Нейронная сеть как зеркало социальных установок: анализ искажений в генеративных изображениях

УДК 316.455 DOI 10.26425/2658-347X-2024-7-4-13-21

Получено 23.10.2024 Доработано после рецензирования 20.12.2024 Принято 25.12.2024

Тертышникова Анастасия Геннадьевна

Канд. социол. наук, ст. преп. каф. социологии

ORCID: 0000-0001-5873-9850

E-mail: tertyshnikova_ag@pfur.ru

Старовойтова Мария Дмитриевна

Стажер-аналитик кафедры социологии

ORCID: 0009-0006-9092-9164

E-mail: mashastar2001@yandex.ru

Павлова Ульяна Олеговна

Магистрант

ORCID: 0000-0003-0437-6438

E-mail: u.pavlova.56@mail.ru

Российский университет дружбы народов имени Патриса Лумумбы, г. Москва, Россия

АННОТАЦИЯ

Работа посвящена рассмотрению нейросетевых генеративных технологий как маркера социальных стереотипов и установок. Цель – апробация генеративного искусственного интеллекта (далее – ИИ) в качестве метода социологического исследования социальных стереотипов, содержащихся в больших данных. Для реализации этой цели первоначально рассмотрены суть ИИ, правовые рамки применения и распространение на данный момент. Результаты апробации показывают, что возвращаемая ИИ информация содержит в себе социальные стереотипы, в первую очередь связанные с гендером и возрастом, а значит, ИИ действительно может использоваться в качестве инструмента для изучения социальных стереотипов. Источник сдвигов в данных в сторону стереотипических образов содержится в информации, на которой обучается ИИ, а также в коде самой программы, то есть в установках и мировоззрении разработчиков, так или иначе влияющих на процесс разработки программ. В подавляющем большинстве случаев (более 80 % от всей сгенерированной информации) ИИ возвращает по запросам, связанным с высокооплачиваемыми профессиями, молодых людей, преимущественно мужчин, что справедливо для гендеризированных и негендеризированных формулировок запроса. ИИ также свойственно приписывать различным социальным группам отдельные черты, например неряшливость и неорганизованность, репрезентовать их в связке с определенным стилем одежды, а также использовать ряд повторяющихся маркеров для обозначения статуса или богатства.

Ключевые слова

Искусственный интеллект, генеративные технологии, стереотипное смещение, дата-анализ, социальные стереотипы, гендерные стереотипы, эксклюзия, предвзятость, профессия

Благодарности. Статья подготовлена в рамках инновационного научно-исследовательского проекта № 100938-0-000 «Использование искусственного интеллекта: перспективы, угрозы, ограничения (на примере представлений студенчества)».

Для цитирования

Тертышникова А.Г., Павлова У.О., Старовойтова М.Д. Нейронная сеть как зеркало социальных установок: анализ искажений в генеративных изображениях//Цифровая социология. 2024. Т. 7. № 4. С. 13–21.

Neural network as a mirror of social attitudes: analysis of distortions in generative images

Received 23.10.2024 Revised 20.12.2024 Accepted 25.12.2024

Anastasiya G. Tertyshnikova

Cand. Sci. (Sociol.), Senior Lecturer at the Sociology Department

ORCID: 0000-0001-5873-9850

E-mail: tertyshnikova_ag@pfur.ru

Maria D. Starovoytova

Trainee Analyst at the Sociology Department

ORCID: 0009-0006-9092-9164

E-mail: mashastar2001@yandex.ru

Ul'yana O. Pavlova

Graduate Student

ORCID: 0000-0003-0437-6438

E-mail: u.pavlova.56@mail.ru

Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow, Russia

ABSTRACT

The article is devoted to the consideration of neural network generative technologies as a marker of social stereotypes and attitudes. The aim of the research – approbation of generative artificial intelligence (hereinafter referred to as AI) as a method of sociological research of social stereotypes contained in big data. To realise this goal, the essence of AI, the legal framework of application and the spread to date are initially considered. The results of approbation show that the information returned by AI contains social stereotypes, primarily related to gender and age, which means that AI can indeed be used as a tool for studying social stereotypes. The source of shifts in data towards stereotypical images is contained in the data on which AI is trained, as well as in the code of the program itself, that is in the attitudes and worldview of developers, which in one way or another influence the process of program development. In most cases (more than 80% of all generated information), the AI returns young people, predominantly men, for queries related to high-paying professions, which is true for both gendered and non-gendered query formulations. AI is also characterised by attributing certain traits to different social groups, such as slovenliness and disorganisation, representing them in connection with a certain style of dress, and using several recurring markers to denote status or wealth.

