети не реже раза в неделю, а 23 % – ежедневно. Ключевыми областями применения технологии стали генерация текстов (71 %) и изображений (51 %). Среди основных преимуществ – ускорение рабочих процессов (51 %) и экономия времени на рутинных задачах (42 %). Эти данные подтверждают стремительное проникновение технологии в медиаиндустрию и актуальность научного исследования. Растущая роль и доступность нейронных сетей для отечественных СМИ подчеркивает актуальность исследований, направленных на изучение возможностей и вызовов применения этих технологий в создании медиаконтента. На фоне все большего включения нейронных сетей в рабочие процессы журналиста (сбор и анализ информации, создание и персонализация контента, распознавание и верификация данных), рассмотрение вопроса использования нейронных сетей в генерации дата-контента представляется особенно актуальным.

Объектом исследования в магистерской диссертации является дата-контент в СМИ.

Предметом – дата-контент, сгенерированный нейронными сетями.

Цель магистерской диссертации: на основе анализа современного дата-дискурса выявить характеристики сгенерированного дата-контента, оценить его с точки зрения роли в решении профессиональных задач, выявить эффективные методы, проблемы и перспективы генерации, а также вызовы и угрозы, связанные с использованием в СМИ сгенерированного контента.

Для достижения поставленной цели необходимо решить следующие задачи:

- систематизировать научные представления о дата-контенте как сегменте медиадискурса, а также об использовании нейросетей для генерации журналистских текстов, основанных на данных;
- оценить современный опыт использования нейронных сетей для создания текстов на основе данных и визуализации больших данных в журналистике;
- на основе анализа современного дата-дискурса в СМИ выявить эффективные и перспективные методы, проблемы и перспективы генерации, потенциальные возможности, риски и ограничения использования нейронных сетей для решения профессиональных задач журналистики;
- разработать рекомендации по выбору методов и подходов к использованию нейронных сетей для анализа и визуализации данных, генерирования дата-текста, а также минимизации возможных рисков и ограничений.

Теоретико-методологическая база исследования включает труды отечественных и зарубежных исследователей, изучающих применение нейронных сетей и искусственного интеллекта в журналистике и журналистике данных. Важной частью для научного обоснования исследования послужили работы Иванова А.Д. [13], [14]. Асташкина А.Г. [3], Аксенова С.В. и Новосельцева В.Б. [1], Шестериных А.М. и Н.О. [28]. Среди зарубежных исследователей разработкой вопроса занимались Р. Чои [35], Ф.М. Саймон [62], А. Кренкер [48], М.Ф. де Лима Сантос [36], Э. Фернандес [39], Л. Илия [42]. Особенно следует отметить вклад в изучаемую область немецкого профессора Ника Ньюмана, который ежегодно подготавливает отчеты о развитии журналистики, включая отдельные обзоры на применение нейросетей в этой области [43], [44], [45].

Для решения задач исследования необходимо были использованы следующие методы:

- общенаучные фундаментальные методы – системный анализ, а также анализ и синтез, которые позволили изучить систему сложных взаимосвязей между журналистикой и нейронными сетями, выделить частности в этой системе, после чего сформулировать общие выводы и рекомендации на основании исходных посылок;
- библиографические методы – для критического изучения источников, отбора и систематизации научных работ, определения степени научной разработанности темы;
- качественный комплексный анализ – для изучения контента, созданного с помощью нейронных сетей;
- структурно-семантический анализ – для выявления диалектической взаимосвязи содержания и формы дата-контента;
- сравнительно-сопоставительный метод для выявления сходств и различий между материалами, созданными журналистами и сгенерированными нейронными сетями;
- прогностический метод – для выявления перспектив и потенциальных возможностей использования нейронных сетей в журналистике.

В качестве эмпирической базы исследования использовались материалы «РБК» – СМИ, открыто заявляющего об использовании нейронных сетей и популяризирующего такие кейсы. В эмпирическую базу также вошел комментарий главного редактора «Краснодарских известий» по результатам использования нейросети в производстве газеты – выводы издательства были проанализированы и классифицированы для подкрепления анализа эмпирического исследования.

Хронологические рамки исследования включают генеративные материалы, подготовленные «РБК»: сетевой выпуск газеты от апреля 2023 года, для производства которого использовались нейросети GigaChat и Kandinsky 2.1; материалы из раздела «Написано ИИ» на «РБК Тренды» за 2024 и 2025 год (в аналогичном количестве и за аналогичный весенний период для корректного сравнения). Эти тексты можно считать редким примером

честности в медиапрактике: сегодня СМИ никак не маркируют материалы, созданные с помощью нейронных сетей, что существенно затрудняет формирование репрезентативной выборки публикаций для анализа.

Гипотеза исследования: нейронные сети, несмотря на их потенциал и скорость развития не способны полностью заменить работу журналиста и могут применяться эффективно только как вспомогательный инструмент в создании медиаконтента с обязательным участием человека в подготовке материалов.

Положения, выносимые на защиту.

Использование нейронных сетей в обработке контента для журналистских материалов трансформирует традиционные методы работы журналистов, и не только создает новые возможности для оптимизации работы и экономии времени, но и требует от специалистов новых профессиональных компетенций, а от редакций – существенных финансовых вложений для внедрения технологии.

Журналистские материалы, созданные с помощью нейронных сетей, уступают по качеству традиционным, и несут в себе риски фактических ошибок, этических нарушений, нехватки уникальности и авторского стиля, требуют дополнительной проверки и редактуры со стороны профессиональных журналистов и редакторов.

Теоретическая значимость исследования заключается в существенном вкладе в развитие вопроса применения нейросетей в журналистике, а также систематизации терминологического аппарата. В работе предпринята попытка объединить разрозненные подходы к определению и соотношению ключевых понятий исследуемой области: предложена целостная понятийная система, которая может использоваться журналистами в работе. Проведен детальный анализ практического использования генеративных моделей в российском СМИ с маркировкой генеративного контента.

Научная значимость определяется критическим анализом возможностей и ограничений нейросетей в создании журналистского контента: исследование

предлагает научно обоснованные рекомендации по составлению промптов и доработке генеративного контента на основе эмпирического анализа генеративных материалов. Результаты формируют методологическую базу для дальнейших исследований интеграции нейросетей в медиапрактику.

Практическая значимость исследования заключается в возможности использования рекомендаций в работе СМИ при использовании готовых решений или внедрении новых разработок нейронных сетей в процессы создания контента. Материалы работы могут быть применены в качестве образовательных для журналистов, которые планируют применять изучаемые технологии в работе.

Магистерская диссертация состоит из введения, двух глав, заключения и списка литературы.

В первой главе «Теоретические основы применения нейронных сетей в генерации дата-контента» проводится комплексный и детальный анализ научных исследований, посвященных теоретическим аспектам использования нейронных сетей в современной журналистике. Особое внимание в разделе уделяется точному определению и концептуальному позиционированию нейронных сетей в понятийном аппарате исследуемой области (искусственный интеллект, машинное обучение, нейросети, глубокое обучение) с основной целью четкого разграничения терминологии, которая в профессиональном и научном дискурсе часто некорректно используется как взаимозаменяемая. В главе также подробно и структурированно рассматриваются потенциальные возможности, существующие проблемы и актуальные вызовы, непосредственно связанные с практическим применением нейросетевых технологий для генерации данных и медиаконтента в современной журналистской практике.

Во второй главе «Анализ опыта, качественная оценка и методические рекомендации по использованию генеративного контента в журналистике» всесторонне исследуется современный практический опыт применения нейронных сетей в различных аспектах журналистской деятельности. Особое

внимание в этом разделе уделяется детальному качественному анализу генеративных медиатекстов на конкретном примере издания «РБК», созданных с использованием современных нейросетевых технологий. Посредством углубленного качественного анализа тщательно исследуются содержательные и формальные аспекты генеративного контента, его функциональная способность эффективно дополнять и грамотно саммаризировать традиционные журналистские материалы, а также разнообразные функциональные возможности такого типа контента в медиасреде. На основе проведенного исследования формулируются конкретные практические рекомендации для профессиональной работы с генеративным контентом, направленные на оптимизацию использования нейросетей в повседневной журналистской практике и минимизацию потенциальных рисков и ошибок.

Глава 1 Теоретические основы применения нейронных сетей в генерации дата-контента

1.1 Место нейронных сетей в понятийном аппарате исследуемой области: ключевые термины и взаимосвязи между ними

В научной литературе, посвященной применению нейросетей в журналистике, можно отметить неточное использование ключевых терминов «искусственный интеллект», «машинное обучение», «искусственная нейронная сеть» и «глубокое обучение» из-за их размытых границ и частой взаимозаменяемости. Использование этих понятий требует более строгого разграничения для формирования четкой единообразной научной базы изучаемой области, так как эти понятия интегрированы в одно поле, но не синонимичны [11], [35].

В некоторых исследованиях практическая журналистика, использующая нейросети, машинные алгоритмы и программное обеспечение, обозначается единым термином «компьютерная журналистика» или «роботизированная журналистика», но также и «автоматизированная журналистика», «вычислительная журналистика» [13], [14]. В работах не всегда обозначаются конкретные аспекты раздела науки о данных, которые исследуются – термины скорее используются как синонимичные для разнообразия текста. Анализируемые методы, предметы и объекты работ из-за этого не интегрированы в общую систему и могут рассматриваться абстрактно, в отрыве от нее – рассматриваться так, как нужно автору исследования. Это усложняет анализ литературы и может представлять сферу искусственного интеллекта в искаженной форме. Такой пробел представляется упущением российской теории массовых коммуникаций, так как единый, корректный и понятный терминологический аппарат для точного описания всех элементов изучаемой системы все еще не закреплён.

Как отмечают А.М. Шестерина и Н.О. Шестерин, проблема размытых границ понятий в контексте искусственного интеллекта вызвана следующими факторами: «1) Отсутствие точных представлений о границах искусственного интеллекта в профессиональном медийном сообществе и сообществе ученых-гуманитариев. Обстоятельства, сформировавшие эту причину, кроются и в сложности самого явления, и в систематически повторяющихся моделях некорректного применения термина, которые со временем в силу их частотности становятся легитимными. <...>. 2) Стремление избежать тавтологии. Такой процесс наблюдается в тех текстах, где изначально искусственный интеллект используется в корректном варианте, но позже подменяется терминами, соотносимыми с близкими процессами. 3) Желание стимулировать интерес аудитории, привлечь ее к материалу. <...> работникам медиасферы важно понимать и подчеркивать различия между искусственным интеллектом и другими формами проявления алгоритмизированных процессов. Только в этом случае мы сможем избежать появления фейковых новостей, путаницы и закрепления лженаучных представлений в сознании аудитории» [28, с. 3].

Журналисту, работающему с нейросетями, необходимы базовые представления о ключевых понятиях и методах этой области. Несмотря на то, что осмысление терминологического аппарата может составлять отдельную исследовательскую задачу, этот вопрос нельзя игнорировать, так как научная работа требует конкретного теоретического обоснования и точной трактовки понятий, исключающей разночтения. Для формирования такого основания нужно обратиться к истокам ключевых концепций.

Считается, что понятие «искусственный интеллект» было введено в 1956 году, когда группа ученых-программистов предположила, что компьютер можно научить мыслить и логически рассуждать по аналогии с человеческим мозгом. При этом как отметил Д. Маккарти на Дартмутской конференции: «каждый аспект обучения или любая другая особенность интеллекта в принципе может быть описана настолько точно, что машина

может имитировать его» [5, с. 256]. Именно этот принцип и был назван «искусственным интеллектом». Эта область ориентирована на автоматизацию интеллектуальных задач, обычно выполняемых человеком.

Понятие «искусственный интеллект» выступает собирательным для систем, выполняющих интеллектуальные задачи, но оно не подразумевает какой-либо формы обучения. Для иллюстрации тезиса можно привести пример символического ИИ, который сфокусирован на явном написании (кодировании) правил для конкретного сценария в ограниченной области – для задачи символического ИИ человеком уже прописаны готовые правила, исходящие из человеческого знания. Искусственный интеллект в этом случае действует только в рамках заданной ограниченной правилами системы. Р.И. Чои отмечает в своем исследовании: «Хотя символический ИИ эффективен при решении четко определенных логических задач, он часто терпит неудачу в задачах, требующих распознавания образов более высокого уровня, таких как распознавание речи или классификация изображений. В этих более сложных задачах хорошо зарекомендовали себя методы машинного обучения и глубокого обучения» [35, с. 1]. Машинное и глубокое обучение в этой системе представляют собой методы достижения обозначенных целей, но они оба находятся в области искусственного интеллекта. Глубокое обучение представляет собой подмножество (тип или вид машинного обучения), использующее нейронные сети с их глубокими слоями.

Ф.М. Саймон предлагает следующее определение искусственного интеллекта для его рассмотрения в контексте журналистики: «процесс компьютерного моделирования человеческой деятельности и навыков в узко определенных областях, как правило, применение подходов машинного обучения, с помощью которых машины обучаются на основе данных и (или) своей собственной производительности» [62, с. 11]. Ф.М. Саймон также отмечает одну очень важную деталь касательно сложностей в определении ИИ: «пока не существует единого мнения о том, что представляет собой ИИ, и не существует его общепринятого определения. Однако есть общее согласие

относительно того, чем ИИ не является: это не сознательный и универсальный интеллект, способный понимать и работать в различных областях. Некоторые исследователи, такие как социолог Елена Эспозито, утверждают, что попытки воссоздать интеллект могут быть вообще не основным назначением ИИ. Она отмечает, что алгоритмы воспроизводят не сам интеллект человека, а информативность коммуникации. Взаимодействие с алгоритмами, особенно с чат-ботами, скорее представляет собой искусственную форму общения, а не интеллекта. В этом контексте вопрос о том, действительно ли система «умна» (какой бы смысл в это не вкладывался), становится второстепенным. Недавние достижения в области генеративного ИИ (например, систем, способных создавать новые, реалистичные данные, такие как текст, изображения и аудио) подкрепляют точку зрения Эспозито. Важно не столько то, является ли ИИ по-настоящему интеллектуальным, сколько то, что мы относимся к нему как к таковому, признавая его способность выполнять задачи, ранее считавшиеся исключительно человеческими» [62, с. 11]. Определение Ф.М. Саймона подчеркивает узкоспециализированный характер ИИ и его зависимость от машинного обучения. Исследователи сходятся во мнении, что современный ИИ не обладает сознанием или универсальностью, присущим человеческому интеллекту. Ключевую ценность представляет именно способность ИИ эффективно решать «человеческие» задачи, что особенно важно в контексте развития генеративного ИИ. Такой подход позволяет рассматривать искусственный интеллект лишь как инструмент, повышающий продуктивность журналистской работы.

Ф.М. Саймон также отмечает, что «Наиболее часто встречающуюся форму ИИ можно условно определить как узкий или слабый ИИ. Эта форма включает широкий спектр приложений и техник с разными уровнями сложности, автономии и абстракции, направленных на решение конкретных задач. Эти системы и программы не могут функционировать за пределами своих заранее заданных возможностей – это верно даже для больших языковых моделей, таких как GPT-4, которые могут работать в различных

текстовых доменах, но не обладают сознанием или комплексной моделью мира. Примеры узких форм ИИ включают приложения машинного обучения (ML) и его подмножество – глубокое обучение, а также различные формы обработки естественного языка (NLP), которые часто опираются на методы ML. Общим для них является то, что компьютерная программа или система обучается непосредственно на примерах, данных и опыте, используя алгоритмы, тренированные на больших объемах данных для улучшения производительности системы при выполнении узко определенных задач с течением времени. Такое обучение происходит в диапазоне от контролируемого до неконтролируемого и может включать дальнейшие шаги, подкрепление на основе обратной связи от человека. В индустрии новостей ИИ часто используется как общее понятие для описания набора технологий в общении с коллегами, партнерами или общественностью <...>» [62, с. 10].

Если говорить о принципах упомянутых подходов, можно отметить следующее: машинное обучение ориентировано на развитие способности ИИ применять новые алгоритмы для оптимальной работы с заданными данными. В отличие от классического программирования, где машине предоставляются входные данные и алгоритм для получения результата, в машинном обучении она получает входные и выходные данные, а алгоритм решения просчитывает самостоятельно. Как отмечает в своей работе Р.И. Чои, «В машинном обучении существуют четыре широко используемых метода обучения, каждый из которых полезен для решения различных задач: контролируемое обучение (с учителем), неконтролируемое обучение (без учителя), полуконтролируемое обучение и обучение с подкреплением» [35 с. 2]. Они находят применение и в журналистике данных: контролируемое обучение используется, например, в задачах прогнозирования и классификации данных, автоматизации написания простых новостей; неконтролируемое помогает в выявлении скрытых структур и зависимостей в больших наборах данных, анализе сложных длинных текстов, поиске новых тем; полуконтролируемое эффективно использует малое количество размеченных данных вместе с

большим объемом неразмеченных – это важно, если возможности для обучения ИИ ограничены финансами или человеческим ресурсом; обучение с подкреплением оптимизирует процессы разработки приложений, автоматизирует фактчекинг, улучшает рекомендационные системы.

Контролируемое обучение осуществляется с применением обучающих, проверочных и тестовых наборов данных: модель использует шаблоны обучающего набора для изучения лучших ответов, после аналогично обрабатывает данные проверочного набора и получает обратную связь с информацией о верности ответов. Тестовые наборы данных применяются в самом конце, для проверки модели: это абсолютно новые для данные, с которыми модель ранее не работала. В своей сути такой метод обучения модели предполагает настраивание точности алгоритма при помощи проверок. Наиболее распространенные задачи этого направления – регрессия (прогнозирование числовых данных) и классификация (отнесение предмета к определенной категории).

Неконтролируемое обучение подразумевает обнаружение связей и закономерностей в наборе данных самой моделью: ей не предоставляются правильные ответы или метки – она обучается без учителя, анализирует входные данные и ищет в них естественную структуру, без эталонного примера решения. Как отмечает Р.И. Чои, «Наиболее распространенными задачами неконтролируемого обучения являются кластеризация, выявление ассоциаций и обнаружение аномалий. Например, кластеризация группирует объекты набора данных в отдельные кластеры на основе определенных комбинаций их характеристик» [35, с. 3]. Ассоциирование работает по аналогии с человеческой памятью – нужно завершить задачу и ассоциировать полученный объект или образ с похожим из памяти. Обнаружение аномалий помогает найти в данных выбросы или нехарактерную новизну.

Полууправляемое обучение подразумевает маркировку некоторого количества вводных данных (обучающих) для последующего анализа моделью немаркированных данных. Тем самым, такое обучение представляет

собой нечто среднее между контролируемым и неконтролируемым обучением.

Обучение с подкреплением, по определению Р.И. Чои, «это метод тренировки алгоритма для выполнения конкретной задачи, где не существует единственного правильного ответа, но требуется достижение определенного результата. Этот подход считается наиболее близким к моделированию процесса человеческого обучения, поскольку он базируется на пробах и ошибках, а не только на анализе данных» [35, с. 3]. Подкрепление буквально означает вознаграждение модели (положительное или отрицательное), благодаря чему модель узнает, какое поведение желательно. Нейросеть наблюдает за окружающей средой, определяет путь движения к цели и порядок действий.

В понятие искусственных нейронных сетей по информации Большой российской энциклопедии вкладывается следующее: «Нейронные сети, искусственные, многослойные высокопараллельные (то есть с большим числом независимо параллельно работающих элементов) логические структуры, составленные из формальных нейронов <...>. Появление формального нейрона во многом обусловлено изучением биологических нейронов. Формальный нейрон <...> является основой любой искусственной нейронной сети. Нейроны представляют собой относительно простые, однотипные элементы, имитирующие работу нейронов мозга. Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены и заторможены. Искусственный нейрон, так же как и его естественный прототип, имеет группу синапсов (входов), которые соединены с выходами других нейронов, а также аксон – выходную связь данного нейрона, откуда сигнал возбуждения или торможения поступает на синапсы других нейронов» [22]. Нейронная сеть является математической моделью, которая моделирует структуру и возможности биологической нейронной сети.

Теория нейронных сетей берет начало в работе американских нейрофизиков У. Маккалоха и У. Питтса, что отмечается в монографии С.В. Аксенова и В.Б. Новосельцева: «В 1943 г. В. Маккаллох и В. Питтс предложили систему обработки информации в виде сети, состоящей из простых вычислителей, созданных по принципу биологического нейрона <...>. Маккаллох и Питтс показали, что такие сети могут производить произвольные вычисления, подобно известной машине Тьюринга» [1, с. 4]. Значительный вклад в развитие направления внес также канадский нейропсихолог Д.О. Хебб, сформулировавший в 1949 году принцип самообучения, согласно которому система самостоятельно осваивает выполнение задачи без внешней помощи. Аксенов С.В. и Новосельцев В.Б. также отмечают, что «В 1962 г. Ф. Розенблатт <...> предложил использовать особый тип искусственной нейронной сети, названный им персептроном <...>. Однако уже в 1969 г. М. Минский и С. Паперт в своей известной работе показали, что персептрон не может решать целый класс простых задач, таких как, например, реализация логической операции XOR (исключающее ИЛИ). Появление этой работы сыграло роковую роль для теории нейронных сетей, все исследования в этой области фактически были приостановлены вплоть до середины 80-х годов. В 1986 г. ситуация изменилась – Д. Румельхарт, Г. Хинтон и Р. Вильямс предложили эффективный алгоритм для обучения более совершенного, так называемого многослойного, персептрона. Алгоритм получил название алгоритма обратного распространения ошибки» [1, с. 5]. Современные нейронные сети развивались именно из идеи персептрона, к которому добавлялось все больше слоев и применялись улучшенные алгоритмы обучения. Персептрон представлял собой наиболее простую нейронную сеть, ее базовый тип или «строительный блок».

Искусственные нейронные сети устроены по аналогии с биологическими нейронными сетями. Вот что отмечает А. Кренкер: «Искусственный нейрон – это основной строительный блок любой искусственной нейронной сети. Его конструкция и функциональные

особенности основаны на наблюдениях за биологическим нейроном, который является базовым строительным блоком биологических нейронных сетей (систем), включая мозг, спинной мозг и периферические ганглии. <...> В случае биологического нейрона информация поступает в нейрон через дендрит, сома обрабатывает информацию и передает ее через аксон. В случае искусственного нейрона информация поступает в его тело через входы, которые имеют вес (каждый вход может быть индивидуально умножен на вес). Тело искусственного нейрона затем суммирует взвешенные входы, смещение и «обрабатывает» сумму с помощью функции передачи. В конце искусственный нейрон передает обработанную информацию через выход(ы). При объединении двух или более искусственных нейронов получается искусственная нейронная сеть. Если один искусственный нейрон почти бесполезен в решении реальных проблем, то искусственные нейронные сети справляются с этим» [48, с. 5]. Исходя из этих положений в области искусственного интеллекта и используются понятия «сеть», «глубокое обучение».

В научной работе Чои Р.И. «Введение в машинное обучение, нейронные сети и глубокое обучение» [35] была предпринята попытка визуализировать в виде схемы взаимосвязи между ключевыми понятиями в области искусственного интеллекта и машинного обучения, аналогичное изображение используется и некоторыми другими исследователями. Предложенная схема отражает иерархию, с которой сложно согласиться: общее поле науки о данных включает искусственный интеллект, в него интегрировано машинное обучение, в него – глубокое обучение, а завершает схему самый небольшой элемент – нейронная сеть. Такая интерпретация представляется методологически некорректной, так как глубокое обучение – не родительская категория для нейросети, а именно метод для тренировки и оптимизации моделей, по аналогии с машинным обучением, которое включается в подобласть искусственного интеллекта. Аналогичная корректная интеграция по смыслу прослеживается в других исследовательских работах [12], [29], [30]:

глубокое обучение рассматривается именно как включенный в область нейросетей элемент. Таким образом, достаточно популярная схема, предложенная в работе Чои Р.И., требует концептуальной доработки для точного отражения взаимосвязей терминов в изучаемой области и может выглядеть следующим образом (рис. 1):



Рисунок 1 – Взаимосвязь ключевых понятий в Data Science

В научной сфере необходимо правильно и точно применять термины искусственный интеллект (ИИ), машинное обучение (МО), искусственные нейронные сети (ИНС) и глубокое обучение (ГО). Нейронные сети представляют собой часть более широкой области ИИ и играют важную роль в его развитии. В качестве вывода по исследованию терминологии можно обобщить полученные данные в виде определений, что и будет сделано далее.

Искусственный интеллект или ИИ (Artificial Intelligence или AI) – самое широкое понятие. Это любая технология, позволяющая компьютерам имитировать человеческое мышление и решать задачи, требующие интеллекта (например, анализ данных, общение, классификация информации), но не

подразумевающие наличия сознания. Включает в себя множество подходов и методов.

Машинное обучение или МО ((Machine Learning или ML) – подраздел ИИ. Это методы, позволяющие компьютерам учиться на данных и улучшать свою работу без явного программирования каждого шага. Включает методы контролируемого, неконтролируемого, полуконтролируемого обучения и обучения с подкреплением, которые применяются, например, для прогнозирования, классификации и кластеризации.

Искусственные нейронные сети, нейронные сети или ИНС (Artificial Neural Network или ANN) – конкретный тип алгоритмов машинного обучения, имитирующий работу человеческого мозга. Нейросети состоят из соединенных узлов (искусственных нейронов), организованных в слои. Формируют основу методов глубокого обучения и решают такие задачи, как, например, распознавание образов или анализ текстов.

Глубокое обучение или ГО (Deep Learning или DL) – направление машинного обучения, использующее сложные многослойные нейронные сети. Термин «глубокое» указывает на большое количество слоев в нейронной сети. Используется для анализа сложных данных и решения высокоуровневых задач: обработки изображений, речи или текстов, благодаря способности нейросетей выявлять сложные зависимости.

Ключевое значение в развитии нейронных сетей и их адаптации к тем или иным задачам также имеют виды обучения – контролируемое, неконтролируемое и полуконтролируемое. Они особенно значимы в способностях нейросетей решать различные задачи, адаптироваться к новым условиям и требованиям. Понимание этих ключевых понятий и их взаимосвязей помогает не только правильно интерпретировать научные исследования, но и эффективно использовать нейросети в журналистской практике.

1.2 Применение нейросетей для генерации данных в журналистике: возможности, проблемы и вызовы

Нейросети стремительно интегрируются в медиасферу и трансформируют привычные журналистские практики. Эта технология, как ключевой инструмент ИИ, приобрела особую популярность благодаря широким возможностям в автоматизации создания и обработки контента. Ведущие мировые медиакомпании – Bloomberg, Associated Press, Reuters, активно внедряют эти технологии в рабочие процессы своих редакций. Как отмечает Давыдов С.Г.: «В отчетах Стэнфордского университета «Индекс искусственного интеллекта», Института изучения журналистики Рейтер при Оксфордском университете, а также консалтинговых и аудиторских компаний Deloitte, McKinsey, Accenture, PwC фиксируется начало стадии долгосрочного роста инновации искусственного интеллекта на высокоразвитых рынках медиатехнологий. Это, в частности, подтверждается рейтингом Глобального индекса развития ИИ Стэнфордского университета, составленным на основе многофакторного анализа открытых данных по теме ИИ 29 стран, в том числе РФ (2017 по 2021 гг.). Уровень развития технологий ИИ определяется следующими индикаторами: количеством научных публикаций, цитирования, материалов конференций, базами данных, числом патентов, объемами инвестиций в технологии ИИ, открытием венчурных компаний, индексом найма на работу специалистов по ИИ, проникновением навыков работы с искусственным интеллектом в бизнес. По совокупности этих показателей с 2017 по 2021 гг. Глобальный индекс развития ИИ отводит России 23-е место. Институт Reuters Оксфордского университета также усматривает косвенное подтверждение стадии роста ИИ-инноваций, когда для повышения конкурентоспособности медиакомпаний выбирают инновационную стратегию внедрения технологий ИИ, в том числе применяют автоматизированную генерацию текстов и преобразование речи в текст, интеллектуальную поддержку редакционных процессов,

персонализированные сервисы дистрибуции, работают с большими данными для мониторинга разных проблемных ситуаций и так далее. По мнению аналитиков Оксфордского университета, системы ИИ претендуют на статус активного субъекта конкуренции в медиа. При этом выражается озабоченность, что применение технологий ИИ может увеличить разрыв между крупными медиакомпаниями и теми изданиями, которые не могут себе позволить долгосрочные инвестиции в инновационные технологии. По данным отчета, большинство опрошенных экспертов считают, что лишь крупные издатели смогут получать выгоду от внедрения технологий ИИ» [11]. Эта тенденция подчеркивает необходимость анализа различных типов нейросетей, используемых в журналистике, и их классификации, что позволит лучше понять текущее состояние и перспективы применения конкретных нейросетевых технологий в медиасфере. С тезисом о невыгодности технологии для небольших агентств сложно согласиться – крупные разработчики уже предлагают рамочные решения с доработкой под любую цель, даже для небольших изданий их внедрение может оказаться выгодным в перспективе.

