

УДК 343.3

DOI: 10.34670/AR.2026.79.95.057

Уголовно-правовая характеристика и проблемы квалификации распространения экстремистских материалов, созданных с использованием технологий искусственного интеллекта

Парфёнов Константин Иванович

Старший преподаватель,
кафедры информационного обеспечения органов внутренних дел,
Уральский юридический институт МВД России,
620057, Российская Федерация, Екатеринбург, ул. Корепина, 66;
e-mail: russia.pki@gmail.com

Аннотация

В статье исследуются проблемы уголовно-правовой характеристики и квалификации деяний, связанных с распространением экстремистских материалов, созданных с использованием технологий искусственного интеллекта. Анализируется вопрос о том, может ли контент, сгенерированный нейросетью, быть признан экстремистским материалом в том значении, которое придаётся этому понятию Федеральным законом «О противодействии экстремистской деятельности». На основе анализа законодательной дефиниции выделены содержательный и целевой признаки экстремистского материала, обосновывается вывод о том, что экстремистский характер представляет собой свойство содержания, а не обстоятельство происхождения материала. Выявляются затруднения, обусловленные стохастической природой генеративных нейросетей, способных создавать неограниченное число уникальных вариаций контента, и обосновывается необходимость перехода от текстуального к содержательному критерию идентификации экстремистских материалов. Особое внимание уделяется проблемам установления субъективной стороны преступления: автором выделены и последовательно проанализированы типичные ситуации, различающиеся по характеру взаимодействия лица с нейросетью и по содержанию умысла. Рассматривается вопрос о разграничении создания и распространения экстремистского материала применительно к нейросетевому контенту, предлагается дифференцированный подход, основанный на оценке степени содержательного контроля лица над результатом генерации. Дается оценка общественной опасности целенаправленного обучения нейросетей на массивах экстремистских материалов, обосновывается вывод о наличии пробела в уголовном законодательстве. Анализируется вопрос о целесообразности введения уголовной ответственности разработчиков и операторов генеративных нейросетей.

Для цитирования в научных исследованиях

Парфёнов К.И. Уголовно-правовая характеристика и проблемы квалификации распространения экстремистских материалов, созданных с использованием технологий искусственного интеллекта // Вопросы российского и международного права. 2026. Том 16. № 7А. С. 551-563. DOI: 10.34670/AR.2026.79.95.057

Ключевые слова

Экстремистские материалы, генеративные нейросети, искусственный интеллект, уголовно-правовая квалификация, распространение экстремистских материалов, субъективная сторона преступления, создание экстремистских материалов, нейросетевой контент, противодействие экстремизму.

Введение

Стремительное развитие технологий генеративного искусственного интеллекта в последние годы привело к возникновению ряда проблем, которые до настоящего времени не получили должного осмысления в уголовно-правовой доктрине. Современные генеративные нейросети, среди которых большие языковые модели и модели генерации изображений, позволяют в течение нескольких секунд создавать текстовый, графический, аудио- и видеоконтент, который по своему качеству практически неотличим от контента, созданного человеком. Данное обстоятельство порождает новые возможности для автоматизированного создания материалов экстремистского содержания, масштабного их тиражирования и распространения в информационно-телекоммуникационных сетях. Между тем нормы действующего уголовного законодательства, устанавливающие ответственность за преступления экстремистской направленности, формировались в условиях, когда единственным субъектом создания информационного контента являлся человек. Появление генеративных нейросетей как инструмента создания экстремистских материалов ставит перед наукой уголовного права и правоприменительной практикой вопросы, которые не имеют однозначного решения и требуют самостоятельного научного анализа.

Прежде чем перейти к существу, оговорим терминологию. Технологии искусственного интеллекта - понятие более широкое, нежели генеративные нейросети; последние представляют собой лишь один, хотя и наиболее значимый для настоящего исследования, класс таких технологий. Именно генеративные модели способны самостоятельно порождать новый информационный контент, тогда как иные системы искусственного интеллекта (например, классифицирующие или рекомендательные) подобной функцией не обладают. По этой причине предметом анализа в настоящей статье выступает преимущественно нейросетевой контент, то есть результат работы генеративных нейросетей.

Общие вопросы уголовно-правового противодействия экстремизму достаточно глубоко разработаны в трудах С.В. Борисова, П.А. Кабанова, А.Г. Хлебушкина, А.С. Ахметшина и ряда других исследователей [Борисов, Жеребченко, 2011; Кабанов, 2008; Хлебушкин, 2007]. Проблемы квалификации преступлений, совершаемых с использованием информационно-телекоммуникационных технологий, также получили значительное внимание в научной литературе [Рускевич, 2022; Ахметшин, 2023]. Отдельные аспекты данной проблематики затрагиваются в работах, посвященных правовому регулированию искусственного интеллекта [Морхат, 2017; Мансуров, Головизнин, 2022]. Вместе с тем вопрос о специфике уголовно-правовой оценки деяний, связанных с распространением экстремистских материалов, сгенерированных нейросетями, остается практически неисследованным, а комплексный уголовно-правовой анализ до настоящего времени не проводился.