Keywords

Artificial intelligence, generative technologies, stereotypical bias, date-analysis, social stereotypes, gender stereotypes, exclusion, bias, profession

Acknowledgements. The article has been prepared as part of the innovative research project No. 100938-0-000 "Use of artificial intelligence: prospects, threats, limitations (on the example of students' ideas)".

For citation

Tertyshnikova A.G., Pavlova U.O., Starovoytova M.D. (2024) Neural network as a mirror of social attitudes: analysis of distortions in generative images. *Digital sociology*. Vol. 7, no 4, pp. 13–21. DOI: 10.26425/2658-347X-2024-7-4-13-21

ВВЕДЕНИЕ / INTRODUCTION

В социальных науках актуализируется дискурс вокруг новых возможностей и рисков применения искусственного интеллекта (далее – ИИ). Одними из важных факторов, воспринимаемых как прямая угроза для общества, становятся усиление социальной предвзятости посредством внедрения алгоритмических систем и, как следствие, возникновение новых барьеров для достижения минимально возможного уровня социального неравенства и эксклюзии.

Стереотипы, циркулирующие в обществе, основаны на предшествующем социальном опыте и способе восприятия, обобщающем представления социума в отношении окружающего мира и формирующем социальную реальность [Липпман, 2004]. Часто они могут быть противоречивыми и разрозненными, однако ситуация меняется. Кроме того, ряд исследователей отмечает тенденцию цифровизации сознания людей, при котором мышление становится сильно подверженным искажениям, возникающим в общем доступе и влияющим на механизмы восприятия [Тунда, Тунда, 2024] Наблюдаемая сегодня тенденция в развитии технологий ИИ может усилить предвзятость и закрепить существующие смещения, создавая обобщенный материал на основе больших данных и статистически значимых закономерностей в них.

Генеративный ИИ путем анализа контента, содержащегося в базах данных, создает единую картину мира и проецирует усредненное мировоззрение, вместо того чтобы представлять разнообразие образов: визуальных идентичностей и культур. Группировка индивидов в категории уменьшает сложность реальности до обозримых размеров и формирует ее новое восприятие [Попков, 2002].

Таким образом, подход, лишенный критического анализа, несет в себе угрозу ввиду того, что данные никогда не бывают нейтральными: они содержат информацию о пересекающихся структурах неравенства и социальном порядке. Решающими факторами становятся качество информации, на которой происходит процесс обучения систем, оснащенных ИИ, а также междисциплинарный подход, включающий разработку этических принципов функционирования таких систем и аудит выдаваемых результатов.

ГЕНЕРАТИВНЫЙ ИИ И ЕГО ПРИМЕНЕНИЕ/ GENERATIVE AI AND ITS APPLICATION

Генеративный ИИ – это система, использующая модели глубокого обучения для создания информации (слов, изображений или предложений, видео), по авторству похожей на антропогенную, в ответ на множество сложных и разнообразных стимулов [Lim, Gunasekara, Pallant, Pallant, Pechenkina, 2023] Стремительное развитие и разработка новых моделей машинного обучения объясняется повышением интереса к генеративному ИИ ввиду простоты пользовательского интерфейса большинства платформ, которые позволяют бесплатно или по доступной цене генерировать самый разнообразный контент, вне зависимости от опыта или цели. При этом генерация занимает короткое время – обычно несколько секунд. Следует понимать, что нейронная сеть (далее – нейросеть) – это вид ИИ, а не он сам в прямом его понимании. Генеративный ИИ и нейросети в контексте данного исследования рассматриваются в качестве инструментов, дающих возможность осуществления творческих задач как программными, так и техническими системами [Малышев, Смирнов, 2024].

