Запрещенный контент социокиберфизических систем: методы нейросетевой обработки информации

[1]В. С. Аверьянов, email: averyanov124@mail.ru
[1,2,3,4]И. Н. Карцан, email: kartsan2003@mail.ru

1Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, Красноярск, Россия
2Морской гидрофизический институт РАН, Севастополь, Россия
3ФГБНУ «Экспертно-аналитический центр», Москва, Россия
4ФГАОУ ВО «Севастопольский государственный университет», Севастополь, Россия

Аннотация. *Современные социокиберфизические системы являются мощным источником информационного воздействия на крупные гетерогенные и географически рассеянные целевые аудитории. Существующий инструментарий виртуальной реальности способен погрузить неподготовленного пользователя в иной мир, оказывая резко негативное влияние на его психоэмоциональное состояние. Формирование необходимой жизненной позиции и алгоритмов поведения одна из основных задач агитационной деятельности запрещенных экстремистских организаций и террористических группировок. Особенно остро данная проблематика затрагивает социально незащищенные слои населения и представителей поколения Z. Транслируемый контент по глобальной информационной сети разнообразен по своему содержанию и формам представления. Настоящее исследование направлено на выявление потенциально опасной информации с деструктивным воздействием, содержащейся в популярных масс-медиа. В статье рассмотрен метод нейросетевой обработки данных как основа противодействия антигуманистических идеологий в сети Интернет.*

Ключевые слова: *запрещенный контент, киберфизические системы, социальные сети, информация, нейросеть, поколение Z.*

Введение

Социокиберфизические системы разнородны по своему внутреннему содержанию и географическому расположению. В них может быть размещена любая информация, в том числе запрещенная, со сценами насилия, призывами к сепаратизму, экстремистской и террористической деятельности. По состоянию на декабрь 2021 года

согласно официальной статистике международной организации ООН (Организация Объединённых Наций) в мире зарегистрировано более 4,2 млрд. аккаунтов в различных мессенджерах и социальных сетях. За последний год наиболее популярные интернет площадки привлекли более полумиллиарда новых пользователей, большая часть из которых относится к поколению Z. При этом 95% из числа зарегистрированных уже имеют два и более аккаунта в социальных сетях. К примеру, в России на одного пользователя приходится семь регистраций, в Японии самое низкое среднее количество учётных записей - 3,8, на первом месте Индия с результатом в 11,5 записей на одного человека в разнообразных масс – медиа структурах.

По современным меркам социальная сеть является Speech платформой, в которой находят свое отражение, представление и визуализацию социальные взаимоотношения [4]. При этом за последнее десятилетие прослеживается негативная тенденция, когда информационно – коммуникационные средства выступают агрегатором всевозможных террористических движений и агитационных компаний. Деятельность данных организаций направлена на деструктивное воздействие на социально незащищенные слои населения, и является запрещенной на территории большинства государств. К числу самых известных международных террористических группировок следует отнести: ИГИЛ (исламское государство Ирака и Леванта), «Джебхат ан - Нусра», «Ахлю Сунна Валь Джамаа», религиозная группа «Джамаат «Красный пахарь»» и многие другие. Всего в Едином федеральном списке, в том числе иностранных и международных организаций, признанных в соответствии с законодательством Российской Федерации террористическими содержатся более 30 особо опасных группировок, многие из которых действуют по всему миру и существуют не один год.

Для распространения своей идеологии лидеры группировок широко применяют информационно - коммуникационную среду. При этом проведение широкомасштабной глобальной цифровизации, существующий политический строй, а также мировая экономическая нестабильность лишь усугубляют ситуацию вцелом. Данные факторы подталкивают организаторов к проведению новых агитационных компаний с призывом к вступлению в свои «священные» ряды, привлекая обширную аудиторию социокиберфизических систем. Так, например террористическая организация ИГИЛ (запрещена на территории Российской Федерации), являясь наиболее опасной и структурированной, для распространения своей идеологии и пропаганды использует модель повсеместного участия в социальных сетях. Любые их действия и акции мгновенно становятся доступны

миллионам пользователей по всему миру. При этом основную - «грязную» работу по распространению запрещенной информации выполняют не лидеры группировок, а их непосредственно вовлеченные сторонники, простые легитимные пользователи или владельцы фейковых аккаунтов. Официальные представители запрещенных организаций за денежное вознаграждение занимаются созданием информационного поля в виде: новостей, монтажа видеосюжетов с мест реальных боевых действий, фото и видео обращений лидеров группировок. Подготовленный материал через защищенные каналы связи [1] попадает к активным членам и сторонникам, которые в свою очередь транслируют запрещенную информацию на большую аудиторию, используя закрытые сообщества и профильные форумы. Именно так выглядит схема распространения запрещенного контента в социальных сетях. В табл.1 представлены выявленные противоправные материалы и акты цензуры за 2021 год.

Социальная сеть	Удалено материалов	Кол – во материалов на 1 пользователя	Аудитория
Facebook	2306	15177	35000000
YouTube	4624	17301	80000000
Twitter	192	41666	8000000
Vkontakte	496	149193	74000000
Tik – Tok	87	344827	30000000
Одноклассники	83	554216	46000000

Табл.1 Выявленный в социокиберфизических системах запрещенный деструктивный материал, 2021 год

При этом наиболее подвержены влиянию, агитационному воздействию и восприимчивые к пропагандистским антигуманистическим взглядам следующие слои населения:

– поколение Z (подростки в возрасте от 9 до 15 лет);

– учащиеся средне – специальных и высших учебных заведений, преимущественно мусульмане;

– безработные, социально незащищенные слои населения;

– граждане, пострадавшие от противоправных действий властей;

– лица без образования, ранее осужденные, трудящиеся на производстве по простым специальностям (в возрасте до 30 лет преимущественно);

– деклассированные личности.