Целью настоящей статьи является выявление и анализ проблем уголовно-правовой характеристики и квалификации распространения экстремистских материалов, созданных с

использованием технологий искусственного интеллекта, а также выработка научно обоснованных предложений по их разрешению.

Материалы и методы исследования

Методологическую основу исследования составляют формально-юридический метод, использованный при анализе законодательных дефиниций и норм Особенной части Уголовного кодекса Российской Федерации; сравнительно-правовой метод, примененный при обращении к зарубежным доктринальным конструкциям; а также метод юридического моделирования типовых ситуаций взаимодействия лица с генеративной нейросетью, посредством которого выделяются и разграничиваются варианты, различающиеся по содержанию субъективной стороны. Необходимо учитывать, что на момент подготовки настоящей работы самостоятельная судебная практика по уголовным делам о распространении экстремистских материалов, созданных именно с использованием генеративных нейросетей, в Российской Федерации фактически отсутствует, что объясняется новизной самой технологии. По этой причине анализ опирается преимущественно на сложившуюся практику применения ст. 280 и 282 УК РФ к материалам, распространяемым в информационно-телекоммуникационной сети «Интернет», а также на метод юридического моделирования. Подходы, выработанные судами применительно к «обычному» интернет-контенту, используются в качестве отправной точки с последующим выявлением тех аспектов, в которых специфика нейросетевой генерации делает их недостаточными.

Результаты и обсуждение

Прежде всего необходимо определить, может ли контент, сгенерированный нейросетью, рассматриваться в качестве экстремистского материала в том значении, которое придается этому понятию действующим законодательством. Постановка данного вопроса обусловлена тем, что генеративная нейросеть по своей природе не является субъектом осмысленной деятельности. Нейросеть не понимает значения генерируемого ею контента, не осознает его социальной опасности и не преследует цели побуждения кого-либо к совершению противоправных действий. Она представляет собой математическую модель, которая на основе статистических закономерностей, извлеченных из массива обучающих данных, генерирует последовательности символов, пикселей или звуковых сигналов, наиболее вероятные с точки зрения заданных параметров. Возникает вопрос о том, может ли продукт деятельности такой модели быть квалифицирован как материал, «призывающий к осуществлению экстремистской деятельности либо обосновывающий или оправдывающий необходимость осуществления такой деятельности», как это сформулировано в ст. 1 Федерального закона от 25 июля 2002 г. № 114-ФЗ «О противодействии экстремистской деятельности» [Федеральный закон № 114-ФЗ, 2002].

Анализ приведенной законодательной дефиниции позволяет выделить два признака экстремистского материала: содержательный и целевой. Содержательный признак выражается в том, что материал призывает к экстремистской деятельности, обосновывает или оправдывает ее необходимость. Целевой признак состоит в предназначенности материала для распространения или публичной демонстрации.

Содержательный признак, как представляется, имеет объективную природу. Призыв к экстремистской деятельности, обоснование или оправдание ее необходимости - это

характеристики содержания материала, которые могут быть установлены путем анализа текста, изображения, аудио- или видеозаписи безотносительно к тому, кем и каким способом данный материал был создан. Призыв остается призывом независимо от того, сформулирован ли он человеком, сгенерирован нейросетью или получен иным техническим способом. Как отмечается в научной литературе, общественная опасность экстремистского материала определяется не личностью его автора, а тем воздействием, которое он способен оказать на сознание и поведение лиц, воспринимающих этот материал. Текст, содержащий призыв к насилию по признаку расовой, национальной или религиозной принадлежности, оказывает одинаковое воздействие на читателя вне зависимости от того, написан ли он человеком или сгенерирован нейросетью. С этой точки зрения способ создания материала не влияет на наличие или отсутствие содержательного признака экстремистского материала.

Мы полагаем, что данный вывод имеет принципиальное значение и должен быть четко зафиксирован в доктрине: экстремистский характер материала - это свойство его содержания, а не обстоятельство его происхождения. Любая попытка увязать признание материала экстремистским с личностью или природой его создателя привела бы к неоправданному сужению сферы действия антиэкстремистского законодательства и создала бы лазейку, позволяющую избежать ответственности путем использования автоматизированных средств генерации контента. Заметим, что данный тезис не является самоочевидным и допускает возражение: можно утверждать, что социальная опасность экстремистского высказывания неотделима от стоящей за ним человеческой воли, а текст, порожденный лишенной сознания моделью, не несет в себе той идеологической интенции, которая и составляет существо экстремизма. Этот довод заслуживает внимания, однако, по нашему мнению, он смешивает два разных уровня анализа - уровень содержания материала и уровень виновности лица. Отсутствие воли у самой нейросети значимо для оценки субъективной стороны деяния конкретного человека, но не для определения того, является ли материал экстремистским по своему содержанию. Воздействие текста на аудиторию не зависит от того, осознавал ли его смысл порождающий его алгоритм.