В связи с этим генеративный ИИ получает широкое применение в различных сферах: в здравоохранении, образовании, социальной политике, сфере технологий и бизнесе[1]. Развитие нейросетей также имеет существенное влияние на рынок труда. С каждым годом все больше компаний внедряют технологии генеративного ИИ для оптимизации бизнес-процессов, улучшения производительности и сокращения издержек. Развитие нейросетей и их внедрение в различные сферы бизнеса даже создает новые профессии, связанные с разработкой и обслуживанием технологий ИИ [Мельникова, Лопаткин, Кожева, 2023].

На государственном уровне также происходят трансформации подходов в отношении генеративного ИИ. 7 апреля 2023 г. Министерство экономического развития РФ на пресс-конференции «ТАСС» определило приоритетными отраслями для внедрения технологий ИИ в Российской Федерации (далее – РФ, Россия) здравоохранение, сельское хозяйство, транспорт, промышленность и строительство[2]. Это доказывает растущий потенциал и перспективы повсеместного внедрения и расширения применения технологий ИИ.

В мировой практике темпы развития и усиления использования технологий ИИ на различных уровнях соизмеримы российским[3]. Согласно

[1] Национальный портал в сфере искусственного интеллекта и применения нейросетей в России. Статистика использования технологий искусственного интеллекта. Режим доступа: https://clck.ru/3DmoY9 (дата обращения: 19.10.2024).

[2] ТАСС. МЭР назвало приоритетные отрасли экономики РФ для внедрения ИИ до 2024 года. Режим доступа: https://tass.ru/ekonomika/17477947 (дата обращения: 19.10.2024).

[3] Бевза Д. Тренды развития искусственного интеллекта и темпы его роста в России и мире максимально сблизились. Режим доступа: https://rg.ru/2023/04/17/trendy-razvitiia-iskusstvennogo-intellekta-i-tempy-ego-rosta-v-rossii-i-mire-maksimalno-sblizilis.html (дата обращения: 19.10.2024).

данным iFORA института статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики», Россия входит в топ-20 стран в области исследований ИИ с долей 2,4 %[4]. Ожидается, что в 2024 г. объем рынка ИИ достигнет 184 млрд долл. США, а среднегодовой темп роста рынка составит 28,46 %, в результате чего к 2030 г. его объем достигнет 826,7 млрд долл. США[5]. За последние 6 лет внедрение ИИ организациями варьировалось на уровне 50 %, а в этом году исследование показало, что уровень внедрения возрос до 72 %[6]. Эти данные свидетельствуют об устойчивом развитии и растущем потенциале технологий ИИ в России и мире.

Несмотря на стремительный рост популярности, согласно закрытым опросам McKinsey, 44 % респондентов заявили, что их организации столкнулись с негативными последствиями использования ИИ-генерации. Чаще всего сообщалось о неточностях в генерации, которые без должного уровня проверки фактов повлияли на их компании. Другой наиболее частой проблемой стало нарушение кибербезопасности и прозрачности данных[7].

На сегодняшний день единого правового регулирования ИИ в России не существует. Единственная принятая мера – реализация законопроекта с целью определить ответственность разработчиков и исключить случаи использования ИИ в мошеннических целях[8]. В отношении этического момента применения ИИ негласно принято решение о том, что это является личной ответственностью компаний, так как не основано на реальных рисках, а лишь несет в себе потенциальные угрозы. Наиболее распространенным кодексом этики в сфере ИИ выступает кодекс, созданный на площадке «Альянса»[9]. Его главная цель – установить общие этические принципы и стандарты поведения, которыми будут руководствоваться в своей деятельности участники отношений в сфере ИИ. Его подписание добровольно и носит исключительно рекомендательный характер для российских компаний и организаций.

Рост популярности и практически полное отсутствие ограничений ведет к тому, что ИИ используется повсеместно для решения широкого спектра рабочих и личных задач. При этом в сфере визуального контента ИИ часто заменяет художников и дизайнеров, поставляя в открытый доступ множество типовых изображений, содержащих обобщенные образы, носящие на себе отпечаток стереотипического «сдвига». Поэтому нам видится важным рассмотреть вопрос репрезентации профессиональных групп.

