



приоритет2030⁺
право для лидерства

АНАЛИТИЧЕСКИЙ ДОКЛАД

*«Современное состояние и перспективы
развития машинного обучения и
искусственного интеллекта для решения
задач судебно-экспертной деятельности»*

*Авторский коллектив: Е.Р. Россинская, Н.С. Неретина,
Е.А. Чубина, А.И. Семикаленова, Г.П. Шамаев, М.С. Чернявская*



Содержание

АННОТАЦИЯ	3
ВВЕДЕНИЕ	5
1. НЕЙРОТЕХНОЛОГИИ В СУДЕБНОЙ ЭКСПЕРТОЛОГИИ И ЭКСПЕРТНОЙ ПРАКТИКЕ: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ	7
2. ОЦЕНКА МЕДИАКОММУНИКАЦИИ В КОНТЕКСТЕ СУДЕБНОЙ ЭКСПЕРТИЗЫ	21
3. ОПЫТ ВЫЯВЛЕНИЯ НЕЙРОТЕХНОЛОГИЙ, ПРИМЕНЯЕМЫХ В МЕДИАКОММУНИКАЦИИ	27
ЗАКЛЮЧЕНИЕ	35

Аннотация

В докладе с позиций теории цифровизации судебно-экспертной деятельности как частной теории судебной экспертологии рассматриваются перспективы внедрения нейронных сетей в судебную экспертизу и возникающие при этом актуальные проблемы. Авторами отмечается изменения в методологии и технологиях разработки экспертных методик в связи с внедрением алгоритмов искусственного интеллекта – нейронных сетей. Обозначены сферы применения нейронных сетей для решения разнообразных задач науки и практической деятельности. На конкретных примерах продемонстрированы возможности использования алгоритмов контролируемого обучения нейросетей в судебно-экспертной практике. Дан подробный анализ причин, по которым использование нейронных сетей в судебной экспертизе может привести к ошибочным заключениям. Особое внимание уделено галлюцинациям нейронных сетей глубокого обучения на больших языковых моделях.

Существует опасность, что эксперт, всецело полагаясь на нейросеть, может дать неправильный вывод, поскольку самообучаемые генеративные нейросети не дают объяснения, почему приняли то или иное решение. Для разработки методик решения типичных экспертных задач на основе нейросетей предлагается создание баз данных (Dataset) для анализа и машинного обучения по различным объектам судебной экспертизы. Для хранения Dataset необходимо организовать репозитории, которые могут содержать наборы данных по родам (видам) судебных экспертиз. Dataset и репозитории обеспечат контроль качества данных и верификацию моделей. В статье обоснована необходимость новых компетенций Data Scientist – специалиста, который формирует инструменты для решения судебно-экспертных задач при внедрении нейросетей и других алгоритмов искусственного интеллекта в судебную экспертологию, а также работающего с ним в контакте инженера по машинному обучению.

Также в докладе представлены отдельные результаты исследования социальных медиа, проведенного лабораторией медиабезопасности кафедры судебных экспертиз. Описаны случаи маскировки деструктивных смыслов и дезориентации потребителя в контенте с помощью жанра дисклеймера, функциональные признаки которого изменяются. Сделан вывод о эффективности использования инструмента медиааналитики для целей экспертного прогнозирования и правового просвещения.

Также рассмотрены имеющиеся инструменты по выявлению контента, созданного при помощи нейротехнологий. Несмотря на то, что в настоящий момент отсутствует технологии и программное обеспечение, которые способны дать однозначные результаты анализа, это лишь говорит о еще более возрастающей роли эксперта, который способен верно интерпретировать полученные результаты и дать корректный вывод.

В заключении работы приводятся следующие выводы и рекомендации:

1. Создании системы искусственного интеллекта судебно-экспертной деятельности в рамках Федеральной программы «Искусственный интеллект» либо инициация принятия новой федеральной программы «Искусственный интеллект в судебной экспертизе».

2. Подготовка Data Scientist – специалистов, которые формируют инструменты для решения судебно-экспертных задач, выбирая ранее созданные алгоритмы и анализируя данные, строят модели и тестируют их.

3. Использование экспертного прогнозирования для превентивных мер в отношении правонарушений в области медиаконтента, созданного при помощи нейротехнологий, а также в целом выявления любого созданного таким образом контента, направленного на дезинформацию или иную деструктивную цель.

4. Обеспечить контроль за привлечением к судебно-экспертной деятельности в отношении медиакоммуникаций только лиц, обладающих необходимой компетенцией, для чего следует внедрить систему сертификацию судебных экспертов или апробацию судебно-экспертных методик.

Введение

По итогам конференции «Путешествие в мир искусственного интеллекта», состоявшейся 23–24 ноября 2022 года в Москве, президент России Владимир Путин дал поручение обеспечить внедрение технологий искусственного интеллекта в каждой отрасли экономики и социальной сферы.

При этом нейротехнологии активно применяются в сфере медиакommunikации. Их внедрение обусловлено автоматизацией ряда процессов. Например, программу анализа больших данных Wordsmith активно использует «Assosiated Press». С помощью специального алгоритма редакция, ранее создававшая усилиями живых специалистов только 300 статей о корпоративных доходах публичных компаний США в квартал, то технологии ИИ позволили изменить это число на 4400 с полным сохранением качества¹. Нейросети позволяют генерировать изображения по запросу, в том числе по относительно короткому описанию (промту), что выводит на принципиально новый уровень дискуссию о статусе документа в СМИ². Отметим, что сейчас на равном технологическом уровне доступно генерирование и видео при помощи нейросетей.

Нейротехнологии позволяют увеличить объем создаваемого контента, особенно в сочетании с различными формами ботов. Данный контент может быть при этом направлен на дезинформацию, мошенничество, на нарушение авторских прав и подключен к совершению иной противоправной деятельности. При том, что контент создается при помощи нейросети, следует отметить, что, как было отмечено выше, его изначальное описание дается человеком, в связи с этим именно является субъектом правонарушения, более

¹ Colford, P.A. (2014). Leap forward in quarterly earnings stories. The Definitive Source. URL: www.blog.ap.org/announcements/a-leap-forward-in-quarterly-earnings-stories (дата обращения 04.05.2024 г.)

² Сурков, А. В. Проблематика использования нейросетей как средства визуальной медиакommunikации / А. В. Сурков // Медиа как фактор адаптации человека к социальным, экономическим и политическим изменениям : сборник материалов Международной научно-практической конференции, Екатеринбург, 20–22 апреля 2023 года / составитель О. Ф. Автохутдинова. – Екатеринбург: Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, 2023. С. 176.

того, сам контент может преподносить то или иное сообщение в завуалированной форме или носить неоднозначный характер, но он создается на базе тех ключевых слов и описаний, которые заданы пользователями. И уже эти слова и описания ввиду необходимости их использования для создания текста, изображения или видео, носят конкретный характер. Также следует учитывать, что нейросетевой контент, вызывающий дезинформацию, может не иметь злонамеренного характера, но быть обусловленным незнанием пользователя об особенностях функционирования генеративных нейросетей и их ограничениях. В случае со СМИ подобные аспекты особенно важны, так как они предоставляют сведения широкому кругу граждан и участвуют в формировании мнений и информационного поля общества в целом.

Ввиду этого важно выявлять сферы внедрения нейротехнологий в медиакommunikацию для сосредоточения усилий по выявлению в них генеративного контента, а также исследования его с точки зрения судебной экспертизы.

1. Нейротехнологии в судебной экспертологии и экспертной практике: проблемы и перспективы

Одними из основных закономерностей судебной экспертологии являются закономерности информационно-компьютерного обеспечения судебно-экспертной деятельности, информатизации и компьютеризации судебной экспертизы, включающие закономерности активного использования информационных компьютерных технологий (процессов, методов поиска, сбора, хранения, обработки, представления, распространения информации), создания и совершенствования информационных систем судебно-экспертного назначения и информационно-телекоммуникационных сетей специального назначения³. Близкой к нашей позиции придерживалась и Т.В. Аверьянова, которая рассматривала использование компьютерных средств и баз данных, а также систем гибридного интеллекта, говоря об информационных процессах в судебно-экспертной деятельности⁴.