Что касается целевого признака - предназначенности для распространения или публичной демонстрации, - то его наличие определяется не свойствами самого материала, а действиями лица, которое осуществляет или намеревается осуществить распространение. Нейросеть не может «предназначить» сгенерированный контент для распространения - эту функцию выполняет человек, который принимает решение о публикации, рассылке или ином доведении материала до третьих лиц. Целевой признак экстремистского материала, следовательно, также не зависит от способа создания контента. Более того, материал, сгенерированный нейросетью и оставшийся исключительно в памяти устройства пользователя, не доведенный до иных лиц, целевым признаком не обладает, что само по себе исключает его признание экстремистским материалом в полном смысле этого понятия.

Вместе с тем, признавая принципиальную возможность квалификации нейросетевого контента в качестве экстремистского материала, мы не можем не обратить внимание на серьезные практические затруднения, которые возникают при применении этого вывода.

Наиболее существенное из них связано с процедурой признания материалов экстремистскими. В соответствии со ст. 13 Федерального закона «О противодействии экстремистской деятельности» информационные материалы признаются экстремистскими судом по месту их обнаружения, распространения или нахождения организации, осуществившей производство таких материалов, на основании представления прокурора или

при производстве по соответствующему делу об административном правонарушении, гражданскому или уголовному делу. Признанные экстремистскими материалы включаются в Федеральный список экстремистских материалов, ведение которого осуществляет Министерство юстиции Российской Федерации.

Указанная процедура была разработана в расчете на материалы, имеющие фиксированное содержание: конкретную книгу, брошюру, листовку, видеозапись, интернет-публикацию. Каждый такой материал является индивидуально определенным объектом, который может быть точно идентифицирован и включен в Федеральный список экстремистских материалов с указанием его реквизитов, позволяющих отличить его от всех иных материалов.

Нейросетевой контент принципиально отличается от «традиционного» экстремистского материала именно в этом отношении. Генеративная нейросеть при каждом обращении к ней создает уникальный результат. Даже при тождественном запросе (промте) нейросеть генерирует текст, который отличается от предыдущего по формулировкам, структуре, выбору аргументов, хотя и может совпадать по общей смысловой направленности. Это свойство обусловлено стохастической (вероятностной) природой процесса генерации и является имманентной характеристикой генеративных нейросетей как класса технологий. Практическое следствие данного свойства для правоприменения весьма значимо. Признание судом экстремистским одного конкретного текста, сгенерированного нейросетью, не влечет автоматического признания экстремистскими иных текстов, сгенерированных той же нейросетью по аналогичному или даже тождественному запросу, поскольку каждый результат генерации является самостоятельным информационным объектом, формально подлежащим самостоятельной правовой оценке.

Масштаб данной проблемы трудно переоценить. Современные генеративные нейросети позволяют создавать значительное число текстов экстремистского содержания за непродолжительное время. При существующем подходе к идентификации экстремистских материалов, основанном на точном совпадении текста с текстом, включенным в Федеральный список, подавляющее большинство таких текстов формально не будут считаться «ранее признанными экстремистскими», несмотря на то что по своему содержанию они могут быть тождественны по смыслу ранее признанным.

Данная проблема, по нашему мнению, свидетельствует о необходимости корректировки самого подхода к идентификации экстремистских материалов. Мы полагаем, что применительно к нейросетевому контенту критерием идентификации экстремистского материала должно выступать не текстуальное (буквальное) тождество с ранее признанным экстремистским материалом, а содержательное (смысловое) соответствие - совпадение смысловой направленности, характера призывов, идеологических установок. Подобный подход не является беспрецедентным для российского права. Так, в практике применения законодательства о защите интеллектуальной собственности давно используется критерий «сходства до степени смешения», позволяющий признать нарушением использование не только тождественного, но и существенно сходного обозначения. Аналогичная логика, адаптированная к специфике антиэкстремистского законодательства, могла бы быть применена и в рассматриваемой сфере.