СТЕРЕОТИПИЧЕСКОЕ СМЕЩЕНИЕ / STEREOTYPICAL BIAS

Предвзятость и стереотипизация – неизбежная реальность для общества. Предубеждения в отношении разных социальных групп позволяют упорядочить знания о мире на индивидуальном уровне. Получая информацию из средств массовой информации или напрямую из окружающей среды, индивид вынужден когнитивно анализировать ее, что невозможно без выявления категорий и закономерностей, которые впоследствии формируют модель познавательного процесса, трансформирующуюся в стереотипы. Предубеждения и установки необходимы для дифференциации информации и ее корректировки в случаях дефицита, переизбытка или искажения [Чвякин, Григорьев, Коноплин, 2023]. Учитывая то, что индивид не может предугадать все варианты развития событий после принятия решения в какой-либо ситуации, часто выбор модели поведения базируется на уже сложившихся в обществе моделях, которые из-за невозможности формального закрепления существуют в виде стереотипов [Питерова, Тетерина, 2016].

Однако стереотипизация, упрощая восприятие, одновременно искажает его. Ограниченное представление о группах ведет к неточностям в суждениях и дискриминации. Формирование объективных знаний требует критической оценки стереотипов, использования многообразных источников информации и осознания влияния когнитивных искажений на процесс познания. Только так можно минимизировать негативное воздействие стереотипного смещения. Ввиду унификации потребляемого цифрового контента возникают массовые когнитивные искажения, которые в итоге становятся обобщенными стереотипами, формирующими ложное или истинное мнение.

[4] Институт статистических исследований и экономики знаний. Пленарная сессия «Искусственный интеллект: тренды, риски, регулирование». Режим доступа: https://issek.hse.ru/announcements/828298542.html (дата обращения: 19.10.2024).

[5] Statista. Artificial intelligence – worldwide. Режим доступа: https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide (дата обращения: 19.10.2024).

[6] QuantumBlack AI. The state of AI in early 2024: gen AI adoption spikes and starts to generate value. Режим доступа: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (дата обращения: 19.10.2024).

[7] Там же.

[8] Афанасьев Н. Интеллект в законе. Режим доступа: https://www.kommersant.ru/doc/6621034 (дата обращения: 19.10.2024).

[9] Альянс в сфере искусственного интеллекта. Кодекс этики в сфере ИИ. Режим доступа: https://ethics.a-ai.ru/ (дата обращения: 19.10.2024).

Одной из общепринятых ценностей является минимизация предвзятости, поэтому повсеместно проводятся попытки сглаживания социальных предубеждений и стереотипов. Их открытое выражение становится социально неприемлемым и вызывает осуждение [Смирнова, 2009]

В сегодняшних реалиях развития технологий ИИ в процесс создания когнитивных установок внедряется ИИ, становясь социальным актором. При этом его влияние на стереотипизацию не ограничено конкретной группой индивидов, а может охватывать широкую общественность и совершенно разные социальные группы, так как на сегодняшний день его использование не лимитировано и сами его технологии доступны без ограничений. Это становится проблемой, так как не только закрепляет уже устоявшиеся стереотипы, но и унифицирует их еще сильнее.

Традиционно социальные установки неоднородны и могут варьироваться от индивида к индивиду. Поэтому происходит естественное общественное регулирование. Этот процесс формирует социальные ценности и нормы, которые определяют пределы допустимого поведения на уровне индивидов, социальных групп или социальных институтов [Баширова, 2000]. Существующие противоречия позволяют избежать радикализма и перегибов в этом вопросе.

Современные модели генеративного ИИ не только становятся универсальным актором, проецирующим свои установки повсеместно, но и персонализировано направляют их пользователям, масштабируя наиболее распространенные стереотипы. Из-за этого происходит смещение и репрезентация социальных групп становится некорректной.

РЕЗУЛЬТАТЫ АНАЛИЗА / ANALYSIS RESULTS

Чтобы рассмотреть, как происходит смещение в генеративном ИИ, был проведен анализ выдачи изображений по текстовому запросу в популярном инструменте ИИ – Kandinsky. Выбор нейросети был обусловлен несколькими параметрами: возможностью создания фотореалистичных изображений, бесплатным доступом, использованием русского языка как основного для обработки запросов.

Временные границы исследования – период с 1 апреля 2024 г. по 30 сентября 2024 г., стартовая точка совпадает с выходом новой модели Kandinsky 2.1.