Глобальные процессы цифровизации во всех сферах жизни социума дали новый импульс развитию ИТ-технологий в судебно-экспертной деятельности, вследствие замены традиционных аналоговых способов отображения объектов судебных экспертиз (следов, фотоснимков, видео и звукозаписей и пр.) качественно новыми объектами, представленными в цифровом виде – цифровыми следами. Это породило целый ряд теоретических, правовых и организационных проблем, решение которых обусловило формирование новой частной теории экспертологии – теории цифровизации судебно-экспертной деятельности⁵, которая включает два направления:

1. Система информационно-компьютерного обеспечения судебно-экспертной деятельности, как методологическая и технологическая основа

³ Россинская Е.Р., Галяшина Е.И., Зинин А.М. Теория судебной экспертизы (Судебная экспертология) : учебник; под ред. Е. Р. Россинской. – 2-е изд., перераб. и доп. – Москва: Норма : ИНФРА-М, 2020. С. 48–50.

⁴ Аверьянова Т. В. Судебная экспертиза: курс общей теории. – Москва: Норма, 2006. – С. 382–391.

⁵ Россинская Е.Р. Концепция частной теории цифровизации судебно-экспертной деятельности // Вестник экономической безопасности. №5, 2022. С.173-176.

использования IT-технологий в экспертных исследованиях, включающая разработку компьютерных систем анализа изображений; создание баз данных по объектам судебной экспертизы, программных комплексов для выполнения расчетов по известным алгоритмам; автоматизацию сбора и обработки экспериментальных данных; создание гибридных человеко-машинных систем автоматизированного решения экспертных задач⁶.

2. Судебно-экспертное исследования цифровых следов, включая их природу, происхождение, механизм слеодообразования, свойства и признаки, информационную составляющую в зависимости о рода (вида) судебных экспертиз и формы представления: на отдельных носителях; в компьютерных системах и сетях, мобильных коммуникаторах, планшетах, на серверах, в облачных хранилищах; в виде электронных документов; образов, отображаемых в цифровом виде с помощью программных продуктов, образов цифровых следов, отображаемых на бумажных носителях⁷.

В XXI веке на смену экспертным системам, описывающим алгоритм действий по выбору решения в зависимости от конкретных условий, пришло машинное обучение – класс методов автоматического создания прогнозных моделей, представляющих собой инструменты для анализа данных и прогнозирования событий или тенденций, которые дают возможность определить поведение системы и принимать обоснованные решения. Выбор алгоритма машинного обучения зависит от типа решаемой задачи, доступных вычислительных ресурсов и характера данных.

Машинное обучение может быть *контролируемым* – это подход к обучению алгоритмов на основе размеченных данных. Он предполагает, что у имеется набор данных, для каждого элемента которого известно, к какому классу он относится. Затем алгоритм изучает признаки, характерные для каждого класса, и строит модель, которая позволяет определять класс нового

⁶ Теория информационно-компьютерного обеспечения криминалистической деятельности. Монография. /Под ред. Е. Р. Россинской. Москва: Проспект, 2022. С. 165–168.

⁷ Россинская Е.Р. Концепция частной теории цифровизации судебно-экспертной деятельности.

объекта на основе его признаков. Эти алгоритмы уже давно реализованы в криминалистических учетах и судебно-экспертных исследованиях⁸, например, автоматизированная дактилоскопическая идентификационная система АДИС «Папиллон».

При *неконтролируемом* обучении машина учится без явного вмешательства человека. Этот метод часто используется для анализа больших объемов данных, когда невозможно вручную определить все зависимости и закономерности.

Третьим вариантом является обучение *с подкреплением* – используются алгоритмы, которые обучаются на результатах и определяют, какое действие следует предпринять. После каждого действия алгоритм получает отзыв, помогающий определить, является ли сделанный выбор правильным, нейтральным или неправильным. Этот метод удобно использовать в автоматизированных системах, которые должны принимать много мелких решений без вмешательства человека.

В связи с увеличением вычислительных возможностей программно-аппаратных комплексов, в том числе в результате использования графических процессоров и распределенных архитектур вычислительных систем, стало доступным широкое применение машинного обучения на базе множества вычислительных систем, что привело к значительному повышению качества разрабатываемых технологических решений⁹.

За последние годы в науке, технологиях и социальной сфере России в полтора раза расширилось использование решений в области искусственного интеллекта. Об это говорил Президент РФ В.В. Путин на пленарном заседании Международной конференции по искусственному интеллекту и машинному

⁸ Россинская Е.Р., Галяшина Е.И., Зинин А.М. Теория судебной экспертизы (Судебная экспертология) : учебник; под ред. Е. Р. Россинской. – 2-е изд., перераб. и доп. – Москва: Норма : ИНФРА-М, 2020. С. 323–359.

⁹ П. 7 Национальной стратегии развития искусственного интеллекта на период до 2030 года Утверждена Указом Президента Российской Федерации от 10 октября 2019 г. № 490.

обучению Artificial Intelligence Journey 2023 на тему «Революция генеративного ИИ: новые возможности» 24 ноября 2023 г.¹⁰

В русле Национальной стратегии был утвержден Паспорт национального проекта Национальная программа «Цифровая экономика Российской Федерации» со сроком реализации до конца 2024¹¹. К сожалению, проблемы судебно-экспертной деятельности не нашли в нем своего отражения, хотя как верно указывают авторы статьи «Искусственный интеллект в судебной экспертологии», судебная экспертиза «является самостоятельной сферой деятельности федеральных органов исполнительной власти, осуществляющих функции по выработке и реализации государственной политики, нормативно-правовому регулированию, надзору и контролю в области внутренних дел в различных сферах»¹², а судебная экспертология – это отрасль науки и технологии весьма важная для теории и практики судопроизводства.

Среди алгоритмов машинного обучения и искусственного интеллекта можно выделить один из подходов – нейронные сети. Нейронная сеть – это математическая модель, её программное и аппаратное воплощение, построенные по принципу организации и функционирования сетей нервных клеток живого организма.

Сферы применения нейронных сетей для решения задач весьма разнообразны это например:

- распознавание образов (изображений, текста, звуков и других типов данных);

¹⁰ <http://kremlin.ru/events/president/news/72811> (дата обращения 28.12.2023).

¹¹ Паспорт национального проекта Национальная программа «Цифровая экономика Российской Федерации» Утвержден протоколом заседания президиума Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам от 4 июня 2019 г. № 7.

¹² Чеснокова Е.В., Усов А.И., Омелянюк Г.Г., Никулина М.В. Искусственный интеллект в судебной экспертологии. // Теория и практика судебной экспертизы Том 18, № 3, 2023. С. 60-77.

- прогнозирование различных тенденций (например, экономических показателей);
- кластеризация данных, разделения их на группы на основе сходства;
- анализ больших объемов данных, прогнозирование и выявления закономерностей и тенденций;
- создание систем управления беспилотными аппаратами, роботами;
- обучение и оптимизация других алгоритмов, оптимизация процессов и систем.

Современные нейросети работают по следующей схеме. В нейросеть загружается некоторое количество конкретных, необходимых для эксперимента или исследования, данных. Информация передается с помощью искусственных синапсов (соединений) от нейрона к нейрону, от слоя к слою, каждый нейрон может иметь несколько входящих синапсов с данными. Данные, полученные каждым нейроном, представляют собой сумму всех данных, умноженных на коэффициент веса каждого искусственного синапса. Полученные значения формируют выходные сигналы, которые передаются до тех пор, пока информация не достигнет конечного выхода.

В последние годы в России и за рубежом появились научные исследования, направленные на применение нейросетей в криминалистике и судебной экспертизе. Причем, как представляется, в частной теории цифровизации судебно-экспертной деятельности должны быть разработаны методологическая и технологическая основа использования нейронных сетей как нового инструментария экспертных исследований, так и как новых экспертных объектов.