Безусловно, реализация такого подхода сопряжена со значительными трудностями. Оценка содержательного соответствия неизбежно предполагает определенную степень усмотрения правоприменителя, что создает риски произвольного расширительного толкования. Эти риски нельзя недооценивать: чрезмерно широкое понимание смыслового соответствия способно

привести к тому, что экстремистским будет признаваться любой материал, лишь затрагивающий определенную тематику, что недопустимо и противоречило бы принципу правовой определенности. Для минимизации этих рисков необходима разработка четких критериев содержательного соответствия, что может быть осуществлено как на уровне разъяснений Пленума Верховного Суда Российской Федерации, так и на уровне методических рекомендаций для экспертных учреждений. По нашему мнению, такие критерии должны опираться на совпадение не темы материала, а конкретной экстремистской смысловой направленности - характера призывов, объекта ненависти или вражды, обосновываемых идеологических установок. Детальная разработка таких критериев выходит за рамки настоящей статьи, однако мы считаем необходимым обозначить данную проблему и указать направление ее возможного решения.

Обратимся теперь к проблеме, которую мы считаем наиболее сложной и наименее разработанной в контексте рассматриваемой темы, - к проблеме установления субъективной стороны преступления при использовании генеративных нейросетей для создания и распространения экстремистского контента.

Статьи 280, 280.1 и 282 Уголовного кодекса Российской Федерации [Уголовный кодекс РФ, 1996], составляющие нормативную основу уголовно-правового противодействия распространению экстремистских материалов, предполагают наличие у виновного прямого умысла. Данное положение является общепризнанным в доктрине уголовного права и подтверждено судебной практикой. Пленум Верховного Суда Российской Федерации в пункте 8 постановления от 28 июня 2011 г. № 11 «О судебной практике по уголовным делам о преступлениях экстремистской направленности» [Постановление Пленума ВС РФ № 11, 2011] указал, что преступления, предусмотренные ст. 280 УК РФ, совершаются только с прямым умыслом и направленностью действий на нарушение основ конституционного строя и безопасности государства. Аналогичный подход применяется и в отношении преступлений, предусмотренных ст. 282 УК РФ.

Применительно к «традиционным» преступлениям данной категории, при которых экстремистский материал создается непосредственно человеком, установление умысла, хотя и представляет определенную сложность, осуществляется по устоявшимся правилам. Следствие и суд оценивают содержание распространенного материала, обстоятельства его создания и распространения, личность обвиняемого, его взгляды и убеждения, наличие или отсутствие экстремистской мотивации. Между намерением лица и конечным результатом - экстремистским материалом - существует прямая и непосредственная связь: лицо само формулирует текст, подбирает изображения, монтирует видеозапись. Содержание материала в полной мере отражает замысел его создателя.

Использование генеративной нейросети как инструмента создания экстремистского контента существенно усложняет эту картину. Между намерением лица и конечным результатом появляется промежуточное звено - нейросеть, которая обладает значительной степенью автономности в формировании выходных данных. Пользователь формулирует запрос, но конкретное содержание ответа определяется нейросетью самостоятельно, на основе статистических закономерностей, извлеченных из обучающей выборки. Пользователь не может с абсолютной точностью предсказать, какой именно текст будет сгенерирован в ответ на его запрос. Это обстоятельство порождает целый спектр ситуаций, существенно различающихся между собой по содержанию субъективной стороны и, следовательно, требующих дифференцированной уголовно-правовой оценки.

Мы считаем необходимым выделить и последовательно проанализировать эти ситуации, поскольку именно через их разграничение может быть выработан научно обоснованный подход к квалификации рассматриваемой категории деяний.

Первая ситуация является наиболее простой с точки зрения уголовно-правовой оценки. Она характеризуется тем, что лицо целенаправленно формулирует запрос к нейросети, прямо направленный на получение контента экстремистского содержания, получает такой контент и осознанно распространяет его. Примером может служить ситуация, при которой лицо вводит запрос с прямым требованием создать текст, призывающий к насилию в отношении представителей определенной национальности, получает соответствующий текст и размещает его в социальной сети. В подобном случае нейросеть выступает не более чем техническим средством, орудием совершения преступления, подобно тому как текстовый редактор или графическая программа являются орудиями при создании экстремистского материала «вручную». Умысел лица здесь очевиден и охватывает как содержание создаваемого материала, так и факт его последующего распространения. Лицо осознает общественную опасность своих действий, предвидит наступление общественно опасных последствий и желает их наступления. Квалификация подобного деяния не вызывает принципиальных затруднений и должна осуществляться по соответствующей статье Особенной части Уголовного кодекса Российской Федерации на общих основаниях. Тот факт, что непосредственным «исполнителем» текста выступила нейросеть, не влияет на оценку субъективной стороны, поскольку использование нейросети охватывалось умыслом виновного и являлось избранным им способом достижения преступного результата.