Теоретическая рамка исследования – подход социологии знания к анализу образа профессий в ИИ с целью рассмотрения стратегий, посредством которых происходит репрезентация профессиональных групп. Появляется новый социальный актор, транслирующий ценности и взгляды сообщества разработчиков на широкую аудиторию, но при этом персонализировано, используя системы генеративного ИИ.

В ходе работы было получено 4 тыс. изображений (по 800 на каждую профессию, 400 из которых по запросу «изобрази представителя профессии [отрасль профессии]» на первом этапе исследования и 400 по запросу «изобрази представителя профессии [отрасль профессии] на рабочем месте» на втором этапе исследования, где [отрасль профессии] – название рассматриваемой профессии) для топ-5 высокооплачиваемых профессий по версии аналитической службы аудиторско-консалтинговой сети FinExpertiza на первую половину 2024 г. (специалист нефтегазовой отрасли, топ-менеджер, программист, финансист, работник авиации)[10]. По каждому изображению было произведено ручное кодирование, чтобы предотвратить смещения, которые потенциально могут быть вызваны дополнительным использованием технических программ.

На первом этапе исследования проводилась категоризация изображений, полученная по запросу «изобрази представителя профессии [отрасль профессии]». В ходе анализа фокус внимания был направлен на пол представителя профессии, предполагаемый возраст и особенность внешнего вида.

В ходе анализа изображений было выявлено, что во всех случаях происходит смещение по гендерному и возрастному признаку. По результатам исследований в каждой из приведенных областей отмечается рост числа женщин. Например, обнаружено, что за первые 6 месяцев 2023 г. число женщин, которые находятся в поисках работы в сфере информационных технологий, выросло на 11 % по сравнению с тем же периодом 2022 г.[11] Также отмечается, что работодатели стали лояльнее относиться к женским резюме, подаваемым на вакансии традиционно «мужских» должностей – их просмотры выросли на 13 %[12]. Таким образом, существующая демография значительно разнообразнее репрезентации в генеративном ИИ.

[10] Finexpertiza. Названы самые высокооплачиваемые профессии в России. Режим доступа: https://finexpertiza.ru/press-service/researches/2024/sam-vysokooplach-prof/ (дата обращения: 19.10.2024).

[11] РБК. Среди россиянок вырос интерес к работе в IT. Режим доступа: https://www.rbc.ru/technology_and_media/31/08/2023/64ef765f9a7947365bd13f45 (дата обращения: 19.10.2024).

[12] Луцюк К. Женщины стали чаще пробовать себя в «мужских» профессиях. Режим доступа: https://lenta.ru/news/2024/07/23/zhenschiny-stali-chasche-probovat-sebya-v-muzhskih-professiyah/ (дата обращения: 19.10.2024).

Таблица 1. **Анализ полученных изображений на первом этапе исследования**

Table 1. Analysis of the obtained images at the first stage of the study

Отрасль/ характеристика	Пол (при изображении людей)	Предполагаемый возраст	Особенности внешнего вида
Специалист нефтегазовой отрасли	Мужской – 100 %	30–50 лет – 100 %	Нейтральное выражение лица, в 13 % изображений присутствует легкая улыбка. Изображены одетыми в специализированную форму
Топ-менеджер	Мужской – 100 %	30–50 лет – 100 %	Базовые эмоции на лице – гнев и презрение. Открытая поза или положение рук, скрещенных на груди. В 100 % случаев одеты в костюм
Программист	Мужской – 94 %. Женский – 6 %	Не удалось установить	В 75,8 % случаев невозможно распознать базовую эмоцию, так как представитель профессии изображен в капюшоне, а взгляд направлен в монитор
Финансист	Мужской – 100 %	30–50 лет – 100 %	Базовые эмоции на лице – гнев и отвращение. Жесты обозначают уверенность (шпилеобразное положение рук; руки, скрещенные в замке). В 100 % случаев одеты в костюм
Сотрудник авиации	Мужской – 100%	30–50 лет – 100 %	На изображении акцент делается на форме и технических сооружениях

Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

Обращая внимание на детали в изображении специалистов нефтегазовой отрасли, стоит отметить, что в абсолютном большинстве сотрудники представлены на фоне буровых установок или нефтяных вышек, лишь 14 % изображений отражают карьерный рост, командную работу или исследовательскую деятельность. В таких представлениях специалисты нефтегазовой отрасли могут быть восприняты не как профессионалы с высоким уровнем квалификации и разными обязанностями, а как физические работники, занятые исключительно в жестких условиях, что может исказить представление о разнообразии задач в этой области.