Для специфических журналистских задач создаются специализированные нейросети, а их растущее разнообразие усложняет систематизацию – многие типы моделей оказываются на стыке технологий и сочетают в себе их особенности. Несмотря на это, можно выделить ключевые характеристики основных классов – такая классификация будет базовой, хотя ее нельзя и назвать исчерпывающей. Важным аспектом для анализа нужно отметить понятие «архитектура нейронной сети» или «топология нейронной сети» – под ним подразумевается способ соединения искусственных нейронов в саму сеть. Исходя из того, что формальные нейроны могут быть соединены между собой разными способами, искусственные нейронные сети можно глобально разделить на два класса, учитывающих архитектуру, как это делает А. Кренкер в исследовании «Введение в искусственные нейронные сети» [48]:

– нейронные сети с прямыми связями,

– нейронные сети с рекуррентными связями.

Схожую классификацию выстраивают в своей монографии также С.В. Аксенов и В.Б. Новосельцев: «Все виды нейронных сетей можно условно разделить на сети прямого распространения и сети с обратными связями. Как следует из названия, в сетях первого типа сигналы от нейрона к нейрону распространяются в четко заданном направлении – от входов сети к ее выходам. В сетях второго типа выходные значения любого нейрона сети могут передаваться к его же входам. Это позволяет нейронной сети моделировать более сложные процессы, например временные, но делает выходы подобной сети нестабильными, зависящими от состояния сети на предыдущем цикле» [1, с. 10]. Так как типов нейронных сетей выделяется много, логично рассмотреть основные их типы, наиболее часто используемые, как и будет сделано далее.

Первый тип – нейронную сеть с прямой связью, или сеть прямого распространения (Feedforward Neural Network, FNN), как отмечает А. Кренкер, можно охарактеризовать следующим образом: «Искусственная нейронная сеть с топологией прямой передачи данных называется искусственной нейронной сетью с прямой передачей <...>. У такой сети есть одно условие: информация должна передаваться от входа к выходу только в одном направлении, без обратных связей. Нет ограничений на количество слоев, тип функции передачи, используемой в отдельных искусственных нейронах, или количество соединений между отдельными нейронами. Самая простая искусственная нейронная сеть с прямой передачей – это одиночный перцептрон, который способен обучаться только линейно разделимым задачам» [48, с. 7], например сжатию файлов. Многослойные перцептроны (MLP) с тремя и более слоями способны решать более сложные задачи. В контексте задач журналистики эти нейросети могут применяться, например, для анализа аудитории, прогнозирования трендов, написания небольших заметок.

Так, Питер Поташ и коллектив авторов в работе «Отслеживание предвзятости в новостных источниках с использованием социальных сетей: Российско-украинский кризис на майдане 2013-2014 гг.» [58] применили нейронную сеть с прямой связью для анализа политической предвзятости информационных текстов на основе пользовательских реакций. Точность нейронной сети составила 93,5 %. Исследование основывалось на реакции читателей на новость определенного источника для оценки его предвзятости. В число источников попали lenta.ru, ria.ru, rbc.ua, forbes.ua и другие крупные СМИ. Работа проводилась с данными 2013-2014 годов, во время обострения кризиса между Россией и Украиной. Данные для анализа собирали из социальной сети «ВКонтакте», как самой популярной в русскоязычном сегменте интернета. Политическая предвзятость источников информации оценивалась на основе реакции пользователей, включала лайки, репосты, распространение ссылок на новости. В ходе работы нейронная сеть проанализировала принадлежность домена источника стране, а также ключевые слова, и разделила пользователей на два лагеря на основе политических убеждений. Нейросеть смогла определить принадлежность новости и источника к стране по пристрастному отношению СМИ к проблеме. Такие данные могут быть использованы в обучении других нейросетей, в особенности, использующих синтез нескольких методов обработки данных, – для повышения точности их работы. Практическое применение таких результатов, например, разметка и классификация СМИ по политическим предпочтениям, могла бы помочь читателям критичнее воспринимать потребляемый контент.

Рекуррентная нейронная сеть (Recurrent Artificial Neural Networks, RNN) отличается от сетей с прямой связью двунаправленной передачей информации – прямой и обратной, что обеспечивает запоминание ранее обработанных данных. Как отмечает А. Кренкер, «Она похожа на нейронную сеть с прямой передачей данных, но без ограничений на наличие обратных связей. В таких сетях информация передается не только в одном направлении,

но также в обратном, что создает внутреннее состояние сети, позволяющее ей демонстрировать динамическое временное поведение. Рекуррентные искусственные нейронные сети могут использовать свою внутреннюю память для обработки любых последовательностей входных данных» [48, с. 8]. Это позволяет модели адаптироваться к входным данным разной длины и временной последовательности. Например, при анализе текста такая модель сохраняет контекст предыдущих слов независимо от длины предложений. Базовым элементом модели выступает рекуррентная структура, где каждый нейрон связан со всеми остальными, благодаря чему формируется сеть, способная к обучению и хранению информации. RNN представлены различными подтипами: сетями Хопфилда, Элмана, Джордана и двунаправленными сетями, каждая из которых оптимизирована для конкретных задач – прогнозирования, рекогнитивных технологий (имитирующих когнитивные процессы человека – восприятие, мышление, память и обучение) или обработки естественного языка.

Пример применения рекуррентной нейросети описан в исследовании «Обнаружение слухов из микроблогов с помощью рекуррентных нейронных сетей» [52]. Исследователи применили нейросети для выявления недостоверной информации в Twitter и Sina Weibo (китайском сервисе микроблогов) путем моделирования распространения информации о событиях в социальных сетях с течением времени. Система автоматически идентифицировала слухи, анализировала особенности распространения и динамику сообщений. Результаты оказались впечатляющими: модель GRU-2 достигла точности 83,9 % для Twitter и 89,0 % для китайской платформы Weibo при анализе постов в первые 12 часов после публикации – это значительно быстрее официальных ответов касательно содержания поста со стороны властей или специализированных компаний. Такая технология позволяет фильтровать дезинформацию задолго до реакции государственных организаций, что существенно снижает негативное влияние слухов на общество.

Рекуррентные нейросети можно включить и в более детализированную классификацию искусственных нейронных сетей следующего уровня, что также отмечается А. Кренкером в его работе «Введение в искусственные нейронные сети» [48]. Основой этой классификации выступают особенности архитектуры и специализированные задачи, под которые оптимизированы конкретные нейросети. Такое деление помогает углубить понимание содержания каждого типа и выбрать подходящую архитектуру для объектов и целей конкретного проекта. Кроме уже рассмотренных рекуррентных нейросетей, в рамках этой более подробной классификации А. Кренкера можно выделить модели, описанные далее.

Многослойные перцептроны (Multi-layer Perceptrons, MLP) представляют собой класс нейросетей с прямым распространением и одним или несколькими скрытыми слоями (информация движется от входного слоя через скрытые слои к выходному слою, без обратных связей), что делает их идеальными для задач классификации и регрессии, так как MLP способны обрабатывать сложные нелинейные зависимости данных в своих слоях [48]. MLP могут быть применены в журналистике для анализа влияния и популярности новостей среди аудитории. Так, в рамках исследования «Определение авторства коротких текстов с использованием многослойного перцептрона» [60], была поставлена задача идентификации авторов на основе анализа стиля их письма. Используемая в эксперименте выборка состояла из 400 твитов от каждого из четырех авторов и была расширена до двадцати авторов в более обширной выборке. MLP обучили на этом наборе данных, чтобы модель могла научиться узнавать уникальные стилистические особенности каждого автора. Нейросеть достигла точности 96,44 % при работе с четырьмя авторами. При работе с двадцатью точность снизилась до 80 %, но это все равно превышает результаты других, классических методов машинного обучения.

Сверточные нейронные сети (convolutional neural networks, CNN) – тип искусственных нейросетей, разработанных для обработки и анализа данных с

сетчатой структурой (изображений и видео). Этот тип подробно рассмотрен в работе З. Ли «Обзор сверточных нейронных сетей: анализ, применение и перспективы» [51]. Основное преимущество этого вида сетей – автоматическое выявление иерархических признаков данных, распознавание объектов, анализ изображений и видео. В журналистике такие нейросети могут использоваться для обнаружения фейковых новостей, анализа изображений или видео, распознавания лиц и объектов на фотографиях, автоматической классификации новостного, и модерации пользовательского контента в социальных сетях, выявления манипуляций.

Такие манипуляции и фейковые новости всегда были важной проблемой, затрагивающей все сферы общества. Их распространение только усилилось с развитием интернета и социальных сетей. Фейки вводят людей в заблуждение и оказывают серьезное влияние на общественное мнение и политику, стабильность социальной ситуации в стране. Считается даже, что на президентских выборах в США в 2016 году фейки сыграли заметную роль, повлияв на исход выборов. В исследовании «Ти-Эн-Эн: сверточные нейронные сети для обнаружения фейковых новостей» [65] для решения проблемы авторы предлагают модель TI-CNN, которая использует текстовую и графическую информацию в анализе и обучении для поиска фейков. Это позволяет эффективно идентифицировать подделки в новостях с учетом явных (кликбейтные названия, избыток эмоциональности) и скрытых признаков (например, нестандартных грамматических конструкций), которые сложно обнаружить традиционными методами. Модель анализирует текстовые данные и выявляет лингвистические особенности, которые указывают на фейк, например, ищет слова и фразы-маркеры. Нейросеть анализирует изображения в новостях, находит несоответствия между текстом и изображением, отслеживает низкое качество изображений (что также часто является признаком фейка). TI-CNN продемонстрировала высокую точность в распознавании фейков на основе анализа 80 % данных для обучения, 10 % для

проверки и 10 % для тестирования. Результаты показали точность модели 92,20 %, а это достаточно высокая эффективность.

Генеративно-состязательные сети (Generative Adversarial Networks, GAN) – искусственные нейронные сети, созданные для генерации реалистичных новых данных. Этот тип подробно рассмотрен в исследовании «Генеративные состязательные сети» [40]. Такие нейросети позволяют улучшать качество данных, создавать синтетические наборы для обучения других моделей и выявлять выбросы. В работе «Генеративные состязательные сети для повышения достоверности больших данных» [37] авторы предложили использовать GAN для предварительного обучения платформы больших данных: материалом послужила база Sketchy, содержащая 75 471 эскиз по 125 категориям, из них 19 633 эскиза были отобраны для обучения модели ACGAN (тип GAN, который определяет реальность изображения и классифицирует его по категориям). Обучение проводилось в течение 100 эпох (весь набор обучающих данных использовался для обновления параметров модели 100 раз, чтобы повысить точность), с размером выборки по 64 эскиза. Для оценки был организован опрос на платформе Amazon Mechanical Turk с классификацией генеративных изображений пользователями. Результаты показали общую точность предсказаний 43,8 %, для некоторых классов – большую эффективность (например, для WineBottle – 100 %, а для Hammer – 92,8 %). Точность классификации изображений, созданных человеком, составила 87,7 %. Учитывая сложность и разнообразие категорий, полученные результаты для ACGAN были признаны достаточно хорошими. Такие сгенерированные изображения могут использоваться для пополнения обучающих наборов данных с целью повышения устойчивости нейросетей. GAN также обеспечивает метод обнаружения выбросов, что снова положительно влияет на качество данных, защищает систему от ошибочных записей.

В работе «Вычислительно-интеллектуальный агент для обнаружения фейковых новостей с использованием генеративных состязательных сетей»

исследуется проблема фейковых новостей и их влияние на общество. Автор работы Хириянна С. отмечает следующее: «Набор данных, использованный для обучения модели классификации, состоял из 105 104 заголовков новостей. Этот набор данных был случайным образом перемешан и разделен на обучающую, проверочную и тестовую выборки: 40 % заголовков для обучения, 35 % для проверки и 25 % для тестирования. Это дало 40 990 заголовков для обучения, 27 327 для проверки и 36 788 для тестирования. Обучающая выборка была передана в классификатор, который затем тестировался на тестовой выборке. Достигнутая точность на этапе проверки составила приблизительно 95,3 %. При тестировании была достигнута точность 94,8 %» [41, с. 93]. Из 15 759 сгенерированных новостей система правильно классифицировала большую часть, ошибочно приняв за реальные только 1600. Из 19 121 реальных новостей лишь 307 были неверно определены как сгенерированные. Несмотря на высокую общую точность, наличие ошибок указывает на ограничения модели, а авторы исследования логично рассматривают эти недостатки как направления для дальнейшего совершенствования системы.

В целом, генеративные нейросети находят широкое применение в журналистике и трансформируют процессы как создания, так и распространения новостей. Это отмечают в своем исследовании авторы Ши Й. и Сун Л.: «Генеративный ИИ в настоящее время используется для поддержки и ускорения определенных процедурных задач, таких как создание контента, проверка фактов, обработка данных, генерация изображений, преобразование речи и перевод. Это снижает нагрузку на человека и повышает его эффективность» [61, с. 582]. Генеративные модели используются для сбора информации: анализа больших объемов данных, выявления значимых показателей, автоматизируют повторяющиеся задачи (например, сбор справочной информации), прогнозируют тенденции. Так, Reuters News Tracer применяет машинное обучение для мониторинга актуальных новостей в твитах, обеспечивая редакцию оперативной информацией. Генеративные

модели также помогают журналистам создавать контент, адаптированный к стилю издания. Во французской газете Le Monde ИИ используется для перевода статей, благодаря чему в англоязычной версии ежедневно публикуется около 30 материалов. Особого внимания заслуживает NewsGPT, полностью автоматизированный новостной сайт, где ИИ самостоятельно анализирует и интерпретирует информацию из разных источников, создает репортажи без участия людей. Генеративные нейросети оптимизируют производство новостей и освобождают журналистов от рутинных задач. Они анализируют поведение аудитории и дают рекомендации по распределению журналистов по задачам. Здесь показательны примеры чат-бота Аделаида от Forbes, предоставляющего персонализированные новости согласно интересам пользователей, и Charlie от OneSub, который снижает уровень стресса и тревожности пользователей, избегая демонстрации негативной и шокирующей информации в новостях. Генеративные сети не только повышают эффективность, качество и точность традиционного процесса создания новостей, но и активно вовлекают аудиторию в производство и обсуждение новостного контента.

Генеративные нейросети играют ключевую роль в создании и улучшении контента, но одной из самых передовых технологий в области обработки естественного языка считаются трансформеры [57]. В работе исследователей из Google Brain, А. Васвани, на момент анонса технологии, отмечает: «Мы предлагаем новую простую архитектуру сети – Transformer, которая основана исключительно на механизмах внимания, полностью исключая рекуррентность и свертки. Эксперименты с двумя задачами машинного перевода показывают, что эти модели превосходят другие по качеству, при этом они лучше параллелизуются и требуют значительно меньше времени на обучение» [63, с. 1]. Трансформеры (Transformers) – специальный класс моделей глубокого обучения, который эффективно обрабатывает текстовые данные и одновременно анализирует весь вводимый текст, благодаря чему лучше отслеживает длинные зависимости по всему

материалу. Благодаря этому такие модели быстрее обучаются. Основным принципом работы трансформеров основан на механизме внимания (attention mechanism) – сеть фокусируется на разных частях текста при формировании ответа и определяет несколько акцентов в материале. Такой подход называют multi-head attention – он делает анализ более точным.

Трансформеры обычно состоят из двух частей – энкодера и декодера. Первый преобразует входные данные в рабочий «внутренний» для нейросети формат, а декодер использует его для создания ответа. Это особенно полезно для задач по переводу текста (так работа идет со всем контекстом материала, сложные и многозначные термины трактуются корректнее, с учетом окружающего материала) и ответов на вопросы по текстам, где требуется найти соответствующие фрагменты в длинных блоках. То же отмечает Попов Т.С.: «Их ключевая особенность заключается в механизме самовнимания (self-attention), который позволяет моделям анализировать зависимости между элементами входных данных независимо от их позиции. Это делает трансформеры особенно эффективными для обработки последовательных данных, таких как текст, и позволяет учитывать долгосрочные взаимосвязи. Дополнительные исследования показали, что трансформеры успешно применяются в задачах, связанных с обработкой временных рядов и графов, что делает их универсальным инструментом в задачах оптимизации <...>. До появления трансформеров основные задачи работы с последовательными данными решались с помощью рекуррентных нейронных сетей (RNNs) и их усовершенствованных версий, таких как LSTM и GRU. Однако их возможности были ограничены проблемами с долгосрочной памятью и высокой вычислительной сложностью. Трансформеры преодолели эти ограничения благодаря возможности параллельной обработки данных и более глубокому учету контекста <...>. Популярность трансформеров растет благодаря их высокой производительности в задачах обработки текста, изображений и временных рядов. Компании, такие как Google, Microsoft, OpenAI, активно используют трансформеры в своих продуктах, включая

поисковые системы, голосовые помощники и генеративные модели, такие как GPT. Их применение позволяет значительно улучшить качество рекомендаций, повысить вовлеченность пользователей и увеличить доход» [26, с. 1]. Наиболее известные модели на основе трансформеров включают BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer) и T5 (Text-To-Text Transfer Transformer). Эти модели уже доказали свою эффективность в задачах обработки текста и нашли широкое применение в практике.

Кроме перевода текста и поиска информации, трансформеры в журналистике выполняют анализ настроений аудитории, мониторинг социальных медиа для выявления ключевых тем и трендов, автоматизацию написания новостных сводок и отчетов, проверку фактов. В исследовании «Основанные на трансформерах модели глубокого обучения для анализа настроений в данных социальных сетей» [47] разрабатывался улучшенный метод анализа эмоциональной окраски текста с помощью продвинутых моделей глубокого обучения. Работа была направлена на решение проблем обработки зашумленных данных с учетом слов, отсутствующих в словарях, а также контекста и эмоций в тексте. Исследователи применили модель на основе BERT от Google в сочетании со специальными нейронными сетями (CBRNN). Для обучения и тестирования использовались разнообразные источники данных: отзывы на авиакомпаниях, самоуправляемые автомобили, материалы о президентских выборах в США и рецензии на фильмы с IMDb. Разработанная модель превзошла уже существующие, показав повышение общей точности на 0,3 %, способности различать положительные и отрицательные классы на 0,4 %, F1-меры (среднего значения между точностью и полнотой) на 0,2 %. Эти улучшения признаны значимыми, несмотря на их кажущуюся скромность, поскольку базовые показатели точности уже были достаточно высокими. При работе с большими данными, сложными задачами и комплексными метриками, даже небольшие улучшения уже существенно повышают общую производительность системы.

На основании рассмотренной классификации и примеров можно отметить, что специализированные нейросети, такие как модели с прямыми связями (FNN), рекуррентные нейронные сети (RNN), многослойные перцептроны (MLP), сверточные нейронные сети (CNN) и трансформеры, уже сегодня демонстрируют высокую эффективность в решении журналистских задач: генерации и анализе текста, автоматизации написания статей, переводах и проверке фактов. В журналистике наиболее активно применяются способности генеративных трансформеров. Они играют важную роль в создании контента – текстов и изображений. Общеизвестным примером здесь может выступить модель ряда GPT (Generative Pre-trained Transformer) от OpenAI, которые широко применяются в генерации текстов. Популярность GPT отмечает Бейненсон В.А.: «В феврале 2019 года лаборатория OpenAI объявила о создании нейросети GPT-2, которая генерирует сообщения, похожие на тексты, написанные людьми. Поэтому полную публикацию ее программного кода (что является распространенной практикой) поначалу посчитали опасной, поскольку технология могла быть использована в интересах злоумышленников (позже ее все же выложили в открытый доступ). Человеку достаточно написать 1-2 предложения («затравку»), и нейросеть дописывала его до связного и логичного текста, приводя правдоподобные доказательства и выдуманные цитаты реальных экспертов. В мае 2020 г. OpenAI разработала модель нового поколения – GPT-3. Она выдавала тексты более высокого качества, «затравкой» могло быть не только предложение, но произвольно сформулированное задание. Кроме того, GPT-3 могла давать рекомендации, например, фильмов или книг. При этом если GPT-2 могла преимущественно сочинять забавные, но бесполезные фейковые тексты, то модель GPT-3 явно имела перспективы для практического применения, в том числе в медиа <...>. 30 ноября 2022 года OpenAI выпустила новую «пишущую» модель ChatGPT на базе нейросети GPT-3.5. Важным отличием для пользователей новой системы является удобный интерфейс в виде чата,

где можно задавать вопросы, давать задания, а также просить скорректировать результат» [4, с. 355].

Трансформеры, как наиболее передовые и современные по архитектуре нейросети, благодаря своей способности эффективно обрабатывать большие объемы текстовых данных и учитывать контекст, становятся в журналистике незаменимыми. Они позволяют улучшить качество контента и глубже понять настроение и реакцию аудитории. Несмотря на разнообразие и сложность нейросетей, выделение их ключевых характеристик и особенностей применения в журналистике позволяет удобно классифицировать модели для понимания механизмов их работы, а значит, эффективнее применять технологии в повседневной практике.

Одной из ключевых проблем применения нейросетей в журналистике является вопрос достоверности генерируемой ими информации. Нейросети обучаются на огромных массивах данных, которые могут содержать неточности, ошибки и даже намеренно искаженные данные. В результате, сгенерированный нейросетью текст может включать в себя недостоверную или вводящую в заблуждение информацию. Так это описывает С.Г. Давыдов: «Пока технологии ИИ не могут полностью заменить человека в проверке информации. Фактчекинг сегодня в большей степени ложится на плечи «журналистов-людей». Данная функция, по мнению экспертов, крайне сложна в автоматизации, так как многие источники все же нуждаются в традиционной, «человеческой» проверке. При этом есть не очень сложные операции по проверке информации, которые можно доверить алгоритмам. В частности, это уточнение имен собственных, географических названий и пр. <...> медиаменеджер, отвечавший на вопросы в ходе индивидуального интервью, считает одной из первостепенных задач ИИ «выявление фейковой информации и противодействие высокотехнологичным фейкам». Нельзя не упомянуть и достоверность выводов, которые могут быть сделаны с помощью ИИ. Независимый медиаэксперт, также индивидуально отвечавший на вопросы, считает это ключевым препятствием на пути внедрения ИИ в

медиапрактику: «Текущие технологии, особенно технологии машинного обучения, часто могут сравнительно вольно трактовать входящие данные. Это прекрасно, когда нам нужно творчество, но очень плохо, если это творчество надо поместить в рамки достоверности, объективности и прочих журналистских принципов» [11, с. 16]. Нейросети пока не способны в полной мере понимать контекст и нюансы человеческого языка, что может приводить к искажению смысла и неверной интерпретации фактов. Эта проблема требует особого внимания и тщательной проверки информации, полученной от нейросетей.

Проблема достоверности информации становится особенно значимой, поскольку алгоритмы генерируют контент, который сложно отличить от созданного человеком. Это требует от журналистов и редакторов особого внимания к фактам и прозрачности процесса создания контента при помощи нейросетей. Легитимность СМИ, использующих нейросети, будет зависеть от того, насколько используемый ими генеративный контент соответствует нормам, ценностям и ожиданиям аудитории, в том числе в отношении его достоверности и надежности [55]. Исследования в области ИИ и нейросетей предлагают методы автоматизации процесса, но необходимы практические инструменты, адаптированные для использования журналистами в реальных условиях. Риски автоматизации проверки фактов включают потенциальное использование технологий для легитимизации фейковых новостей, недостаточный учет контекста и увеличение нагрузки на журналистов. Непонимание машиной как контекста, так и сатиры, иронии или сарказма, а также в целом общей культурной, политической и социальной картины, может привести к неверной интерпретации информации и ошибочной оценке достоверности.

Можно сформулировать конкретные проблемы применения нейросетей в генерации новостного контента на основании замечаний исследователя С. Г. Давыдова:

- «сложность понимания логики, стоящей за генерацией текста;

- склонность к «галлюцинациям», генерация правдоподобной и похожей на подготовленную человеком, но фактически неверной информации;
- предвзятость, воспроизведение предубеждений, присутствующих в обучающих данных;
- токсичность, воспроизведение и приумножение токсичного языка данных, на которых обучалась модель» [11].

Для корректной проверки достоверности тезисов и подтверждения фактов, необходимо разбираться в механизмах генерации контента, критически анализировать и проверять текст на галлюцинации, предубеждения, токсичные высказывания, которые засоряют материал и снижают его достоверность. Особенности устройства нейросетей приводят к тому, что они склонны генерировать обобщенные, однотипные и похожие друг на друга новости, а уникальные и по-настоящему разнообразные сюжеты им создать достаточно сложно именно из-за механизма работы: модель предсказывает следующее вероятное слово на основе комбинации предыдущих, воспроизводит форму по примеру, но критически не анализирует содержание, что лишает материалы уникальности и глубины.

Проблема приобретает особую актуальность в современных СМИ, активно использующих пользовательский контент (UGC) социальных сетей: с развитием мощных генеративных нейросетей, которые создают тексты, изображения и видео, почти неотличимые от созданных человеком, возникает серьезная угроза распространения «искусственного» пользовательского контента на этих платформах и его использование в производстве новостей. Если раньше основную опасность представляли дезинформация и фейки, созданные людьми, то сегодня растет количество примеров реалистичного контента, полностью сгенерированного нейросетями. Такой UGC может быть использован для манипуляции общественным мнением, он создает серьезные репутационные риски для СМИ, которые интегрируют пользовательские заметки, тексты и фото в свои материалы. Издания, полагающиеся на

пользовательский контент как на источник оперативной информации и различных точек зрения, оказываются уязвимыми перед этой новой угрозой. Для журналистов и редакций важно уметь отличать настоящий UGC от генеративного, а это требует не только развития традиционных навыков проверки фактов, но и понимания того, как именно работают современные генеративные модели, где они ошибаются, и чем генеративный контент отличается от пользовательского. Можно также использовать специальные инструменты для выявления такого контента, если у издания есть такая возможность. Проблема усугубляется постоянным совершенствованием технологии нейросетей, что делает генерируемые материалы все более трудноотличимыми от созданных человеком. Для противостояния угрозе медиаорганизациям понадобятся обученные в этой области сотрудники или даже специализированные команды для проверки подлинности контента, а также собственные методы и технологии для выявления подмены. То же отмечает в своем исследовании С.А. Хан: «цифровые каналы наряду с простыми в использовании инструментами редактирования, к сожалению, стали средством распространения ложной/дезинформации, содержащей визуальный контент. Специалисты в области СМИ и специалисты по проверке фактов продолжают сталкиваться с трудностями при быстром и тщательном изучении и опровержении визуального пользовательского контента (UGC), поскольку проверка визуального контента требует высокого уровня квалификации и может быть чрезвычайно сложной с учетом существующих вычислительных инструментов, используемых в новостных агентствах» [46, с. 1]. Очевидно, что ответственность за минимизацию рисков дезинформации ложится не только на сами медиаорганизации, но и на аудиторию. Именно поэтому важно не только внедрять автоматические методы проверки контента, но и в целом повышать осведомленность аудитории о потенциальных рисках, связанных с потреблением генеративных материалов.

Как отмечалось ранее, несмотря на то, что генеративные нейросети создают правдоподобный фейковый контент, они также могут быть использованы для борьбы с распространением дезинформации и ложных новостей [33], [46]. Машинное обучение и искусственный интеллект предоставляют новые возможности для выявления и фильтрации недостоверной информации в интернете и социальных сетях. Одним из подходов является использование алгоритмов машинного обучения для анализа лингвистических особенностей текстов, таких как морфологические закономерности и читабельность. Эти методы, не зависящие от темы и событий, освещаемых в материале, позволяют обнаруживать характерные паттерны, присущие фейковым новостям. Яркий пример – использование глубокого обучения для выявления специфических сигналов естественного языка, характерных для кликбейтных заголовков и манипулятивных текстов. Другой перспективный подход основан на анализе моделей распространения фейков в социальных сетях: используя теорию графов, или математический инструмент для описания сетей, где узлы (сайты и аккаунты) соединены связями (репостами, лайками, комментариями) можно проанализировать структуру сетей, и такие метрики, как центральность (важность узла в сети), изучить паттерны распространения фейковых новостей и выявить ключевые узлы в сетях дезинформации. Машинное обучение на основе графов позволяет анализировать связи между различными новостными сайтами и оценивать их надежность исходя из структуры их взаимодействий между собой и с пользователями.

Для эффективной борьбы с фейками необходимо междисциплинарное сотрудничество специалистов в области социальных и компьютерных наук: совместными усилиями можно разработать универсальные механизмы, способные работать в масштабных системах, как в сети в целом, так и в конкретных социальных медиа в частности. Технологичные методы будут полезны для моделирования и анализа сложных информационных систем, выявления закономерностей и взаимосвязей между надежными и

ненадежными источниками. Это позволяет уже не просто определить отдельные тезисы или целые материалы как фейковые, но и понять общую структуру и динамику распространения фейковых материалов в сети. Очевидно, что в этом контексте качественная журналистика, основанная на объективных и беспристрастных фактах, разнообразии мнений и ответственности перед общественностью, становится как никогда важной для противодействия обману и укрепления доверия общества к СМИ. Это особенно актуально в условиях, когда индустрия столкнулась с серьезными вызовами, среди которых изменение привычек потребления новостей у молодого поколения, недоверие к традиционным медиа из-за предвзятости и распространение фейков. Эти факторы побуждают качественные медиа искать новые бизнес-модели, основанные на платном контенте и открытых отношениях с аудиторией. Так, страны с сильным сектором качественных медиа, например Норвегия, демонстрируют высокий уровень доступа к платному контенту и доверия к новостям, что подчеркивает уже важность инвестиций в качественную журналистику и построения лояльных отношений с читателями [33], [46].

