УДК 343.98:004.8

Колотовкина А.А.
магистрант 3 курса
Красноярский государственный аграрный университет
(г. Красноярск, Россия)

ИСПОЛЬЗОВАНИЕ АНАЛИТИЧЕСКИХ ТЕХНОЛОГИЙ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ВЫЯВЛЕНИИ И ПРОФИЛАКТИКЕ ЭКСТРЕМИСТСКОЙ АКТИВНОСТИ В СОЦИАЛЬНЫХ СЕТЯХ

Аннотация: *в статье анализируются возможности применения аналитических технологий и искусственного интеллекта при выявлении и профилактике экстремистской активности в социальных сетях. Рассматриваются сетевой, временной и содержательный анализ, методы машинного обучения, классификация публикаций и обнаружение аномалий. Выявляются риски ложноположительных и ложноотрицательных результатов, недостаточной объяснимости моделей и неверного толкования цифровых связей. Предлагается многоуровневая модель, сочетающая автоматизированный отбор данных, экспертную проверку, междисциплинарное взаимодействие и обязательное принятие решений человеком при соблюдении требований законности и соразмерности.*

Ключевые слова: *искусственный интеллект, социальные сети, экстремистская активность, криминалистическая профилактика, машинное обучение, сетевой анализ, цифровой мониторинг, защита персональных данных.*

Рост объема информации, создаваемой и распространяемой в социальных сетях, затрудняет ее исследование исключительно ручными методами. Пользовательская активность включает публикации, комментарии, изображения, аудио- и видеоматериалы, контакты, реакции и сведения о взаимодействии с другими учетными записями. Значимые признаки могут быть распределены между множеством объектов, а распространение материалов происходит быстро и нередко одновременно на нескольких площадках. Эти

обстоятельства обусловливают интерес к автоматизированному поиску, аналитике больших данных и технологиям искусственного интеллекта (далее — ИИ) при выявлении и профилактике преступлений экстремистской направленности.

Вместе с тем применение алгоритмов создает самостоятельные риски. Автоматическая классификация может ошибочно отнести нейтральное сообщение к подозрительной категории, не распознать значимый материал или не учесть контекст цитирования, сатиры, критики и дискуссии. Сетевое совпадение и статистическая аномалия также не равнозначны установлению противоправной деятельности. Цель статьи — определить возможности аналитических технологий, выявить ограничения автоматизированной оценки и предложить организационные условия их ответственного применения.

Технологический процесс исследования социальных сетей включает определение источников и правовых пределов получения информации, фиксацию и копирование доступных данных, приведение материалов к формату, пригодному для анализа, исследование содержания и связей между объектами, а также проверку полученных закономерностей по независимым источникам. Автоматизация может ускорить отдельные этапы, но не отменяет необходимость сохранять исходные данные и документировать процедуру их обработки [4, с. 132–139, 8, с. 248–251].

Одним из направлений выступает сетевой анализ. Социальная сеть представляется в виде графа, узлами которого являются пользователи, сообщества, публикации и иные объекты, а ребрами — подписки, сообщения, упоминания, реакции и пересылки. Такая модель позволяет исследовать структуру контактов, выделять группы интенсивного взаимодействия и выявлять аккаунты, связывающие разные сообщества. Показатели центральности могут указывать на количество непосредственных связей, посредническое положение узла или его близость к другим участникам. Они помогают определить направления дальнейшей проверки, но не позволяют

автоматически сделать вывод о руководящей роли лица или преступном характере его действий.

Алгоритмы кластеризации применяются для выделения групп аккаунтов, которые чаще взаимодействуют друг с другом, чем с остальной сетью. Их использование может помочь обнаружить повторяющееся распространение сходных материалов или устойчивые коммуникационные группы. Однако кластер может быть сформирован вследствие общих интересов, учебы, профессиональной деятельности или иных законных причин. Поэтому статистическая близость аккаунтов не должна рассматриваться в качестве правового вывода. Она имеет лишь ориентирующее значение и нуждается в объяснении на основе содержания сообщений и иных обстоятельств.