В сфере топ-менеджмента и управления персоналом лишь в 10 % случаев показаны процессы командной работы, составлений стратегий управления, планирования, которые бы более полно отразили реальную суть их профессиональной деятельности. При этом большинство изображений фокусируется на стереотипах о роли консультанта, а именно как о человеке в костюме, который лишь обеспечивает отчетность, в то время как реальная работа включает элементы креативности, коммуникации и взаимодействия с клиентами.

Образ представителя профессии финансиста также стереотипизирован: сотрудники изображены строго в костюмах, галстуках и в очках, у многих из них есть отличительная особенность – брошь или часы, что не может не подчеркивать высокий статус и достаток.

Программисты в 100 % случаев представлены в темноте, за компьютером, в неформальной одежде. Это может создать впечатление, что данная работа не требует серьезного подхода или усилий, тогда как на самом деле она достаточно напряженная и для нее необходима высокая квалификация.

Примечательно, что 67 % изображений демонстрируют работников авиационной отрасли в идеализированной обстановке, ухоженных и улыбающихся, что формирует ложное представление о реальных условиях труда.

На втором этапе исследования, используя тот же инструмент генеративного ИИ (Kandinsky 2.1), было получено 2 тыс. изображений, репрезентирующих сотрудников из вышеупомянутых отраслей.

Мы сформулировали запрос следующим образом: «Изобрази сотрудников из сферы [название отрасли] на рабочем месте». Использование множественного числа минимизировало смещение, которое потенциально могло быть вызвано особенностями русского языка в названиях профессий, а фраза «на рабочем месте» позволила отразить рабочую среду. Для каждой профессиональной группы было создано 400 изображений.

Они проанализированы по следующим параметрам:

– пространственное расположение – определение местоположения представителей профессиональной группы в кадре (передний план, задний план);

– демографические характеристики– оценка возраста и пола представителей профессии;

– деятельность и рабочая среда – анализ выполняемых задач и особенностей рабочей среды на изображении.

Анализ визуального контента в 5 профессиональных группах выявил значительную гендерную диспропорцию. Мужчины доминируют в подавляющем большинстве изображений (94 %), в то время как женщины представлены незначительно (6 %). Более того, изображения женщин часто ограничиваются возрастной группой до 35 лет.

Что касается возрастного распределения, то изображения преимущественно сосредоточены на мужчинах в возрасте до 40–50 лет (78 %), в то время как представление лиц старше 50 лет практически отсутствует. Эта недопредставленность создает искаженное понимание демографического состава рабочей силы в этих профессиональных областях.

Так, в обоих случаях ввиду чрезмерной генерализации в ИИ наблюдаются несоответствие и стереотипизация образов: генеративный ИИ на основе значимых статистических закономерностей приписывает целой группе определенные характеристики или поведение, увековечивая и масштабируя существующие предрассудки. Результаты выдачи могут быть связаны с тем, что модели обучаются на огромных объемах данных из различных источников, но все эти источники происходят из «необъективного» мира. Кроме того, они могут отражать предубеждения разработчиков.

Отметим также, что в случаях с нейросетью проблему составляет не столько механизм ее работы, сколько исходные данные. Если установленные стереотипы присутствуют в сгенерированных изображениях, они присутствуют и в социальной реальности. Вопрос заключается в том, каким образом возможны расширение стереотипического «репертуара» нейросетей и последующее устранение вредоносных паттернов генерации.

Одним из перспективных методов является использование синтетических данных для дополнения существующих наборов изображений. Такие компании, как Generated Media и Qoves Lab, применяют архитектуры машинного обучения для создания новых портретов, представляющих широкий спектр рас и этнических групп. Это позволяет создавать «по-настоящему справедливые» наборы данных, которые более полно отражают разнообразие человеческого опыта.