Международные издания, например, The Guardian, успешно адаптируются к цифровой эпохе, делая акцент на честных и открытых отношениях с аудиторией, основанных на доверии и высоком качестве контента. Такой подход помогает им оставаться независимыми и жизнеспособными в условиях меняющихся условий медиа. Как отмечает в своем исследовании Лукина М.М.: «Так как применение технологий искусственного интеллекта до сих пор не стало общепринятым явлением в большинстве редакций, вопрос о внесении изменений в этические кодексы и редакционные стандарты пока активно не поднимается профессиональным медийным сообществом. Однако размышления о возможных профессиональных этических конфликтах, связанных с применением ИИ в редакциях и журналистами, и о потенциальных моральных рисках можно встретить и в экспертных оценках, и в академическом дискурсе. Так,

например, обсуждается вопрос о влиянии алгоритмов ИИ на одну из фундаментальных ценностей в сфере журналистской этики, которой является объективность и непредвзятость. Эти категории, закрепленные в кодексах журналистской этики всех стран мира, подразумевают умение журналиста быть над схваткой, освещать все точки зрения, не занимать чью-либо сторону, уйти от субъективных оценок, что в жизни достичь крайне трудно. Эту задачу смогут решить системы искусственного интеллекта, а в редакциях смогут избежать субъективной оценки, так как для сбора информации и генерации текстов были применены алгоритмы, а не труд человека, наделенного личными взглядами и стереотипами. В этом случае можно говорить о так называемой «механистической объективности». Технооптимисты считают, что ИИ может обладать универсальной объективностью, так как исключается человеческий фактор. Однако многие ученые не принимают такой подход и считают его слишком идеалистичным. Так, Карлсон отмечает, что алгоритмам не стоит приписывать универсальную объективность, так как на практике они также могут допускать ошибки, однако природа этих промахов иная по сравнению с автором-человеком. Исследователь считает, что наделение алгоритмов некой механистической объективностью несет в себе определенные риски, может привести к тому, что само понятие профессиональной объективности в деятельности журналиста начнет размываться. В отличие от технооптимистов приверженцы более критического отношения к технологическим новациям предлагают концепцию так называемой «алгоритмической предвзятости» (по англ. *algorithmic bias*). Согласно этой концепции, алгоритмические решения могут способствовать проявлению субъективности в силу особенности работы того или иного алгоритма» [20, с. 688]. Концепция «алгоритмической предвзятости» на современном этапе развития технологии представляется более корректной. Очень важно в этом вопросе все же учитывать и влияние данных, используемых для обучения ИИ и нейронных сетей: если они содержат стереотипы, предвзятость, предубеждения, ошибки, неточности,

неверные метки – эти проблемы будут только усиливаться на этапе обучения технологии. Оптимистичный взгляд на «объективный ИИ» представляется утопичным, поскольку фактически требует, чтобы обучающие данные были абсолютно нейтральными, безошибочными, «идеальными», но обработка данных всегда требует участия человека в их подготовке, а это неизбежно вносит в них субъективность.

Заслуживающая доверия журналистика с поддержкой ИИ и нейросетей требует баланса между работой человека и автоматизацией. Технологии трансформируются из инструмента в фактор, меняющий новостное производство. Нейросети повышают точность, актуальность и разнообразие новостей, укрепляют честную журналистику, но они же имеют и двойственный потенциал: могут как бороться с фейками, так и создавать их, в зависимости от целей человека. Симбиоз качественной журналистики и нейронных сетей открывает новые возможности для медиа, но требует ответственного подхода и регулирования этой области для предотвращения неэтичного использования технологий [55].

Дискуссионными вопросами остаются авторское право на результаты работы нейросети, конфиденциальность обрабатываемых данных (их защита) и конкуренция между компаниями во внедрении инноваций [54]. Навыки аналитиков больших данных и инфраструктура СМИ, техническое оснащение организации также играют немаловажную роль: для успешной работы с данными требуется квалифицированный персонал из различных областей – компьютерных наук, математики, экономики, а также экспертной области, в которой обучается нейросеть. Программное обеспечение, необходимое для анализа больших данных, также требует значительных инвестиций, что создает проблемы для малого и среднего бизнеса, особенно в странах с низким и средним уровнем дохода. Еще одной важной проблемой остается сложность обработки больших объемов данных – они могут быть многомерны, разнообразны, сложны, могут включать шумы и ошибки, быть неполными. Это создает трудности на этапе предварительной обработки – очистки,

выборки, нормализации. Такого рода трудности могут влиять на эффективность работы моделей машинного обучения, которые требуют высокой точности данных на входе для точного результата.

С.В. Водопетов в рамках изучения вопроса этики для связи «журналистика – искусственный интеллект» в своей работе отмечает риски использования искусственного интеллекта в работе журналиста и в подкрепление приводит мнение Е.А. Зверевой: «Во-первых, как уже отмечалось ранее, есть риск превращения журналистики в маркетинг, что в целом будет противоречить научной и моральной этике этой сферы. Контент станет максимально кастомизированным, а, следовательно, та самая общая история будет утеряна. Во-вторых, искусственные интеллектуальные системы, лишённые чувств нравственности и морали, просто не способны «мыслить» соответствующим образом. Так, в рамках своего исследования Зверева Е.А. выделяет следующие этические проблемы применения ИИ в журналистике:

- более простой и удобочитаемый контент, создаваемый нейронными сетями, может быть использован как инструмент манипуляции;
- активно развивающийся персонализированный контент создает «эхо-камеры»; создаваемое замкнутое информационное поле для довольно ограниченного круга людей может привести к тому, что конкретно здесь не будут приняты общечеловеческие ценности;
- фейковые новости, а также неверные выводы и суждения, которые способны сформировать роботы, не контролируемые или слабо контролируемые человеком.

Вполне вероятно, что частично данные проблемы будут решены путем включения положений о применении ИИ в Кодекс журналистской этики, но в любом случае ответственность будет лежать на самом человеке» [9, с. 59].

Применение нейросетей в журналистике создает новые вызовы для редакций – СМИ должны адаптировать сгенерированные тексты под свою стилистику, тематическую направленность и ценности. Результаты работы

технологии из-за этого требуют постоянной доработки. Использование моделей также поднимает ключевые этические вопросы: определение авторства текстов, обеспечение прозрачности процесса создания и ответственность за контент. Необходимы четкие принципы, гарантирующие соответствие генеративных материалов профессиональным и этическим стандартам журналистики. Текущее обсуждение этих проблем недостаточно глубоко рассматривает как технические аспекты, так и этические нормы применения. Причина может заключаться в недостаточной осведомленности журналистов о принципах работы нейросетей, этических аспектах внедрения этих технологий в работу.

Для решения социальных, этических и политических вопросов, связанных с нейросетевыми технологиями и искусственным интеллектом, необходим многосторонний подход. По мнению Л. Уччи, он может включать:

- «обеспечение корректной информации для общественности о сути технологии и принципе работы ИИ, особенностях генерируемого контента, возможных рисках в потреблении и использовании такого рода контента;
- привлечение экспертов по этике и ИИ к научным исследованиям и общественной дискуссии для более глубокого освещения темы;
- разработку четкой государственной политики и нормативной базы для технологий ИИ с учетом государственной политики» [56, с. 927].

Это позволит минимизировать риски негативных социальных последствий, связанных с искусственным интеллектом, и выработать сбалансированный подход к освещению и регулированию этой технологии с учетом интересов всех вовлеченных в процесс сторон (разработчиков нейронных сетей, редакций СМИ, журналистов, читательской аудитории). Одним из примеров, демонстрирующих необходимость такого подхода, может служить стремительное развитие нейросетей для генерации текста (или автоматической текстовой генерации – automated text generation или ATG). Эта тема во взаимосвязи со взглядом человека на технологию подробно освещена

в работе Лерман-Аэнестрога А. и Киммерле Й. «Понимание и восприятие автоматического генерирования текстов среди населения: два исследования с репрезентативными выборками в Германии» [50]. Генеративные нейросети способны в больших масштабах создавать высококачественный текстовый контент, а его трудно отличить от журналистских материалов. В настоящее время отсутствует единое мнение о том, какие тексты, сгенерированные ИИ, можно считать приемлемыми, а какие не допускать к распространению. Цифровые платформы пока не выработали четкой политики в отношении использования нейросетей для генерации текста, а уже действующие нормативные акты и законы не в полной мере учитывают риски, связанные с технологией. Чтобы решить эти проблемы, необходимо глубже изучать особенности генеративных текстов, определять правила применения технологии и совершенствовать законодательство для должного контроля. Доказано, что люди воспринимают сгенерированные ИИ тексты как заслуживающие доверия (наравне с текстами, созданными человеком), а это только подчеркивает важность разработки мер по повышению осведомленности общественности о возможностях и ограничениях генеративного контента, а также необходимость его четкой маркировки, чтобы читатели могли принимать осознанные решения о том, насколько они доверяют материалу и источнику.

Широкое распространение генеративных текстов уже вызывает беспокойство у экспертов. Л. Иллиа отмечает, что «Очевидная способность новых агентов искусственного интеллекта генерировать текст, похожий на человеческий, наряду с низкими затратами и потенциалом масштабного распространения текстов, создаваемых ИИ в режиме реального времени, вызывает споры в технологической индустрии. Например, Тимнит Гебру, один из руководителей команды Ethical AI Google, был недавно уволен после того, как отправил внутреннее письмо, в котором поднимались вопросы о предвзятости, встроенной в будущее, гиперсовременное ИИ Google (ключевая область инвестирования), и утверждалось, что требуется

значительная работа для выявления и решения этических проблем, связанных с распространением агентов ИИ для автоматического генерирования текста» [42]. Учитывая способность передовых моделей, таких как GPT-4, создавать тексты, которые все сложнее отличимы от написанных человеком, возникают опасения, что нейросети могут начать играть все более активную роль в формировании контента онлайн-коммуникаций. Это приведет к значительному влиянию на общественное мнение и потребует переосмысления этических последствий использования нейронных сетей. Когда модели смогут массово производить убедительный правдоподобный контент, который будет восприниматься людьми как заслуживающий доверия, это приведет к серьезным последствиям для общества. Модели могут быть использованы для распространения дезинформации и манипулирования общественным мнением. Широкое использование генеративного контента может привести к снижению роли человека в создании и распространении информации, а это уже имеет негативные последствия для журналистики как профессии.

Для решения этических проблем предлагаются два основных подхода: контроль над ИИ и нейросетями и ограничение их использования различными организациями. В контексте вопроса Л. Илия отмечает: «Существуют два основных подхода: контроль самих ИИ-агентов или ограничение их использования различными организациями. В рамках первого подхода разработчики ИИ могут включить в системы специальную «программу честности», которая исключает возможность искажения или намеренного введения в заблуждение. <...> Также можно встраивать определенные механизмы «суждения» для предотвращения проблемы минимального знаменателя, ограничивая источники, из которых ИИ обучается. Однако такая система контроля может быть спорной, поскольку сложно провести грань между исключением недостоверных данных и цензурой. Этот вопрос требует выработки руководящих принципов, ответственность за создание которых могла бы взять на себя независимая организация, состоящая из экспертов и

назначенная государством. <...> Однако интеграция «программы честности» не является единственным решением. Другой подход – ограничение доступа к мощным ИИ-агентам для недобросовестных акторов или обучение пользователей правильному использованию технологий. Роль в регулировании этого процесса могли бы взять на себя государственные учреждения. <...> Важно учитывать, что сами правительства тоже могут быть недобросовестными пользователями. В этом контексте стандарты «образования» ИИ должны регулироваться аналогично школьным программам, чтобы избежать манипуляций. Еще одним решением проблемы посредничества может стать разработка рекомендаций для корпораций и организаций, касающихся использования ИИ-агентов. Например, применение ИИ для создания корпоративной коммуникации с заинтересованными сторонами должно быть дополнено участием человека. Это могло бы стать частью практики корпоративного управления и помочь снизить вероятность недоразумений. В заключение, учитывая, что контент, создаваемый ИИ-агентами, становится все более сложным и труднодоступным для выявления, решить проблему будет непросто. Этот вопрос требует скоординированных усилий со стороны множества организаций, включая рекомендации, подобные указанным выше. Только комбинированный подход позволит эффективно противостоять вызовам, связанным с развитием ИИ» [42].

Одним из заметных шагов в этом направлении является разработка этических принципов и норм использования искусственного интеллекта. В 2021 году в России был предложен «Кодекс этики в сфере ИИ» в качестве устанавливающего документа в этой области. Документ во многом схож с «Азиломарскими принципами» 2017 года от группы экспертов в области ИИ, права и этики – «Азиломарские принципы» представляют собой рекомендации по ограничению работы нейросетей и использованию умных технологий во благо человечества. Касательно содержания самих принципов А. Р. Акуличева пишет следующее: «Основные принципы были сведены к тому, что технологии должны быть безопасны для людей, они не должны

использоваться во вред людям и человек должен иметь возможность контролировать ИИ, а не наоборот» [2]. Схожие идеи были сформулированы еще в 1942 году Айзеком Азимовым в его трех законах робототехники.

Кодекс и принципы призваны обеспечить безопасное и ответственное развитие технологий ИИ, но имеют лишь рекомендательный характер. На сегодняшний день законы, полноценно регулирующие использование этих технологий в медиасфере, отсутствуют. Это создает риски злоупотребления, распространения фейков, дипфейков, манипулирования мнением общества. Нужны законы, устанавливающие четкие правила использования ИИ в журналистике и ответственность за их нарушение. Только сочетание этического саморегулирования с правовыми механизмами обеспечит безопасное применение этой технологии в медиа. Для эффективного решения этических проблем потребуются в том числе практические механизмы их реализации: системы мониторинга использования ИИ и нейронных сетей, образовательные программы для журналистов, правовые инструменты.

Серьезной проблемой в журналистике остается недопонимание возможностей и рисков технологии в профессиональной деятельности. Стремительное развитие этой области опережает темпы изучения новых инструментов и методов журналистами, из-за чего специалисты могут не в полной мере осознавать специфику работы нейросетей, их сильные и слабые стороны, области применения и сопутствующие риски. Отсутствие необходимых знаний и навыков может приводить к некорректному использованию нейросетей, снижать качество материалов и доверие аудитории, как отмечает Феликс М. Саймон в своем отчете «Искусственный интеллект в новостях. Как инструменты ИИ рационализируют и трансформируют журналистику и публичное пространство» [62]. Проблема становится особенно актуальной в свете того, что в последние годы новостные агентства по всему миру активно внедряют решения на основе ИИ в своих редакциях. Дополнительными препятствиями становятся конкуренция за таланты и «утечка мозгов» в технологическую отрасль.

ИИ и нейросети также часто используются для повышения вовлеченности аудитории с помощью рекомендательных систем и адаптации бизнес-стратегий изданий под запросы читателей, например, для прогнозирования отмены подписки или создания персонализированных платных подписок. Эти алгоритмы помогают изданиям увеличивать доходы, но некоторые из технологических решений создаются сторонними организациями и крупными платформами, что поднимает еще один вопрос – о безопасности передачи личных данных компании в модель, поддерживаемую сторонней компанией, а также заинтересованности третьей стороны в таких данных.

Интересно, что модели обработки естественного языка (NLP) используются в издательствах реже, возможно, из-за сложности их воспроизведения на разных языках. Часто автоматизация в СМИ сводится к базовому применению простых алгоритмов для написания шаблонных историй или применению компьютерного зрения как инструмента для журналистских расследований и проверки фактов в социальных сетях. Из-за необходимости в построении инфраструктуры, найме квалифицированного персонала и инвестиций, изучаемые технологии чаще используются крупными новостными агентствами. Последнее поднимает вопрос состязательности между СМИ и проблему необходимости достаточного финансирования для внедрения новых технологий в небольшие издания. Многие ИИ-приложения в новостной индустрии разрабатываются на гранты крупных технологических компаний, что создает препятствия также и для независимости разработок.

Многие проекты пока полагаются на простую автоматизацию, заполняющую пробелы в шаблонных материалах, но некоторые уже используют более продвинутые подходы, такие как обработка естественного языка и машинное обучение. Например, в Бразилии новостные агентства используют ИИ в Twitter, применяя технологию для обработки больших объемов данных и анализа взаимодействий пользователей на цифровых

платформах. Несмотря на ожидание снижения затрат, модели ИИ обычно создаются под конкретные проекты и цели, и каждый раз требуют новых инвестиций. Нейронные сети могут быть использованы также и в негативном ключе. Исследователи отмечают фальсификацию выборов США в 2016 году – в своем материале Заркадакис Г. указывает следующее: «почему традиционные СМИ на выборах в США в 2016 году, несмотря на то, что почти все они были настроены против кандидата Трампа, не смогли помешать ему стать президентом? Ответ: потому что в 2016 году в либеральных демократических странах развитого мира политический дискурс изменился, и алгоритмы искусственного интеллекта взяли на себя роль традиционных СМИ в формировании общественного мнения и влиянии на избирателей» [66]. В период предвыборной кампании были зафиксированы случаи использования ботов в социальных сетях, особенно Twitter и Facebook, для распространения ложных новостей, пропаганды и дезинформации. Более того, модели часто обучаются на старых предвзятых данных, а это приводит к проблемам с этикой и достоверностью уже без злого умысла, воспроизведению стереотипов и предубеждений в новых текстах: если исходные данные содержат в себе искаженные представления о социальных группах или отражают дискриминационные установки – обученные на них модели ИИ будут принимать решения, основанные на этих предубеждениях.

М.-Ф. де Лима-Сантос и Уилсон Керон отмечают, что: «Традиционные новостные издания, такие как The New York Times, The Washington Post и Associated Press, успешно внедрили проекты на основе ИИ в свои редакционные процессы. Однако даже для таких организаций использование ИИ остается затратным. Более того, ведущие новостные компании должны существенно продвинуться, чтобы полностью понять потенциал ИИ в журналистике. Например, выход GPT-3 в 2020 году поднял модели машинного обучения текста на новый уровень. Эта «авторегрессивная языковая модель с 175 миллиардами параметров, что в 10 раз превышает показатель любых предыдущих несжатых моделей способна выполнять множество задач, таких

как генерация статей, перевод, создание резюме текстов и прогнозирование, при этом используя меньше вычислительных ресурсов. Тем не менее, эта применимость несет за собой и риски, такие как создание искаженного контента, который может вводить аудиторию в заблуждение» [36, с. 5]. Достижения в области нейросетей существенно изменяют взаимодействие специалистов с технологиями в журналистике, но требуют не только ответственного подхода к их использованию и подготовке специалистов, но и осознания долгосрочных последствий, включая риск искажений и манипуляций, что подчеркивает важность изучения этих процессов.

Таким образом, интеграция искусственного интеллекта в журналистскую практику представляет собой двойственное явление: с одной стороны, различные типы нейросетей (FNN, RNN, MLP, CNN, генеративные модели и трансформеры) значительно повышают скорость и эффективность создания контента в журналистике, а с другой – вызывают серьезные этические и правовые вызовы, связанные с достоверностью информации. Корректный подход к использованию и контролю использования ИИ и нейронных сетей в журналистике должен быть комплексным, включать как технические аспекты (совершенствование алгоритмов верификации), так и организационные меры (повышение квалификации специалистов, междисциплинарное сотрудничество, более точечная переработка стандартов этики). Только сбалансированное сочетание этих подходов способно обеспечить конструктивное развитие журналистики в эпоху цифровизации, сохранить доверие аудитории к СМИ как фундаментальную ценность всей медиасферы.

Выводы главы первой.

Для предметной области исследования важны такие термины как «искусственный интеллект», «машинное обучение», «нейронные сети» и «глубокое обучение», содержание которых можно определить следующим образом: искусственный интеллект – технология, позволяющая компьютерам имитировать человеческое мышление без наличия сознания; машинное

обучение – подраздел ИИ, позволяющий компьютерам обучаться на данных без явного программирования каждого шага; нейронные сети – конкретный тип алгоритмов машинного обучения, имитирующий работу человеческого мозга через систему соединенных узлов (искусственный аналог биологических нейронов и их связей); глубокое обучение – метод, который использует сложные многослойные нейронные сети для анализа комплексных данных. Существенное значение в технологии применения нейросетей имеют различные виды обучения (контролируемое, неконтролируемое, полуконтролируемое) и типы нейросетей (сети прямого распространения, рекуррентные, сверточные, многослойные перцептроны, генеративные модели, трансформеры), которые повышают производительность создания журналистского контента, но также порождают разные этические и правовые вызовы.

Сбалансированная интеграция ИИ-технологий в журналистику требует комплексных решений, которые будут сочетать технические аспекты (совершенствование алгоритмов), организационные меры (повышение квалификации журналистов, междисциплинарное сотрудничество), адаптацию этических стандартов и правовое регулирование, направленное на защиту достоверности информации. Комплексный подход будет способен обеспечить конструктивное развитие журналистики в цифровую эпоху при сохранении доверия аудитории как фундаментальной ценности информационной сферы.

Глава 2 Анализ опыта, качественная оценка и методические рекомендации по использованию генеративного контента в журналистике

2.1 Анализ современного опыта использования нейронных сетей в сфере журналистики

Технологии ИИ и нейронных сетей все глубже проникают в журналистскую практику, трансформируя традиционные подходы в создании медиаконтента. Актуальность эмпирического исследования обусловлена тем, что стремительное внедрение нейросетевых технологий в процессы редакций значительно оторвано от систематизации практических подходов к их использованию. Разрыв между теорией и практикой, особенно в области работы с данными, требует изучения кейсов применения нейросетей в российских и англоязычных СМИ для выработки рекомендаций по оптимизации взаимодействия журналиста и нейросетевых технологий, а также минимизации связанных с этим рисков.

Основная проблема эмпирического исследования заключается в противоречии между потенциалом нейросетей с одной стороны, а также необходимостью постоянного контроля качества генеративного контента человеком – с другой стороны. Нейросети, призванные улучшать и упрощать работу журналиста, для эффективного внедрения в практику, требуют системного подхода, включая критическую доработку контента, проверку фактов и адаптацию материалов под стандарты редакции. Качество такого контента напрямую зависит от грамотной работы с нейросетью и качеством подготовки вводных и очистки выводных данных.

Исследование включает систематический обзор научных публикаций и оригинальный анализ генеративного контента. На подготовительном этапе отбирались релевантные научные работы, содержащие данные о практическом применении нейросетей в редакционных процессах. В исследованиях

эффективность различных архитектур нейросетей оценивается по унифицированным метрикам: точность (accuracy), прецизионность (precision), полнота (recall) и F-мера, что обеспечивает объективное сопоставление результативности моделей. Точность измеряет долю правильных прогнозов среди всех сделанных прогнозов, прецизионность определяет, насколько нейросеть точна в положительных прогнозах, полнота – какую часть всех реальных положительных случаев нейросеть смогла корректно идентифицировать, а F-мера представляет собой среднее между прецизионностью и полнотой. Для сравнимости результатов анализировались работы со стандартизированными наборами данных (о выборах, отзывы на авиалинии, автомобили, фильмы). Систематизация публикаций по результатам применения нейросетей в журналистике легла в основу первого пункта текущей главы.

Эмпирической базой для исследования генеративных материалов послужили материалы сетевой версии газеты «РБК» – «Выпуск газеты от 28 апреля 2023 № 061 (3731) (2804)» [10] был выпущен с использованием нейросетей GigaChat и Kandinsky 2.1; а также материалы из раздела «Написано ИИ» на «РБК Тренды» [21]. Тексты из раздела отбирались по порядку для достоверности исследования (не фильтровались ни по каким другим признакам). Для сопоставимости в сравнении год к году анализировались весенние материалы за 2024 и 2025 годы. В ходе работы был проведен качественный контент-анализ: генеративные тексты оценивались с точки зрения содержания и структуры, лингвистических особенностей, фактологической достоверности, функциональности и применимости для аудитории. После оценки каждой группы текстов (по годам) формулировались выводы по ограничениям моделей и проблемам проанализированного контента. Для сравнения результатов анализа были дополнительно проанализированы проблемы, выделенные газетой «Краснодарские известия» после работы редакции с нейросетью.

Работа проводилась поэтапно: сначала были осмыслены теоретические представления по применению технологии в СМИ, далее проводился качественный анализ генеративных материалов, по итогам анализа были разработаны практические рекомендации. Результаты проведенной работы позволяют не только расширить теоретические представления о возможностях и ограничениях нейросетей в журналистике, но и предложить конкретные инструменты для повышения качества и эффективности их применения в редакционной практике.

Интеграция нейросетевых технологий в современное медиапространство открывает новые возможности для оптимизации журналистских задач. Особенно существенно трансформируют и упрощают работу журналиста генеративные модели. Анализ применения искусственного интеллекта в медиасфере выявил несколько ключевых направлений его использования. Во-первых, нейросети эффективно выполняют функцию мониторинга и анализа трендов, систематизируют массивы информации, выделяя значимые темы, что позволяет редакциям оперативно реагировать на изменения информационной повестки. Это отмечает Акуличева А. Р в работе «Нейросеть и журналистика: этический вопрос использования умных технологий в СМИ» [2]. Здесь же важно понимать, что полностью автоматизировать аналитические процессы невозможно – нейросеть должна выступать вспомогательным инструментом, оптимизирующим, но не заменяющим аналитическую работу журналиста. Во-вторых, алгоритмические решения существенно расширяют возможности персонализации медиаконтента – изучение пользовательских предпочтений позволяет создавать индивидуальные подборки материалов и повышать вовлеченность аудитории, что описано Асташкиным А.Г. в его исследовании «Нейросети в структуре журналистской профессии» [3]. Здесь нейросети выполняют роль посредника между журналистами и потребителями информации: для медиапрофессионалов они автоматизируют рутинные

процессы, для пользователей формируют персонализированный опыт взаимодействия с контентом.

Особую ценность представляет способность нейросетей автоматизировать создание шаблонного контента на основе данных, например, небольших спортивных сводок, метеорологических прогнозов, финансовых отчетов [3], [13], [23]. Это освобождает журналистов от механической работы и позволяет сосредоточиться на материалах, требующих творческого подхода и глубокого анализа. Нейронные сети также обладают значительным потенциалом в области фактчекинга и верификации данных – такая автоматизированная проверка фактов повышает точность материалов, хотя и требует последующего контроля со стороны человека.

Нейросети демонстрируют высокую эффективность в анализе текстов, выявлении трендов и паттернов, а также в классификации информации, что позволяет журналистам быстрее обрабатывать большие объемы данных и генерировать сводки новостей [13], [23]. При внедрении нейросетевых решений в редакционные процессы необходимо учитывать технические ограничения, связанные с объемом обрабатываемых данных и доступными вычислительными мощностями. Для эффективного использования моделей в специализированных журналистских задачах может потребоваться разработка специализированных решений, адаптированных под потребности конкретной редакции.

Масштабный анализ публикаций, проведенный Э. Фернандес, С. Моро и П. Кортес в исследовании «Наука о данных, машинное обучение и большие данные в цифровой журналистике: обзор современных достижений, вызовов и возможностей» [39], подтверждает стремительное расширение сферы применения нейросетей в журналистике. Показательно, что половина всех исследований в этой теме опубликована за последние три года из рассмотренного авторами десятилетнего периода. Исследователи выделяют шесть ключевых направлений интеграции искусственного интеллекта в медиапрактику: интеллектуальный анализ текста, анализ онлайн-

комментариев, системы рекомендаций, автоматизированная журналистика, анализ данных и алгоритмические рекомендации. Такая динамика свидетельствует о формировании устойчивого тренда на внедрение нейросетевых технологий в редакционную практику с целью повысить конкурентоспособность медиакомпаний.

Нейронные сети эффективно обрабатывают и анализируют большие объемы данных, особенно из финансовой области, данных о погоде и спортивной статистике. Журналист в этом случае сосредотачивается на творчестве и анализе, а рутинный сбор и первичная обработка передаются в модель. Асташкин А.Г. в своем исследовании показывает следующие примеры изданий, которые пошли по такому пути: «ТАСС: использование нейросетей для автоматического сбора и первичной обработки данных о международных событиях, что позволяет журналистам оперативно получать актуальную информацию без необходимости вручную мониторить все источники.

РИА Новости: разработка систем на основе искусственного интеллекта для анализа больших данных, что позволяет автоматически выявлять новостные тренды и значимые события на региональном и глобальном уровне.