Рассмотрим некоторые примеры применения нейросетей в судебной экспертизе. Одним из вариантов является использование алгоритмов контролируемого обучения нейросетей для автоматизации этапа работы, традиционно считавшегося «ручным» – просмотра экспертом рекомендательных списков, сформированных в результате автоматических поисков АДИС. Нейросеть, внедренная в ЭКЦ МВД по Республике Крым в 9-

ую версию АДИС (AFIS) «Папиλλον» («Папилон-АДИС (AFIS)-9-НейроЭксперт») позволяет: в десятки раз сократить трудозатраты специалистов на просмотр рекомендательных списков; уменьшить вероятность пропуска цели; идентифицировать следы по «родному» кандидату, находящемуся в недоступной для просмотра области результатов поисков, и за счет этого увеличить результативность. Внедрение данного программного обеспечения способствовало раскрытию преступлений прошлых лет на полуострове¹³.

Другим примером использования нейросетей для дактилоскопических исследований является сверточная нейронная сеть на основе глубокого обучения должны, которая используется для обнаружения отпечатков пальцев на изображениях. Она включает в себя два пути: доменно-адверсивный и состязательный. Первый путь помогает сети распознавать объекты независимо от их позиции, масштаба или ориентации. Однако внутренняя среда может измениться, и ранее созданный объект может оказаться недействительным для изменившейся среды. Чтобы адаптироваться к изменившейся среде, требуется собрать огромное количество размеченных выборок данных и заново провести обучение, что трудоемко и требует много времени. Чтобы преодолеть этот недостаток, предлагается второй путь – доменная состязательная нейронная сети (DANN), которой нужны только немаркированные образцы данных из измененной среды для обновления отпечатка пальца до адаптации к изменившейся среде¹⁴.

В судебной экспертизе могут найти применение исследования по использованию нейросетей для компьютерного зрения (computer vision) – области искусственного интеллекта и компьютерных наук, которая занимается разработкой алгоритмов и систем, способных распознавать и

¹³ <https://www.papillon.ru/about/blog/v-9-uju-versiju-adis-afis-papilon-vnedrena-nejroset/> (дата обращения 29.12.2023).

¹⁴ Wei Liu, Zhiqiang Dun. D-Fi: Domain adversarial neural network based CSI fingerprint indoor localization // Journal of Information and Intelligence, Volume 1, Issue 2, July 2023, Pages 104-114.

интерпретировать объекты на изображениях и видео для извлечения из них полезной информации. Нейросети при обучении на больших наборах данных, содержащих изображения с метками, учатся распознавать объекты на изображениях и определять их класс, что затем может использоваться для анализа новых изображений. Компьютерное зрение уже применяется в различных приложениях, включая распознавание объектов, обнаружение и отслеживание объектов, обработка медицинских изображений, изображений автономных транспортных средств и робототехники, систем видеонаблюдения и безопасности, анализе документов и многого другого.

Благодаря развитию инструментов для редактирования видео каждый может легко им манипулировать, не оставляя заметных визуальных следов. Повышение частоты кадров (FRUC), основанное на схемах интерполяции кадров, используется злоумышленниками для подделки видео, например для создания видео с фальшивой частотой кадров. Артефакты, которые остаются в интерполированных кадрах, часто трудно различимы. Нейросеть использует стек последовательных кадров в качестве входных данных и эффективно изучает артефакты интерполяции, используя сетевые блоки для изучения пространственно-временных особенностей¹⁵. Для проверки подлинности цифровых изображений и преодоления сомнений в их редактировании¹⁶, дипфейков¹⁷ также предлагаются многочисленные алгоритмы нейронных сетей. Нейронные сети могут быть использованы для распознавания лиц на фотографиях или видеозаписях, для определения возраста, пола и других характеристик человека.

¹⁵ Minseok Yoon, Seung-Hun Nam, In-Jae Yu, Wonhyuk Ahn, Myung-Joon Kwon, Heung-Kyu Lee. Framerate up-conversion detection based on convolutional neural network for learning spatiotemporal features. // Forensic Science International, Volume 340, November 2022, 111442.

¹⁶ См, например, Yuxue Zhang, Yunfeng Yan, Guorui Feng. Feature compensation network based on non-uniform quantization of channels for digital image global manipulation forensics // Signal Processing: Image Communication, Volume 107, September 2022, 116795.

¹⁷ Samuel Henrique Silva, Mazal Bethany, Alexis Megan Votto, Ian Henry Scarff, Nicole Beebe, Peyman Najafirad. Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models // Forensic Science International: Synergy, Volume 4, 2022, 100217.

В литературе описаны многочисленные исследования документов для выявления признаков подделок, распознавания текстов, добавления новых страниц, идентификации почерка и подписей¹⁸, исследования следов оружия на пулях и гильзах¹⁹.

Интересны разработки использования нейросетей в автотехнических исследованиях. Авторами разработан алгоритм машинного обучения позволяющий решить обратную диагностическую задачу на основе остаточной поврежденной геометрической конфигурации автомобиля (его обломков) и восстановить предаварийное исходное состояние автомобиля²⁰.

Нейронные сети могут служить эффективным инструментарием в компьютерно-технической экспертизе, например, сверточная нейросеть используется для распознавания сетевых атак путем анализа собранного сетевого трафика на уровне пакетов²¹. Авторы отмечают высокую эффективность и производительность предлагаемой ими структуры, когда она классифицирует отдельные классы атак с точностью 99,4%, 99,0% и 90% на каждый набор данных соответственно.

Этот краткий обзор показывает новые возможности в развитии технологий, которые использование нейросетей и других алгоритмов искусственного интеллекта открывают в судебной экспертизе. Заманчивы

¹⁸ Бахтеев Д.В. Особенности распознавания подлога подписи человеком как первичные критерии для разработки системы искусственного интеллекта // Сибирское юридическое обозрение. – 2020. – № 17 (4). С. 514–522.; Wuyang Shan, Yaohua Yi, Ronggang Huang, Yong Xie. Robust contrast enhancement forensics based on convolutional neural networks. // Signal Processing: Image Communication, Volume 71, February 2019, Pages 138-146.

¹⁹ Федоренко В.А., Сорокина К.О., Гиверц П. В. Классификация изображений следов бойков по экземплярам оружия с помощью полносвязной нейронной сети // Известия Саратовского университета. Новая серия. Серия: Экономика. Управление. Право, 2022. Т. 22, вып. 2. С. 184–190; Giverts P., Sorokina K., Fedorenko V. Examination of the possibility to use Siamese networks for the comparison of firing marks // Journal of Forensic Sciences. 2022. 00. P.1-9.

²⁰ Yuxi Xie, C.T. Wu, Boyuan Li, Xuan Hu, Shaofan Li. Computer Methods in Applied Mechanics and Engineering. A feed-forwarded neural network-based variational Bayesian learning approach for forensic analysis of traffic accident. // Computer Methods in Applied Mechanics and Engineering, Volume 397, 1 July 2022, 115148.

²¹ Sonam Bhardwaj, Mayank Dave. Enhanced neural network-based attack investigation framework for network forensics: Identification, detection, and analysis of the attack. // Computers & Security, Volume 135, December 2023, 103521.

перспективы быстрого получения новой доказательственной информации, расширения круга решаемых диагностических и идентификационных задач при минимизации затрат и ресурсов.