Вторая ситуация существенно сложнее. Она возникает тогда, когда лицо формулирует запрос, прямо не направленный на получение экстремистского контента, однако в силу содержания запроса или особенностей нейросети получает результат, содержащий признаки экстремистского материала. Предположим, лицо обращается к нейросети с просьбой подготовить эмоциональный текст о некотором межнациональном конфликте, не имея явного намерения получить именно экстремистский материал, а нейросеть генерирует текст, содержащий призывы к насилию. Если лицо, осознав экстремистский характер полученного результата, тем не менее распространяет его, возникает вопрос о содержании умысла на момент распространения. Полагаем, что в подобной ситуации определяющим является психическое отношение лица к материалу именно в момент его распространения, а не в момент формулирования запроса. Если лицо осознавало экстремистский характер материала и желало его распространить, налицо прямой умысел на распространение экстремистского материала, и деяние подлежит квалификации на общих основаниях. Способ получения материала и первоначальная направленность запроса в данном случае значения для квалификации не имеют, хотя и могут учитываться при характеристике личности виновного и назначении наказания.

Третья ситуация представляет, на наш взгляд, наибольшую теоретическую сложность. Она характеризуется тем, что лицо распространяет сгенерированный нейросетью контент, не осознавая его экстремистского характера. Подобное возможно, например, в случаях, когда экстремистский смысл материала является неочевидным, замаскированным, требующим специальных познаний для своего выявления, либо когда лицо в силу невнимательности не ознакомилось должным образом с содержанием сгенерированного материала перед его распространением. Поскольку рассматриваемые составы преступлений предполагают прямой умысел, отсутствие осознания лицом экстремистского характера распространяемого материала исключает уголовную ответственность. Деяние, совершенное по неосторожности либо при

отсутствии вины в отношении экстремистского характера материала, не образует состава преступления, предусмотренного ст. 280 или 282 УК РФ. Данный вывод закономерно вытекает из принципа субъективного вменения, закрепленного в ст. 5 УК РФ, согласно которой лицо подлежит уголовной ответственности только за те общественно опасные действия и наступившие общественно опасные последствия, в отношении которых установлена его вина. Вместе с тем практическая реализация этого положения сопряжена со значительными трудностями доказательственного характера, поскольку разграничение действительного отсутствия осознания и ссылки на такое отсутствие как на способ защиты требует тщательной оценки всей совокупности обстоятельств дела. Подробное рассмотрение этих доказательственных аспектов выходит за рамки уголовно-правовой характеристики деяния и составляет самостоятельный предмет исследования.

Четвертая ситуация связана с так называемым косвенным умыслом и сознательным допущением. Она возникает, когда лицо, формулируя запрос к нейросети, предвидит возможность получения экстремистского контента и сознательно допускает такую возможность либо относится к ней безразлично, после чего распространяет полученный результат. Сложность данной ситуации обусловлена тем, что, как отмечалось выше, рассматриваемые составы преступлений предполагают наличие прямого умысла, тогда как при косвенном умысле лицо не желает, а лишь сознательно допускает наступление общественно опасных последствий. Формально-юридический анализ приводит к выводу о том, что при распространении экстремистского материала с косвенным умыслом состав преступления, предусмотренного ст. 280 или 282 УК РФ, отсутствует. Однако такой вывод порождает очевидное противоречие с потребностями практики. Лицо, которое многократно обращается к нейросети с провокационными запросами, заведомо способными повлечь генерацию экстремистского контента, и распространяет полученные результаты, относясь безразлично к их характеру, представляет несомненную общественную опасность, но при буквальном толковании закона может избежать уголовной ответственности. Данное противоречие, по нашему мнению, отражает более общую проблему недостаточной приспособленности действующего уголовного законодательства к реалиям использования технологий искусственного интеллекта и заслуживает самостоятельного законодательного решения.

Анализ зарубежной доктрины обнаруживает институт, который мог бы оказаться полезным при осмыслении третьей и четвертой ситуаций, - концепцию «намеренного неведения» (*willful blindness*), известную уголовному праву ряда стран англо-американской правовой семьи. Сущность данной концепции состоит в том, что лицо, сознательно уклоняющееся от получения сведений об обстоятельствах, делающих его поведение противоправным, приравнивается к лицу, осведомленному о таких обстоятельствах. В отечественной доктрине идея, близкая по смыслу, обсуждается в связи с проблемой разграничения форм вины; так, А.И. Рарог подробно анализирует содержание интеллектуального и волевого моментов умысла и критерии их установления [Рарог, 2003]. Прямое заимствование конструкции «намеренного неведения» в российское уголовное право представляется нам сомнительным, поскольку оно вступило бы в противоречие с устоявшимся учением о формах вины и принципом субъективного вменения. Тем не менее сама постановка проблемы - о юридической оценке поведения лица, сознательно уклоняющегося от ознакомления с содержанием распространяемого им контента, - сохраняет актуальность и применительно к российскому правопорядку.