Ведомости: применение алгоритмов машинного обучения для сбора и агрегации экономических показателей, что ускоряет процесс создания аналитических материалов по финансовым темам» [3].

Самое популярное направление в использовании нейросетей – текстовая генерация. Генеративные модели, работающие с текстами, могут создавать как небольшие сообщения или несложные отчеты, так и полноценные статьи. Модели особенно полезны в случаях, когда генерация основана на фактах и не требует творчества, уникальности или оригинальности подачи материала. Например, если нужно по шаблону создать репортаж. В своем исследовании А.Г. Асташкин приводит примеры и для этого направления:

«Lenta.ru: эксперименты с генерацией новостных сводок по предварительно заданным параметрам, например, для автоматического создания текстов о погоде, спортивных результатах или курсах валют.

Коммерсантъ: использование нейросетей для создания первичных версий репортажей о культурных событиях, которые затем дорабатываются журналистами, добавляется анализ и экспертные комментарии.

Mail.ru News: применение искусственного интеллекта для автоматической генерации новостного контента на основе анализа тенденций в социальных сетях и интернете, что позволяет быстро реагировать на интересы аудитории.

РБК: активное использование нейросетей на всех этапах создания медиа продукции: контент, дизайн, верстка. В 2023 году был представлен номер газеты РБК, полностью созданный нейросетями» [3].

Нейросети анализируют огромные объемы информации для выявления общественных, экономических и политических трендов, помогают журналистам создавать глубокие и всесторонние материалы. Здесь А.Г. Асташкин отмечает СМИ: «RT (Russia Today): Анализ социальных медиа и онлайн-форумов с использованием нейросетей для выявления настроений и мнений по актуальным темам, что помогает формировать редакционную политику и выбор тем для освещения.

РБК: Применение алгоритмов машинного обучения для анализа финансовых рынков и прогнозирования экономических тенденций на основе больших данных, что используется при создании глубоких аналитических материалов.

Известия: Использование искусственного интеллекта для мониторинга и анализа публичных высказываний политических лидеров и общественных деятелей, что позволяет выявлять тренды и изменения в политическом дискурсе» [3].

Нейросети также способны качественно персонализировать новостной поток для пользователей, анализировать их предпочтения и поведение,

подбирать наиболее релевантный и интересный контент, тем самым вовлекая читателя. Асташкин А. Г. отмечает следующее: «Яндекс.Новости разрабатывают и используют алгоритмы машинного обучения для персонализации новостных лент пользователей на основе их интересов, предпочтений и истории просмотров, что улучшает пользовательский опыт, повышает вовлеченность аудитории.

«ВКонтакте» интегрирует нейросетевые решения для персонализации ленты новостей в социальной сети, позволяя пользователям видеть более релевантный контент из различных медиаисточников.

Rambler&Co применяет ИИ для адаптации контента на основе поведенческих факторов аудитории, способствуя более глубокому погружению пользователей в медиапространство компании.

В контексте журналистской профессии, искусственный интеллект и нейросети играют роль катализаторов для радикальных изменений, внося инновации в процессы сбора, обработки и распространения информации. Эти технологии предлагают решения для автоматизации рутинных операций, позволяя журналистам выделять больше времени на творческие и аналитические задачи, одновременно повышая точность и объективность представляемых данных» [3]. По мнению А. Р. Акуличевой, на данном этапе журналист обладает следующим неоспоримым преимуществом перед нейросетями: «профессиональным журналистам нечего опасаться и нейросеть не сможет заменить живого журналиста с его восприятием мира, событий, которое отражается в эмоциональной, яркой, творческой подаче информации о событиях и проблемах» [2].

Особенно ценны преимущества моделей при работе с большими массивами информации, когда человеческих возможностей оказывается недостаточно для эффективного анализа. Современная журналистика все чаще сталкивается с необходимостью обработки таких данных перед созданием аналитических материалов и расследований, а нейронные сети способны анализировать их и делать прогнозы на основе выявленных шаблонов и

взаимосвязей. Для достижения точных, или близких к точным, прогнозов, данные, которые подаются на вход, должны быть надлежащим образом обработаны и подготовлены в несколько шагов – можно систематизировать на основании исследований Кириллова Д.С. и Молостова Д.Д., описанных в работе «Способы предварительной обработки данных для прогнозирования с применением нейронных сетей» [17].

Первый – очистка данных: включает в себя устранение ошибок и несоответствий в наборе данных (дублирующиеся записи, пропущенные значения или логические ошибки). Цель такой очистки – подготовить данные во избежание искажений в процессе обучения. Некачественные данные могут привести к ошибкам, снижению точности прогнозов и общему ухудшению работы модели. Очистка данных включает исправление неправильных значений, проверку корректности формата данных, удаление нерелевантной информации. Этот критически важный этап закладывает основу для дальнейшей обработки.

Второй – извлечение признаков: включает выбор наиболее релевантных переменных. Процесс позволяет выделить ключевые параметры из больших массивов данных, наиболее значимых для анализа и в обучении модели. Извлечение признаков также уменьшает размерность данных, что упрощает обработку и повышает эффективность работы модели. На этом этапе происходит идентификация тех характеристик, которые напрямую влияют на анализируемую задачу. Лишние и малозначимые данные исключаются из обработки.

Третий – нормализация данных: выступает заключительным этапом подготовки данных. Это процесс масштабирования данных для приведения их к единому диапазону, формату и масштабу. Так набор данных становится единообразным. Нормализация помогает избавиться от возможных искажений, возможных из-за разницы в масштабах переменных. Данные должны быть чистыми, чтобы их можно было проанализировать без какой-либо путаницы.

На точность и ценность результатов обработки данных также большое влияние оказывает и выбранная архитектура нейросети: различные архитектуры могут иметь свои преимущества в зависимости от характера данных и целей анализа. Для крупномасштабного прогнозирования и децентрализованной классификации данных могут быть использованы эхо-сети (ESN, Echo State Network) – подтип рекуррентных нейронных сетей. Чирома Х., Али У., Абдулахамид Ш.М. подробно рассматривают эту и описанные ниже разновидности нейросетей в своей работе «Прогресс в области искусственных нейронных сетей для анализа больших данных: обзор» [34]. Эхо-сети особенно полезны для анализа временных рядов и распределенной между разными устройствами. В журналистике такие сети могут быть применены для выявления трендов в больших массивах данных, например, при подготовке экономических обзоров или анализе социальных явлений, где информация поступает из множества источников.

Нейросети с обратным распространением ошибки (BPNN) представляют собой эффективный инструмент для работы с большими данными благодаря интеграции с платформами для их обработки, такими как Hadoop или Spark. Они позволяют ускорить обработку информации и снизить время от сбора данных до публикации материала, оптимизируя использование вычислительных ресурсов (стандартных компьютеров или серверов). Для журналистских задач это ценно благодаря возможности оперативно анализировать массивы информационных потоков с минимальными затратами времени.

Глубокие нейронные сети (DNN, Deep Neural Networks) демонстрируют высокую точность в анализе временных данных, анализе, обработке и классификации текстовых данных – их способность самостоятельно находить скрытые связи без предварительных указаний от пользователя делает их особенно полезными для обнаружения неочевидных тенденций в новостных потоках или социальных медиа. Благодаря послойной неконтролируемой предварительной обработке данных каждая часть этой сети обучается

отдельно, а это значительно повышает общую эффективность. Такие модели помогают в анализе сложных текстовых массивов и поиске закономерностей для создания аналитических материалов.

Сверточные нейронные сети (ConvNet) оптимальны для анализа изображений и создания интерактивных визуализаций. Их применение охватывает обработку визуального контента, анализ графики, фотографий и создание инфографик для иллюстрации материалов. Такие инструменты особенно полезны для повышения наглядности данных, аналитики, статистики.

Роевые и эволюционные нейронные сети могут выступать как вспомогательные механизмы для повышения точности и скорости обработки больших данных. Они пригодны для оперативной подготовки журналистских расследований, в которых требуется анализировать большие массивы информации. Эти сети работают как поддержка основной модели, улучшают ее производительность в работе со сложными задачами.

В исследовании «Мини-обзор машинного обучения в аналитике больших данных: приложения, проблемы и перспективы» [54] применение нейронных сетей в работе с большими данными было измерено количественно: большинство научных статей (58 %) основаны на прогнозной аналитике больших данных. Многие из них используют методы машинного обучения, включая нейронные сети. Глубокие нейронные сети (DNN) и сверточные нейронные сети (CNN), были применены в 25 из 66, или примерно в 38% анализируемых работ. Модели показали свою эффективность благодаря способности запоминать важную информацию и использовать ее для дальнейшего анализа, сохранять достоверность на протяжении всей аналитической цепочки, автоматически распознавать и извлекать значимые данные из набора временных рядов. Использование методов глубокого обучения в аналитике больших данных обеспечивает значимые и ценные результаты, в целом формирует взаимовыгодную связь между этими областями.

В работе «Co-LSTM: Сверточная модель LSTM для анализа настроений в больших социальных данных» [32] показано использование методов глубокого обучения для анализа настроений на основе данных из социальных сетей. Социальные сети, такие как Twitter и Facebook, генерируют огромные объемы данных, содержащих ценную информацию о настроениях пользователей. Эти данные создают как возможности, так и проблемы для журналиста из-за высокой скорости производства такого контента, его большого объема и неоднородности, сложностей семантики. Но журналистский материал может быть значительно обогащен благодаря анализу пользовательских комментариев и общественного мнения. Быстро провести сбор и обработку данных можно с помощью нейросети: модель обрабатывает огромные объемы данных из социальных сетей, выявляет ключевые темы, настроения и тенденции, что позволяет журналистам точнее понимать интересы, мнение и настроение аудитории, повышать релевантность и привлекательность контента. Так можно выявить темы, которые вызывают наибольший резонанс, вопросы, которые остаются без ответа и аспекты темы, требующие более глубокого освещения. Модель позволяет оперативнее реагировать на общественные настроения и даже предсказывать возможные изменения в восприятии событий или новостей. Глубокая нейронная сеть, разработанная под обозначенную задачу, обрабатывает большие объемы данных из социальных сетей без ущерба производительности. Co-LSTM (Convolutional Long Short-Term Memory – сверточная модель с долгой краткосрочной памятью) сочетает сверточную (CNN) и рекуррентную нейросеть (LSTM), учитывает контекст и долгосрочные зависимости между словами в предложениях. Для эксперимента были отобраны четыре набора данных: обзоры фильмов, авиакомпаний, президентских выборов в США и самоуправляемых автомобилей – их объединили для тренировки модели, а затем оценили раздельно. Все наборы данных были сбалансированы по количеству положительных, отрицательных и нейтральных отзывов – оно было примерно одинаковым. Производительность модели оценивалась с

помощью параметров точности, прецизионности, запоминания и F-меры. Как отмечалось ранее, схожие наборы данных используются не только в исследовании «Co-LSTM: Сверточная модель LSTM для анализа настроений в больших социальных данных» [32]. Авторы прибегают к анализу одинаковых массивов из-за их доступности и репрезентативности, разнообразия, а также универсальности и возможности сравнения изучаемой нейронной сети с другими для доказательств результатов работы. Изучаемая модель превзошла другие алгоритмы машинного обучения, например, точность для набора данных о фильмах составила 83,13 %, что выше, чем у алгоритма машинного обучения для задач классификации (83,11 %) и сверточной нейронной сети (82 %). Для набора данных об авиакомпаниях модель Co-LSTM также показала лучшие результаты, с точностью 94,96 %. Аналогичные результаты были получены для набора данных о самоуправляемых автомобилях, где была продемонстрирована высокая точность, f-мера (баланс между точностью и отзывом) и отзыв (доля из положительных отзывов, которые модель смогла правильно выявить). В случае набора данных о президентских выборах в США точность составила 90,45 %. Основным преимуществом предложенной модели является ее универсальность: она может быть обучена на данных из разных доменов без ущерба производительности, эффективно извлекает важные признаки из данных с учетом контекста отзывов, улучшая качество анализа, особенно для отзывов с отрицанием или сарказмом, которые достаточно сложны в классификации для моделей. Такая архитектура эффективно использоваться в анализе и прогнозировании настроений, что делает ее универсальным инструментом для работы с большими данными из социальных сетей. Аналитика настроений при помощи нейросети в целом значительно экономит время журналиста на оценку отношения аудитории к любому материалу.

В работе «Data2Vis: Автоматическая генерация визуализаций данных с использованием рекуррентных нейронных сетей Sequence to Sequence» [38] исследователи подробно описывают применение нейросети в визуализации

данных: в исследовании представлена модель на основе архитектуры кодер-декодер с механизмом внимания (фокуса на значимых частях данных), которая обучается на спецификациях визуализации Vega-Lite – инструмента для описания графиков, диаграм и других визуальных представлений данных на языке JSON. Такие спецификации упрощают описание графических элементов и представляют собой готовые шаблоны для визуализации. Кодер и декодер этой модели – двунаправленные рекуррентные нейронные сети (RNN) с ячейками долгой краткосрочной памяти (LSTM), которые показали лучшие результаты в экспериментах. Кодер считывает входные данные и переводит их во внутреннее представление (скрытое состояние) нейросети, декодер берет это скрытое состояние и шаг за шагом генерирует выходные данные на основе входных. В контексте работы над визуализацией кодер преобразует данные в форму, удобную для обработки, а уже декодер генерирует наглядные графики и диаграммы. Для проверки универсальности модели визуализации автоматически создавались из разных наборов данных, которые случайно отбирались из общей коллекции. Так авторы проиллюстрировали способность Data2Vis работать с любыми данными, что значительно упрощает процесс создания визуализаций и делает его доступным даже для начинающих пользователей. Для создания графиков модель выбирает наиболее вероятные и качественные результаты из множества возможных, исследуя при этом несколько комбинаций данных, выделяет значимые части данных, благодаря чему верно подбирает типы графиков и их настройки. Для технологии разработано веб-приложение, интегрирующее Data2Vis, оно позволяет импортировать данные, создавать и обновлять графические элементы. Важно отметить, что аналитика и визуализация данных значительно облегчается и автоматизируется благодаря использованию таких моделей. Инструменты, подобные Data2Vis, помогают быстро и эффективно представлять любые данные в наглядной форме.

Как отмечается в научной статье «NeuroDAVIS: нейросетевая модель для визуализации данных» [53] уменьшение размерности и визуализация

многомерных наборов данных остается сложной проблемой в изучаемой области на протяжении довольно длительного времени, но современные технологии с высокой пропускной способностью позволяют создавать большие наборы данных в самых разных формах и видах. Для их визуализации требуются мощные сложные алгоритмы, которые выявляют скрытые закономерности, при этом не нарушая взаимосвязей в данных. Таких методов сейчас очень мало, но одним из вариантов решения проблемы может стать предложенная новая глубокая нейросеть без учителя – NeuroDAVIS, предназначенная для визуализации данных. Модель извлекает важные характеристики из данных и визуализирует их, не теряя взаимосвязей даже между многомерными данными при их переводе в более простую для анализа форму. Эффективность модели продемонстрирована на разных наборах данных, включая специальные тестовые наборы, числовые, текстовые, графические и биологические данные. NeuroDAVIS превосходит классические методы обработки данных в сохранении формы, размера и структуры кластеров данных, визуализации данных с предварительно уменьшенной размерностью (это делает сама модель). Нейросеть универсальна для уменьшения размеров больших данных и их корректной визуализации, не имеет ограничений на количество измерений. Эта модель требует значительного объема памяти и времени на настройку параметров, но технология такого типа действительно может стать единым решением для проблемы визуализации наборов данных из самых разных категорий.

Систематизация исследований по практическому применению нейросетей в генерации дата-контента показывает, что мнения исследователей имеют как точки пересечения, так и расхождения. Большинство авторов, в том числе Акуличева А.Р. [2], Асташкин А.Г. [3], сходятся во мнении, что нейросетевые технологии играют важную роль в медиапроизводстве, но представляют собой только вспомогательный инструмент, который не способен полностью заменить профессионального журналиста. Основной ценностью моделей здесь выступают автоматизация трудозатратных

процессов и работы с большими данными. Журналисты в этом случае сосредотачиваются на анализе и интерпретации, обеспечивают яркую и творческую подачу информации. Достаточно полно в научной литературе освещаются технические аспекты применения моделей: Кириллов Д.С. и Молостов Д.Д. [17] подчеркивают важность качественной подготовки данных для эффективной работы нейросетей, отмечают, что некачественная подготовка снижает точность результатов. Однако качество результата напрямую зависит от человека уже на этапе ввода. Исследователи Чироме Х., Али У. и Абдулахамид Ш.М. [34] также исследуют техническую сторону вопроса и показывают разнообразие архитектур нейросетей в связке с необходимостью их корректного применения в зависимости от типа выполняемой задачи и соответствия ей класса модели.

Можно отметить небольшие отличия в оценке уровня автономности нейросетей: если отечественные исследователи, например, А. Р. Акуличева [2], считают, что эмоциональная и творческая составляющие контента доступны только человеку, то иностранные работы, например, исследующие модели Co-LSTM [32], Data2Vis [38], NeuroDAVIS [53], показывают рост и развитие возможностей нейросетей в анализе эмоций, визуализации и работе со сложными информационными структурами. Здесь модели выглядят более обособленно от человека. Однако модели и здесь не оказываются полностью самостоятельными элементами, более того, можно заметить, что они сосредоточены на одном типе задачи, достигая именно успеха в ней. Испанские исследователи Э. Фернандес, С. Моро и П. Кортес систематизируют шесть ключевых направлений применения нейросетей в медиа, показывая рост тренда на внедрение этих технологий в журналистскую практику. То же подтверждается и наблюдаемым ростом публикаций по теме. Таким образом, научный дискурс показывает двойственность взглядов исследователей: признается значительный вклад моделей в оптимизацию работы издательств, но большинство публикаций все же показывает необходимость сохранения человеческого участия в создании контента,

особенно в задачах, требующих творческого подхода, эмоциональной глубины и соблюдения этики.

2.2 Результаты качественного анализа генеративного контента в актуальном медиадискурсе

Конкретные примеры использования нейросетей, рассмотренные ранее, демонстрируют сегодняшнюю разностороннюю практику применения моделей в визуализации и классификации данных, анализе социальных сетей и других задачах. Серьезной проблемой остается то, что российские СМИ достаточно редко говорят открыто об использовании нейросетей и не описывают эти кейсы для науки. Если упоминания о применении нейронных сетей можно найти в исследованиях, то конкретные примеры с явными метками генерации остаются достаточно редкими, что ограничивает возможности для научного анализа. Работа моделей скрыта от публичного взгляда, а это подчеркивает важность и ценность анализа любых доступных материалов. Их важно изучать исходя из того, что несмотря на активное развитие и применение моделей, технология все еще сталкивается с рядом фундаментальных проблем, которые продолжают оказывать влияние на качество ее работы и требовать дальнейшего развития направления. Достижения последних версий нейронных сетей включают увеличение количества обрабатываемых токенов; расширение контекста, который нейросеть способна удерживать в работе; улучшение работы с источниками (сегодня нейросети не придумывают их, а действительно с ними работают); более точное извлечение данных и так далее. То есть, нейросети становятся все более удобными и доступными для человека с технической точки зрения, но сам принцип организации моделей подразумевает некоторые ограничения, которые могут нивелироваться с их развитием, но вряд ли в ближайшем будущем могут полностью сойти на нет. Стоит также учитывать, что быстрые темпы развития нейросетей и значительные достижения в этой области

особенно характерны для американских продуктов, в России же нейросетевые технологии развиваются значительно медленнее.

Эмпирическими материалами для исследования выступили, во-первых, специальный выпуск газеты «РБК» – «Выпуск газеты от 28 апреля 2023 № 061 (3731) (2804)» [10]. Часть текстовых и все визуальные материалы в номере были подготовлены с использованием нейросетей GigaChat первого поколения (вторая версия вышла только в марте 2025 года) и Kandinsky 2.1. Для выпуска были сгенерированы саммари и краткие справки, которые не редактировались журналистами. Анализ этих текстов не проводился исследователями ранее, он позволил выявить и изучить ограничения и проблемы, с которыми сталкивались и продолжают сталкиваться как журналисты, так и простые пользователи генеративных нейросетей. Эти проблемы все еще актуальны на сегодняшний день, несмотря на все технологические усовершенствования моделей, а ранняя версия генеративной модели особенно ярко демонстрирует такие ограничения. Во-вторых, эмпирической базой также выступили материалы, размещенные в разделе «Написано ИИ» на «РБК Тренды», эти тексты также ранее не анализировались исследователями. Для подтверждения результатов анализа эмпирических материалов также были систематизированы и классифицированы в список проблем результаты применения нейронной сети в работе «Краснодарских известий» [19].

Выпуск сетевой версии газеты «РБК» от 28 апреля 2023 года включает 7 материалов (один из них – слово главного редактора, этот материал не содержит генеративных текстов). Категории материалов выпуска – «Бизнес», «Политика и экономика», рубрики – «Рынок труда», «ИТ», «Инновации», «Транспорт», «Предпринимательство» и «Образование».

Первый текст, «Нейрообъятия для человечества» Марии Решетниковой, представляет собой качественный научно-популярный материал, логично структурированный по принципу от простого к сложному. Автор последовательно раскрывает понятие нейросетей, историю их развития,

принципы функционирования и области применения, завершает работу анализом подходов к созданию общего искусственного интеллекта. Материал отличается объективностью – прогнозы не выглядят излишне оптимистичными или пессимистичными. Специфические термины адаптированы и доступны для широкой аудитории, что позволяет тексту выполнять просветительскую функцию, развенчивать распространенные мифы, и помогать читателю ориентироваться в сложной технологической теме.

Генеративный текст «О чем эта статья» демонстрирует точное понимание структурно-содержательных элементов исходного материала. Нейросеть корректно определила центральные темы публикации и сохранила логическую последовательность изложения. Однако в саммари заметна тенденция к упрощению некоторых деталей и нюансов оригинального текста, что сигнализирует об обратной стороне способности нейронной сети быстро создавать обобщения – модель ограниченно сохраняет смысловую глубину исходного материала.

Второй генеративный текст «Что такое GigaChat» функционально выполняет роль самопрезентации технологии, участвующей в создании публикации. С точки зрения информационной ценности этот фрагмент сочетает фактические данные о технической архитектуре модели с элементами маркетинговой риторики, что особенно проявляется в цитате Германа Грефа о «прорывном» характере разработки и ее «уникальности» по сравнению с зарубежными аналогами. Такое сочетание технического описания и маркетинговых утверждений поднимает вопрос объективности самопрезентации нейросети и демонстрирует характерное ограничение генеративных моделей – они воспроизводят существующие утверждения из своих баз данных без независимой верификации или критического анализа информации.

Лингвистический анализ выявил в генеративных текстах определенные шаблоны. Заметны упрощенные обобщенные формулировки: «Этот текст о

том, что такое нейронная сеть», «становится все более вероятным, что мы станем свидетелями», «не только массовые пользователи... но и студенты и даже исследователи». Характерно тяготение к механическому перечислению без акцентов и конкретных примеров: «включая медицину, финансы, производство, транспорт и многое другое», «таких как классификация изображений, распознавание речи, анализ текста, прогнозирование временных рядов и многое другое». Это снижает выразительность текста и создает впечатление шаблонности изложения.

Фактологическая составляющая генеративных текстов в целом корректна, но выявлена проблема использования технических терминов без требуемых пояснений: «мультимодальная нейросеть», «нейросетевой ансамбль модели NeONKA», «метод supervised fine-tuning, reinforcement learning with human feedback». Для последних терминов было бы уместно использовать локализованные версии, уже принятые в русскоязычной практике: «методика адаптации предварительно обученных языковых моделей» и «обучение с подкреплением на основе отзывов».

С этической точки зрения анализируемый материал демонстрирует прозрачный подход к использованию искусственного интеллекта в журналистской работе. Генеративные фрагменты четко обозначены как продукт нейросети и выполняют вспомогательную роль. Интеграция генерированного текста в журналистский материал создает интересный зеркальный эффект – технология, о которой рассказывает статья, сама становится ее соавтором. Это иллюстрирует формирующуюся модель гибридного авторства в современной журналистике, где граница между человеческим и машинным вкладом становится более размытой, но при этом требует ответственного подхода, фактчекинга и редактирования.

В целом, сгенерированные тексты демонстрируют как возможности, так и ограничения современных моделей. Основные проблемы связаны с тяготением к шаблонному, обобщенному изложению и упрощению сложных концепций, что снижает информационную глубину материала. Несмотря на

эти ограничения, генеративные фрагменты здесь успешно выполняют свою вспомогательную функцию, дополняя основной журналистский материал с интересным эффектом самопрезентации технологии.

Второй текст «От дипфейков до пиратства» Людмилы Клейменовой представляет собой прогностический обзорный материал, исследующий перспективы развития российского интернет-сегмента. Его основная цель – определение ключевых трендов и возможностей для индустрии в условиях существующих экономических вызовов. Статья основана на исследовании Российской ассоциации электронных коммуникаций (РАЭК) и выстроена вокруг конкретных тематических направлений: технологии, экономика, государство и общество. Текст логичен и последователен – от обзора конкретных технологических инноваций (генеративный ИИ, дипфейки, биометрия, беспилотники) автор переходит к экономическим аспектам (концентрация бизнеса, экосистемная модель), государственным инициативам (импортозамещение, цифровизация управления) и завершает работу анализом общественных трендов (самозанятость, пиратский контент, маркетплейсы). Стилль изложения доступен для профессиональной аудитории и читателей, имеющих базовые представления в области. Материал сбалансирован: наряду с оптимистичными прогнозами о развитии технологий присутствует и описание ограничивающих факторов (экономический спад, санкции, нехватка оборудования).

В GigaChat для второго материала были сгенерированы три информационных блока: справка о РАЭК, общее саммари и историческая справка о Рунете. Генеративные тексты лишь частично передают основную идею и ключевые акценты оригинала. В саммари расставлены иные приоритеты – упущены важные элементы, включая общественные тренды (самозанятость, пиратство, маркетплейсы). Логическая последовательность нарушена – информация представлена фрагментарно, без органичных переходов между тематическими блоками. Наблюдается упрощение контента: вместо детального анализа представлены обобщенные тезисы без

необходимой конкретики («продолжат изменять ландшафт», «будут необходимы»).

Функционально генеративные тексты выполняют базовую справочную роль, предоставляя дополнительную контекстную информацию, однако их информационная ценность снижена из-за поверхностности и отсутствия глубины анализа. Наблюдается смещение в сторону технической информации в исторической справке, при практически полном отсутствии аналитической составляющей. В отличие от оригинального материала, где прослеживается аналитический взгляд на описываемые тенденции, нейросетевой текст скорее нейтрален и безоценочен. Тексты не предлагают новых ракурсов или неочевидных выводов, которые могли бы обогатить восприятие основного материала.

Тексты показывают лингвистические особенности, характерные для языка нейросетей: прослеживается шаблонная организация материала, особенно заметная в первом и третьем блоках, где используются списки (они часто встречаются в генеративном текстовом контенте – нейросети всегда отдают предпочтение явному структурированию контента). Информация представлена в виде справок с формальным делением на пункты, маркированными списками и хронологической последовательностью. Такая структурная стерильность контрастирует с более органичным и «живым» основным журналистским материалом. Генеративная лексика унифицирована и лишена стилистического разнообразия – часто используются клишированные формулировки («Этот текст является», «является одной из крупнейших», «Необходимость импортозамещения привела к тому, что», «ожидается, что»).

Особенно показательным признаком нейросетевого происхождения являются фактические неточности в третьем блоке (галлюцинации нейросети): неверные даты создания поисковых систем Nigma (работа над ней началась в 2010 году, а стала доступна для широкой аудитории она лишь в 2017 году, а в материале указан 1997 год) и Яндекс (годом создания можно считать 1997, в

материале – 2003), упоминание несуществующего закона «О национальной компьютеризированной сети России» от 1993 года. При обсуждении потенциально дискуссионных тем (импортозамещение, санкции) тексты демонстрируют нейтральность без явных искажений, однако эта нейтральность создает лишь видимость объективности при фактическом отсутствии критического анализа. Во втором блоке можно отметить информационную разреженность при кажущейся содержательности: текст наполнен обобщениями без конкретики, тогда как оригинальный материал насыщен фактами.

С этической точки зрения важно отметить, что, хотя в выпуске и указано генеративное происхождение текстов, проблемой остается их публикация с фактологическими ошибками. Читатель, который доверяет изданию и не осознает рисков генеративного контента, может воспринимать всю информацию как достоверную и не видеть необходимости в дополнительных проверках.

Сильной стороной генеративных текстов является их способность предоставить базовую справочную информацию в структурированном виде. Однако слабые стороны существенны: фактические неточности, шаблонность изложения, информационная поверхностность и отсутствие аналитической глубины. Для повышения качества необходимы: тщательный фактчекинг, стилистическое разнообразие, углубление аналитической составляющей и обязательное указание на генеративное происхождение текста. В настоящем виде материалы могут служить лишь вспомогательным информационным дополнением при условии редакторской проверки.