В XXI веке вычислительные мощности, позволяющие оперировать большими массивами данных (BigData) позволяют создавать большие языковые модели глубокого обучения, которые предварительно обучены на огромных объемах данных²². Лежащий в основе трансформер – это набор нейронных сетей, каждая из которых состоит из кодера и декодера с возможностью самонаблюдения. Кодер и декодер извлекают значения из последовательности текста и понимают отношения между имеющимися в ней словами и фразами. Глубокие нейронные сети или сети глубокого обучения имеют несколько скрытых слоев с миллионами связанных друг с другом искусственных нейронов. Теоретически глубокие нейронные сети могут сопоставлять любой тип ввода с любым типом вывода. Они способны осуществлять самообучение, усваивать знания. Однако стоит учитывать, что им требуется гораздо более сложное обучение, чем другим методам машинного обучения. Таким узлам нужны миллионы примеров обучающих данных, а не сотни или тысячи, как в случае с простыми нейронными сетями. Но в отличие от описанных нами выше рекуррентных нейронных сетей, которые последовательно обрабатывают входные данные, трансформеры обрабатывают целые последовательности параллельно, что значительно сокращает время обучения. Архитектура нейронной сети трансформера позволяет использовать очень большие модели, часто с сотнями миллиардов параметров. Такие сверхбольшие модели могут получать огромные объемы данных из Интернета и, например, из таких источников, как индекс Common Crawl – более 50 миллиардов веб-страниц или Википедии – около 57 миллионов страниц. Большое количество исследований проведенных за

²² Бахтизин А. Р. , Брагин А. В. , Макаров В. Л. Большие языковые модели четвёртого поколения как новый инструмент в научной работе // Искусственные общества, 2023. – Т. 18, Выпуск 1. <https://artsoc.jes.su/s207751800025046-9-1> (дата посещения 02.01.2024).

последнее время показало, что предварительно обученные (pre-trained models – РТМ) модели нейронных сетей обработки естественного языка позволяют избежать обучения новой модели с нуля²³.

Последнее четвёртое поколение РТМ моделей такие как GPT-4, имеют значительно большее количество параметров по сравнению с их предшественниками, что позволяет им улавливать более широкий спектр языковых моделей. Они проявляют лучшее понимание контекста, грамматики и семантики, что приводит к более связному и контекстуально точному созданию текста, что позволяет с большой производительностью эффективно настраивать РТМ модели для решения широкого спектра задач, таких как перевод, реферирование и ответы на вопросы из различных областей знания.

Внедрение нейросетей в различные сферы науки и практической деятельности для междисциплинарных исследований, общедоступность существующих алгоритмов и лавинообразное нарастание разработки все новых и новых затрагивает и сферу производства судебных экспертиз. Этому может способствовать соблазн упрощения поиска, быстроты получения результатов, возможности получения новых данных. Пока, как показывают приведенные выше примеры, используются нейросети с контролируемым обучением и с проверкой на традиционных признаках объектов, выявленных вручную.

Мы полагаем, что использование нейросетевых алгоритмов в судебно-экспертных технологиях должно базироваться на методологии теории цифровизации судебно-экспертной деятельности, поскольку при этом открываются новые возможности, но и предвидятся серьёзные ограничения применения подобного инструментария.

Для разработки технологии экспертного исследования с применением нейросети необходима определить следующее:

²³ Qiu X. et al. Pre-trained models for natural language processing: A survey // Science China Technological Sciences. 2020, vol. 63, №. 10.

Задача, которую надо решить	
модель, которая может эту решить задачу,	какую точность считать приемлемой для модели, если ее применять на практике
данные на которых можно обучиться (провести эксперименты)	достаточно ли данных, чтобы при верификации модели сказать, что модель можно применять для решения задачи, результаты работы модели необходимо модерировать
данные на которых эта модель верифицируется	
математический аппарат, который показывает статистическую значимость того или иного объема данных.	работа с моделями – это работа с вероятностями, результат – это всего лишь его очень высокая вероятность.

Причины, по которым использование нейронных сетей в судебной экспертизе может привести к ошибочным заключениям:

- некорректность или недостаточная точность моделей, не подходящих для использования в судебной экспертизе;
- недостаток данных или их некорректная разметка. Нейронные сети требуют большого количества данных для обучения. Если данных недостаточно или они некачественные, то сеть может обучиться неправильным зависимостям и делать ошибки;
- предвзятость данных, поскольку нейросети обучаются на тех данных, которые им предоставляют. Если данные содержат предубеждения, то и нейросеть будет их воспроизводить. Например, если в обучающей выборке большинство дактилоскопических отпечатков были оставлены мужчинами, то нейросеть может начать ассоциировать преступность с мужским полом;
- неправильный алгоритм обучения нейросети, который может быть связан с выбором архитектуры сети, обучающих данных, использованием неподходящего оптимизатора, выбором функции ошибки, неправильной настройкой гиперпараметров, отсутствием проверки на переобучение и т.д.;

- неправильная интерпретация результатов, поскольку нейросети не дают объяснения, почему они приняли то или иное решение. Эксперты, полагаясь всецело на нейросеть, не понимают, на чем основаны ее выводы;
- подмена понятий, работа с моделями – это всегда работа с вероятностями, но эксперты забывают о том, что категорический вывод нейросети – это всего лишь очень высокая вероятность результата. Истина на 90% сознании людей подменяется на безусловную истину.

Возможны и некоторые другие причины ошибок, например галлюцинации нейросети, когда нейросеть «видит» или «слышит» то, чего на самом деле нет, выдает за факты придуманную ею же информацию, причем очень убедительно. Галлюцинации опасны тем, что искусственный интеллект упрямо настаивает на своём. Если уличить нейросеть в ложности ответа, она продолжает отстаивать свою правоту.

Галлюцинации нейросетей могут представлять собой:

- создание реалистичных изображений, которые кажутся настоящими, но на самом деле не существуют, например вставление в видео несуществующих людей, создавая иллюзию, что они были там изначально;
- неправильное распознавание объектов или представление несуществующих объектов, например, беспилотный автомобиль может увидеть несуществующие препятствия или пешехода на пустой дороге; робот думает, что он держит инструмент, когда в действительности инструмент отсутствует;
- восприятие прочитанных или произнесенных слов или фраз, когда на самом деле их не было, что может привести к неправильному пониманию или интерпретации текста;
- ошибочно интерпретировать результаты инструментальных исследований и анализов, например, симптомы для нейросетей, обученных на медицинских данных;
- выдача нерелевантных или нежелательных рекомендаций, вызванных некорректной обработкой входных данных.

Приведем несколько конкретных примеров. Во время одного эксперимента чат-бот ChatGPT написал текст на медицинскую тематику. В нём встретился неверный факт, и исследователи попросили чат-бот подтвердить информацию научными источниками. Модель выдала названия пяти работ, но ни одной из них не существовало в реальности. При этом нейросеть дала реальные ID исследований, но все они принадлежали статьям, которые не имели отношения к теме²⁴.

Примером галлюцинации нейросети в области юриспруденции является первый иск в федеральный суд США из-за галлюцинаций нейросети к компании OpenAI, который подал радиоведущий из штата Джорджия (США) Марк Уолтерс из-за того, что ChatGPT обвинил его в мошенничестве и растрате средств некоммерческой организации²⁵. Ответ ChatGPT на запрос, сгенерированный журналистом Фредом Рилом гласил, то Уолтерс незаконно присвоил средства некоммерческой организации по защите прав на оружие на сумму более 5 млн долларов. Журналист решил перепроверить всю информацию и убедился в том, что никаких подобных дел против Уолтерса не существует.

Галлюцинации нейронных сетей зависят от их архитектуры и того, как они обучаются. Например, на вопрос, какое число больше –999 934 или –32 323 232 332? Нейронная сеть от отвечает, что –32 323 232 332 больше, поскольку в обучающий алгоритм было заложено, чем больше цифр в числе, тем оно больше²⁶.

Резюмируя вышеизложенное, полагаем, что в настоящее время нейронные сети и другие алгоритмы искусственного интеллекта следует внедрять в судебно-экспертные технологии с большой осторожностью.

Для разработки экспертных методик на основе нейросетей необходимо создание баз данных (Dataset) для анализа и машинного обучения –

²⁴ <https://vc.ru/future/918465-opasnye-fantazii-chto-takoe-gallyucinacii-ii-i-kak-ih-obnaruzhit> (дата обращения 02.01.2024).

²⁵ <https://www.pvsm.ru/news/385293> (дата обращения 02.01.2024).