Отдельного рассмотрения заслуживает вопрос о разграничении создания и распространения экстремистского материала применительно к нейросетевому контенту. В «традиционной»

ситуации создание и распространение экстремистского материала представляют собой относительно самостоятельные деяния, граница между которыми достаточно очевидна. Создание материала - это его изготовление, придание ему законченной формы; распространение - доведение материала до сведения третьих лиц. Применительно к нейросетевому контенту разграничение этих деяний приобретает новые грани.

Возникает вопрос: кого следует считать создателем экстремистского материала, сгенерированного нейросетью? Формально материал создан нейросетью, однако нейросеть, как уже отмечалось, не является субъектом права и не может нести юридической ответственности. Лицо, сформулировавшее запрос, инициировало процесс создания материала, но не определяло его конкретное содержание во всех деталях. На наш взгляд, при решении данного вопроса следует исходить из степени содержательного контроля лица над результатом генерации. Если лицо посредством детализированного запроса, корректировки промежуточных результатов, целенаправленного отбора из нескольких вариантов фактически определило содержание итогового материала, есть основания рассматривать его как создателя этого материала. Если же роль лица свелась к формулированию общего запроса, а конкретное содержание было полностью определено нейросетью, квалификация деяния как создания экстремистского материала становится сомнительной, и на первый план выходит оценка действий лица по распространению такого материала. Предложенный критерий степени содержательного контроля носит, разумеется, оценочный характер и требует дальнейшей разработки, однако он, как нам представляется, позволяет наметить разумную границу между двумя формами преступного поведения.

Наконец, нельзя обойти вниманием вопрос об общественной опасности целенаправленного обучения и дообучения (файнтюнинга) нейросетей на массивах экстремистских материалов. Современные технологии позволяют дообучить общедоступную нейросеть на специально подобранной выборке текстов экстремистского содержания, в результате чего нейросеть приобретает устойчивую способность генерировать экстремистский контент даже в ответ на нейтральные запросы. Лицо, осуществившее такое дообучение, создает, по существу, инструмент массового производства экстремистских материалов. Общественная опасность подобных действий, на наш взгляд, чрезвычайно высока и в ряде случаев превосходит опасность создания отдельного экстремистского материала, поскольку речь идет о создании средства неограниченного тиражирования такого контента. Между тем действующее уголовное законодательство не содержит нормы, прямо предусматривающей ответственность за подобные действия. Само по себе дообучение нейросети, не сопровождающееся распространением конкретных экстремистских материалов, не охватывается диспозициями ст. 280 и 282 УК РФ. Налицо, по нашему мнению, пробел в уголовно-правовом регулировании, который требует законодательного устранения. Возможным направлением решения могло бы стать дополнение Уголовного кодекса нормой об ответственности за создание программных средств, заведомо предназначенных для генерации экстремистских материалов, по аналогии с уже существующей в ст. 273 УК РФ конструкцией ответственности за создание вредоносных компьютерных программ. Мы сознаем дискуссионность такого предложения и не настаиваем на его безусловном принятии, однако полагаем, что обозначенный пробел не может быть оставлен без внимания законодателя.

С рассмотренной проблемой тесно связан и вопрос о возможной уголовной ответственности разработчиков и операторов генеративных нейросетей за создание их моделями экстремистского контента. Данный вопрос приобретает особую остроту в связи с обсуждением

в зарубежной литературе и правоприменительной практике общих подходов к ответственности за вред, причиненный системами искусственного интеллекта. Применительно к российскому уголовному праву постановка вопроса об ответственности разработчика или оператора нейросети за сгенерированный ею экстремистский контент наталкивается на фундаментальное препятствие в виде принципа личной и виновной ответственности. Разработчик, создавший нейросеть общего назначения и принявший разумные меры для предотвращения генерации противоправного контента, не может нести ответственность за то, что пользователь посредством специально подобранных запросов обошел установленные ограничения и добился создания экстремистского материала. В подобной ситуации отсутствует вина разработчика в отношении конкретного экстремистского материала. Иной подход означал бы возложение на разработчика объективного вменения, что противоречит основополагающим принципам уголовного права. Вместе с тем вопрос о пределах ответственности оператора нейросети, который, будучи осведомленным о систематической генерации его моделью экстремистского контента, не принимает мер к устранению этой возможности, представляется менее однозначным и заслуживает дальнейшего исследования.

Заключение

Проведенный анализ позволяет сформулировать следующие выводы.

Во-первых, контент, сгенерированный нейросетью, может быть признан экстремистским материалом по действующему законодательству, поскольку закон не устанавливает требований к способу создания экстремистского материала, а экстремистский характер является свойством содержания, а не обстоятельством происхождения материала. Способ создания контента (человеком либо нейросетью) не влияет ни на содержательный, ни на целевой признак экстремистского материала.