Временной анализ позволяет сопоставлять последовательности событий: размещение и редактирование публикаций, сообщения, изменения профилей, появление новых контактов и повторное распространение материалов. Резкое увеличение частоты публикаций, согласованное размещение схожих текстов или одновременное изменение нескольких страниц может быть признано аномалией относительно обычной активности. Однако аномальность представляет собой статистическую характеристику. Ее причины могут быть как противоправными, так и нейтральными, поэтому аналитический результат должен становиться основанием для проверки, а не заменой такой проверки.

Содержательный анализ включает обработку текстов, изображений, аудио- и видеоматериалов. Для текстов применяются поиск по ключевым словам, лемматизация, токенизация, классификация, анализ тональности и стилометрические методы. Исследование может помочь выявить повторяющиеся формулировки, тематические связи и вероятные признаки единого источника текстов. Однако короткие сообщения, сленг, ирония, цитирование, региональные особенности речи и контекст общения способны влиять на результат. Поэтому автоматическая классификация текста не должна подменять лингвистический анализ и оценку всех обстоятельств публикации.

Анализ изображений и видеозаписей может включать распознавание текста, объектов и сцен, поиск дубликатов, выявление признаков редактирования и сопоставление визуальных материалов. При этом цифровая обработка способна изменять исходные свойства файла, а технологии генерации изображений и видео усложняют проверку происхождения контента. Алгоритмический вывод о сходстве или возможной модификации объекта следует проверять с учетом оригинального файла, метаданных, условий получения и иных сведений. Внешняя реалистичность материала не является гарантией его подлинности.

ИИ и методы машинного обучения применяются для предварительной классификации публикаций, группировки сходных объектов, поиска аномалий и ранжирования материалов по заданным критериям. В контексте профилактики это позволяет быстрее выделять большие массивы данных, требующие внимания специалиста. Исследователи отмечают, что автоматизированные системы способны стать инструментом выявления и предотвращения распространения экстремистской и террористической пропаганды, но эффективность таких систем зависит от корректности задач и исходных данных [1, с. 113–128].

Качество модели определяется составом и репрезентативностью обучающей выборки, выбранными признаками и настройками алгоритма. Если в наборе данных недостаточно нейтральных текстов, материалов с цитированием или иных контекстуально сложных сообщений, модель может воспроизводить систематические ошибки. Ложноположительный результат возникает, когда допустимый материал классифицируется как подозрительный, ложноотрицательный — когда значимый материал не выявляется. Поэтому при оценке инструмента важны не только общие показатели точности, но и анализ ошибок по отдельным категориям контента [2, с. 218–221, 7].

Дополнительную проблему составляет объяснимость. Модель может присвоить публикации высокий показатель вероятности принадлежности к заданной категории, однако не предоставить достаточного описания признаков, повлиявших на результат. Для криминалистических целей этого недостаточно:

должно быть понятно, какая версия программного обеспечения применялась, какие данные анализировались, какие параметры использовались и как получен вывод. Необходимо сохранять исходный объект, результаты промежуточной обработки и сведения о версии модели. Формулировка «материал выявлен искусственным интеллектом» сама по себе не обеспечивает проверяемости результата [12, с. 273–278].

Алгоритмы также не обладают самостоятельной компетенцией в юридической оценке содержания. Значение сообщения может зависеть от предыдущих реплик, адресата, характера дискуссии, цитируемого источника и иных обстоятельств. Один и тот же термин может использоваться в научном, журналистском, историческом или сатирическом контексте. Система, обученная преимущественно на ключевых словах, рискует не учитывать смысловую структуру высказывания. Поэтому автоматическое выявление должно сопровождаться изучением полного контекста и, при необходимости, привлечением специалистов соответствующего профиля.

Использование ИИ в криминалистике следует рассматривать как вспомогательный аналитический механизм. Он может ускорять поиск, структурировать массив данных, выявлять повторяемость и подсказывать направления проверки. Однако решение о дальнейших процессуальных действиях, вмешательстве в права пользователя, передаче информации или правовой оценке должно приниматься уполномоченным человеком. Данный подход согласуется с научной постановкой вопроса о пределах применения ИИ и необходимости соблюдения профессиональных и этических требований [3, с. 331–337].

Профилактический мониторинг, оперативно-розыскная деятельность и доказывание в уголовном процессе имеют различную правовую природу и не должны смешиваться. Автоматическое обнаружение публикации или сетевой связи может послужить поводом для дополнительной проверки, но не становится само по себе доказательством совершения преступления. Переход от

аналитической информации к процессуальному использованию требует соблюдения установленного порядка получения и закрепления сведений.