Третий материал – «Быстрее, бюджетнее, практичнее» Андрея Абрамова представляет собой обзорный текст о перспективах отечественного EdTech-рынка. Три дополнительных блока сгенерированы нейросетью GigaChat – историческая и информационная справка, а также саммари. Журналистский материал опирается на экспертные мнения и отраслевые данные, выстроен логически и последовательно – от ретроспективного анализа 2022 года к

прогнозам на 2023 год. Текст ориентирован как на профессиональную аудиторию (представителей EdTech-индустрии, инвесторов), так и на широкий круг читателей с базовыми знаниями в области, интересующихся образовательными технологиями и экономическими трендами.

Генеративные тексты с точки зрения структуры и содержания представляют собой три информационных блока: «С чего начинался сегмент», «Что такое EDTECH» и «О чем эта статья». Они демонстрируют характерные признаки машинного происхождения и имеют шаблонную организацию, типичную для справочных материалов: первый блок кратко перечисляет компании-пионеры и лидеров отрасли, второй дает определение термину и перечисляет основные компоненты с областями применения, а третий излагает содержание статьи с отсылками к экспертам. Журналистский текст в сравнении с генеративными выглядит более органично и менее шаблонно.

В целом справки и саммари предоставляют собой базовую справочную информацию без какой-либо интерпретации, критического анализа данных и приведенных мнений. Тексты решают именно задачу информирования, не предлагая никаких ценных выводов.

Генеративные тексты выстроены лексически нейтрально и унифицировано, в них, например, встречаются следующие шаблонные стандартные формулировки: «является термином, который относится», «может включать в себя различные технологии», «согласно прогнозам экспертов», обобщенные фразы без конкретики: «значительные средства», «широкий спектр ресурсов», «необходимых для достижения успеха». В текстах использованы преимущественно простые однотипные синтаксические конструкции. Особенно заметна тенденция к перечислению в первом и втором блоках («такие компании, как», «такие технологии, как»). В третьем блоке механически излагаются мнения экспертов – между ними нет логических связей, что создает эффект искусственного соединения цитат.

В части фактов присутствуют неточности и упрощения. В первом блоке перечислены лидеры области, где в один ряд с компаниями Google, Amazon,

Microsoft и Apple поставлен «Сбер», что может некорректно отражать специфику мирового рынка.

С этической точки зрения проблематичным является отсутствие редакторской адаптации генеративных текстов. Явных фактологических ошибок нет, авторство нейросети указано открыто, но присутствуют смысловые перекосы (например, в первом блоке «Сбер» поставлен в один ряд с мировыми технологическими гигантами, что может исказить представление о рынке).

Сильной стороной генерации текстов здесь можно отметить способность нейросети кратко предоставить базовую справочную информацию в структурированном виде. Слабые стороны такой работы существенны: фактические неточности, шаблонность изложения, поверхностность и отсутствие аналитической глубины. Для повышения качества необходимы: тщательный фактчекинг (в том числе для искажения информации, а не очевидных фактических или логических ошибок), стилистическое разнообразие, углубление аналитической составляющей. В настоящем виде материалы могут служить вспомогательным дополнением при условии редакторской правки.

Журналистский текст «Венчуры фокусируются на конкретном» Марии Решетниковой представляет собой обзорный материал о состоянии и перспективах рынка стартапов в России с акцентом на разных отраслях. Автор выстраивает материал логично и последовательно, переходит от общего анализа рынка к примерам конкретных компаний, а в конце к прогнозам развития. Особая ценность текста – база из экспертных мнений и статистических данных. Так материал сохраняет информативность для специалистов области и широкой аудитории и получается более достоверным.

Саммари по материалу, сгенерированное GigaChat («О чем эта статья»), в части структуры и содержания демонстрирует признаки машинного происхождения. В отличие от структурированного и логически выстроенного журналистского материала с конкретными примерами и экспертными

мнениями, генеративный текст представляет собой шаблонное обобщение с механическим перечислением компаний и сфер без необходимых акцентов. В оригинальном тексте проанализированы отраслевые направления (медтех, фитнех, зеленые технологии), однако в саммари вместо этого обобщенно перечисляются «программное обеспечение для бизнеса, электронная коммерция и логистика, промышленность, технологии, образование и инновации», упускаются ключевые области областей, выделенные автором.

С функциональной точки зрения генеративное саммари не справляется с задачей адекватного отражения основных идей и ключевых акцентов оригинала. В тексте отсутствуют упоминания большинства компаний, которые автор приводит в качестве примеров (за исключением QBoard, информация о котором при этом искажена). Такой подход к обобщению подменяет основной фокус статьи, заменяя его общими фразами, из-за чего теряется изначальная аналитическая ценность материала, которая заключается именно в отраслевой специфике венчурного рынка России и в конкретных примерах.

В саммари обнаруживаются шаблонный нейтральный язык («Согласно отчету компании», «которая предполагает») и обобщенные утверждения без конкретики: «Наиболее востребованным продуктом на российском рынке». Текст состоит из механических перечислений компаний и сфер стартапов без каких-либо акцентов, особенно в первом и третьем абзацах. Синтаксические конструкции однотипны и лишены стилистического разнообразия, которое присутствует в оригинальном журналистском материале.

Генеративное саммари содержит явные фактологические искажения. В основном материале упоминается QBoard – стартап в сфере квантовых вычислений, о котором нейросеть существенно искажает информацию, называя его «наиболее востребованным продуктом» и добавляя вымышленные данные о сотрудничестве бренда с IBM и Intel: «Наиболее востребованным продуктом на российском рынке стартапов оказалась платформа квантовых вычислений QBoard, которая получила инвестиции от

нескольких крупных корпораций, включая IBM и Intel». Информацию не удастся найти в основном материале или подтвердить в открытых источниках. Вместо указанных в оригинальном тексте партнеров Газпромбанка и лаборатории Genotek, искусственный интеллект заменил компании на западные.

С этической точки зрения проблематичной является публикация материала с очевидными фактологическими искажениями. Несмотря на указание генеративного происхождения текста, издание, размещая подобный контент, рискует распространять недостоверную информацию. Упоминание несуществующих связей QBoard с IBM и Intel может ввести читателей в заблуждение относительно действительного положения компании на рынке и повлиять на восприятие этого бизнеса заинтересованными сторонами.

Сгенерированное GigaChat саммари демонстрирует характерные проблемы генеративного контента: фактологические искажения (галлюцинации), шаблонность изложения, отсутствие аналитической глубины и неспособность передать ключевые акценты оригинального материала. В настоящем виде такой текст не только не выполняет свою функцию (качественное обобщение основных положений статьи), но и вводит читателя в заблуждение. Для повышения качества подобных материалов необходима тщательная редакторская проверка и доработка, особенно в части фактчекинга и сохранения ключевых смысловых акцентов оригинального текста.

Обзорный текст Татьяны Колосовой «Расставание с рутиной» содержательно исследует трансформацию рынка труда под влиянием искусственного интеллекта. Автор раскрывает тему через факты и историю развития искусственного интеллекта, а также дает прогнозы для рынка профессий. Текст опирается на экспертное мнение специалистов из различных областей (от Александра Ведяхина из Сбербанка до исследователей Оксфордского университета). Профессии рассматриваются взвешенно — приведены примеры областей, которые находятся под угрозой автоматизации, и направления, где человеческий труд ожидаемо сохранит свою ценность.

Статья характеризуется журналистской конкретикой, четкими временными ориентирами прогнозов на 10-15 лет и сбалансированной позицией, избегающей избыточной тревожности или необоснованного оптимизма в прогнозах.

Саммари, сгенерированное GigaChat для этого материала, значительно отличается от оригинала по структуре и содержанию. Если журналистский текст имеет четкую фокусировку на трансформации рынка труда под влиянием ИИ, то в саммари происходит размывание этого фокуса и смещение к обобщенным рассуждениям о том, как ИИ «изменит многие отрасли». Нейросеть вводит темы, отсутствующие в исходном материале – вопросы конфиденциальности, «угрозу для человечества» (основной текст фокусируется на конкретном влиянии ИИ на работу человека), разработку систем, «лучше понимающих потребности людей». Конкретика оригинального текста утрачивается – имена экспертов не упоминаются, исчезают временные прогнозы и перечень уязвимых профессий.

Генеративное саммари не справляется с основной функцией – адекватным отражением ключевых идей и смысловых акцентов оригинала. Основная мысль журналистского текста, которая должна была быть передана в саммари, – оценка влияния искусственного интеллекта на рынок труда сейчас и в будущем, с опорой на мнение экспертов области. Модель предлагает набор общих и типовых рассуждений, очень отдаленно отражающих действительное содержание оригинального текста. В результате саммари может служить обобщением практически любой статьи об искусственном интеллекте, так как в нем отсутствуют конкретные черты исходного материала.

Модель наполняет текст обезличенными общими конструкциями («имеет потенциал изменить многие отрасли и повлиять на жизнь людей во многих аспектах», «некоторые утверждают», «другие считают», «были достигнуты значительные успехи»). Язык саммари демонстрирует типичные для генеративных текстов признаки: обобщенность, шаблонность, отсутствие

конкретики. В отличие от журналистского текста с его разнообразием стилистических приемов и живым языком, саммари выглядит безлико и стандартно. Синтаксические конструкции однотипны и предсказуемы, лишены выразительности оригинала.

В саммари отсутствуют явные фактические ошибки, однако происходит существенное искажение акцентов исходного материала. Введение тем, не затронутых в оригинале (вопросы конфиденциальности, «угроза для человечества»), можно рассматривать как фактологическую неточность – приписывание тексту содержания, которого в нем нет. Отсутствие упоминаний конкретных экспертов, временных ориентиров прогнозов и перечня профессий, подверженных автоматизации, делает саммари фактологически размытым.

С этической точки зрения проблематичной является подмена основного содержания и фокуса оригинального текста. Несмотря на указание генеративного происхождения саммари, читатель получает искаженное представление о содержании и направленности журналистского материала. Это может рассматриваться как форма дезинформации, даже если она непреднамеренна. Публикация такого саммари без редакторской проверки и корректировки может подрывать доверие к изданию.

Анализ саммари к статье «Расставание с рутиной» выявляет характерную проблему современных генеративных систем – тенденцию к обобщению в ущерб точному отражению уникального содержания контента. Текст утрачивает специфику оригинала, подменяя конкретику общими рассуждениями и вводя темы, отсутствующие в исходном материале. Это демонстрирует ограничения использования нейросетей (на текущем этапе их развития) для создания саммари без редакторского вмешательства.

Для повышения качества генеративных текстов необходимо совершенствовать работу моделей для более точного отражения специфики оригинальных текстов и обязательно проводить редакторскую проверку и

доработку. В текущем виде генеративные саммари могут служить лишь вспомогательным материалом с низкой ценностью.

Текст Игоря Епанешникова «Карма электрокара» включает в себя анализ статистики, на который накладываются особенности отрасли и геополитический контекст. Автор выстраивает логически стройное повествование о трансформации российского рынка электромобилей, начиная с обозначения исходной точки (утвержденная в 2021 году государственная концепция) и последовательно показывает, как события 2022 года скорректировали планы. Работа сочетает макроэкономическую статистику, дает примеры конкретных инвестиционных проектов и детальный анализ инфраструктурных барьеров. Текст сбалансирован – в нем описаны как ключевые факторы торможения развития (санкции, нехватка комплектующих, слабая инфраструктура, отсутствие профессионального сервиса), так и стимулирующие элементы (государственные субсидии, пониженные таможенные пошлины, растущий потребительский интерес). Материал завершается прогнозом временных горизонтов электромобильного перехода для Европы и России, что позволяет читателю соотнести ситуацию в стране с глобальным трендом. Материал получился достаточно подробным и всесторонним.

Саммари, сгенерированное для этого текста, в части структуры и содержания, демонстрирует умеренную аналитическую точность в передаче содержания оригинала: алгоритм сохраняет хронологическую и логическую структуру статьи, правильно приводит ключевую информацию о концепции 2021 года, коррективах после 2022 года, основных проектах (описаны «Рэнера» и «Автотор», КамАЗ и «Москвич»). В генеративном тексте воспроизводятся точные числовые данные по инвестициям (5 млрд рублей) и временным горизонтам электромобильного перехода для Европы и России. Генеративная справка «Электромобили XXI века» демонстрирует полное игнорирование российского контекста исходной статьи в пользу шаблонного изложения истории глобальной электромобильной отрасли. Нейросеть

воспроизводит фактически корректную, но оторванную от контекста исходной статьи информацию об истории компании Tesla и ее влиянии на мировой рынок электромобилей. Это создает заметный дисбаланс между основным журналистским материалом и генеративным текстом.

При функциональном анализе видно, что саммари сохраняет хронологическую и логическую структуру оригинальной статьи, что помогает передать основные тезисы. Идея текста в целом сохраняется, хотя аналитическая глубина заметно снижается. Справка, призванная дополнить тему, фактически размывает аналитическую специфику рассматриваемого материала и подменяет значимую информацию о российском рынке универсальными сведениями о глобальных игроках и трендах. При этом она все же остается в рамках анализируемой темы, из-за чего эта особенность не так заметна при беглом прочтении. Такое смещение фокуса снова отражает общую проблему современных генеративных моделей – склонность к созданию обобщенного, «усредненного» контента на основе доминирующих в обучающем датасете нарративов.

С лингвистической точки зрения саммари демонстрирует тенденцию к генерализации. Это проявляется в гиперболизации информации о международном сотрудничестве. В оригинале: «Закономерным сейчас кажется сотрудничество с Китаем», «как ключевая задача государства в этой области отмечено «работать в связке с частным бизнесом». В саммари мысль трансформируется в «работа «правительств многих стран» над привлекательностью отрасли для инвесторов», что искажает исходный смысл из-за формулировки.

Эта проблема еще сильнее проявляется в дополнительном тексте, где алгоритм воспроизводит текст по аналогии с информацией из обучающих сетей, но фактически игнорирует особенности анализируемого материала. Это может быть не очевидно при поверхностном рассмотрении, так как модель все же остается в рамках общей темы электромобилей.

В фактологическом плане наблюдается двойственная картина. С одной стороны, саммари сохраняет ключевые фактологические элементы исходного текста, включая информацию о государственной концепции, основных проектах в области электромобилей и временных горизонтах перехода. С другой стороны, значительный объем фактологического материала упущен: в саммари не достаёт данных о количестве электромобилей, статистике по авторынку в России, информации о государственных субсидиях. Показательным дефектом саммари является гиперболизация информации о международном сотрудничестве, что искажает исходный смысл статьи. Еще более заметная проблема наблюдается в справке «Электромобили XXI века», который содержит фактически корректную, но оторванную от контекста исходной статьи информацию об истории компании Tesla и ее влиянии на мировой рынок электромобилей, что полностью игнорирует российский контекст основного материала.

С этической точки зрения, саммари не содержит явных грубых нарушений, но гиперболизация информации о международном сотрудничестве может вводить читателя в заблуждение. Более серьезную этическую проблему представляет именно справка, которая создает значительный дисбаланс между основным журналистским материалом и генеративным дополнением. Такая подмена специфического российского контекста универсальными глобальными данными может существенно исказить восприятие читателем основного материала.

Анализ генеративных текстов выявляет две характерные проблемы современных нейросетей. Во-первых – тенденцию к упущению значительного объема фактологического материала и генерализации данных в саммари, что приводит к искажению некоторых аспектов исходного текста. Во-вторых – игнорирование специфического контекста в пользу шаблонного изложения. Оба случая отражают проблему склонности нейросетей к обобщению и «дописыванию» контента на основе обучающего датасета, а не точному воспроизведению специфики материала. Этот дисбаланс между основным

журналистским материалом и генеративными текстами подчеркивает ограничения современных нейросетей и необходимость более тщательной настройки генеративных моделей на точное отражение фактологического содержания и контекстуальных особенностей оригинальных текстов.

На основе анализа генеративных текстов можно выделить конкретные характерные особенности, говорящие об искусственном происхождении текстового контента:

Во-первых, структурные и языковые особенности:

- шаблонность текста с излишней формализацией, частые механические перечисления;
- клиширование («некоторые утверждают», «согласно мнению ученых» и так далее);
- упрощение синтаксических конструкций, преобладание простых предложений;
- разреженность фактов и ценной информации – текст кажется содержательным, но в нем много общих фраз без конкретики.

Во-вторых, искажение фактов и смещение смысловых акцентов:

- неточности в фактах и галлюцинации, вымышленные несуществующие данные, ошибочные даты;
- гиперболизация, преувеличение значимости отдельных фактов, изменение масштабов явлений, из-за чего искажается исходный смысл;
- смещение контекста, например, переход с анализа локального рынка на глобальный;
- игнорирование ключевой информации, пропуск важных данных, статистики.

В-третьих, аналитические ошибки и неточности:

- замена глубокого анализа обобщениями, потеря аналитической ценности оригинального материала;

- тяготение к доминирующим нарративам из обучающего сета в ущерб специфике конкретного материала;
- потеря смысловых акцентов оригинального текста, потеря его уникальности;
- подмена экспертного мнения универсальными суждениями без подкрепления информации источниками.

Наиболее показательные проблемы включают:

- дисбаланс между оригинальным и генеративным контентом. Сгенерированный контент будет тематически связан с оригиналом, но вероятнее всего, будет игнорировать его специфический контекст;
- универсализация уникального журналистского текста в типовой и обезличенный;
- «дописывание» и «додумывание» контента, то есть генерация информации, отсутствующей в исходном материале, на основе обобщенных паттернов из обучающих данных;
- избирательное воспроизведение фактов. Одни фактические элементы, например, даты и статистическая информация, будут воспроизведены верно, а другие будут упущены или искажены.

Некоторые из выделенных особенностей текста подтверждаются и другими исследователями, так, Соколова М.Е. отмечает, что: «журналисты приходят к выводу, что использование нейросетей без контроля и редактуры со стороны профессионала непродуктивно. Например, согласно данным студенческого исследования по направлению «Журналистика» в Тамбовском государственном университете имени Г.Р. Державина в апреле 2023 г., хотя нейросети и способны вполне качественно и профессионально создавать или дописывать новости, авторские тексты, следуя авторскому замыслу и верно излагая, и описывая фактологию, – получившиеся в результате тексты лишены авторской точки зрения, экспрессивности, эмоционально окрашенной лексики. Публикация таких текстов оправдана в условиях нехватки времени и человеческих ресурсов в редакционно-издательском процессе. В некоторых

же текстах были искажены факты, привнесен домysel. Примерно такие же результаты получились и в случаях использования нейросетей, ориентированных на написание текстов по ключевым словам, где был допущен наибольший процент искажения фактов и неправдоподобности. Исследователями был сделан вывод о том, что бесконтрольное использование нейросетей на основе искусственного интеллекта в деятельности журналиста и редакции в целом недопустимо, и необходим контроль со стороны сотрудника редакции» [27].

Следующим этапом исследования стал анализ генеративных текстов за 2024 год. После изучения ранних версий моделей, представленных на российском рынке в 2023 году, их было бы логично сравнить с более совершенными моделями, которые интегрировались в медиаиндустрию в течение 2024 года. Это позволяет отследить актуальные изменения в качестве работы моделей и дополнительно проверить гипотезу о необходимости включения журналиста в процесс производства контента нейросетью, особенно в контексте развития технологии. В качестве эмпирического материала выступили 6 публикаций, но уже из специализированного раздела «Написано ИИ» информационного портала РБК, созданных виртуальным сотрудником «Иваном Кремневым» [21]. Особенность нейросети, используемой РБК – ее гибридная архитектура: это совокупность моделей, оптимизированных под решение своих специфических задач. Для хронологической сопоставимости отбирались публикации весны 2024 года, что позволило провести корректное сравнение материалов год к году и проследить эволюцию генеративных технологий.

Первый материал «Google I/O 2024: кратко обо всех анонсах конференции» размещен в рубрике «Инновации». Он представляет собой информационный текст в формате новостной сводки. В содержательном плане материал показывает отличную способность нейросети систематизировать и структурировать информацию. Текст разделен на две логических части «Что происходит» и «Что это значит», что в целом характерно для классической

структуры новостных материалов, где читателю сначала предлагаются факты, а после дается их интерпретация. Нейросеть успешно определила ключевые анонсы конференции. Они удобно сгруппированы в тематические блоки, но содержание материала остается преимущественно поверхностным и перечислительным, ему не хватает контекста анонсов, например, информации о том, как новые решения соотносятся с предыдущими продуктами Google или предложениями конкурентов, что снижает информационную ценность для читателя. Следует учитывать, что текст все же представлен как «краткий анонс», что делает проблему не такой критичной.

С лингвистической точки зрения текст демонстрирует стилистическую однородность и шаблонность. Особенно заметна одна синтаксическая модель, которая повторяется в описании большинства технологических новинок: «[Продукт или функция] теперь или будет [действие]», например: «Google Lens теперь позволяет», «Функция Ask Photos позволит», «Модель Gemini 1.5 Pro будет интегрирована», «Поиск Google также получит обновления». Монотонное сухое изложение создает механический эффект и требует усилий читателя в восприятии информации. В тексте практически отсутствуют стилистические приемы качественной журналистики – метафоры, аналогии, образные сравнения, которые могли бы сделать описанные новинки более понятными и доступными для любой аудитории. Лексическое разнообразие ограничено, что особенно заметно в аналитической части – выводы формулируются по шаблону типовых фраз без оригинальных речевых конструкций.

Фактологическая достоверность удовлетворительна, текст не содержит очевидных ошибок и неточностей, соответствует официальным анонсам Google. Материалу не хватает ссылок на источники информации, кроме общего упоминания видеоматериалов от Google, которые размещены на платформе YouTube и недоступны на момент анализа. Некоторые технические термины (например, SynthID) вводятся без пояснений, что может затруднить понимание для неподготовленного читателя. Отсутствуют важные

контекстуальные детали (даты выпуска анонсированных продуктов, информация по ценовой политике, сравнение с предыдущими версиями технологий), что опять же может объясняться и краткостью формы материала.

С точки зрения функциональности текст успешно выполняет базовую задачу – дает читателю обзор ключевых анонсов. Можно отметить, что в разделе «Что это значит», в котором должно быть аналитическое осмысление, содержит скорее поверхностные наблюдения и очевидные выводы. Например, утверждение «Новые функции, такие как поиск по видео и Ask Photos, упрощают взаимодействие с информацией и делают поиск более интуитивным» не раскрывает потенциал инноваций для пользователя или рынка, а скорее тезисно обозначает нововведение. В тексте нет критического анализа технологий, не упоминаются оценки их потенциальных недостатков или ограничений. Нейросеть сфокусировалась на технических характеристиках и возможностях. Качественный классический журналистский текст, вероятнее всего, предоставил бы «человеческую» информацию: как новые решения повлияют на повседневную жизнь пользователей, какие этические или социальные вопросы могут возникнуть в связи с внедрением представленных технологий.

В целом текст показывает значительный прогресс в создании структурированных и фактологически корректных информационных материалов. В сравнении с ранними примерами однозначно следует отметить улучшение в построении логичной структуры текста и соблюдения структурных требований к жанру новостной заметки. Материал по-прежнему обнаруживает ограничения в аналитической глубине, стилистическом разнообразии и оригинальности в описании событий. Информация агрегирована корректно, но не показывает признаков критического мышления или творческого подхода к интерпретации фактов, что остается прерогативой журналистов.

Следующий генеративный текст «Apple добавила новые функции для управления iPhone и iPad глазами» также размещен в разделе «Инновации».

Он представляет собой краткую новостную заметку об анонсе новых функций универсального доступа от компании Apple. В структурном отношении текст снова следует стандартной формуле новостной заметки с разделением на факты («Что происходит») и интерпретацию («Что это значит»). Организация материала говорит о способности нейронных сетей соблюдать базовые принципы построения журналистских произведений. Объем материала при этом крайне ограничен для темы, которая затрагивает технологические инновации, социальные аспекты доступности и стратегические направления развития компании Apple. Фактически, в первой части приводится тезисное перечисление трех новых функций без какого-либо углубления в технические детали или практические сценарии использования. Важно отметить, что короткий формат текста может быть обусловлен особенностями промпта издательства, ограничивающего формат вывода.

С точки зрения информационной насыщенности текст содержит минимум фактов для новостного сообщения: анонс функций, приблизительные сроки их появления (в iOS 18 и iPadOS 18 «позже в этом году») и краткое описание. Описания функций поверхностны и не раскрывают важные аспекты инноваций, например, информации о точности работы технологии, поддерживаемых моделях устройств, ограничениях использования – для получения очевидно актуальной для статьи информации придется обращаться к дополнительным источникам. Упоминание о том, что «настройка и калибровка происходит с помощью фронтальной камеры за несколько секунд» и что «все данные остаются на устройстве» выглядят как воспроизведение маркетинговых формулировок, которые, явно, не подкреплены критическим анализом и не проверены.

Лингвистический анализ текста обнаруживает несколько проблемных аспектов.

В материале присутствуют стилистические и грамматические неточности, например: «Apple объявила о новых функциях универсального доступа, которые появятся в iOS 18 и iPadOS 18 которые выйдут позже в этом

году, до мероприятия WWDC» (отсутствует запятая после упоминания операционных систем, есть лексический повтор) – само предложение слишком длинное и сложное для восприятия, особенно в части «до мероприятия WWDC»; «это в свою очередь в некотором роде формирует тренд на инклюзивность ны рынке носимых устройств» (опечатка «ны» вместо «на»); «Музыкальная тактильность» другая новая функция» (упущено тире). Встречается некорректная формулировка «навигации по устройствам», правильнее было бы сказать «навигация в устройствах». Такие ошибки могут быть свидетельством галлюцинаций нейросети или особенностей использования модели, обученной на англоязычных дата-сетах. Последнее объяснило бы сложное предложение о функциях универсального доступа, которое на английском языке бы звучало лаконичнее и не требовало бы запятой. Если этот тезис – только предположение, то однозначно можно отметить недостаточность редактирования и проверки генеративного текста.

Вторая часть материала («Что это значит») вместо содержательного анализа влияния новых функций на пользователей с ограниченными возможностями дает общие поверхностные фразы о «стремлении компании использовать достижения в области искусственного интеллекта», утверждение, что «будущее технологий будет тесно связано с ИИ». Выводы не находят связей с конкретными фактами из первой части и могут быть применены к любому технологическому анонсу. Особенно показательна неуверенная формулировка с упущенными знаками препинания «в некотором роде формирует тренд на инклюзивность ны рынке носимых устройств» и фактической неточностью (iPhone и iPad, которые рассматривались в первой части материала, – не являются носимыми устройствами).

В контексте фактологической достоверности материал не содержит явных фальсификаций, но информация фрагментарна, два блока текста слабо связаны между собой.

С точки зрения публицистической ценности текст не предлагает уникальной перспективы или экспертной оценки, в нем нет комментариев

специалистов или представителей сообществ людей с ОВЗ, аналитиков рынка. Материал не содержит сравнений с аналогичными решениями конкурентов или предыдущими инициативами Apple, что лишает материал необходимого контекста и глубины.

Особого внимания заслуживает заключительный абзац, где упоминается Всемирный день осведомленности о равном доступе (16 мая) – в корректном переводе на русский язык «Всемирный день осведомленности о доступности». В положительном ключе это демонстрирует способность нейросети связывать новостные события с более широким социальным контекстом, но эта информация выглядит изолированно от текста.

В целом генеративный текст демонстрирует связность повествования и соответствие минимальным требованиям журналистского формата, но имеет свои недостатки в части причинно-следственных связей, глубины анализа, критического осмысления информации и внимания к деталям. Материал успешно информирует об анонсе, но не обозначает его значения или последствия для человека, скорее напоминает расширенный пресс-релиз, чем полноценный связный журналистский текст.

Анализ генеративного текста «Тепловое воздействие на пожилых людей в мире удвоится к 2050 году», сгенерированного в жанре новостной заметки на базе научного исследования из журнала Nature Communications, показывает в структурном отношении соответствие вероятно заданному формату с разделением на две основные части – «Что происходит» и «Что это значит».

Информационная составляющая фокусируется на ключевых выводах исследования: прогнозируется увеличение воздействия высоких температур на пожилых людей к 2050 году, подробно указываются региональные отличия с акцентом на Азию и Африку. Текст содержит конкретику в отношении количественных данных, что усиливает фактологическую базу. Такая детализация показывает, что нейросети способны выделять и систематизировать ключевые показатели исходных текстов.

Аналитическая часть материала («Что это значит») снова демонстрирует поверхностный подход к интерпретации. Вместо анализа социальных, экономических и политических последствий явления текст ограничивается очевидными выводами о влиянии жары на здоровье пожилых людей и поверхностными рекомендациями по смягчению последствий. Следует отметить, что раздел включает определение гипертермии, что является положительным элементом с точки зрения доступности дополнительной полезной информации, но при этом не содержит конкретики в отношении практических шагов или, например, политических решений.