Другой подход заключается в применении методов постобработки для уменьшения предвзятости в сгенерированных изображениях. Алгоритмы могут быть разработаны для выявления и удаления нежелательных признаков, таких как цвет кожи, пол или возраст. Например, проект FairFace использует генеративно-состязательные сети (англ. generative adversarial network) для удаления предвзятости из наборов данных изображений лиц, обеспечивая более справедливое представление.

Кроме того, междисциплинарный подход имеет решающее значение для эффективного противодействия предубежденности в генеративном ИИ. Сотрудничество между социологами, специалистами по компьютерным наукам, психологами и другими экспертами позволяет получить более глубокое понимание природы предвзятости и разработать комплексные решения.

***Таблица 2.* Анализ полученных изображений на втором этапе исследования**

Table 2. Analysis of the obtained images at the second stage of the study

Отрасль/характеристика	**Пространственное расположение**	**Демографические характеристики**	**Деятельность и рабочая среда**
Специалист нефтегазовой отрасли	На переднем плане изображен мужчина или группа мужчин (94 %)	Изображены мужчины до 40–50 лет (78 %), женщины до 35 лет (22 %)	На фоне буровых установок и нефтяных вышек (100 %)
Топ-менеджер	На переднем плане изображен мужчина или группа мужчин (94 %)	Изображены мужчины до 40–50 лет (81 %), женщины до 35 лет (19 %)	В офисе за письменной работой (73 %), за компьютером (27 %)
Программист	На переднем плане изображен мужчина или группа мужчин (94 %)	Изображены мужчины до 40–50 лет (92 %), женщины до 35 лет (8 %)	В темном офисе за компьютером (100 %)
Финансист	На переднем плане изображен мужчина или группа мужчин (94 %)	Изображены мужчины до 40–50 лет (88 %), женщины до 35 лет (12 %)	В офисе за письменной работой (73 %), за компьютером (27 %)
Сотрудник авиации	На переднем плане изображен мужчина или группа мужчин (94 %)	Изображены мужчины до 40–50 лет (7%), женщины до 35 лет (29 %)	На фоне самолетов и технических сооружений (100 %)

Составлено авторами по материалам исследования / *Compiled by the authors on the materials of the study*

ЗАКЛЮЧЕНИЕ / CONCLUSION

Таким образом, созданный посредством генеративных технологий контент демонстрирует явный сдвиг репрезентации в полученных образах. В первую очередь он связан с полом и возрастом изображаемых людей. Так, заметно смещение в возрастном распределении: женщины представлены практически исключительно молодыми, а мужчины – не старше среднего возраста. Это говорит о том, что социальный медийный стереотип строится вокруг людей трудоспособного возраста. При этом ценность женщины определяется через ее молодость в большей степени, чем ценность мужчины. В некотором смысле также можно говорить о том, что молодой внешний вид является таким же маркером высокого социального статуса, как упомянутые часы или брошь. Соответственно, исходя из данных, на которых обучалась использованная нейросеть, у концепции высокооплачиваемого работника есть не только конкретный пол и возраст, но и внешний вид.

Стоит учитывать общую гендеризированность русского языка, не позволяющую задать полностью гендерно-нейтральный запрос.

Одновременно тот же самый факт, очевидно, имеет место и в данных, на которых обучалась нейросеть, а значит, этот сдвиг свойственен не столько ввиду запроса, сколько ввиду общей архитектуры исходных данных, которая также сильно гендеризирована.

Изображаемые окружение и образ жизни сотрудников тоже вызывают вопросы: во многих случаях передаваемый нейросетью «медийный образ» содержит неверные представления о предложенных профессиях. Интересно, что чаще всего нейросеть испытывает трудности именно с изображением креативной работы и социальных взаимодействий, которые в «реальном мире» являются ключевыми практически в любой человеческой деятельности.

Соответственно, получаемые с помощью генеративных технологий образы могут служить индикатором, позволяя выделить базовые элементы в стереотипической картине мира, репрезентованной в открытых данных. Стоит помнить, что эти образы не самостоятельны, а являются гиперболизированным отражением существующих в обществе установок, за которыми, как и за разработками технологии, стоят люди-носители элементов социальной реальности.

СПИСОК ЛИТЕРАТУРЫ

Баширова Л.С. Социальная норма и девиация. Психопедагогика в правоохранительных органах. 2000;2(14):97–100.