Во-вторых, стохастическая природа генеративных нейросетей, способных создавать практически неограниченное число уникальных вариаций контента сходной смысловой направленности, делает малоэффективной существующую процедуру идентификации экстремистских материалов, основанную на текстуальном тождестве с записью в Федеральном списке экстремистских материалов. Применительно к нейросетевому контенту критерием идентификации целесообразно признать содержательное (смысловое) соответствие ранее признанному экстремистским материалу.

В-третьих, установление субъективной стороны преступления при использовании генеративных нейросетей требует дифференцированного подхода в зависимости от характера взаимодействия лица с нейросетью и содержания его умысла. Определяющим для квалификации является психическое отношение лица к материалу в момент его распространения. Отсутствие осознания лицом экстремистского характера распространяемого материала исключает уголовную ответственность по ст. 280 и 282 УК РФ, а распространение такого материала с косвенным умыслом при буквальном толковании закона не образует состава соответствующего преступления, что обнаруживает недостаточную приспособленность действующего законодательства к реалиям использования технологий искусственного интеллекта.

В-четвертых, разграничение создания и распространения экстремистского материала применительно к нейросетевому контенту целесообразно проводить на основе критерия степени содержательного контроля лица над результатом генерации.

В-пятых, целенаправленное обучение нейросетей на массивах экстремистских материалов обладает высокой общественной опасностью, однако не охватывается действующими нормами Особенной части Уголовного кодекса Российской Федерации, что свидетельствует о наличии пробела в уголовно-правовом регулировании, требующего законодательного устранения.

Изложенное свидетельствует о том, что распространение технологий генеративного искусственного интеллекта требует переосмысления ряда базовых для уголовно-правового противодействия экстремизму положений. Сформулированные в настоящей статье выводы относительно содержания экстремистского материала, критериев его идентификации и особенностей субъективной стороны деяния образуют необходимую материально-правовую основу для последующего исследования процессуальных и криминалистических особенностей доказывания по уголовным делам рассматриваемой категории.

Библиография

1. Ахметшин А.С. Уголовно-правовое противодействие преступлениям с признаками экстремистской деятельности, совершаемым с использованием информационно-телекоммуникационных сетей, в том числе сети Интернет // Вестник Казанского юридического института МВД России. 2023. № 2 (52). С. 158–165.
2. Борисов С.В., Жеребченко А.В. Квалификация преступлений экстремистской направленности: учебное пособие. Москва: Волтерс Клувер, 2011. 304 с.
3. Кабанов П.А. Политическая преступность: понятие, сущность, виды, причины, личность политического преступника, меры противодействия (криминологическое исследование): автореф. дис. ... д-ра юрид. наук: 12.00.08. Екатеринбург, 2008. 60 с.
4. Мансуров Г.З., Головизнин А.В. Робоправо: правовые основы искусственного интеллекта. Екатеринбург: Уральский государственный экономический университет, 2022. 117 с.
5. Морхат П.М. Искусственный интеллект: правовой взгляд. Москва: Буки Веди, 2017. 257 с.
6. О противодействии экстремистской деятельности: Федеральный закон от 25 июля 2002 г. № 114-ФЗ // Собрание законодательства Российской Федерации. 2002. № 30. Ст. 3031.
7. О судебной практике по уголовным делам о преступлениях экстремистской направленности: постановление Пленума Верховного Суда Российской Федерации от 28 июня 2011 г. № 11 // Бюллетень Верховного Суда Российской Федерации. 2011. № 8.
8. Рарог А.И. Квалификация преступлений по субъективным признакам: монография. Санкт-Петербург: Юридический центр Пресс, 2003. 304 с.
9. Русскевич Е.А. Уголовное право и «цифровая преступность»: проблемы и решения. 2-е изд. Москва: ИНФРА-М, 2022. 351 с.
10. Уголовный кодекс Российской Федерации от 13 июня 1996 г. № 63-ФЗ // Собрание законодательства Российской Федерации. 1996. № 25. Ст. 2954.
11. Хлебушкин А.Г. Экстремизм: уголовно-правовой и уголовно-политический анализ: монография. Саратов: Саратовский юридический институт МВД России, 2007. 159 с.