С точки зрения стилистики и языкового оформления, текст написан преимущественно сухим, информационным языком, что соответствует жанру новостной заметки. Ему недостает выразительности и уникальности для привлечения внимания читателя к серьезности проблемы. Отсутствуют яркие примеры или истории конкретных людей, которые могли бы проиллюстрировать абстрактные статистические данные и сделать материал более запоминающимся и убедительным. В тексте есть проблема с построением предложений: «и адаптировать медицинские системы и информационные политики в области общественного здравоохранения» – часть сложной для восприятия длинной конструкции с механическим перечислением содержит неточную формулировку «информационные политики», которая выглядит неестественно для русского языка; в предложении также есть проблемы с пунктуацией. Можно отметить некорректную формулировку в части «экстремальные климатические события», где вернее было бы сказать о погодных явлениях. В одном из тезисов можно заметить упущенный союз «Тем не менее ростом температур столкнутся все регионы»

Существенным ограничением текста выступает отсутствие критического осмысления или обсуждения возможных альтернативных точек зрения. Генеративный текст принимает выводы исследования как данность и сухо передает в виде кратких тезисов, не рассматривая вопросы надежности

прогнозов, погрешностей, ограничений использованных моделей. В качественной журналистике такие аспекты традиционно освещаются для более полного представления о достоверности и контексте научного исследования.

В тексте не представлен более широкий контекст проблемы в виде сопоставления с другими исследованиями в этой теме или связи с актуальными изменениями в вопросах климата. Такие связи могли бы усилить значимость материала и продемонстрировать место приведенной работы в общественном дискурсе. Отмеченная проблема не критична, так как материал все же представлен в виде краткой сводки.

В целом, анализируемый генеративный текст показывает способность нейросетей агрегировать и структурировать информацию в том числе научного характера, соблюдать базовые требования к новостным материалам. Текст успешно выполняет информационную функцию и сообщает читателю о существовании проблемы и ее основных параметрах, но не обеспечивает нужного уровня контекста (при том, что приводит дополнительное определение для лучшего понимания вопроса), интерпретации и критического анализа несмотря на то, что тема научного исследования имеет серьезный социальный вес. Нейросетевые технологии достаточно развиты для создания базового информационного контента, но все еще ограничены в воспроизведении более сложных аспектов журналистской работы.

Следующий анализируемый генеративный текст, «Джеймс Уэбб» подтвердил наличие атмосферы на каменистой экзопланете», посвящен научному открытию в области астрономии – обнаружению атмосферы на экзопланете 55 Cancri e с помощью телескопа Джеймса Уэбба. Структурно текст организован в соответствии с форматом, делящим материал на факты («Что происходит») и интерпретацию («Что это значит»).

В информационной части текста нейросеть демонстрирует умение структурировать и передавать читателю научные данные в доступной форме: материал содержит ключевую информацию о самом открытии – обнаружении

атмосферы на экзопланете, характеристиках объекта исследования (размере, массе, температурных условиях, удаленности от Земли) и источнике данных. В материале есть и исторический факт – упоминается, что планета была открыта в 2004 году и с тех пор активно изучается.

В части «Что это значит» нейросеть пытается дополнить открытие контекстом и объяснить его значимость – здесь можно отследить способность модели связывать конкретные события с более широким научным контекстом и выделять эти элементы из текста: потенциальная ценность исследования упоминается в связи с пониманием ранней истории Земли и поиском внеземной жизни. Интерпретации при этом все еще даются на достаточно общем уровне, без конкретных примеров или мнения экспертов, что могло бы углубить понимание открытия.

Лингвистический анализ текста показывает, что материал написан доступным языком, с умеренным использованием специализированной терминологии («супер-Земля», «экзопланета», «инфракрасные спектрометры»), которую невозможно убрать из текста – она необходима для адекватного освещения темы. Лексически и стилистически текст соответствует жанру научно-популярной журналистики, он не излишне сложный, но и не слишком упрощен. В материале можно отметить некорректные формулировки, например «атмосфера 55 Cancri e богата углекислым газом или угарным газом» – это разные вещества, и здесь нужно пояснение, каким из веществ богата атмосфера, или указание на то, что у исследователей нет точных данных по ее составу, или «55 Cancri e имеет значительную атмосферу».

Текст не имеет глубины или журналистской оригинальности, предлагает линейное и одномерное изложение информации без критического осмысления или приведения альтернативных точек зрения. Отсутствуют прямые цитаты исследователей или комментарии экспертов, которые могли бы добавить материалу авторитетности и глубины. Это ограничение может быть как

характерным для генеративных текстов, так и результатом конкретного требования промпта к формату вывода.

В отношении достоверности и объективности текст типичен: представляет информацию нейтрально, без явного авторского мнения или оценочных суждений. Такая нейтральность может рассматриваться как преимущество, непредвзятость, так и недостаток, или отсутствие критического осмысления. Материал не пытается оценить качество или надежность исследования, выводы передаются как данность без сравнения со схожими результатами из предметной области.

В целом, анализируемый генеративный текст представляет собой компетентную, но лаконично сжатую реализацию научно-популярного журналистского материала. Он успешно выполняет базовую функцию информирования, сообщает о научном открытии и предоставляет основной контекст для его понимания.

Следующий текст «Смартфоны – самый популярный способ выхода в сеть, а доля ПК снижается» представляет собой новостной материал, построенный на статистических данных НИУ ВШЭ о цифровизации российского общества. В структурном отношении текст классически подразделяется на фактологическую («Что происходит») и интерпретационную («Что это значит») части. Первая, значительно более объемная, часть содержит основную статистику и динамику, а вторая, более лаконичная, – предлагает обобщенные выводы по представленным тенденциям.

Особенностью материала является интеграция визуальных элементов – инфографик из исследования НИУ ВШЭ, очевидно, проведенная человеком, что существенно повышает информативность текста и облегчает восприятие статистических данных.

С точки зрения содержания текст демонстрирует компетентную работу со статистикой, приводит конкретные цифры для иллюстрации динамики цифровизации: процента домохозяйств с доступом в интернет,

распространенности устройств для выхода в сеть, частоты использования интернета, отношения пользователей к цифровым технологиям и других показателей. Для большинства из них приведены сравнительные данные за несколько лет, что позволяет отслеживать динамику процессов. Последнее приведено в оригинальном исследовании и корректно выделено из всего объема информации нейросетью.

Стилистически текст соответствует жанру информационной заметки – он написан нейтральным, деловым языком без эмоциональной окраски или оценочных суждений. Структура предложений простая, что облегчает восприятие информации, изобилующей цифровыми данными. Текст построен логично: от общих показателей проникновения интернета, к детализации, субъективной оценке, проблемам, и уровню цифровой грамотности.

В интерпретационной части («Что это значит») обнаруживаются общие выводы, которые могли бы относиться практически к любому исследованию цифровизации: «страны с высоким уровнем цифровизации показывают большее экономическое развитие» и «важно развивать механизмы защиты и повышать уровень осведомленности населения». Такие выводы не опираются на конкретные показатели из первой части и не показывают глубокого анализа или учета причинно-следственных связей между данными. Наблюдаются некорректные формулировки, например «рассказал о росте проникновения интернета и повышении цифровых угроз» – необходимы уточнения, куда проникает интернет – речь идет про гаджеты или города, села.

Отсутствует контекст изложения для представленных тенденций – не показано, как они соотносятся с социально-экономической ситуацией в России или с аналогичными показателями других стран (кроме общего утверждения о соответствии «глобальным трендам» без конкретизации последних). В журналистском материале можно было бы привести информацию о влиянии снижения использования настольных ПК и роста популярности смартфонов на характер производства и потребления контента.

В отношении достоверности и точности представленных данных, текст опирается на официальное исследование авторитетного института и корректно передает основные статистические показатели, что обеспечивает фактологическую надежность материала. В тексте не удается найти очевидных логических противоречий или неправдоподобных утверждений.

По сравнению с ранее проанализированными генеративными текстами РБК, данный материал демонстрирует особенно высокую компетентность в работе со статистикой. Здесь данные представлены ясно, точно и структурированно.

В целом, генеративный текст представляет собой качественный информационный материал, компетентно представляющий основные тенденции цифровизации российского общества. Он успешно информирует аудиторию о результатах исследования НИУ ВШЭ. Для повышения качества текста следовало бы увеличить глубину аналитической составляющей и предоставить экспертные комментарии. Эти аспекты все еще остаются сложными для воспроизведения с помощью современных нейросетей. В жанре новостной заметки, ориентированной на представление статистических данных, эти ограничения мало заметны, что делает применение генеративных моделей в этом кейсе достаточно эффективным.

Генеративный текст «Neuralink ищет второго добровольца для имплантации после неудачи с первым» структурирован согласно стандартному формату публикаций РБК с разделением на факты («Что происходит») и их интерпретацию («Что это значит»).

Заголовок материала информативен и указывает на ключевую информацию текста – проблему с первым пациентом и текущие действия компании. Подзаголовок конкретизирует характер проблемы («разрыв электродов в мозгу»), что помогает читателю сразу понять суть ситуации. Такая структура заголовочного комплекса демонстрирует компетентность генеративной системы и в создании информативных, цепляющих заголовков.

В фактологической части материал сжато и последовательно излагает основные события. Текст конкретизирует проблему, ссылаясь на источник информации (Wall Street Journal), упоминает риски, формируя этический контекст.

Текст содержит важную персонализированную информацию: упоминается имя первого пациента (Ноланд Арбо), причина его паралича (несчастный случай в 2016 году), а также его личный опыт использования имплантата и последующее разочарование. Такой подход к подаче информации через личную историю делает материал более интересным и понятным для читателя. Такой прием эмоционального усиления характерен для классической журналистики.

В интерпретационной части текст предлагает более общую информацию о технологии нейрокомпьютерных интерфейсов, их потенциальном применении и роли Neuralink в изучаемой области. Эта часть одновременно имеет и научно-просветительский характер, и помогает дать дополнительный, более обширный контекст событиям.

Стилистически текст сбалансирован – сочетает фактическую точность и доступность, в нем нет излишнего упрощения или усложнения терминологией. Термин «нейрокомпьютерный интерфейс» объяснен простым языком, приведены примеры, что особенно ценно для читателя. Наблюдается упрощение формулировок, клишированное изложение, например, в блоке «Но после того, как гаджет начал терять функциональность из-за разорванных электродов, Арбо выразил свое разочарование. Neuralink внесла изменения в алгоритм, чтобы восстановить часть функциональности для Арбо, но его долгосрочные перспективы с имплантатом остаются неясными», где снова повторяется схема [Кто или что] теперь или будет [действие], из-за чего текст становится механическим: «гаджет начал терять», «Арбо выразил разочарование», «Neuralink внесла изменения». Встречаются и некорректные связки, например, «Они сделали значительный прогресс» или «привести к созданию нового уровня взаимодействия между человеком и машиной, что

будет иметь глубокие последствия»: новый уровень скорее не создается, а достигается, а «будет иметь последствия» логичнее заменить на «приведет к последствиям». Такие ошибки могут указывать как на галлюцинации нейросети, так и на использование англоязычных моделей (особенно для связки «будет иметь», так как «will have» – привычная форма для будущего времени в английском языке).

Текст демонстрирует нейтральность в освещении этически сложной темы: упоминает как положительные аспекты технологии («имеют потенциал значительно улучшить качество жизни людей»), так и проблемные моменты («недостаточная прозрачность и этические вопросы»). Это снова указывает на ограничения модели в выражении личного мнения, что можно оценить как отрицательный механический подход в изложении, или же как положительную непредвзятость.

В целом текст выглядит более сбалансированным, глубоким и персонализированным, но из-за перехода от краткого изложения фактов к попытке структурировать изложение в виде истории, материал обнаруживает большее количество ошибок и неточностей, чем прежние короткие информационные сводки.

Качественный анализ генеративных текстов, подготовленных в рамках эксперимента сетевого издания РБК с нейросетями, показал определенные проблемы и ограничения.

Во-первых, структурные и языковые особенности:

- грамматические, пунктуационные и синтаксические проблемы, ошибки в согласовании слов, неправильный выбор предлогов и падежных форм, некорректная расстановка знаков препинания;
- шаблонность структуры, часто по формуле [кто или что] теперь или в будущем [действие], которая создает ощущение механического изложения, перечисления действий;

- неестественный порядок слов и нетипичные для русского языка синтаксические конструкции, очень схожий с дословным переводом с английского языка.

Последняя проблема может указывать как на галлюцинации нейросети, так и на использование англоязычных моделей, обученных на английских дата-сетах. Такая модель готовит ответ на русском языке для промпта на русском языке, но данные, которые она использует в расчетах представлены на английском, и их нужно дополнительно переводить на язык запроса.

Во-вторых, достоверность фактов и корректность смысловых акцентов:

- дисбаланс в представлении точек зрения, который объясняется абсолютной нейтральностью нейросети. В целом доминирование позитивной или нейтральной подачи информации, которая создает эффект одностороннего рассмотрения проблемы;
- отсутствие критической оценки, недостаточность скептических мнений и критических позиций;
- склонность к упрощению сложных научных концепций в ущерб достоверности и точности.

Последнюю особенность можно также объяснить промптом, используемым РБК: информация может быть упрощена целенаправленно, так как общий формат всех текстов достаточно краткий, обзорный.

В-третьих, особенности анализа исходной информации:

- поверхностность аналитики: обобщенные, универсальные выводы, сухо излагающие тезисы исходного текста;
- отсутствие экспертной глубины, в некоторых случаях даже потеря связей между причиной и следствием из-за их неочевидности;
- ограниченная контекстуализация, недостаточное встраивание событий в более широкий контекст, при этом в материале, где была предпринята попытка сделать обобщение, наоборот, обнаружилось большее количество проблем;

- слабая работа с источниками: неспособность критически оценивать достоверность и авторитетность источников информации;
- склонность к использованию общих и очевидных выводов – эта проблема может вытекать из предыдущей, так как провести более глубокую оценку материала исходной статьи позволили бы именно перекрестная проверка информации и анализ полярных мнений.

Наиболее показательные проблемы включают:

- лингвистические признаки машинного происхождения текста, возможно, использования моделей, которым требуется промежуточный перевод;
- односторонние, нейтральные, ограниченные оценки, отсутствие экспертных оценок и цитат, замена реальной экспертизы обобщенными формулировками.

Фактически материалы представляют собой краткие саммари на основе текстов, подаваемых на ввод (например, научных исследований, публикаций). Проведенный анализ подтверждает роль журналиста в работе с генеративными технологиями – текстовый контент необходимо проверять на достоверность, а также вычитывать во избежание грамматических, пунктуационных и синтаксических ошибок. Нейросети могут выступать вспомогательным инструментом, но не способны заменять человеческую экспертизу в проверке фактов, обеспечении баланса мнений, установке логических связей и адаптации контента к целевой аудитории. Журналисту необходимо дополнять генеративные тексты экспертными оценками, корректировать стилистику и язык текста. Нейросеть, в отличие от журналиста, не способна придать тексту уникальный голос, который не только влияет на интерес и вовлеченность аудитории, но и транслирует ценности самого издания. Технологические возможности должны обязательно сочетаться с профессиональным журналистским подходом для создания вовлекающего качественного контента.

Для сравнительного анализа отбирались материалы рубрики «Написано ИИ» из «РБК Тренды» [21] в сопоставимом количестве, но уже за весенний период 2025 года. Тексты отбирались по порядку, без какой-либо сортировки или фильтрации для непредвзятого сравнения поведения и уровня развития моделей в работе с текстовым контентом.

Анализ генеративного текста «Исследование ВШЭ зафиксировало рост числа молодых ученых в России» показал в части содержания корректное изложение исследования НИУ ВШЭ о росте числа молодых ученых. Контент материала соответствует заголовку и раскрывает суть новости.

Структура выдержана в классическом формате «Что происходит» (факты) и «Что это значит» (их значение). Информация представлена логично, от фактов к их интерпретации.

Тема раскрыта поверхностно. Отсутствует контекст для оценки значимости роста на 1,2 тыс. человек (непонятно какой это процент). В тексте не достаёт конкретных примеров лабораторий, грантов, деталей программы «Приоритет-2030». Информация просто передана, в тексте нет качественной интерпретации цифр. Показательно, что без процентного выражения цифра «1,2 тыс. человек» может выглядеть впечатляюще, но выраженная в процентах (2,3 %) она уже не выглядит значительной.

Текст написан в типичном стиле с маркерами «Что происходит» и «Что это значит» (вероятнее всего, спустя год издательство использует все тот же промпт). Присутствуют клишированные выражения («играет ключевую роль», «на фоне», «становится важным фактором»).

Язык статьи довольно бедный, преобладают простые конструкции. Отсутствуют образные выражения, метафоры или иные стилистические приемы, нехарактерные для нейросетей, которые могли бы оживить текст. Материал нейтрален, он сухо передает исходные факты.

Терминология корректна и доступна широкой аудитории. Специфические термины («мегагранты», «молодежные лаборатории») использованы уместно, но без пояснений – это и не требуется для

неподготовленной аудитории, так как смысл написанного достаточно доступен.

Основные цифры представлены точно, но в тексте упускается несколько важных деталей, которые были отсеяны из исходного материала. Например, не достаёт информации о численных оценках, динамике прошлых периодов (исследование проводится с 2017 года), различиях между типами организаций, а особенно о проблемных аспектах (международное сотрудничество и зарплаты) и экспертном комментарии. В исходном материале Леонид Гохберг даёт подробную интерпретацию полученных данных, обозначает проблемы и прогнозы. Отсутствие этой важной информации может создать поверхностное, излишне оптимистичное представление о ситуации и мешает составить объективное мнение на основе полных данных. Заметна тенденция к обобщениям без подтверждающих фактов, особенно в разделе «Что это значит». Например, утверждение о «мировой практике» и опыте «стран ЕС и США» выглядит подвешено и не подкреплено примерами.

Контекст представлен недостаточно полно, что может быть снова связано с некорректной работой с источником, упущен исторический контекст, проблематика, прогнозы – эти данные не потребовали бы обращения к дополнительным источникам, они есть в исходном материале. Как новостная заметка текст информативен, но недостаточно глубок, из-за чего ему сложно удерживать внимание и вовлекать (в отличие от исходного материала НИУ ВШЭ). Текст требует редакторского вмешательства для углубления контекста путем добавления конкретики и экспертных комментариев. В итоге читатель получает базовое представление о тенденции, но не действенную информацию, она неприменима на практике. В целом материал все же отличается четкой структурой корректной передачей отобранных фактов, но он достаточно шаблонный и поверхностный, в нем нет живых примеров и конкретики (несмотря на то, что нужные детали все же были в материале НИУ ВШЭ).

Анализ текста об ИИ-помощниках «Яндекс Образование» запускает ИИ-помощников для преподавателей» в части структурных характеристик показывает классическую структуру с разделами «Что происходит» и «Что это значит», ключевая информация из исходного текста представлена в начале, далее идут дополнительные факты и хронология внедрения технологий.

Материал описывает конкретные инструменты (Нейроквизы, Нейродетектор, ИИ-помощник), дает достаточно полную информацию о запуске и применении продуктов. В тексте отсутствуют комментарии экспертов, упоминания потенциальных рисков или ограничений, критический анализ или какое-либо сравнение с конкурентами, нет примеров использования инструментов на практике. Факты излагаются нейтрально и тезисно.

Отсутствие глубины создает однобокое, оптимистичное представление о технологии. Читатель получает только позитивную или нейтральную информацию о новых инструментах, но в итоге не понимает их реальные возможности, ограничения и потенциальные проблемы использования. Журналистская заметка в том же объеме будет гораздо полезнее и информативнее. Генеративный текст не позволяет сформировать объективное мнение о целесообразности и эффективности использования описанных решений в образовании, не дает никакого понимания, насколько успешно эти инструменты уже работают в реальных условиях.

Текст «Женщины в науке: Россия среди мировых лидеров по числу исследовательниц» передает актуальные статистические данные НИУ ВШЭ об участии женщин в российской науке. Описаны общая численность (130,5 тыс.), доля от общего числа исследователей (38,5 %), приведена даже сравнительная статистика с другими странами, что расширяет контекст данных. Контент соответствует заявленной теме и опирается на реальное исследование НИУ ВШЭ.

Текст имеет четкую структуру с разделами «Что происходит» и «Что это значит», информация излагается логично: сначала факты, потом их

интерпретация. Внутри каждого раздела информация организована по принципу от общего к частному, что способствует ясности восприятия.

Тема раскрыта поверхностно. В тексте нет контекста изменения показателей во времени, не раскрыты причины такого положения дел, не проанализированы проблемы, с которыми сталкиваются женщины-исследователи. Отсутствуют экспертные мнения и комментарии самих женщин-ученых, из-за чего страдает глубина тезисов – они сухо, механически и нейтрально излагаются на основе исходного текста. При этом в материале есть обобщение, которого нет в исходном тексте: «Тенденция увеличения числа женщин в исследованиях соответствует мировому тренду на гендерное равенство» – это достаточно спорная формулировка, которая требует подкрепления.

Статья написана в нейтральном инфостиле с преобладанием клишированных конструкций, характерных для новостных текстов, используются стандартные переходы и шаблонные фразы («По последним данным», «Традиционно высокая доля»). Отсутствуют стилистические приемы, которые могли бы сделать текст более выразительным. Более обобщенный и менее точный язык генеративного текста приводит к размыванию формулировки «согласно последним данным официальной статистики, доступной по итогам 2023 г.», которая переходит в простое «по последним данным» без указания года. Нейронные сети обучаются на массивах с определенными временными промежутками, поэтому для человека такая формулировка в 2025 году может выглядеть сомнительно, а для модели, обучающие данные для которой ограничены, допустим, 2023 годом, будет выглядеть логично. Для таких элементов текста требуется внимание журналиста, который сможет отследить и скорректировать этот момент.

Лексический запас материала ограничен, наблюдаются повторы слов и конструкций. Синтаксис преимущественно простой, с преобладанием повествовательных предложений средней длины. Метафоры, эпитеты и

другие выразительные средства отсутствуют. Язык изложения точный и доступный для широкой аудитории, с этим не возникает сложностей.

В основном статистика выглядит адекватно и достоверно, данные согласуются между собой. Формулировка «Подробнее о результатах исследования можно узнать здесь» с размещением гиперссылки может указывать на редактуру текста человеком. В тексте не удается найти явные фактические ошибки или противоречия.

Контекст представлен ограниченно. Генеративный материал уступает оригиналу в полноте из-за отсутствия визуализации и примеров конкретных научных достижений женщин (несмотря на то, что в более ранние материалы редакторы переносили визуализации источника).

Статья в целом справляется с базовой информационной задачей и сообщает ключевые факты, но ни в одном тезисе не встречается более глубокого анализа. Читателю такой заметки будет сложно комплексно представить ситуацию и критически ее осмыслить из-за обрывчатости и несвязности тезисов. Текст требует дополнения редактором – включения экспертных мнений, возможно, исторического контекста, анализа тенденций, описания проблем и вызовов. Необходимо стилистическое обогащение и устранение лексических повторов. Материал может выступить хорошей основой для журналистской доработки.

Текст «В России сложилась постоянная аудитория научно-популярного медиаконтента» основан на исследованиях НИУ ВШЭ в рамках проекта «Делаем науку в России». Большинство ключевых статистических данных из исходного текста передано точно, но можно заметить фактическую ошибку в оценке уровня коммуникаций с обществом: в генеративном тексте указано 3,79 балла, в оригинальном исследовании НИУ ВШЭ – 3,9 балла.

Текст структурирован по той же схеме с разделами «Что происходит» и «Что это значит». Такая структура логична, но тезисное изложение и отсутствие визуального представления данных значительно упрощает богатый материал оригинального исследования. В первой части анализируемого текста

последовательно представлены основные результаты исследования, во второй – их интерпретация, но без опоры на все данные из оригинала.

Тема раскрыта поверхностно и выборочно. Из обширного исследования отобраны отдельные данные, упущены аспекты оценок по типам организаций (показатели ВУЗов выше коммерческих организаций); статистика по форматам потребления контента (телепередачи и видео в интернете оказываются предпочтительнее), разница между поколениями; комментарий эксперта с интерпретацией числовых данных к практике упрощен до пары слов.

Текст написан в стандартном инфостиле, использует шаблоны и клише. Заметно стремление к нейтральности, из-за чего язык более формализован, чем в оригинале. Эмоциональный комментарий эксперта опущен («Среди ученых все больше тех, кто видит смысл в том, чтобы делиться своими знаниями и экспертизой не только с академическим сообществом, но и с широкой общественностью, и готов это делать <...>»), из-за чего материал становится безэмоциональным и механическим.

Лексика ограничена, с заметными повторами ключевых слов («наука», «опрос», «научно-популярный»). Синтаксические конструкции однотипны, преобладают простые или сложносочиненные предложения среднего размера, отсутствуют образные средства и выразительные обороты, которые могли бы оживить материал.

Терминология корректна и понятна для широкой аудитории. Термины, например, «мониторинг инновационного поведения», «научно-популярный медиаконтент», не мешают восприятию материала и используются без излишних усложнений. В этой части генеративная статья удачно адаптировала материал для широкой аудитории.

Большинство фактов материала соответствует оригинальному исследованию, но встречаются отдельные расхождения, например уже упомянутая оценка уровня коммуникаций с обществом (3,79 вместо

3,9 балла); отсутствие размера выборки опроса (6 946 респондентов) и упоминания оценки уровня коммуникаций с обществом выше других аспектов

Текст выполняет базовую информационную функцию, сообщая об исследовании и предоставляя некоторые его результаты, но он значительно менее плотный в сравнении с оригиналом при примерно сопоставимом объеме текстового контента.

Анализ текста «Microsoft разрабатывает свой ИИ для конкуренции с давним партнером OpenAI» о стратегическом повороте Microsoft в области ИИ представляет собой обратный пример – нейросеть развернула краткое новостное сообщение в небольшой информационный материал, изменив подачу информации.

Основные факты из оригинала воспроизведены корректно: разработка моделей с возможностью рассуждать, создание семейства MAI, тестирование альтернативных решений, упомянуты также инвестиции и привлечение Мустафы Сулеймана. Интересно отметить небольшое изменение в модальности сообщения – в исходном тексте расставлены маркеры «сообщается», «как сообщается», а генеративная версия более уверенно преподносит факты как общеизвестные и достоверные.

В тексте присутствуют и орфографические ошибки: «наем» вместо нормативного «найм» в части «включая наем соучредителя DeepMind и Inflection Мустафы Сулеймана»; «внедерния инноваций», что указывает на неустойчивость модели при работе с нормами русского языка. Есть проблемы с пунктуацией, например, упущенная запятая в предложении «Компания разрабатывает в том числе ИИ-модели <...>».

Генеративная статья существенно перестроила оригинал, фактически создав на основе фактов новый блок с интерпретацией. Рассуждения в этой части текста довольно обобщенные, не подтверждаются источниками или экспертным мнением. Этот блок очень показателен именно из-за генерации модели без подкрепления данными (здесь она работает самостоятельно) – он изобилует типичными клише без конкретики: «может дать преимущество на

рынке», «компании стремятся создать уникальные решения», «становится ключевой технологией». Показательна и незавершенная логическая цепочка: «Усиление конкуренции среди крупных игроков, таких как Microsoft, может привести к ускорению внедрения инноваций <...>» – где перечисление обрывается на единственной компании, хотя формулировка «таких как» очевидно подразумевает продолжение списка. Текст показывает упрощение и стандартизацию языка, использует шаблонные конструкции с типичными оборотами. Структура «один факт – один абзац» облегчает восприятие, но лишает текст индивидуальности.

Генеративная статья представляет не просто перевод или адаптацию оригинала, а заметно трансформирует его. Основные проблемы текста связаны даже не с прямыми фактическими ошибками, а с более тонкими искажениями, влияющими на восприятие достоверности информации: самостоятельная генерация блока «Что это значит» показывает иллюзию аналитического анализа при отсутствии подкрепляющих фактов. Эта проблема в формате тезисного изложения исходного текста становится заметна именно в момент, когда нейросеть опирается не на данные промпта, а на свои обучающие дата-сеты.

Следующая генеративная статья «Индикаторы образования в России: статистика и тренды на 2025 год» основана на тезисах из статистического сборника «Индикаторы образования: 2025» и показывает заметную селективность в отборе информации: последний тезис о программах профиля ИИ полностью проигнорирован в основной части статьи. Парадоксально, но именно эта тема получила развитие в разделе «Что это значит». Интересно, что нейросеть корректно суммирует данные о выпускниках техникумов и колледжей, верно объединяя их в выпускников СПО. В оригинале: «выпущены 208,7 тыс. бакалавров, специалистов и магистров, 73,2 тыс. специалистов среднего звена и 14,5 тыс. – квалифицированных рабочих и служащих», после переработки: «численность выпускников вузов

составила 208,7 тыс. человек, выпускников программ СПО – 87,7 тыс. человек».

В большинстве случаев модель опускает сравнение с предыдущими годами, несмотря на то, что данные есть в оригинале, а это лишает читателя возможности оценить тренды.