²⁶ <https://dzen.ru/a/ZFDU0hv67xCPOATH> (дата обращения 02.01.2024).

совокупностей данных по различным объектам судебной экспертизы, для решения типичных экспертных задач. Это могут быть наборы верифицированных структурированных (например, таблицы), или неструктурированных (например, текстовые документы или изображения) данных. Для их хранения необходимо организовать репозитории (хранилища) наборов данных по родам (видам) судебных экспертиз – централизованные места для хранения с регламентированной системой доступа, которые можно использовать как основу для создания судебно-экспертных методик. Dataset и репозитории для хранения и систематизации данных обеспечат возможность сравнения результатов различных исследований, контроль качества данных и верификацию (воспроизводимость) моделей. Если модель была обучена год назад, то факты которые имели место за год ей недоступны для оперирования, поэтому процесс обучения и верификации должен осуществляться на постоянной основе с разумной периодичностью.

Для создания судебно-экспертных репозиториев можно использовать уже существующие наиболее известные платформы²⁷:

- Kaggle для соревнований по машинному обучению и анализу данных, где участники могут загружать свои наборы данных и делиться ими с другими пользователями.
- UCI Machine Learning Repository – содержит наборы данных для различных задач машинного обучения.
- OpenML – репозиторий для машинного обучения, который предоставляет доступ к наборам данных и моделям для научных исследований.
- Figshare – хранилище для научных данных, включая наборы данных, алгоритмы и программное обеспечение
- Многие другие.

²⁷ <https://paperswithcode.com/datasets> (дата обращения 02.01.2024).

Создание Dataset и репозиторий по родам (видам) объектов возможно только после согласования методических подходов к исследованию объектов судебных экспертиз, разработанных различными ведомствами.

Изменения в методологии и технологиях судебно-экспертных исследований требует новых экспертных компетенций. Как и в других отраслях науки внедрение нейросетей и других алгоритмов искусственного интеллекта в судебной экспертологии обуславливает появление Data Scientist – специалиста, который формирует инструменты для решения судебно-экспертных задач, выбирая ранее созданные алгоритмы и анализируя данные, строит модели и тестирует их. Он работает на стыке статистики, машинного обучения, программирования и специальных знаний в данном роде (виде) судебных экспертиз. С ним в контакте работает инженер по машинному обучению, в обязанности которого входит автоматизировать работу моделей, следить, чтобы они работали качественно, устранять ошибки. Если точность модели падает, инженер выявляет причины и переобучает алгоритм²⁸. Все модели на чем-то учатся, а значит отстают от актуального состояния ноосферы.

В заключении необходимо подчеркнуть, что целесообразной представляется постановка вопроса о создании системы искусственного интеллекта судебно-экспертной деятельности в рамках Федеральной программы «Искусственный интеллект» либо инициация принятия новой федеральной программы «Искусственный интеллект в судебной экспертизе», как это предлагают авторы РФЦСЭ при Минюсте России²⁹.

2. Оценка медиакommunikации в контексте судебной экспертизы

²⁸ Humberto Ferreira, Pedro Ruivo, Carolina Reisb. How do data scientists and managers influence machine learning value creation. // *Procedia Computer Science*, Volume 181, 2021, Pages 757-764.

²⁹ Чеснокова Е.В., Усов А.И., Омелянюк Г.Г., Никулина М.В. Искусственный интеллект в судебной экспертологии. // *Теория и практика судебной экспертизы*, 2023. Т. 18. № 3. С. 60-77.

Развитие нейротехнологий требует оперативной реакции на способы их противоправного применения. При этом прогресс в данной сфере будет происходить именно от правонарушителей, которые ищут и внедряют новые способы совершения преступлений, правоприменители же расследуют результаты их деяний. Поэтому экспертная наука всегда уделяла внимание учению об экспертном прогнозировании³⁰. Сегодня, когда мир так стремительно меняется, значение прогнозирования многократно возрастает, но столь же многократно усложняется сам процесс его проведения, в т.ч. определение исходных данных для прогнозирования, особенно если речь идет об экспертном прогнозировании в рамках судебных лингвистических экспертиз.

В современном мире медийной коммуникации происходит самая настоящая *дезорIENTATION потребителя в контенте*. Для иллюстрации: информационные (в т.ч. новостные) платформы, чтобы монетизироваться и конкурировать с эмоционально заряженным маркетинговым контентом, делают новостные статьи похожими на рекламу, привлекая кликбейтами, заголовками-приманками (*Вы не поверите, что...; Вот почему...; Все знают, что* и пр.), реклама же, напротив, работает «под прикрытием». Появился скрытый маркетинг, внедренный маркетинг, маркетинг слухов, бренд-маркетинг, здесь реализуется не прямое воздействие, цель которого – сделать так, чтобы адресат вообще не заподозрил призыва купить и не увидел связи с компанией, создавшей или спонсировавшей материал, напротив, восхитившись необычной историей или забавным фактом, начал делиться с друзьями полученной информацией³¹. Граница между непредвзятым контентом и маркетинговой коммуникацией размывается, как размывается

³⁰ Белкин Р.С. Курс криминалистики. В 3т. Т.2. Частные криминалистические теории. – М., 1997. С.420-440; Аверьянова Т.В. Судебная экспертиза. Курс общей теории. М., 2007. – С.43; Россинская Е.Р., Галяшина Е.И., Зинин А.М. Теория судебной экспертизы (Судебная экспертология): учебник. – 2-е изд., перераб. и доп. – М.: Норма: Инфра-М, 2016. С.75-77.

³¹ Эйнштейн, М. Реклама под прикрытием. Нативная реклама, контент-маркетинг и тайный мир продвижения в интернете / Эйнштейн М. - М.: Альпина Паблишер, 2017. - 301 с. Электронный ресурс. Код доступа: URL: <https://znanium.com/catalog/product/1003009> (дата обращения: 28.11.2023).

граница между стилями и жанрами.

Кроме того, речевая коммуникация стремительно усложняется технологиями дополненной реальности. Информационное сетевое пространство заполняется пользовательским контентом, когда «каждый пользователь также является реальным соучастником процесса производства публикуемой и распространяемой в Сети информации»³². Развертывающийся процесс коммуникационного взаимодействия предполагает в каждом его участнике соактора, причем это может быть не только единичный, но и коллективный субъект, а может и вовсе «гибрид человека и машины», как нередко определяют искусственный интеллект.

Формирование системы методических рекомендаций для решения задач, только появляющихся в сфере судебной лингвистической экспертизы, невозможно без должного уровня экспертного прогнозирования, в противном случае нельзя говорить о работе на опережение, о предупреждении деструктивных проявлений. Экспертное прогнозирование является одной из задач Лаборатории медиабезопасности, созданной на кафедре судебных экспертиз Университета имени О.Е. Кутафина (МГЮА).

Лаборатория изучает возможности использования технологий автоматизированного анализа текстов для нужд экспертного прогнозирования. Хорошие результаты дает медиааналитика социальных медиа.

Социальные медиа – это публичный контент, размещенный пользователями на различных интернет-площадках. Важно то, что это именно *пользовательский, не редакторский* контент.

Аналитика соцмедиа позволила выявить, например, какие жанры используются для *маскировки деструктивного контента*. Неожиданно для автора статьи это стал жанр *дисклеймера*.

Дисклеймер (от англ. disclaimer) – заявление об отказе от

³² Дружинин А.М., Гуров Ф.Н. Цифровизация медиапроизводства: адаптация, целеполагание, аксиологические аспекты. Гуманитарный вестник, 2022, вып. 2. Электронный ресурс. Код доступа: <http://dx.doi.org/10.18698/2306-8477-2022-2-772> (дата выхода 18.10.2023).