Мониторинг должен учитывать баланс между предупреждением преступлений и конституционными гарантиями свободы мысли, слова и неприкосновенности частной жизни. Общедоступность отдельной публикации не означает возможности бесконтрольно формировать массивы персональных данных о пользователях. Сбор должен быть связан с конкретной целью, ограничен необходимым объемом и сопровождаться правилами хранения, доступа, обезличивания и удаления сведений, не подтвердивших значимости. Такой подход соответствует положениям законодательства о противодействии экстремистской деятельности, информации и персональных данных [6, 9–11].

Необходимо также учитывать целесообразность многоуровневой модели проверки результатов мониторинга. На первом уровне автоматизированная система выделяет материалы по формальным и смысловым признакам. На втором специалист проверяет источник, контекст, дату и связь публикации с другими объектами. На третьем уровне, если этого требуют свойства материала, проводится экспертное исследование текста, изображения или видеозаписи. На четвертом уровне уполномоченное лицо принимает решение о дальнейшем использовании сведений. Такая последовательность снижает риск того, что машинное срабатывание будет воспринято как окончательный вывод.

Для обеспечения проверяемости автоматического мониторинга необходимо документировать не только сам материал, но и процедуру его выявления. В аналитической карточке целесообразно отражать адрес страницы, идентификатор публикации, сведения об учетной записи, дату и время обнаружения, полный контекст, связанные комментарии и предыдущие либо последующие публикации. Дополнительно должны сохраняться версия модели, параметры анализа, исходный файл и результаты обработки. Это позволяет воспроизвести ход анализа и оценить, какие именно признаки повлияли на автоматическое предупреждение.

Отдельное значение имеет регулярный аудит качества алгоритмов. Тестовые наборы должны включать не только потенциально опасные материалы, но и нейтральные публикации, научные тексты, цитаты, сатиру и критические высказывания. Это позволяет оценивать не только способность системы обнаруживать целевые объекты, но и вероятность ошибочного отнесения допустимых материалов к подозрительной категории. При изменении социальной платформы, языка коммуникации или структуры данных модель должна проходить повторную оценку.

Организационным условием результативного применения ИИ выступает междисциплинарное взаимодействие. В работе могут участвовать криминалисты, специалисты по информационным технологиям, лингвисты, психологи, религиоведы и специалисты по защите персональных данных. Каждый участник оценивает те признаки, для интерпретации которых обладает соответствующей компетенцией. Техническое сходство файлов, лингвистический смысл высказывания и уголовно-правовая оценка не должны подменять друг друга. Совместная работа позволяет отделять аналитическую гипотезу от подтвержденного факта [5, с. 92–100].

Для устойчивой организации мониторинга представляется целесообразным формировать специализированные подразделения или центры анализа, объединяющие автоматизированную обработку, экспертную проверку и взаимодействие с администрациями платформ. Их деятельность должна сопровождаться едиными стандартами описания цифровых объектов, защищенной инфраструктурой, разграничением доступа, журналированием операций и независимым аудитом алгоритмов. Взаимодействие с платформами следует строить через официальные каналы, указывая конкретные объекты и правовые основания обращения. Сохранение данных необходимо отделять от их предоставления, поскольку предотвращение утраты объекта не означает автоматического раскрытия всего объема сведений о пользователе.

Перспективным направлением является развитие специализированных языковых моделей и корпусов данных, учитывающих особенности русского

языка и контекст цифровой коммуникации. При подготовке таких наборов данных должны учитываться требования обезличивания, защита персональных данных и включение нейтральных материалов, позволяющих системе распознавать контекст. Без достаточного представления нейтральных и контекстуально сложных сообщений алгоритм может чрезмерно реагировать на отдельные слова или устойчивые выражения. Поэтому повышение качества моделей связано не только с техническим совершенствованием, но и с прозрачностью процедур их обучения и тестирования.

Аналитические технологии расширяют возможности выявления и профилактики экстремистской активности в социальных сетях за счет обработки больших массивов информации, исследования сетевых связей, временных закономерностей и содержания публикаций. Наибольшую практическую ценность имеет сочетание сетевого, временного, технического и содержательного анализа.