Наиболее проблематичный аспект генеративной статьи – разрыв логической связи между разделами «Что происходит» и «Что это значит». Третий абзац части второй части «Включение модулей по искусственному интеллекту в образовательные программы отражает глобальную тенденцию по интеграции ИИ в образование, что способствует подготовке студентов к современным вызовам» – не имеет фактологического основания в первой части и никак не связан с текстом. В оригинале для этого вывода приведен целый абзац с конкретными цифрами о программах по профилю и образовательных модулях. Модель, очевидно, использует эти данные для вывода, но теряет их в указании фактов.

Аналогичная проблема с причинно-следственными связями наблюдается в выводе о «росте числа студентов в сфере IT», который не имеет фактологического подтверждения в первой части текста (там приводится только общее количество выпускников, динамика упущена). В оригинале данные сравниваются год к году и показывают неоднозначную динамику: снижение выпуска в вузах, но рост в СПО.

Генеративная статья незначительно упрощает стиль и структуру оригинала. Исходный текст плотнее насыщен фактами и сравнениями данных, а генеративная версия разбивает материал на короткие, изолированные фрагменты, каждый такой тезис содержит лишь одну основную мысль. Это с одной стороны облегчает чтение и восприятия текста, но с другой – приводит к потере значительной части аналитики и статистики.

Этот текст особенно ярко демонстрирует типичные проблемы генеративного контента: селективный отбор информации, логические ошибки, упрощение стиля, обобщение информации с потерей фактов. Особенно

показательно игнорирование некоторых элементов исходной информации при построении текста, но построение выводов на эту тему в заключении.

Качественный анализ текстов подтверждает ранее выявленные системные проблемы, которые существенно влияют на достоверность, информационную ценность и этическую корректность текстового контента. Эти проблемы отражают фундаментальные ограничения современных генеративных моделей.

Во-первых, проблемы со структурными и языковыми особенностями:

- шаблонность изложения, использование преимущественно простых синтаксических конструкций, тяга к типовым оборотам и клише, особенно если нейросеть генерирует информацию с нуля, а не обрабатывает предоставленные данные;
- фрагментация – текст разбивается на короткие, слабо связанные между собой блоки информации, что создает ощущение механического изложения, предложения просто выстраиваются одно за другим, связи между ними и удаленными друг от друга абзацами могут быть слабыми или отсутствовать вообще (например, интерпретация факта, который нейросеть не взяла в повествование, неверный вывод из факта);
- грамматические, пунктуационные и синтаксические ошибки, неустойчивость модели в работе с русским языком (эта проблема может быть обусловлена галлюцинациями моделей или проблемами с промежуточным переводом при использовании англоязычной модели);
- логические ошибки и разрывы, незавершенные по смыслу предложения (например, предложение «среди крупных игроков, таких как Microsoft» без продолжения списка);

Во-вторых, недочеты в достоверности фактов и смысловых акцентах:

- селективность отбора фактов, игнорирование важных фрагментов исходных текстов при сохранении второстепенных деталей (человек

мог провести фильтрацию совершенно иначе, нейросеть не расставит акценты, может убрать из исходной информации любой факт, рассматривая все их равноценно);

- потеря контекста, пропуск пояснений, экспертных мнений, сравнительной динамики, исторического контекста – всего того, что делает факты значимыми и наполняет их смыслом для читателя;
- фактические искажения из-за некорректной работы с числовыми данными.

В-третьих, ошибки в анализе информации:

- необоснованные выводы, заключения и интерпретация, не опирающиеся на конкретные факты, из-за чего нарушается причинно-следственная связь, ломается логика текста;
- использование обобщенных фраз и утверждений, создающих видимость экспертной оценки (например, «По последним данным», «Традиционно высокая доля»);
- поверхностная интерпретация фактов без учета скрытых взаимосвязей.

Среди наиболее показательных проблем можно отметить:

- искажение информации, которая влияет на достоверность – даже при отсутствии грубых фактических ошибок смысловые акценты могут быть изменены, что ведет к снижению достоверности. Особенно критично здесь обобщение, которое не опирается на источники, а выглядит как очевидная, всем известная истина. Это подменяет достоверные утверждения и факты, которые можно рассмотреть и с другой точки зрения;
- логические ошибки;
- создание внешне схожих с журналистскими, но содержательно неполноценных материалов.

Выявленные проблемы снова показывают, что работа с генеративными текстами требует обязательного человеческого участия не только для

стилистической правки, но и для критической проверки фактов, логических связей и этической корректности информации. Несмотря на быстрый темп развития моделей, они все так же воспроизводят ограничения своей технологии и показывают те же проблемы. Ограничения могут выглядеть менее критично благодаря общему повышению уровня генерации, но они все так же обнаруживаются в контенте.

Тезис о том, нейросети должны выступать только лишь вспомогательным инструментом журналиста, но не могут его заменить из-за своих ограничений, подтверждается и результатами практического опыта использования нейросетей в существующей редакции «Краснодарские известия», которая 1 апреля 2025 года выпустила газету с визуальными и текстовыми материалами, полностью сгенерированными нейросетями [19]. По словам главного редактора Владимира Вербицкого, «В эксперименте были задействованы такие нейросети, как GigaChat, GPT-4o, YandexGPT, Rytr, AI Search. Для генерации картинок и графиков привлекли Sora AI, GigaChat, Midjourney, Reve и Kandinsky» [8]. Материалы выпуска редактировали профессиональные журналисты и редакторы, поэтому они имеют ограниченную ценность для анализа: в этих публикациях недостатки «сырых» генеративных текстов уже устранены. В этом случае были проанализированы выводы самого СМИ (слова главного редактора) о проблемах, с которыми они столкнулись в процессе работы. Оценка нейросетей действующей редакцией в данном случае намного ценнее в понимании проблем и ограничений технологии, чем анализ отредактированных материалов.

Из официального комментария главного редактора «Краснодарских известий» можно выделить ряд проблем.

Во-первых, достоверность фактов – обязательность фактчекинга для генеративных текстов, так как нейросеть может использовать неактуальную информацию, особенно при работе над обобщающими запросами, требующими аналитики (например, с материалами для темы «главные факты о Краснодарском крае за последние десять лет» – то есть массивом

информации, распределенной по годам, где нужно верно выстроить связи, расставить акценты и сделать выводы).

Во-вторых, искажение цитат – отмечается, что нейросеть может внести «свои допущения» в цитату, а это уже создает серьезную проблему достоверности и журналистской этики.

В-третьих, необходимость человеческого контроля – несмотря на заявления о «полностью сгенерированном» номере, редакции потребовалась существенная работа человека – журналистов, редакторов, фотожурналистов. Специалисты проводили проверку фактов, цитат, обеспечивали доработку материалов.

Опыт «Краснодарских известий» демонстрирует, что даже при полномасштабном использовании разного типа моделей в создании целого выпуска газеты, человеческий контроль над производством генеративного контента на сегодняшний день остается обязательным условием качественной журналистики. Несмотря на развитую способность воспроизводить человеческую письменную речь, генеративные тексты страдают от фактологической ненадежности, склонности к галлюцинациям и искажениям. Генеративные тексты достаточно поверхностны и нейтральны, не передают глубину человеческого анализа и экспертного мнения. Специфический контекст журналистского исследования может быть проигнорирован в пользу универсальных шаблонов, а язык генеративного текста размыт и стандартизирован. Даже при условии умеренной точности в передаче основных фактов, нейросеть может упустить исходные смысловые акценты, нюансы и полутона человеческой речи, выровнять информационный рельеф, исключить из текста оценочные личные мнения, упростить глубокие аналитические выводы. Для редакционной практики это означает, что использование генеративных текстов без тщательной проверки и обработки может снизить ценность и достоверность публикаций: каждый генеративный текст нужно критически оценить, выявить и устранить общие дефекты генерации перед использованием.

На основании выявленных ограничений возникает практическая необходимость повысить полезность нейронных сетей в редакционной работе и снизить влияние их недостатков. Использование моделей требует не только понимания их возможностей, но и соблюдения стратегий эффективного взаимодействия с ними. Ключевыми аспектами такой работы являются создание корректных промптов, проверка достоверности генерируемого контента и соблюдение этических стандартов. Эти направления определяют качество, точность и достоверность результатов работы нейросети. Особое внимание нужно уделить промптам: от корректности передачи входных данных напрямую зависит верность интерпретации запроса, содержания и структуры выходных данных. Подходы к взаимодействию с моделями можно условно разделить на два основных типа:

- нейросеть обрабатывает уже структурированные данные, подготовленные человеком заранее;
- модель выполняет задачу на основе естественно-языкового промпта.

Прямая обработка данных (без необходимости подготовки текстовых инструкций) используется при работе со значительной частью моделей, в том числе со сверточными нейросетями, осуществляющими классификацию и сегментацию изображений, рекуррентными нейронными сетями, анализирующими временные ряды и последовательные данные, системами автоматического распознавания речи, алгоритмами машинного обучения в рекомендательных системах и другими. На вход подаются данные определенного формата, а результат генерируется согласно предварительно заданным параметрам обучения модели. Альтернативный подход к работе с моделями осуществляется через естественно-языковые промпты. Он характерен, например, для генеративных нейросетей (GPT) и генеративных моделей изображений (DALL-E, Midjourney и других), мультимодальных моделей, способных интерпретировать различные типы входных данных (таких, как GPT-4, Claude, Gemini). Очевидно, что в контексте данного исследования такие нейросети представляют особый интерес. Открытый

доступ к генерации текстового и визуального контента может быть использован любым специалистом из области журналистики, в отличие от закрытых корпоративных решений издательств, которые требуют специальных разрешений и финансовых затрат от организации для получения доступа.

Промпт представляет собой не просто набор инструкций, а специфическую форму программирования, позволяющую задавать контекст и правила работы нейросети, ограничить ожидаемые результаты под потребности пользователя. Это отмечает в своей работе Д. Уайт: «Промпт инжиниринг становится все более важным навыком для эффективного взаимодействия с большими языковыми моделями (LLM), такими как ChatGPT. Промпт – это инструкция, которая дается LLM для обеспечения соблюдения правил, автоматизации процессов и обеспечения определенного качества (и количества) выходных данных. Промпты также являются формой программирования, которая позволяет настраивать выходные данные и взаимодействие с LLM. <...> Промпты – это метод передачи знаний, аналогичный паттернам (шаблонам программирования), поскольку они являются многократно используемыми решениями для общих проблем, с которыми сталкиваются в определенном контексте, например, для генерации выходных данных и взаимодействия при работе с LLMS» [64]. Разработка эффективных промптов требует учета их универсальности, адаптивности и ясности, чтобы они могли решать широкий спектр задач. Д. Уайт в той же работе «Каталог промптов для улучшения промпт-инжиниринга с использованием ChatGPT» [64] выделяет ряд аспектов, которым нужно следовать при создании промптов для работы.

Структура промпта должна включать фундаментальные утверждения (основные идеи), которые нужно передать модели. Они должны быть изложены четко и просто, чтобы их можно было адаптировать под различные задачи. Структура должна оставаться гибкой, позволяя формулировать запросы в разных формах и с использованием различных подходов.

Промпт должен быть универсальным, то есть подходить для множества областей применения, включая генерацию текста, автоматизацию процессов и визуализацию данных. Универсальные промпты легче адаптируются к разным задачам и могут использоваться повторно для новых задач.

Промпт должен быть понятен и для технических специалистов, и для журналистов, не обладающих глубокими знаниями в программировании, и для самой нейросети. Простые формулировки без двусмысленных трактовок позволяют быстрее добиться желаемых результатов и корректно интерпретируются как пользователями, так и моделью.

Промпт должен быть гибким, адаптивным. Важно, чтобы он учитывал особенности взаимодействия с моделью, но и имел возможность меняться в зависимости от контекста. Переписывание и улучшение промптов помогает быстрее и эффективнее достигать результатов в новых или уточненных сценариях.

Каждый промпт должен сопровождаться примером практической реализации, чтобы пользователи и модель могли понять, каким должен быть результат. Пример реализации для нейросети также показывает ожидаемый формат выходных данных. Сами промпты могут включать указания для модели задавать пользователю уточняющие вопросы по задаче – как именно она должна быть выполнена, как расставить акценты и так далее, они помогут дополнительно уточнить запрос.

Необходимо контролировать последствия промптов, то есть учитывать их преимущества и недостатки, чтобы минимизировать риски и выбрать оптимальный подход. Предварительная оценка и проверка промптов помогает корректировать его под конкретные задачи и избегать ошибок.

Эффективный промпт универсален, понятен для нейросети и человека, а также адаптивен. Корректные промпты значительно расширяют возможности использования языковых моделей, помогая автоматизировать задачи и упрощать взаимодействие с моделью, особенно журналистам без специальных навыков работы с технологией. Д. Уайт предлагает применять такой подход к

разработке универсальной базы промптов: специалист может выбрать нужный промпт под конкретную задачу, минимально изменить входные данные, и получить корректный ответ. Этот подход, очевидно, может применяться и в журналистике – используя описанные требования к промпту, команда может создать каталог промптов для работы с нейросетью, который будет использоваться всеми сотрудниками для упрощения своей работы.

Пример корректного промпта, учитывающий отмеченные требования, может выглядеть следующим образом:

«Проанализируй влияние искусственного интеллекта на трансформацию рынка труда в банковском секторе России за период 2018-2023 гг.

В анализе должны быть:

- количественные показатели сокращения рабочих мест в конкретных профессиональных категориях;
- новые специальности, сформированные благодаря цифровизации;
- оценка экономического эффекта внедрения ИИ для банковского сектора.

Результаты представь в формате структурированного материала. Объем текста от 2000 до 2500 знаков. Обязательно выдели ключевые тенденции и сценарии дальнейшего развития ситуации. Учти фактор экономических санкций и их влияние на отрасль».

Методологически неверный промпт на аналогичную тему может выглядеть так: «Напиши текст про влияние технологий на экономику».

Такой вариант не ограничивает тему, в нем нет временных и географических границ, требований по формату представления результатов, объему и глубине анализа. Вероятнее всего, модель выдаст в ответ на такой промпт общий, поверхностный текст, не соответствующий задачам пользователя. Полезно уточнять промпт в процессе работы, так как обычно при обращении к модели человек имеет хотя бы примерные представления об ожидаемом результате работы. Таким образом промпт удастся

конкретизировать уже с учетом ответа нейронной сети, даже если изначально нет понимания, как сузить границы запроса.

Ученые Корнельского университета предложили свою классификацию промптов, которая может помочь определиться с нужной структурой. В рамках исследовательской работы рассмотрена возможность применения таких промптов в журналистской деятельности, соответственно, для каждого типа промпта приведены авторские примеры. Классификация включает шесть ключевых категорий [16], [64]:

Семантика ввода – этот тип промптов позволяет задавать правила для модели, которые она сохраняет в памяти и использует в дальнейшей работе. Эти промпты оптимизируют синтаксис и помогают значительно экономить время.

Пример промпта: При обработке запросов о политических событиях укажи три вида источников: официальные заявления, экспертные комментарии, данные социологических опросов. Если данные относятся к выборам, приведи для сравнения информацию о результатах предыдущего периода.

Настройка вывода – такие промпты позволяют запросить результат в конкретном формате или шаблоне. Перед этим необходимо предоставить модели набор правил и затем уточнить, когда и как их использовать

Пример промпта: Сгенерируй заголовок, лид и три ключевых тезиса для материала о влиянии социальных сетей на распространение фейковых новостей. Используй формат:

[ЗАГОЛОВОК] – не более 10 слов, содержит интригу.

[ЛИД] – раскрывает суть материала, 30-40 слов.

[ТЕЗИС 1] – статистический факт.

[ТЕЗИС 2] – мнение эксперта.

[ТЕЗИС 3] – практический вывод.

Обнаружение ошибок – промпты из этой категории направлены на выявление источников данных и фактов, на которых строится ответ модели.

Они позволяют проверить ответ, чтобы не копировать контент в том же виде с возможными ошибками. Можно попросить модель сгенерировать список фактов для проверки или задать уточняющие вопросы, чтобы подкорректировать запрос и получить более точный результат.

Пример промпта: «Проанализируй фактические данные в этом тексте о климатических изменениях в России: [текст]. Выдели все числовые данные, укажи их источник и год получения. Отметь устаревшую информацию и предложи актуальные данные из официальных источников за последний год».

Улучшение запросов – если модель отказывается выдавать ответ, пользователь может попробовать обойти ограничения. Особенно полезно поговорить с моделью, позволить ей задать уточняющие вопросы, которые помогут более эффективно сформулировать запрос. Они очертят поле, в котором нейронная сеть готова работать и позволят найти более корректную формулировку.

Пример промпта: «Мне нужно разработать структуру расследования о коррупции в строительстве социальных объектов. Задай мне 5 уточняющих вопросов о том, какие аспекты темы следует раскрыть, какие источники использовать, и как структурировать повествование, чтобы материал был максимально информативным».

Взаимодействие – эта категория охватывает задачи, требующие многократной генерации данных (например, заголовков или идей) по шаблону. Чтобы избежать постоянного ввода однотипных запросов, можно настроить модель на непрерывную генерацию в соответствии с установленным правилом.

Пример промпта: «Сгенерируй 15 вариантов заголовков для новостной статьи о [внедрении системы распознавания лиц в общественном транспорте Москвы], они должны отражать [технологический прогресс и проблемы приватности]. Один твой ответ – один заголовок для статьи. После каждого заголовка я буду отвечать «далее» для генерации следующего варианта или «замени», если вариант не подходит – тогда предложи другой заголовок».

Контроль контекста – такие промпты позволяют ограничить контекст, в котором модель работает. Это помогает отсеять нерелевантный контент и сосредоточиться на важных данных.

Пример промпта: «При подготовке справочного материала о системе здравоохранения используй только данные по России и странам СНГ за период 2021-2024 гг. Исключи информацию о западноевропейских и североамериканских медицинских системах, сосредоточься исключительно на СНГ».

Рассмотренная классификация не только упрощает процесс взаимодействия с моделями, но и помогает быстрее определить подходящий подход к решению задачи. Примеры, иллюстрирующие виды промптов призваны показать возможности их применения: от настройки формата вывода до контроля контекста. Такая удобная систематизация может помочь журналистам повысить эффективность работы с моделями для создания качественного контента, соответствующего профессиональным стандартам. На основе рассмотренных требований к промптам и их типологии, которые уже излагались в научных работах, целесообразно разработать каталог промптов для практикующего журналиста, с адаптированными для его практики промптами. Цели каталогизации промптов: экономия времени, повышение качества итогового материала и демонстрация возможностей нейросетей с открытым доступом в изучаемой области. Каталог представлен в таблице 1.

Таблица 1 – Каталог возможных журналистских задач для нейросетей

Задача	Нейронная сеть	Пример промпта на естественном языке (если он необходим)
Транскрипция интервью	Whisper (OpenAI) с открытым исходным кодом на GitHub (для работы нужны Python и дополнительные библиотеки FFmpeg, Git, Torch (PyTorch), сам пакет Whisper)	Нейросеть работает с аудио-файлами без запросов, проводя их транскрипцию. Промпт может уточнять задачу, например тайминг и формат вывода: «Транскрибируй запись с 5 по 35 минуты. Каждый вопрос интервьюера должен начинаться с нового абзаца»

Продолжение таблицы 1

Генерация текста на основе данных	ChatGPT (OpenAI), Claude, Copilot, YaGPT	«Напиши пресс-релиз на основе данных: [данные]. Сохрани все ключевые факты и цифры. Объем текста должен быть не более 500 слов»; «Сгенерируй пять заголовков для статьи о [тема статьи], интересные для аудитории 18-35 лет»; «Создай короткую заметку объемом до 1000 слов о влиянии искусственного интеллекта на журналистику, основываясь на этих данных: [данные]. Отдельно в виде списков отметь ключевые преимущества и проблемы»
Классификация изображений	Google Vision AI (доступ через Google Cloud Platform, требуется настройка API и регистрация на платформе), Azure ML (доступ через Microsoft Azure и API)	Промпты не требуются, инструменты работают через API – задачи структурируются в виде параметров кода
Анализ тональности текстов	Hugging Face (доступна на Hugging Face Hub, библиотеки интегрируются в Python), AWS Comprehend (доступен через AWS Console с API для интеграции)	Запросы формируются в виде структурированных данных или параметров при использовании API
Подбор видеоматериалов	RunwyML, Pictory.ai	«Подбери видео на тему изменение климата. Материалы должны иметь явный смысл, показывать наводнения, таяние ледников, лесные пожары. Видео должны быть в качестве Full HD и выше, продолжительностью менее 5 минут. Укажи источник, из которого берешь материал»
Автоматизация фактчекинга	GhatGPT (OpenAI), Hugging Face Transformers, DeepMind RETRO, другие модели, которые могут обращаться к поиску в сети	«Проверь на фактические и логические ошибки: [текст или утверждение]. Найди три независимых источника для подтверждения или опровержения информации»
Суммаризация документов	Claude, Bing AI	«Суммаризируй документ: [документ]. Выдели ключевые идеи и тезисы в виде логичного связного текста объемом до 200 слов, исключи из него подзаголовки и списки»

Продолжение таблицы 1

Визуализация данных	DALL-E, Kandinsky 3.0, YandexART	«Изображение в минималистичном стиле на тему развития городской среды Тольятти. Элементы: современные жилые кварталы, завод АвтоВАЗ, Волга и окружающая природа (лес и горы). Цветовая палитра: серые, голубые и зеленые тона. Атмосфера: индустриальный город, окруженный природой. Мне нужно будет несколько изображений в единой концепции для лонгрида, они должны быть последовательными»
Создание мультимедийных лонгридов	ChatGPT (или другая генеративная нейросеть для текстов) и DALL-E (или другая аналогичная модель для графики)	«По моим черновым наброскам создай план мультимедийной статьи о проблемах урбанизации. Выдели тезисы или текстовые фрагменты, к которым стоит добавить инфографику и тематические изображения»; «Разработай концепцию мультимедийной статьи о проблемах развития городской среды с фокусом на взаимодействии пользователя. Предложи идеи для анимации, интерактивных карт и таймлайнов на основании данных: [данные с географической и временной информацией], которые помогут лучше раскрыть тему. Определи ключевые блоки, где нужна визуализация данных для максимального раскрытия темы»

Каталог демонстрирует широкий спектр задач, которые журналисты без специальных навыков могут решать с помощью нейросетей. Такой путь упрощает сложные процессы и делает более доступными современные технологии, что особенно актуально для региональных СМИ с ограниченными человеческими и денежными ресурсами. Каталог охватывает множество задач: от автоматизации транскрипции текстов и работы с мультимедийным контентом до несложной аналитики и фактчекинга. Технических навыков, очевидно, может потребовать интеграция методов API для некоторых

моделей, но этот путь можно заменить платформами-посредниками, которые предоставляют нужные доступы онлайн. Каталог позволяет использовать технологии для более визуально насыщенных, интерактивных и информативных материалов, вовлекать аудиторию, что критически важно для медиа с небольшим финансированием и возможностями.

Кроме необходимости корректной проработки промпта, журналисты должны понимать важность обязательного фактчекинга любой сгенерированной информации, поскольку даже самые совершенные модели не гарантируют ее достоверность. Журналисты могут применять ряд подходов для верификации точечных данных, суммаризаций, выводов или полноценных текстовых материалов, сгенерированных при помощи нейронных сетей. Для проверки достоверности информации могут использоваться как технологичные новые методы, так и классические журналистские методы фактчекинга. М.С. Борзова, основываясь на исследованиях М.С. Корнева, лаконично описывает традиционный фактчекинг в несколько шагов: «Во-первых, искать первоисточник или подтверждение информации из нескольких независимых источников. Во-вторых, рассматривать противоположные точки зрения на событие. В-третьих, учиться распознавать фейки. Причем под фейками исследователь подразумевает не только осознанные подделки и симуляции, но и небрежность, опечатки и технические ошибки. Говоря о борьбе с фактоидами, журналист отдельно упоминает об осторожном использовании информации, полученной из социальных сетей, а также о том, что подлинный факт важнее сенсации» [7]. Для упрощения проверки генеративного материала можно использовать специальные ресурсы для фактчекинга, например, [Factcheck.org](https://factcheck.org), [Politifact.com](https://politifact.com) и [FactCheckEU.org](https://factcheckeu.org).

Для улучшения качества генеративного сырого материала текст можно подвергать переработке путем редактуры, дополнения собственными исследованиями, комментариями, контекстуальной информацией. Если генеративный контент не проверяется и нейросеть выступает скорее соавтором контента, а не технологичным помощником в сборе и обработке

информации, следует также четко указать на использование модели в процессе создания контента и проинформировать аудиторию о потенциальных рисках. Это обеспечит прозрачность отношений между журналистом и читателем, а также поможет последнему критически оценить информацию. О необходимости маркировки контента рассуждает исследователь Лукина М.М.: «Еще одна проблема, которую необходимо решить при разработке этического кодекса, связанного с применением ИИ в современной журналистике, – это маркировка текстов и продуктов, созданных с помощью алгоритмических решений. Проведенные российскими учеными исследования доказали, что не все редакции в этом плане транспарентны, не все из них считают необходимым знакомить своих читателей с тем, что они потребляют новости, написанные с использованием автоматических генераторов контента. Необходимо также обратить внимание на потенциальную возможность диффамации в тексте, который написан алгоритмом. О такой угрозе задумываются специалисты из разных стран. Например, группа исследователей из США рассуждает о том, кто должен нести ответственность в том случае, если в материале, созданном машиной, содержится информация, порочащая честь и достоинство одного человека или группы людей. Такие рассуждения находятся в области теоретических дискуссий, так как в реальной практике пока подобные прецеденты не встречались. Однако в недалеком будущем специалистам по медиаправу, возможно, придется столкнуться и с такими казусами» [20]. С высказыванием автора сложно согласиться полностью: журналисты ведь не обязаны указывать все технологические средства, которые использовались в подготовке контента (например, программы Microsoft Office или Photoshop), поэтому здесь поднимается сразу несколько дискуссионных вопросов – какой процент контента, сгенерированного нейросетью, указывает на необходимость метки «сгенерировано при помощи нейросети», нужно ли такое указание, если контент полностью проверен и переработан человеком или если нейросеть работала с узкой задачей (например, подбирала видео для материала или очищала массив статистических данных перед его оценкой

человеком), то есть обрабатывала конкретную техническую задачу, но сам контент создавался человеком.

Процесс фактчекинга может быть частично автоматизирован с помощью моделей, что повышает эффективность и скорость проверки информации и особенно больших данных (при этом любое окончательное решение касательно этих данных все равно должно приниматься человеком). Например, для проверки результатов работы одной нейронной сети можно использовать другую нейронную сеть, обученную на выявление недостоверной или ошибочной информации. В целом, технические методы машинной проверки удобно классифицируются исследователями (например, Б. Лакзеи выводит классификацию на основе обзора публикаций Шу К., Сливы А., Ван С., Тан Ц., Лю Х., Юй И. [49], аналогичную структуру дают и авторы Али Р., Го Ч., Шлихткрулл М., Торн Дж. [59], и Л. Азеведо [31]) в зависимости от типа данных, используемых в качестве входных:

- основанные на неструктурированном контенте,
- основанные на структурированном контенте,
- основанные на контексте.

Методы проверки фактов, основанные на неструктурированном контенте, используют лингвистические, визуальные и семантические признаки для определения наличия в тексте недостоверной информации. Как указывает Б. Лакзеи, «Контентные признаки извлекаются непосредственно из содержания публикации. Большинство методов обнаружения дезинформации основаны именно на анализе контентных характеристик, особенно текстовых данных <...>. В целом контентные признаки можно разделить на три основные категории.

Лингвистические признаки – являются одними из наиболее часто используемых характеристик в детекции дезинформации. <...>. Фейковые новости, как правило, имеют характерный стиль написания и используют привлекающие внимание заголовки. Авторы дезинформации используют особые фразы для эмоционального воздействия на аудиторию (лексические

признаки – включают униграммы, биграммы и структуру слов в тексте. Один из простых подходов – анализ наличия ключевых слов и значимых фраз <...>. Также учитываются такие факторы, как наличие специфических терминов, подозрительные слова, длина слов, число предложений, частотность отдельных слов и другие параметры, что улучшает методы выявления дезинформации; синтаксические признаки связаны со структурой текста и стилем написания. Среди таких характеристик: количество существительных, глаголов, наречий и прилагательных, а также использование пунктуационных маркеров, таких как точки, вопросительные и восклицательные знаки. Также анализируется использование местоимений, отрицательных глаголов и наречий).

Визуальные признаки – играют важную роль в обнаружении дезинформации. Люди обладают естественной восприимчивостью к визуальному контенту и могут подвергаться манипуляциям через эмоционально окрашенные изображения <...>. Среди таких характеристик: разрешение, гистограммы, пропорции изображений и обнаружение объектов.

В некоторых методах используются <...> семантические характеристики, связанные с текстовым или визуальным контентом» [49].

Некоторые из таких методов используются поисковыми системами для сбора информации об упоминании проверяемого факта, благодаря чему можно сделать вывод о его достоверности. Этот подход ценен из-за того, что в ходе проверки дополнительно отбираются и подтверждающие документы.