ответственности за возможные негативные последствия («Мы слагаем с себя ответственность, и не говорите потом, что не были предупреждены»)³³. В странах англосаксонской правовой системы используется широко и повседневно, в разных сферах, в т.ч. в сфере фанфикшн (разг. фанфик, фик, фф) (от англ. fan fiction – фанатская литература, фанатская проза), так называют литературный жанр, в котором авторы пишут произведения, основанные на персонажах, сюжетах или мирах, созданных другими авторами³⁴. Наличие дисклеймера – устоявшееся правило сетевого этикета, обычай в интернет-среде, где распространяются произведения такого рода. Автор вторичного произведения указывает, кому принадлежат права на оригинал³⁵.

Несмотря на то, что в России, как в целом в странах романо-германской правовой системы, дисклеймер не «блокирует» действие общеправового принципа «незнание закона не освобождает от ответственности за его нарушение», следовательно, не имеет юридической силы и не исключает ответственности сделавшего его, этот жанр тем не менее знаком многим, поскольку применяется в книгоиздательстве (*«Издательство не несет ответственности за правильность приведенных в Произведении сведений, их интерпретацию и комментарии»*), СМИ (*«Редакция не несёт ответственности за содержание авторских материалов»*), кинематографе (*«Все персонажи являются вымышленными, и любое совпадение с реально живущими или жившими людьми случайно»*), иных сферах.

Анализ соцмедиа показал, что дисклеймеры авторами используются активно, часто с целью *создания иллюзии легитимации коммуникации, иллюзии законности и правомерности информации по содержанию и по форме*. Возникла мода на дисклеймер, как следствие появились оплачиваемые услуги

³³ Лисаченко А.В., Маштакова А.Р. Фанаты или пираты?// Российский юридический журнал, 2016, №3. Электронный ресурс. // СПС «КонсультантПлюс».

³⁴ Словари и энциклопедии на Академике. Электронный ресурс. Код доступа: <https://dic.academic.ru/dic.nsf/ruwiki/1161387>

³⁵ Лисаченко А.В., Маштакова А.Р. Фанаты или пираты?// Российский юридический журнал, 2016, №3. Электронный ресурс. // СПС «КонсультантПлюс».

по созданию дисклеймеров. Появились своеобразные клише, которые используются в дисклеймерах тех авторов, которые размещают деструктивный контент, например: *«Я не имею какой-либо специальной подготовки по проведению экспертиз, направленных на выявление в текстах признаков экстремистской деятельности, следственную и судебную практику по этой категории дел не отслеживаю, не обобщаю и не анализирую. Таким образом, Я объективно лишена возможности самостоятельно оценить то или иное своё/чужое высказывание на предмет наличия или отсутствия в нём признаков экстремизма»; «В случае, если на Моей странице будут выявлены признаки экстремизма, прошу незамедлительно уведомить об этом Меня для немедленного удаления данного контента (текст, фото, видео, аудиофайлов), ПРЕДОСТАВИВ надлежащим образом оформленное и заверенное ЭКСПЕРТНОЕ ЗАКЛЮЧЕНИЕ в отношении конкретной публикации (высказывания, действия)»* и др.

Дезориентация потребителя информации в контенте явление, на наш взгляд, *опасное*, поэтому тенденция к увеличению количества таких дисклеймеров, а это видно из аналитики, видится нам тревожной.

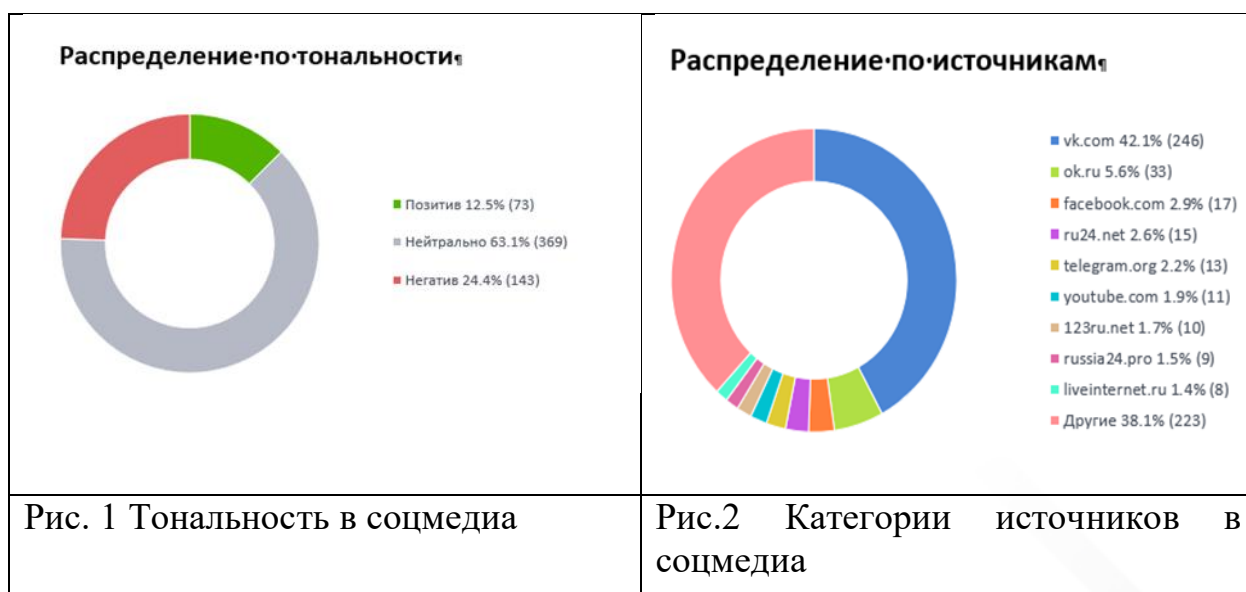
Более того, анализ показал, что функциональные признаки дисклеймера как текста, содержащего отказ от ответственности, не ограничиваются информированием. На первый план выходит привлечение внимания к транслируемым смыслам, демонстрация собственной позиции в соотношении свой-чужой.

Аналитика соцмедиа позволила выявить иные жанры, которые используются для маскировки деструктивного контента. Дезориентация в контенте – угроза не только когнитивной сфере отдельной личности, но и когнитивному суверенитету страны, под которым понимается «наличие собственных аналитических способностей, которые позволяют отделять то,

что нужно человеку, обществу от того, что ему навязывается»³⁶.

Разрешение этих вопросов возможно при помощи лингвистических исследований, основанных на научных методиках и проводимых компетентными экспертами. Однако отсутствие обязательной сертификации экспертов и апробации методик лишает возможности оценки лиц, привлекаемых в качестве специалистов и судебных экспертов. В связи с этим даже судебные лингвистические экспертизы проводятся без надлежащей научно-методической базы.

Судебная лингвистическая экспертиза как объект в СМИ продолжает оцениваться отрицательно, примечательно, что объем таких оценок значительно вырос³⁷. Негативные оценки почти в 2 раза превышают позитивные оценки (Рис.1).



Если говорить о распределении публикаций по источникам, то применительно к соцмедиа проблематика, связанная с судебными лингвистическими экспертизами, актуализирована максимально с большим отрывом в Вконтакте (42,1%), меньшее количество приходится на соцсети,

³⁶ Игнатьева О.А. Формирование когнитивного и технологического суверенитета в контексте концепции островизации // Политические вызовы и политический диалог в условиях глобальной турбулентности. Материалы Всероссийской конференции РАПН с международным участием. Москва, 2022. С. 182-183.

³⁷ В процессе мониторинга соцсетей с помощью системы Brand Analytics по отчету за неделю 30.07.2023 – 09.08.2023 получили 109 сообщений.

блоги, микроблоги.

Проблематика, связанная с судебными лингвистическими экспертизами, в большей мере интересует мужчин, они чаще становятся авторами сообщений, это характерно для соцмедиа. Для понимания того, с какой аудиторией необходимо работать, важны данные возраста авторов, показательно, что это, как правило, люди старше 55 лет (на них приходится 36% сообщений в соцсетях), вторая по численности группа – 35–44 (на них приходится 28% сообщений в соцсетях). Анализ соцмедиа дает возможность определения авторов, причем, что важно, с маркировкой тональности их сообщений.