Липпман У. Общественное мнение. М.: Институт фонда «Общественное мнение»; 2004. 382 с.

Малышев И.О., Смирнов А.А. Обзор современных генеративных нейросетей: отечественная и зарубежная практика. Международный журнал гуманитарных и естественных наук. 2024;1–2(88):168–171. http://doi.org/10.24412/2500-1000-2024-1-2-168-171

Мельникова Д.А., Лопаткин Д.С., Кожева А.А. Искусственный интеллект как способ создания нового контента. Успехи в химии и химической технологии. 2023;1(263(37):43–47.

Питерова А.Ю., Тетерина Е.А. Социальные стереотипы: особенности формирования и изучения. Наука. Общество. Государство. 2016;1(13).

Попков В.Д. Стереотипы и предрассудки: их влияние на процесс межкультурной коммуникации. Журнал социологии и социальной антропологии. 2002;3(5):178–191.

Смирнова Ю.С. Ослабление предубеждений и развитие толерантности как проблема формирования профессионально значимых качеств будущего специалиста. В кн.: Принцип толерантности и его применение в современном образовательном процессе: тезисы 6-й научно-методической конференции, Минск, 24 марта 2009 г. Минск: Белорусский государственный университет; 2009. С. 55–58.

Тунда Е.А., Тунда В.А. Сознание и цифровизация человека. Системный анализ в проектировании и управлении. 2024;1:120–130. http://doi.org/10.18720/SPBPU/2/id24-27

Чвякин В.А., Григорьев Н.Ю., Коноплин Ю.С. Когнитивный смысл социальных стереотипов. Гуманитарий Юга России. 2023;4(12):94–103. https://doi.org/10.18522/2227-8656.2023.4.5

Lim W.M., Gunasekara A., Pallant J.L., Pallant J.I., Pechenkina E. Generative AI and the future of education: Ragnarok or reformation? A paradoxical perspective from management educators. The International Journal of Management Education. 2023;2(21). http://dx.doi.org/10.1016/j.ijme.2023.100790

REFERENCES

Bashirova L.S. Social norm and deviation. Psychopedagogy in law enforcement agencies. 2000;2(14):97–100. (In Russian).

Chviakin V.A., Grigoriev N.Yu., Konoplin Yu.S. Cognitive sense of social stereotypes. Humanitarian of the South of Russia. 2023;4(12):94–103. (In Russian). https://doi.org/10.18522/2227-8656.2023.4.5

Lim W.M., Gunasekara A., Pallant J.L., Pallant J.I., Pechenkina E. Generative AI and the future of education: Ragnarok or reformation? A paradoxical perspective from management educators. The International Journal of Management Education. 2023;2(21). http://dx.doi.org/10.1016/j.ijme.2023.100790

Lippman U. Public opinion. Moscow: Institute of the Public Opinion Foundation; 2004. 382 p. (In Russian).

Malyshev I.O., Smirnov A.A. Review of modern generative neural networks: domestic and foreign practice. International Journal of Humanities and Natural Sciences. 2024;1–2(88):168–171. (In Russian). http://doi.org/10.24412/2500-1000-2024-1-2-168-171

Melnikova D.A., Lopatkin D.S., Kozheva A.A. Artificial intelligence as a way to create new content. Advances in chemistry and chemical technology. 2023;1(263(37):43–47. (In Russian).

Piterova A.Yu., Teterina E.A. Social stereotypes: features of formation and study. Science. Society. State. 2016;1(13). (In Russian).

Popkov V.D. Stereotypes and prejudices: their influence on the process of intercultural communication. Journal of Sociology and Social Anthropology. 2002;3(5):178–191. (In Russian).

Smirnova Y.S. Relaxation of prejudices and development of tolerance as a problem of formation of professionally significant qualities of the future specialist. In: Principle of tolerance and its application in the modern educational process: Proceedings of the 6th Scientific and Methodological Conference, Minsk, March 24, 2009. Minsk: Belarusian State University; 2009. Pp. 55–58. (In Russian).

Tunda E.A., Tunda V.A. Human consciousness and digitalisation. System analysis in design and management. 2024;1:120–130. (In Russian). http://doi.org/10.18720/SPBPU/2/id24-27