Ограничения методов такого типа включают недоступность учета контекста, многозначности слов и выражений. Для повышения точности проверки исследователю также потребуется запас времени в несколько дней для того, чтобы информация физически могла накопиться в сети.

Методы, основанные на структурированном контенте, предполагают преобразование предложений на естественном языке в структурную форму с помощью извлечения информации. Полученные структурированные данные сравниваются с информацией из базы знаний с историческими и

официальными данными, выводами экспертов, для определения достоверности утверждения. Этот метод имеет и недостатки:

- риск потери информации при ее преобразовании из одной формы в другую;
- недоступность некоторых данных в базах (для некоторых областей они физически могут быть еще не созданы);
- сложность создания точных запросов для базы данных;
- слабость аргументов (их может быть мало или они могут вовсе отсутствовать) для проверки достоверности фактов.

Методы проверки фактов, основанные на контексте, особенно подходят для коротких текстов (например, постов в блогах или социальных сетях). Вместо анализа текстового содержимого, эти методы, как пишет Б. Лакзеи, работают следующим образом: «Контекстные характеристики извлекаются из социальных сетей и включают анализ информации о пользователях, источников публикаций, структуры распространения информации и реакций пользователей» [49].

Некоторые другие подходы, основанные на обработке естественного языка и машинном обучении, могут достигать высокой эффективности благодаря тому, что игнорируют семантику и сосредотачиваются только на грамматической и (или) структурной информации текста. В контексте такого подхода проверка фактов считается выполненной даже без понимания смысла самого утверждения, то есть для проверки достаточно будет только определенной совокупности технических характеристик текста.

После обязательной процедуры фактчекинга необходимо найти и скорректировать все содержательные искажения и проверить общее соответствие генеративной информации контексту всего материала, так как нейросети склонны заменять специфические нюансы и аспекты глобальными обобщениями. Нужно дополнить общие формулировки конкретными примерами, выявить и восполнить информационные пробелы – важные сведения, упущенные нейросетью. Особое внимание здесь нужно уделить

корректировке смысловых акцентов, приведению их в соответствие с исходным журналистским замыслом и фактическим положением дел. Стилистическая и структурная обработка также требует особого внимания, так как машинные тексты часто страдают от однообразия и шаблонности: журналисту необходимо самостоятельно варьировать длину и структуру предложений, добавлять уместные стилистические приемы (метафоры, сравнения, полутона, при необходимости шутки или сарказм, на которые нужно обратить особое внимание из-за их недоступности для нейросети – модель воспроизводит форму шутки, но не наделяет ее содержанием). Лучше отследить в тексте клишированные формулировки вроде «согласно отчету», «некоторые утверждают», «ученые считают», и заменить выражениями на живом человеческом языке. Обязательно нужно проверить и указать для них конкретные источники, проверить соответствие фактов исходной достоверной информации.

Если рассматривать применение нейронных сетей и технологий ИИ в проверке достоверности фактов журналистских материалов, можно отметить ряд особенностей и возможных направлений развития.

Во-первых, признаки для проверки текста должны быть изначально определены верно, так как в зависимости от того, на какие характеристики будет сделан упор нейронной сетью при анализе, будет зависеть результат проверки. Для проверки нужно использовать только информацию высокого уровня (не ограниченные отдельные слова и фразы, но и тематику, тональность, структуру текста, логические связи между предложениями, аргументацию). Во-вторых, лучше использовать мультимодальные подходы и анализировать не только сам текст, но и его структуру, а также особенности самого СМИ, нетекстовые данные публикации (например, используемые эмодзи). В-третьих, для дальнейшего улучшения проверки достоверности нужно развивать аналитику сложных случаев, например, текстов, которые содержат множество сокращений, эмодзи, иносказаний или метафор. В-четвертых, применение нейросетей и других технологий ИИ в журналистике

должно быть ответственным и этичным, учитывать экспертизу со стороны человека, его конкретные выводы по анализу текста – такую вспомогательную роль моделей отмечает и А.Г. Асташкин в своей работе «Нейросети в структуре журналистской профессии» [3].

Несмотря на прогресс технологий, роль журналиста в качестве эксперта остается критически важной – на данном этапе развития нейросетей исключить человека из цикла производства и проверки генеративного контента для СМИ невозможно без потери качества. М.М. Лукина отмечает в том числе необходимость изменения или дополнения устоявшихся этических норм журналиста в связи с применением в работе технологий ИИ: «Освоение редакциями и журналистами технологий искусственного интеллекта ставит перед профессиональным сообществом задачи этического регулирования этих инновационных практик, вплоть до внесения изменений или дополнений в устоявшиеся этические кодексы. Это обусловлено социальной значимостью профессии, ее общественной миссией, особенностями работы журналиста с источниками информации, рекомендательными сервисами, таргетированной дистрибуцией контента, а также возможными эффектами воздействия на аудиторию» [20].

Предложенные в этой главе методика конструирования запросов и адаптация научной классификации промптов к журналистским задачам, могут представлять собой полезные инструменты для профессионального сообщества журналистов, а исследованные алгоритмы фактчекинга и этические принципы взаимодействия с нейросетями позволят сформировать общую методологическую основу для сохранения достоверности контента и профессиональных стандартов работы журналиста в контексте работы с нейронными сетями. Важно подчеркнуть во всех изученных процессах критическую роль журналиста как субъекта, осуществляющего контроль над технологией и результатами ее работы – нейросетевые технологии могут быть эффективно интегрированы в журналистскую практику только при условии

соблюдении баланса между инновационными возможностями и традиционными ценностями профессии.

Выводы главы второй.

Нейросети интегрируются в журналистскую деятельность на самых разных этапах: от последовательного мониторинга новостей с помощью рекуррентных сетей с эхо-состоянием (ESN), до ускоренной обработки информации моделями с обратным распространением ошибки (BPNN), выявления неочевидных связей глубокими нейронными сетями (DNN) и автоматизации визуализации данных отдельными инструментами (например, Data2Vis и NeuroDAVIS). Для эффективной интеграции нейросетей в работу журналиста необходимо использовать специализированную методику конструирования запросов (промптов) и верно классифицировать промпты по типам задач. На основании классификации промптов, описанной в англоязычной научной литературе, предложены авторские примеры каждого типа, адаптированные для журналистики, что существенно упрощает применение методологии в редакционной практике. Эффективный промпт должен иметь четкую структуру с изложением основных идей; универсально применяться для разных задач; быть простым и понятным в формулировках, доступным и техническим специалистам, и журналистам; адаптивно меняться с учетом контекста. Теоретические принципы создания качественных промптов дополнены авторскими практическими примерами «хорошего» и «плохого» запросов, которые показывают конкретные ошибки и способы их устранения при работе с нейросетями в контексте журналистских задач.

Классификация из англоязычного исследования, изначально ориентированного совершенно на другую область знаний, но адаптированная к журналистской практике с приведением авторских примеров, подразделяет промпты на следующие типы: 1) семантика ввода – задающие правила для модели, которые она сохраняет в памяти; 2) настройка вывода – определяющие формат вывода; 3) обнаружение ошибок – направленные на выявление источников данных и фактов; 4) улучшение запросов –

позволяющие преодолеть ограничения модели через уточняющие вопросы; 5) взаимодействие – обеспечивающие многократную генерацию данных по шаблону; 6) контроль контекста – ограничивающие информационное поле работы модели.

Значимым вкладом исследования можно считать каталог промптов, для широкого спектра журналистских задач. Для каждого промпта указана типовая, конкретная и доступная для работы нейросеть. Каталог наглядно демонстрирует, что применение нейросетей в журналистике не ограничивается популярными чат-ботами, но включает множество специализированных узких решений, которые СМИ могут внедрять самостоятельно, без обращения к IT-компаниям. Подход актуален для региональных СМИ с ограниченными человеческими и денежными ресурсами, что обеспечивает им доступ к современным технологиям на уровне крупных медиакорпораций, позволяет создавать конкурентноспособные, вовлекающие материалы, дополнять качественную журналистику визуальными и интерактивными материалами. Центральная роль в использовании нейросетей принадлежит журналисту как субъекту, контролирующему технологию и критически оценивающему результаты ее работы, поскольку генеративные тексты могут быть фактологически ненадежны и содержать галлюцинации.

Генеративные тексты отличаются поверхностностью анализа и шаблонностью языка: даже при точной передаче основных фактов контент часто теряет исходные смысловые акценты и полутона, сглаживает информационный рельеф и позицию автора. Для сохранения достоверности и качества публикаций необходимы стандартизированные алгоритмы фактчекинга и строгое соблюдение этических принципов, формирующих основу для соответствия стандартам журналистики. Новые технологии дают преимущества для медиа, которые становятся более оперативными и конкурентоспособными, но не заменяют глубину человеческого анализа и экспертного мнения.

Заключение

Генеративный дата-контент стал новым этапом цифровой трансформации журналистики: его появление открыло перед медиа инновационные подходы в создании материалов, ускорило процессы производства контента, облегчило анализ больших данных и предложило оригинальные формы визуализации. Проведенное исследование показало, что генеративный контент одновременно создает и множество вызовов, требующих критического осмысления и проработки комплексного подхода для работы с ними – от технических, до этических проблем. Ответственность здесь ложится на журналиста, который должен как самостоятельно переработать генеративный контент (если он используется в работе редакции), так и проследить за соблюдением стандартов профессии, обеспечить прозрачность отношений между редакцией и аудиторией, а также достоверность публикуемых сведений. Редакциям следует рассматривать генеративные технологии как инструменты помощи журналисту, а не его замены: итоговый материал всегда должен проходить через критическое осмысление и профессиональную обработку. Оптимальными сценариями использования нейросетей в журналистской практике могут быть:

- создание первичных набросков для дальнейшей углубленной проработки;
- генерация справочных блоков, заголовков или лидов (с обязательным фактчекингом);
- обработка больших массивов данных для дальнейшего анализа;
- адаптация контента под различные платформы.

Анализируя текущее состояние вопроса можно выделить несколько тенденций на ближайшие годы. Во-первых – это переход от экспериментального использования нейросетей в журналистике к системному. Как отмечалось ранее, статистика последних лет указывает на то, что модели перестали восприниматься как эксперимент, и уже внедряются в

редакционные процессы. Крупные медиахолдинги разрабатывают собственные продукты, средние и малые компании могут использовать общедоступные решения с дообучением под свои нужды. Таких моделей может быть достаточно для неспецифичных и несложных задач – небольшие издания смогут сэкономить, используя технологию, не требующую разработки.

Во-вторых, с ростом внедрения нейросетей в журналистскую практику, очевидно можно будет ожидать разделения обязанностей между специалистами и моделями: рутинные задачи могут быть переданы нейросети (черновые наброски текстовых и визуальных элементов, первичная обработка данных, генерация простых материалов о погоде и спорте, не требующих глубокой аналитики), а журналисты смогут сконцентрироваться на творческих сложных и многоэтапных задачах с глубоким погружением в тему.

В-третьих, повсеместное внедрение технологии требует развития специальных навыков: специалисты, умеющие выгодно применять нейросети в своей работе, могут стать более ценными для изданий. В новых условиях на первый план выходят технологические навыки, а особенно значимым оказывается промпт-инжиниринг – умение формулировать точные запросы к нейросети для более качественных результатов. Возрастает значимость углубленного фактчекинга и критической оценки источников. Анализ информации с учетом контекста, способность интерпретировать данные в широком социальном, историческом и культурном контексте, на текущий момент вообще остаются конкурентным преимуществом журналиста перед нейросетью. Интересно, что революция нейросетей в целом усиливает ценность человеческих качеств – эмоциональности, эмпатии, рефлексии, этичности, что не отменяет необходимости исследовать основы технологии и понимать ее слабые места для эффективного применения в работе.

В-четвертых, очевидно, что на смену общим моделям (ChatGPT, YandexGPT, Gemini, Claude) приходят специализированные решения, которые интегрируются в работу под требования и стандарты конкретного издания.

Существующие инструменты, например, для генерации лидов, или анализа настроения аудитории, позволяют упростить работу издания в конкретной узкой области, освободить ресурсы для развития, особенно в контексте многоканальности современной журналистики.

Усиление внедрения нейросетей может в целом изменить основы медиаиндустрии: СМИ смогут оптимизировать расходы за счет автоматизации и привлечь дополнительные средства благодаря монетизации аналитики поведения пользователей с персонализацией контента на основе машинного обучения.

Особенно важными и открытыми остаются вопросы этики и права в изучаемой области, которые все еще требуют регулирования, и могут лечь в основу следующих исследований. Наилучший путь для развития в этой области могут задать крупные редакции и СМИ, которые дорожат своими отношениями с аудиторией: реакциям нужно стараться больше раскрывать информацию об использовании нейросетей в создании контента, ведь такая прозрачность выступает фундаментом для доверия читателя. В правовом поле на данный момент недостаточно инициатив и контроля: вопросы регулирования нейронных сетей должны учитывать и авторское право, и обеспечение прозрачности в использовании технологии.

Профессия журналиста редакции, работающей с нейросетями, должна эволюционировать в сторону куратора в создании контента, эксперта, который отвечает за достоверность и точность информации. Нельзя не согласиться с исследователями Акуличевой А.Р. [2] и Лукиной М.М. [20] в вопросах ценности человеческих качеств в работе с моделями и соблюдения журналистской этики. Как отмечает Лукина М.М., «Это обусловлено социальной значимостью профессии, ее общественной миссией, особенностями работы журналиста с источниками информации, рекомендательными сервисами, таргетированной дистрибуцией контента, а также возможными эффектами воздействия на аудиторию» [20, с. 690]. Эта область очевидно требует четких и отработанных на практике границ, которые

не будут мешать журналисту в работе, но и смогут очертить нормы в работе с технологией.

Исследование позволяет сделать принципиально важный вывод: генеративные технологии не могут заменить журналиста, они представляют собой лишь инструмент, потенциал которого раскрывается только при условии правильного использования. Успешное применение нейросетей в журналистике возможно только при условии контроля человека: модель берет на себя базовые рутинные задачи, а журналист их критически осмысливает, проверяет, контекстуализирует и комплексно оценивает. Несмотря на скорость развития моделей, даже самый совершенный алгоритм сейчас не способен создать полноценный, действительно качественный и социально значимый материал. Будущее направление развития журналистики, очевидно, будет зависеть от возможности изданий и отдельных специалистов адаптироваться к технологическим изменениям, сохранить ключевые принципы профессионализма, точности и достоверности контента. В качестве ключевых направлений для развития отрасли можно отметить развитие стандартов использования нейросетей (верификацию, переработку, подготовку генеративного контента перед использованием); развитие технологических навыков журналиста; формирование норм этики и права, учитывающих реалии новой медиасреды. Как показывает исследование, игнорирование этих факторов может привести к снижению качества материалов, размыванию профессиональных стандартов и, в конечном итоге, подрыву доверия к СМИ со стороны аудитории.

Список используемой литературы и используемых источников

1. Аксенов С. В., Новосельцев В. Б. Организация и использование нейронных сетей (методы и технологии) / Под общ. ред. В.Б. Новосельцева. Томск: Изд-во НТЛ, 2006. 128 с.
2. Акуличева А. Р. Нейросеть и журналистика: этический вопрос использования умных технологий в СМИ // Молодой ученый. 2023. № 51 (498). С. 115-117.
3. Асташкин А. Г. Нейросети в структуре журналистской профессии // Цифровая журналистика: технологии, смыслы и особенности творческой деятельности: сб. материалов Междунар. науч.-практ. конф. Екатеринбург, 2024. С. 193-196.
4. Бейненсон В. А. Применение генеративных нейросетей в журналистике: проблемы и перспективы // Динамика медиасистем. 2023. Т. 3, № 1. С. 352-359.
5. Белоедова А. В., Романова В. П. Искусственный интеллект в рекламе: технологические и правовые вызовы // Вопросы журналистики, педагогики, языкознания. 2024. Т. 43. №. 3. С. 255-265.
6. Большинство работников СМИ положительно оценили использование ИИ в работе [Электронный ресурс] : Lenta. URL: <https://lenta.ru/news/2024/10/31/bolshinstvo/> (дата обращения: 01.11.2024).
7. Борзова М. С. Роль фактчекинга в современной журналистике // Проблемы массовой коммуникации : материалы Всерос. науч.-практ. конф. Воронеж : Факультет журналистики ВГУ, 2017. Ч. I. С. 5-7.
8. В Краснодаре выпущена первая в России газета, полностью созданная нейросетью [Электронный ресурс] : СППИ ГИПП. URL: <https://www.gipp.ru/news/technology-internet-it/v-krasnodare-vypushchena-pervaya-v-rossii-gazeta-polnostyu-sozdannaya-nejrosetyu> (дата обращения: 09.04.2025).

9. Водопетов С. В., Махмуд К. Н. М. Развитие инструментария искусственного интеллекта в журналистике: история, проблемы и перспективы // Успехи гуманитарных наук. 2022. №. 6. С. 56-61.

10. Выпуск газеты от 28 апреля 2023 № 061 (3731) (2804) [Электронный ресурс] : Газета «РБК». URL: <https://www.rbc.ru/newspaper/2023/04/28> (дата обращения: 16.03.2025).

11. Давыдов С. Г., Замков А. В., Крашенинникова М. А., Лукина М. М. Использование технологий искусственного интеллекта в российских медиа и журналистике // Вестник Московского университета. Серия 10. Журналистика. 2023. № 5 С. 3-21.

12. Жиленков А. А., Силкин А. А., Серебряков М. Ю., Колесова С. В. Сравнительный анализ систем глубокого обучения с подкреплением и систем обучения с учителем // Известия Тульского государственного университета. Технические науки. 2022. №. 10. С. 109-112.

13. Иванов А. Д. Роботизированная журналистика и первые алгоритмы на службе редакций международных СМИ // Знак: проблемное поле медиаобразования. 2015. №. 2 (16). С. 32-40.

14. Иванов А. Д. Транспарентность роботизированной журналистики: как новые технологии угрожают принципам профессии // Вестник Челябинского государственного университета. 2017. №. 8 (404). С. 28-33.

15. Как нейросети повлияли на работу создателей контента, авторов, маркетологов и пиарщиков [Электронный ресурс] : Pressfeed. URL: <https://news.pressfeed.ru/kak-nejroseti-povliyali-na-rabotu-sozdatelej-kontenta-avtorov-marketologov-i-piarshhikov/> (дата обращения: 01.11.2024).

16. Как писать промты, чтобы нейросеть выдавала нужный результат [Электронный ресурс] // Яндекс Образование. URL: <https://education.yandex.ru/journal/kak-pisat-prompty-chtoby-nejroset-vydavala-nuzhnyj-rezultat> (дата обращения: 16.03.2025).

17. Кириллов Д. С., Молостов Д. Д. Способы предварительной обработки данных для прогнозирования с применением нейронных сетей //

Фундаментальная и прикладная наука: актуальные вопросы теории и практики: сборник статей Междунар. научно-практ. конф. 2023. С. 45-47.

18. Кисляков А. Н., Поляков С. В. Иерархические методы кластеризации в задаче поиска аномальных наблюдений на основе групп с нарушенной симметрией // Управленческое консультирование. 2020. №. 5 (137). С. 116-127.

19. Краснодарские известия : газета / гл. ред. В.В. Вербицкий. 2025. 1 апр. № 36 (7115).

20. Лукина М. М., Замков А. В., Крашенинникова М. А., Кульчицкая Д. Ю. Искусственный интеллект в российских медиа и журналистике: к дискуссии об этической кодификации // Вопросы теории и практики журналистики. 2022. Т. 11. № 4. С. 680-694.

21. Написано ИИ на РБК Тренды [Электронный ресурс] : РБК Тренды. URL: https://trends.rbc.ru/trends/tag/by_ai (дата обращения: 09.04.2025).

22. Нейронные сети // Большая российская энциклопедия : [Электронный ресурс] URL: <https://bigenc.ru/c/neironnye-seti-e734b3> (дата обращения: 23.02.2025).

23. Неренц Д. В. Способы применения искусственного интеллекта в журналистской деятельности // MEDIAОбразование: медиа как тотальная повседневность: материалы V Межд. науч. конф. Т. Часть 1. Челябинск: ЧелГУ, 2020. С. 334–341.

24. Облачные технологии для работы с медиаконтентом [Электронный ресурс] : Yandex Cloud. URL: <https://yandex.cloud/ru/isv/media-entertainment> (дата обращения: 01.11.2024).

25. От генерации рекламного поста до озвучивания книг: как в медиаиндустрии используют нейросети [Электронный ресурс] : Yandex Cloud. URL: https://yandex.cloud/ru/blog/posts/2024/05/neuronets-in-media?utm_referrer=https%3A%2F%2Fyandex.ru%2F (дата обращения: 01.11.2024).

26. Попов Т. С. Трансформерные модели в монетизационных стратегиях // Современные инновации: теория и практика современной науки : сб. статей

XLVII Международной заочной научно-практической конференции. 2025. № 1(47). С. 11-15.

27. Соколова М. Е. ChatGPT и промпт-инжиниринг: о перспективах внедрения генеративных нейросетей в науке // Научно-исследовательские исследования. 2024. №. 1. С. 92-109.

28. Шестерина А. М., Шестерин Н. О. О корректности использования термина «искусственный интеллект» в медиасфере // Ученые записки Новгородского государственного университета. 2020. №. 4 (29). С. 1-5.

29. Широков И. Б., Колесова С. В., Кучеренко В. А., Серебряков М. Ю. Анализ технологий глубокого обучения с подкреплением для систем машинного зрения // Известия Тульского государственного университета. Технические науки. 2022. №. 10. С. 118-120.

30. Юлкова В. М., Шиловский Г. В. Искусственные нейронные сети. Глубокое обучение // Мониторинг. Наука и технологии. 2019. №. 4. С. 68-72.

31. Azevedo L. Truth or lie: Automatically fact checking news // Companion Proceedings of the The Web Conference. 2018. С. 807-811.

32. Behera R. K., Jena M., Rath S. K., Misra S. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data // Information Processing & Management. 2021. Т. 58. №. 1. С. 1-29.

33. Blanke T., Venturini T. A network view on reliability: using machine learning to understand how we assess news websites // Journal of Computational Social Science. 2022. № 5. С. 69-88.

34. Chiroma H., Abdullahi U. A., Alarood A. A., Gabralla L. A., Rana N., Shuib L., Hashem I. A. T. Progress on artificial neural networks for big data analytics: a survey // IEEE Access. 2018. Т. 7. С. 70535-70551.

35. Choi R. Y., Coyner A. S., Kalpathy-Cramer J., Chiang M. F., Campbell J. P. Introduction to machine learning, neural networks, and deep learning // Translational vision science & technology. 2020. Т. 9. №. 2. С. 1-12.

36. De-Lima-Santos M.-F., Ceron W. Artificial Intelligence in News Media: Current Perceptions and Future Outlook // Journalism and Media. 2022. № 3(1). C. 13-26.

37. Dering M. L., Tucker C. S. Generative adversarial networks for increasing the veracity of big data // 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017. C. 2595-2602.

38. Dibia V., Demiralp C. Data2vis: Automatic generation of data visualizations using sequence-to-sequence recurrent neural networks // IEEE computer graphics and applications. 2019. T. 39. №. 5. C. 33-46.

39. Fernandes E., Moro S., Cortez P. Data science, machine learning and big data in digital journalism: A survey of state-of-the-art, challenges and opportunities // Expert Systems with Applications. 2023. T. 221. C. 1-20.

40. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. Generative adversarial networks // Communications of the ACM. 2020. T. 63, №. 11. C. 139-144.

41. Hiriyannaiah S., Srinivas A. M. D., Shetty G. K. A computationally intelligent agent for detecting fake news using generative adversarial networks // Hybrid computational intelligence. Academic Press, 2020. C. 69-96.

42. Illia L., Colleoni E., Zyglidopoulos S. Ethical implications of text generation in the age of artificial intelligence // Business Ethics, the Environment & Responsibility. 2022. T. 32, № 1. C. 201-210.

43. Journalism and Technology Trends and Predictions 2025 [Электронный ресурс] : Reuters Institute for the Study of Journalism. URL: <https://genevacom.org/2025/03/11/reuters-journalism-media-and-technology-trends-and-predictions-2025/> (дата обращения: 17.03.2025).

44. Journalism, media, and technology trends and predictions 2023 [Электронный ресурс] : Reuters Institute for the Study of Journalism. URL: <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2023#header--8> (дата обращения: 01.11.2024).

45. Journalism, Media, and Technology Trends and Predictions 2024 [Электронный ресурс] : Reuters Institute for the Study of Journalism. URL:

<https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2024> (дата обращения: 17.03.2025).

46. Khan S. A., Sheikhi G., Opdahl A. L., Rabbi F., Stoppel S., Trattner C., Dang-Nguyen D. T., Visual User-Generated Content Verification in Journalism: An Overview // IEEE Access. 2023. T. 11. C. 1-22.

47. Kokab S. T., Asghar S., Naz S. Transformer-based deep learning models for the sentiment analysis of social media data // Array. 2022. T. 14. C. 100157.

48. Krenker A., Bešter J., Kos A. Introduction to the artificial neural networks // Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech. 2011. C. 1-18.

49. Lakzaei B., Haghiri Chehrehghani M., Bagheri A. Disinformation detection using graph neural networks: a survey // Artificial Intelligence Review. 2024. T. 57, №. 3. C. 52.

50. Lermann H. A., Kimmerle J. Understanding and perception of automated text generation among the public: two surveys with representative samples in Germany // Behavioral Sciences. 2024. T. 14. №. 5. C. 353-377.

51. Li Z., Liu F., Yang W., Peng S. et al. A survey of convolutional neural networks: analysis, applications, and prospects // IEEE transactions on neural networks and learning systems. 2021. T. 33, №. 12. C. 6999-7019.

52. Ma J., Gao W., Mitra P., Kwon S., Jansen B. J., Wong K. F. Detecting rumors from microblogs with recurrent neural networks // Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016. C. 3818-3824.

53. Maitra C., Seal D. B., De R. K. NeuroDAVIS: A neural network model for data visualization // Neurocomputing. 2024. T. 573. C. 1-21.

54. Nti I. K., Quarcoo J. A., Aning J., Fosu G. K. A mini-review of machine learning in big data analytics: Applications, challenges, and prospects // Big Data Mining and Analytics. 2022. T. 5. №. 2. C. 81-97.

55. Opdahl A. L., Tessem B., Dang-Nguyen D. T., Motta E., Setty V., Throndsen E., Tverberg A., Trattner C. Trustworthy journalism through AI // Data & Knowledge Engineering. 2023. T. 146. C. 1-17.

56. Ouchchy L., Coin A., Dubljevic V. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media // AI & SOCIETY. 2020. T 35. C. 927-936.

57. Peng X., Xu Q., Feng Z., Zhao H., Tan L., Zhou Y., Zhang Z., Gong C., Zheng Y. Automatic news generation and fact-checking system based on language processing // arXiv preprint arXiv:2405.10492. 2024. URL: <https://zenodo.org/records/11179548> (дата обращения: 16.03.2025).

58. Potash P., Romanov A., Gronas M., Rumshisky A. Tracking bias in news sources using social media: The Russia-Ukraine maidan crisis of 2013–2014 // Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism. 2017. C. 13-18.

59. Rami A. Zhijiang G., Sejr S.M. The fact extraction and VERification over unstructured and structured information (FEVEROUS) shared task // Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER). Association for Computational Linguistics, Dominican Republic, 2021. C. 1-13.

60. Saha N., Das P., Saha H. N. Authorship attribution of short texts using multi-layer perceptron // International Journal of Applied Pattern Recognition. 2018. T. 5, №. 3. C. 251-259.

61. Shi Y., Sun L. How generative AI is transforming journalism: development, application and ethics // Journalism and Media. 2024. T. 5. №. 2. C 582-594.

62. Simon F. M. Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena [Электронный ресурс]. 2024. URL: https://towcenter.columbia.edu/sites/default/files/content/Tow%20Report_Felix-Simon-AI-in-the-News.pdf (дата обращения: 23.02.2025).

63. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in neural information processing systems. 2017. T. 30. C. 1-15.

64. White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J. A prompt pattern catalog to enhance prompt engineering with

chatgpt // arXiv preprint arXiv:2302.11382. 2023. URL: https://www.researchgate.net/publication/368716245_A_Prompt_Pattern_Catalog_to_Enhance_Prompt_Engineering_with_ChatGPT (дата обращения: 16.03.2025).

65. Yang Y. et al. TI-CNN: Convolutional neural networks for fake news detection [Электронный ресурс] // arXiv. 2018. URL: <https://arxiv.org/abs/1806.00749> (дата обращения: 23.02.2025).

66. Zarkadakis G. How AI «rigged» the US election [Электронный ресурс] // HuffPost: официальный сайт. 2016. 11 ноября. URL: https://www.huffpost.com/entry/how-ai-rigged-the-us-election_b_5825dc9be4b0852d9ec214a9 (дата обращения: 01.11.2024).