Итак, необходимо сделать вывод о том, что при распространении нейросетевых технологий, их повсеместного внедрения в жизнь, одновременно происходит усиление недоверия к человеку в качестве субъекта, который может дать объективную оценку контенту в медиакommunikации, его содержанию и направленности. Причем данная позиция также транслируется в медиа. Стоит отметить, что отчасти причиной этого является слабое информирование о сущности лингвистических проверок, а также недостаточное регулирование экспертной деятельности, что позволяет проникать в нее лицам, называющих себя специалистами и экспертами, не обладая при этом нужными навыками и компетенцией. Исправить подобную ситуацию может сертификация экспертов, которая должна быть основана на знании и умении применять апробированные и верифицированные методики. С учетом увеличенного времени на подготовку экспертов при сертификации, необходимо для сохранения оперативности в реагировании на правонарушения, совершенные с применением нейросетевых технологий, развивать экспертное прогнозирование в этой сфере и превентивно изучать возможности нейросетей.

3. Опыт выявления нейротехнологий, применяемых в медиакommunikации

Одновременно с тенденциями, указанными выше, стоит указать на то,

что увеличивается объем контента, созданного при помощи нейросетевых технологий. По прогнозам аналитиков Института Reuters Оксфордского университета, чтобы удовлетворить новые потребности аудитории, медиакомпаниям скорее всего сделают ставку на технологии искусственного интеллекта, которые создают персонализированные рекомендации для дистрибуции, автоматизируют редакционные процессы, автоматически генерируют тексты и выполняют другие операции, которые повышают эффективность медиапроизводства. Фактически системы искусственного интеллекта претендуют на то, чтобы стать активным субъектом конкуренции в медиа³⁸.

По мнению Хомского, все наиболее известные системы ChatGPT (OpenAI), Бард (Гугл) и Сидни (Майкрософт), которые характеризуются как наиболее развитые и способные действительно имитировать некоторые аспекты человеческой деятельности, просто опираются на поиск шаблонов в огромных массивах входных данных. В результате такого «обучения» они становятся все более опытными в генерации внешне подобных человеческому языку и суждениям статистически значимых выходных данных.

Генеративные системы, по мнению Хомского, можно назвать первыми признаками становления так называемого «общего ИИ», т. е. качественно сопоставимого с человеческим интеллектом, имитирующего признаки инсайта, креативности и др. При этом Хомский оценил горизонт наступления этого события как весьма отдаленную перспективу, на фоне которой громкие заголовки и солидные венчурные инвестиции выглядят как непомерно преувеличенные. По его мнению, если такие механистические системы самообучения, как ChatGPT, продолжат доминирование на рынке интеллектуальных инноваций, то «озарение понимания» так никогда и не наступит. Как бы ни были полезны подобные системы в решении специальных задач узких предметных областей, принцип их действия принципиально

³⁸ См.: URL: <https://ict.moscow/static/pdf/files/ai-media-2307.pdf> (дата обращения 13.09.2024)

отличен от человеческого употребления языка и мышления³⁹.

Модели искусственного интеллекта, применяемые в медиа, коммуникациях и журналистике:

- Модели нейронных сетей (ANN = artificial neural networks);
- Машинное обучение (ML=machine learning);
- Глубокое обучение (DL=deep learning);
- Модели естественного языка, или технологии обработки естественного языка (NLP=natural language processing), которые включают генерацию текстов на естественных языках (NLG=natural language generation) + понимание текстов (NLU= natural language understanding);
- Системы машинного перевода (MT = machine translation). Данные модели искусственного интеллекта сегодня применяются практически во всех медийных процессах, в том числе в управлении и продюсировании, профессиональной работе медиакоммуникаторов, журналистов, копирайтеров, smm-специалистов, редакторов и других работников СМИ и медиакомпаний.

По данным совместного исследования РАЭК и Microsoft «Подходы к разработке этических решений в области искусственного интеллекта», вышедшего в 2021 году, использование ИИ несет за собой ощутимые риски, связанные с дискриминацией, необъективностью, предвзятостью, угрозой конфиденциальности и возможными ошибками. При этом выделяется несколько уровней, на которых могут возникнуть подобные риски – это разработка, обучение алгоритма, наполнение данными, внедрение.

Технологии дипфейков создают возможности для социального инжинерии, т.е. для манипулирования большими общественными группами посредством широкого применения как технических, так и психологических механизмов. Здесь дипфейки выступают в качестве эффективного

³⁹ Давыдов С. Г., Замков А. В., Крашенинникова М. А., Лукина М. М. Использование технологий искусственного интеллекта в российских медиа и журналистике // Вестн. Моск. ун-та. Сер. 10: Журналистика. 2023. № 5. С. 3–21. DOI: 10.30547/vestnik.journ.5.2023.321

инструмента воздействия на социальное поведение людей, инструмента внушения и внедрения ожидаемых установок и реакций. По мнению экспертов, последние успехи в области технологий выводят исторические методы социальной инженерии на новый уровень как по масштабу, так и по сложности⁴⁰.

Как было, указано в разделе 2, существующий на данный момент медиаконтент, созданный человеком, все больше обладает манипулятивными свойствами, и именно этот контент используется в том числе и для обучения нейросетевых технологий, которые затем используются для помощи в работе средств массовой информации, журналистов, редакторов и т.д. Это влечет за собой проблему установления авторства текста. При этом стоит учитывать, что даже полностью сгенерированный текст, как правило, проходит через обработку корректором. Таким образом, возникает гибридный объект, созданный одновременно человеком и нейросетью.

Тельпов Р.Е., Ларцина С.В., исследуя типовые различия естественных и сгенерированных текстов, отмечают, что в сгенерированных текстах слова, включённые в заголовок, встречаются по тексту значительно чаще, чем в естественных текстах⁴¹. Данный признак четко указывает на то, что нейросеть создает текст ориентируясь на промт пользователя, таким образом даже полностью сгенерированный текст может указывать на цель сообщения. Сгенерированный текст не раскрывает вопрос в полной мере, включает общую информацию, имеет более простую структура и поверхностные примеры.

Стоит отметить исследование И.А. Туркулец, в котором автор анализирует композиционные особенности текстов, сгенерированных ChatGPT. Автор обращает внимание, что сгенерированный текст включает нумерованный список, конструкция которого имеет вид «Название.

⁴⁰ См.: URL: <https://mysecuritymarketplace.com/reports/social-engineering-blurring-reality-and-fake-a-guide-for-the-insurance-professional/> (дата обращения: 23.09.2024)

⁴¹ Тельпов Р.Е., Ларцина С.В. Типовые различия естественных и сгенерированных нейронной сетью текстов в количественном аспекте // Научный диалог. 2023. Т.12. № 7. С. 47-65.

Развернутое определение». Предшествует списку абзац с вводной информацией, подытоживает список короткое резюме⁴². Важно понимать, что сгенерированный текст не содержит научной новизны, не имеет теоретической и практической значимости. Нелишне отметить, что тексты могут содержать недостоверные факты и литературные источники. Между тем, существуют различные методы выявления сгенерированного текста: анализ статистических характеристик текста (частота встречаемости слов, длина предложений, распределение слов в предложении), использование алгоритма машинного обучения, анализ структуры текста⁴³.

Текстовые отрывки, сгенерированные ChatGPT. В эксперименте Гао и др. (2022) Plagiarismdetector.net распознал почти все из 50 научных тезисов, сгенерированных ChatGPT, как полностью оригинальные. Халил и Эр (Khalil and Er 2023) загрузили 50 эссе, сгенерированных ChatGPT, в две программные системы поиска текстовых совпадений (25 эссе в iTenticate и 25 эссе в систему Turnitin), хотя это просто разные интерфейсы одного и того же механизма. Они обнаружили, что 40 (80%) из них были признаны оригинальными, хотя они определяли оригинальность как сходство на 20% или менее. Халил и Эр (Khalil and Er 2023) также попытались проверить возможности ChatGPT, чтобы определить, были ли эссе сгенерированы ChatGPT, и заявили, что точность составляет 92%, поскольку 46 эссе предположительно были признаны случаями плагиата. По состоянию на май 2023 года ChatGPT теперь выдаёт предупреждение на такие вопросы, как: «Как языковая модель на основе ИИ, я не могу проверить конкретный источник или происхождение предоставленного вами абзаца». Авторы этой статьи считают исследование Халила и Эра (Khalil and Er 2023) проблематичным по двум причинам. Во-

⁴² Туркулец И.А. Композиционные особенности текстов, сгенерированных chatgpt, как маркер несамостоятельности выполнения работ студентами // Правовая реальность в условиях цифровизации общества: материалы Всероссийской научно-практической конференции (г. Хабаровск, 9-11 ноября 2023 г.). Хабаровск: Дальневосточный государственный университет путей сообщения, 2023. С. 59-68.