Criminal-Legal Characterization and Problems of Qualifying the Distribution of Extremist Materials Created Using Artificial Intelligence Technologies

Konstantin I. Parfenov

Senior Lecturer,
Department of Information Support of Internal Affairs Bodies,
Ural Law Institute of the Ministry of Internal Affairs of Russia,
620057, 66, Korepina str., Yekaterinburg, Russian Federation;
e-mail: russia.pki@gmail.com

Abstract

The article examines the problems of the criminal-legal characterization and qualification of acts related to the distribution of extremist materials created using artificial intelligence technologies. The question is analyzed of whether content generated by a neural network can be recognized as extremist material in the meaning given to this concept by the Federal Law "On Countering Extremist Activity." Based on the analysis of the legislative definition, the substantive and purposive features of extremist material are identified, and the conclusion is substantiated that an extremist character represents a property of the content, rather than a circumstance of the material's origin. Difficulties are identified that are caused by the stochastic nature of generative neural networks, capable of creating an unlimited number of unique variations of content, and the necessity of transitioning from a textual to a substantive criterion for identifying extremist materials is substantiated. Particular attention is paid to the problems of establishing the subjective element of the crime: the author identifies and consistently analyzes typical situations that differ in the nature of the person's interaction with the neural network and in the content of intent. The question of distinguishing between the creation and distribution of extremist material in relation to neural network content is examined, and a differentiated approach is proposed based on assessing the degree of the person's substantive control over the result of generation. An assessment is given of the public danger of purposefully training neural networks on arrays of extremist materials, and the conclusion is substantiated that there is a gap in criminal legislation. The question of the expediency of introducing criminal liability for developers and operators of generative neural networks is analyzed.

For citation

Parfenov K.I. (2026) *Ugolovno-pravovaya kharakteristika i problemy kvalifikatsii rasprostraneniya ekstremistskikh materialov, sozdannykh s ispol'zovaniem tekhnologiy iskusstvennogo intellekta* [Criminal-Legal Characterization and Problems of Qualifying the Distribution of Extremist Materials Created Using Artificial Intelligence Technologies]. *Voprosy rossiiskogo i mezhdunarodnogo prava* [Matters of Russian and International Law], 16 (7A), pp. 551-563. DOI: 10.34670/AR.2026.79.95.057

Keywords

Extremist materials, generative neural networks, artificial intelligence, criminal-legal qualification, distribution of extremist materials, subjective element of a crime, creation of extremist materials, neural network content, countering extremism.

References

1. Akhmetchin, A.S. (2023). *Ugolovno-pravovoe protivodeystvie prestupleniyam s priznakami ekstremistskoy deyatel'nosti, sovershaemym s ispol'zovaniem informatsionno-telekommunikatsionnykh setey, v tom chisle seti Internet* [Criminal law counteraction to crimes with signs of extremist activity committed with the use of information and telecommunication networks, including the Internet]. *Vestnik Kazanskogo yuridicheskogo instituta MVD Rossii*, (2(52)), 158–165.
2. Borisov, S.V., & Zhrebchenko, A.V. (2011). *Kvalifikatsiya prestupleniy ekstremistskoy napravlenosti* [Qualification of extremist crimes]. Moscow: Volters Kluver.
3. Federal Law No. 114-FZ of July 25, 2002 "On Countering Extremist Activity". (2002). Collection of Legislation of the Russian Federation, (30), Art. 3031.
4. Kabanov, P.A. (2008). *Politicheskaya prestupnost': ponyatie, sushchnost', vidy, prichiny, lichnost' politicheskogo prestupnika, mery protivodeystviya (kriminologicheskoe issledovanie)* [Political crime: concept, essence, types, causes,

- the personality of a political criminal, countermeasures (a criminological study)] (Author's abstract of Doctor of Law dissertation). Yekaterinburg.
5. Khlebushkin, A.G. (2007). Ekstremizm: ugovno-pravovoy i ugovno-politicheskiy analiz [Extremism: a criminal law and criminal policy analysis]. Saratov: Saratovskiy yuridicheskiy institut MVD Rossii.
 6. Mansurov, G.Z., & Goloviznin, A.V. (2022). Robopravo: pravovye osnovy iskusstvennogo intellekta [Robot law: the legal foundations of artificial intelligence]. Yekaterinburg: Ural'skiy gosudarstvennyy ekonomicheskiy universitet.
 7. Morkhat, P.M. (2017). Iskustvennyy intellekt: pravovoy vzglyad [Artificial intelligence: a legal view]. Moscow: Buki Vedi.
 8. Rarog, A.I. (2003). Kvalifikatsiya prestupleniy po sub"ektivnym priznakam [Qualification of crimes on subjective grounds]. Saint Petersburg: Yuridicheskiy Tsentr Press.
 9. Resolution of the Plenum of the Supreme Court of the Russian Federation No. 11 of June 28, 2011 "On Judicial Practice in Criminal Cases Concerning Crimes of an Extremist Nature". (2011). Bulletin of the Supreme Court of the Russian Federation, (8).
 10. Russkevich, E.A. (2022). Ugolovnoe pravo i "tsifrovaya prestupnost": problemy i resheniya [Criminal law and "digital crime": problems and solutions] (2nd ed.). Moscow: INFRA-M.
 11. The Criminal Code of the Russian Federation of June 13, 1996 No. 63-FZ. (1996). Collection of Legislation of the Russian Federation, (25), Art. 2954.