– Факторы, способствующие к вступлению в ряды «священного джихада»:

– религиозный фактор;

– бандитские побуждения;

– борьба за равенство и социальную справедливость;

– возможность «легкого» заработка;

– иное, случайное событие.

Как отмечает советник Председателя Национального антитеррористического комитета, генерал – лейтенант Е.П. Ильин «противодействие идеологии насилия в современных условиях не может быть задачей только одного государства. Необходимо приложить значительные усилия к тому, чтобы привлечь к участию в антитеррористической работе все здоровые силы общества, в том числе научное, бизнес – сообщества, образовательные структуры, средства массовой информации, общественные, религиозные объединения и организации» [2,3].

1. Архитектура нейронной сети

Целью настоящего исследования является анализ социокиберфизических систем на предмет запрещенных агитационных материалов, выявление возможных связей зарегистрированных пользователей социальных сетей с экстремистскими организациями. Алгоритмы поиска, установление семантических связей, а также сопоставление слов и терминов с экстремистской антигуманистической идеологией содержащейся в аккаунтах необходимо выстроить на нейросетевых технологиях. Нейронная сеть является легко обучаемой, позволяя эксперту в автоматическом режиме подбирать искомое слово или термин по вводным данным. Все современные социальные сети широко применяют технологии хранения и обработки BigData, а также содержат объемные тексты на одном из естественных государственных языков. Для оптимальной работы нейросетевого алгоритма, слова необходимо представлять в векторном формате, что позволит проводить:

– сопоставление аккаунтов с реальными личностями;

– выявлять агитационные идеологические материалы;

– сопоставлять экстремистские организации с реальными аккаунтами пользователей;

– определять географическое местоположение запрещённого сообщества, аккаунта пользователя;

– иные, определяющие значения интересующие эксперта/ов.

Основной принцип работы алгоритма заключается в отображении слова v в векторное пространство отправленного на вход системы.

$V : x_i = (x_i^1, x_i^2, ..., x_i^V), i = 1...z$, где V – общее количество слов в словаре [6], $x_i^j = 1$ при условии что, слово является j-ым словом из V, $x_i^f = 0$, для $f \neq j$. При этом на выходе получим векторное пространство следующей размерности: $V : y = (y^1, y^2, ..., y^V)$. Обучающая выборка имеет вид: $y^j = 1$, при условии что, искомое слово является j-ым словом из V, [5] для всех остальных случаев $y^f = 0$, где $f \neq j$.

Таким образом, результатом работы нейросетевого алгоритма является векторное пространство $V : y = (y^1, y^2, ..., y^V)$ бесконечно большой размерности. Компоненты $y^1, y^2, ..., y^V$ в составе данного пространства служат параметрами контекстного смысла в исследуемом корпусе текстов. Требуемый экспертной группой набор искомых данных возможно получить с помощью динамичного обучения определенно заданной нейросети за короткий временной интервал. Здесь работает следующая аксиома: чем дольше обучение, тем быстрее каждый последующий поиск запрещенного контента либо социально опасного сообщества.

Заключение

За прошедшие два десятилетия внешний облик и внутреннее содержание международного террористического движение изменилось до неузнаваемости. Стремительное развитие социокиберфизических систем придало большое, невиданное доселе распространение агитационным компаниям запрещенных организаций, а о совершенных деструктивных воздействиях мы узнаем быстрее, чем данная информация появляется в официальных источниках СМИ (средства массовой информации). Применение нейросетевых технологий, а также технологий искусственного интеллекта призвано значительно сократить время на поиск запрещенного контента в социальных сетях – например, сцен насильственного характера, призывам к самоубийству и употреблению наркотических веществ, помочь с выявлением сообществ с агитацией сепаратизма, экстремизма и призыву к террористическим действиям. При этом определяющее решение по блокировке и удалению запрещенного контента, а также каждому из выявленных признаков нарушения закона следует отдать экспертной группе, на основании всей собранной информации.

Благодарности

Работа выполнена в рамках государственного задания Минобрнауки России по теме «Концептуальное моделирование информационно-образовательной среды воспроизводства человеческого капитала в условиях цифровой экономики» (Шифр FNRN – E). Работа выполнена в рамках государственного задания Минобрнауки России по теме «Разработка новых методов автономной навигации космических аппаратов в космическом пространстве» (Шифр FNRN – S). Работа выполнена в рамках государственного задания по теме № 0555-2021-0005.

Литература

1. Аверьянов В.С. К вопросу создания защищенного канала связи // Решетневские чтения: Материалы XXIII Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно – космических систем ак. М.Ф.Решетнева, 2019. – С. 409-411.
2. Ильин Е.П. Средства массовой информации в системе противодействия идеологии терроризма. Состояние и перспективы развития // Вестник Национального антитеррористического комитета, 2013, №2 (09). С. 8-13.
3. Ильин Е.П. Об оценках террористических угроз и подходах Российской Федерации к противодействию терроризму // Вестник национального антитеррористического комитета, 2012, №2 (07). С. 10-15.
4. Казнова Н.Н., Овчинникова И.Г. Специфика коммуникации в социальных сетях по сравнению с блогосферой. Вопросы психолингвистики, 2014, №21. С. 86-97.
5. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionally [Электронный ресурс] – Режим доступа : https://papers.nips.cc/paper/5021-distributed-representation-jf-words-and-phrases-and-their-compositionally.pdf (дата обращения 11.12.2021)
6. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097 – 1105.