⁴³ Айдагулова А.Р. Особенности текстов, сгенерированных искусственным интеллектом // Вестник Башкирского государственного педагогического университета им. М. Акмуллы. 2023. С. 155.

первых, стоит отметить, что применение программных систем сопоставления текста для обнаружения текста, созданного большими лингвистическими моделями, не имеет смысла из-за стохастического характера выбора слов. Во-вторых, поскольку большая лингвистическая модель будет «галлюцинировать», нельзя спросить, является ли он автором текста.

Результаты вышеупомянутых исследований показывают, что существующие инструменты для обнаружения текста, созданного искусственным интеллектом, по-прежнему с трудом распознают его, в то время как тексты, написанные человеком, обычно распознаются довольно точно (с точностью выше 80%). Однако способность инструментов распознавать текст, созданный искусственным интеллектом, ставится под сомнение, поскольку во многих исследованиях их точность составляла всего около 50% или чуть выше. В зависимости от инструмента при распознавании текста, созданного ChatGPT или написанного человеком, может наблюдаться предвзятость. Кроме того, инструменты с трудом определяют источник текста, если ChatGPT преобразует текст, написанный человеком, или генерирует текст в определённом стиле (например, объяснение ребёнка). Более того, производительность инструментов обнаружения значительно снижается, когда тексты намеренно изменяются путём перефразирования или переписывания. Обнаружение текста, сгенерированного ИИ, остаётся сложной задачей для существующих инструментов обнаружения, но обнаружить код, сгенерированный ChatGPT, ещё сложнее⁴⁴.

Корпорация Microsoft в 2022 году отметила, что распространение синтетических медиа станет еще изощреннее, дипфейки будут вполне убедительно чередоваться в Сети с реальными событиями, разворачивающимися в мире. В перспективе человечеству предстоит

⁴⁴ Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S. *et al.* Testing of detection tools for AI-generated text. *Int International Journal for Educational Integrity* **19**, 26 (2023). <https://doi.org/10.1007/s40979-023-00146-z> P. 7.

столкнуться с проблемой синтетических фальсификаций людей, поддельными беседами с ними по аудио- и визуальным каналам в режиме реального времени⁴⁵.

Изучение опыта Экспертно-криминалистического центра МВД России (далее – ЭКЦ) показало, что модифицированный с помощью нейронных сетей видеоконтент имеет комплекс характерных диагностических признаков. Перечислим часть из них: выраженные контрастные различия или неестественные границы смежных областей изображения; несоответствие шумовых шаблонов смежных областей изображения; наличие похожих участков на изображении; наличие характерных для работы нейронных сетей частот в спектре изображения; сжатие видеоизображений с применением алгоритма JPEG в два раза и более; наличие области изображения, сжатого в формате JPEG, мера отклонения матрицы квантования которого отличается от остальной области изображения (покадрово для видеоизображения); наличие аномалий в видеопоследовательности; резкие перепады освещенности или контрастности; различные направления (отсутствие) теней от одного источника освещения; несоответствие метаданных EXIF предполагаемым условиям съемки; скачкообразные или неестественные положения, размеры или формы объектов; наличие в изображении объектов с нарушением линейной перспективы.

Данные признаки помимо классических методов выявляются также и при помощи нейросетевых: извлечение набора компактных признаков с применением представлений BOW; извлечение и анализ характеристических признаков с помощью CFFN и CNN; использование матриц различий между соседними кадрами для обнаружения фейков с помощью RNN и LSTM; анализ пространственно-временных особенностей видеопотока с помощью CNN и LSTM; поиск несоответствия разрешений различных областей изображения; применение спектральных данных и классификаторов на основе CNN;

⁴⁵ См.: URL: <https://blogs.microsoft.com/on-the-issues/2022/05/03/artificial-intelligence-department-of-defense-cyber-missions/> (дата обращения: 12.09.2024)

классификация с помощью капсульных сетей⁴⁶.

Также можно выделить возможность искусственного интеллекта распознавать частоту сердечных сокращений по так называемому обесцвечиванию лица человека на видео или изображении. Когда в разных частях лица частота не совпадает, специалист делает вывод, что перед ним ложный контент. Также в этой связи стоит обратить внимание на технологию блокчейн. Используя цепочки блоков данных в децентрализованной сети и сопоставляя информацию на выходе с определенным стартовым неизменным ключом, она может отследить путь цифрового контента вплоть до его создателя. Многие эксперты даже предлагают сформировать на государственных уровнях органы, которые будут отслеживать дипфейки посредством технологии блокчейн⁴⁷.

При этом исследователи отмечают, что при использовании всевозможных технологий сегодня точность обнаружения, например, дипфейковых изображений все-таки пока не на уровне идеала. Предположительно, она составляет 99,3% для необработанного контента и 81% для более сложных задач с низким качеством⁴⁸.

Таким образом на данный момент отсутствует в достаточной мере точная технология для выявления контента, сгенерированного при помощи нейросетевых технологий. При этом ее разработка требуется с точки зрения обеспечения безопасности на государственном уровне. Дополнительная сложность заключается в исследовании текстов, в связи с теми изменениями, которые произошли в медиакommunikации в целом.

⁴⁶ Батоев В.Б., Юмोजапов Р.С. Использование технологий искусственного интеллекта в выявлении видеодипфейков // Вестник Краснодарского университета МВД России. 2023. № 3 (61). С. 77.

⁴⁷ См.: URL: <https://cgirc.ru/upload/iblock/23f/67eesk6aig8owuk7brnrgwospccquaec.pdf> (дата обращения: 12.09.2024)

⁴⁸ См.: URL: www.arxiv.org/abs/1901.08971 (дата обращения: 12.09.2024)

Заключение

Таким образом, нейротехнологии находят активное применение в медиакommunikации, что требует соответствующего развития и в области судебно-экспертной деятельности. С учетом того, что нейросети позволяют автоматизировать создание большого объема контента различного содержания, необходимо развивать применение нейросетей для автоматизации экспертных исследований, в том числе направленных на выявление сгенерированного контента. Для разработки экспертных методик на основе нейросетей необходимо создание баз данных (Dataset) для анализа и машинного обучения – совокупностей данных по различным объектам судебной экспертизы, для решения типичных экспертных задач.

Изменения в методологии и технологиях судебно-экспертных исследований требует новых экспертных компетенций. Как и в других отраслях науки внедрение нейросетей и других алгоритмов искусственного интеллекта в судебной экспертиологии обуславливает появление Data Scientist – специалиста, который формирует инструменты для решения судебно-экспертных задач, выбирая ранее созданные алгоритмы и анализируя данные, строит модели и тестирует их.

Создание баз данных должно также способствовать экспертному прогнозированию, в задачи которого должно входить нахождение превентивных мер в отношении выявления контента, созданного при помощи генеративных нейросетей, а также его исследование с целью выявления промта, заданного пользователем, что позволит оценить целевую направленность созданного медиаконтента.

Создание Dataset и репозиторий по родам (видам) объектов возможно только после согласования методических подходов к исследованию объектов судебных экспертиз, разработанных различными ведомствами. Также следует внедрить процедуру сертификации судебных экспертов для повышения уровня доверия к ним как субъектам оценки контента в медиакommunikации.