Результаты машинного обучения и ИИ имеют ориентирующий характер. Аномалия, сходство публикаций, принадлежность к кластеру или автоматическая классификация не являются самостоятельной юридической оценкой и нуждаются в проверке с учетом полного контекста и независимых источников. Для проверяемости необходимо сохранять исходные данные, описание модели, параметры анализа и историю преобразований.

Предлагается применять многоуровневую модель, в которой автоматизированный поиск дополняется проверкой специалистом, экспертным исследованием при необходимости и решением уполномоченного лица. Ее организационными гарантиями должны стать минимизация обрабатываемых данных, регулярный аудит ошибок, междисциплинарное взаимодействие, защищенное хранение и прозрачный порядок обмена информацией с платформами. Участие человека и соблюдение принципов законности, необходимости и соразмерности являются обязательными условиями использования ИИ в криминалистической профилактике.

СПИСОК ЛИТЕРАТУРЫ:

1. Баранов В.В. Использование искусственного интеллекта для выявления и предотвращения распространения экстремистской и террористической пропаганды в Интернете // Труды Академии управления МВД России. 2024. № 4 (72). С. 113–128
2. Гринько М.С. Проблемы и перспективы использования искусственного интеллекта в современной криминалистике // Пробелы в российском законодательстве. 2024. № 2. С. 218–221
3. Исаева Е.Д. Этические и правовые пределы использования цифровых технологий и искусственного интеллекта в работе следователя // Научный вестник Орловского юридического института МВД России им. В.В. Лукьянова. 2026. № 3. С. 331–337
4. Канузель А.Е. Использование информации социальных сетей при выявлении, расследовании и предупреждении преступлений // Криминалистика: вчера, сегодня, завтра. 2024. № 2. С. 132–139 https://doi.org/10.55001/2587-9820.2024.35.67.013
5. Коломинов В.В. Криминалистическая профилактика преступлений экстремистской направленности, совершенных с помощью современных технологий // Сибирский юридический вестник. 2023. № 2 (101). С. 92–100 https://doi.org/10.26516/2071-8136.2023.2.92
6. Конституция Российской Федерации (с изменениями, одобренными в ходе общероссийского голосования 1 июля 2020 г., с изменениями, внесенными ФКЗ от 04.10.2022 № 8-ФКЗ) // Официальный интернет-портал правовой информации. 06.10.2022. Ст. 0001202210060013
7. Сметанин С.И. Анализ тональности текстов из социальных сетей на основе методов машинного обучения для мониторинга общественных настроений: дис. ... канд. техн. наук. Москва, 2022. 162 с.
8. Титова Д.Д., Седова Д.Д. Анализ социальных сетей в профилактике уголовных преступлений: алгоритмы и проблемы применения // Институты и механизмы инновационного развития: мировой опыт и российская практика: сборник статей 13-й Международной научно-практической конференции. Курск, 2023. С. 248–251

9. Федеральный закон от 25.07.2002 № 114-ФЗ «О противодействии экстремистской деятельности» // Собрание законодательства РФ. 29.07.2002. № 30. Ст. 3031
10. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» // Российская газета. 29.07.2006. № 165
11. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» // Российская газета. 29.07.2006. № 165
12. Шурдумов А.А. Искусственный интеллект как криминалистический инструмент // Теория и практика общественного развития. 2025. № 6 (206). С. 273–278 https://doi.org/10.24158/tipor.2025.6.34

Kolotovkina A.A.
master's degree student of 3rd year
Krasnoyarsk State Agrarian University
(Krasnoyarsk, Russia)

USE OF ANALYTICAL TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE IN DETECTING AND PREVENTING EXTREMIST ACTIVITY ON SOCIAL MEDIA

Abstract: *the article examines how analytical technologies and artificial intelligence can be used to detect and prevent extremist activity on social media. It considers network, temporal, and content analysis, machine learning methods, post classification, and anomaly detection. Risks identified include false positives and false negatives, insufficient model explainability, and misinterpretation of digital connections. A multilevel model is proposed that combines automated data screening, expert review, interdisciplinary collaboration, and mandatory human decision-making while observing requirements of legality and proportionality.*

Keywords: *artificial intelligence, social media, extremist activity, forensic crime prevention, machine learning, network analysis, digital monitoring, personal data protection.*