Полицейская деятельность
Правильная ссылка на статью:
Баумтрог В.Э., Еськов А.В., Смирнов Ю.А. Прототип системы поиска и обнаружения экстремистских сообщений в социальной сети ВКонтакте // Полицейская деятельность. 2024. № 5. С. 98-109. DOI: 10.7256/2454-0692.2024.5.71460 EDN: FLLZDX URL: https://nbpublish.com/library_read_article.php?id=71460

Прототип системы поиска и обнаружения экстремистских сообщений в социальной сети ВКонтакте

Баумтрог Виктор Этмонтович

кандидат физико-математических наук

профессор; кафедра информатики и специальной техники; Барнаульский юридический институт Министерства внутренних дел Российской Федерации

656052, Россия, Алтайский край, г. Барнаул, ул. Чкалова, 49, каб. 432

barnaul@list.ru

Еськов Александр Васильевич

доктор технических наук

Начальник кафедры информационной безопасности ; Краснодарский университет МВД России

350005, Россия, Краснодарский край, г. Краснодар, ул. Ярославская, 128

alesc72@mail.ru

Смирнов Юрий Александрович

Научный сотрудник отдела передачи данных центра средств и систем связи; Научно-исследовательский институт специальной техники ФКУ НПО "СТиС" МВД России

111024, Россия, г. Москва, ул. Прудские Ключики, 2

yra.smirnov01@mail.ru

Статья из рубрики "Информационное обеспечение деятельности полиции"

DOI:

10.7256/2454-0692.2024.5.71460

EDN:

FLLZDX

Дата направления статьи в редакцию:

12-08-2024

Аннотация: Объектом исследования являются нейронные сети, платформа ВКонтакте,

мессенджер Telegram, язык программирования Python и его библиотеки, структурная схема модели компьютерной системы. Предметом исследования является компьютерная технология обнаружения экстремистского контента в текстовом виде и конкретных групп его содержащих в социальной сети ВКонтакте. Авторы подробно рассматривают структурную схему модели компьютерной системы, входящие в неё функциональные модули, иллюстрируют их взаимодействие. В работе используется предварительно обученная модель, предназначенная для обработки русского языка, приводятся условия обеспечения при помощи неё высокой точности распознавания неправомерного контента без признаков переобучения. В работе приводятся результаты проверки тестовых данных, подтверждающих работоспособность компьютерной системы. Предлагаемый прототип компьютерной системы обеспечивает его интеграцию с мессенджером Telegram, что повышает удобство использования и облегчает процесс формирования запросов и отчётов. В ходе исследования использовались общенаучные методы, анализ предметной области, создание модели компьютерной системы, бинарная классификация, эмпирическое тестирование прототипа, систематизация сведений. Новизна исследования заключается в создании прототипа компьютерной системы поиска и обнаружения экстремистских сообщений в социальной сети ВКонтакте, использующей язык программирования Python и программный интерфейс ВКонтакте API (VK API). Основой прототипа компьютерной системы является нейронная сеть, работающая с библиотеками Transformers (предоставляет инструменты и интерфейсы для их простой загрузки и использования) и Torch (современная библиотека глубокого обучения). Особенностью компьютерной системы является возможность анализировать сообщения в социальной сети и подвергать их бинарной классификации на предмет содержания или не содержания в сообщениях противоправной информации. Основные выводы исследования показывают работоспособность системы, простоту и удобство её использования, возможность обнаружения неправомерного текстового контента. Отличительной особенностью прототипа является возможность обнаруживать неправомерный контент, изложенный с использованием сленговых выражений.

Ключевые слова:

ВКонтакте, экстремистский контент, прототип компьютерной системы, нейроная сеть, библиотека, неправомерный, бинарная классификация текста, глубокое обучение, бот, платформа

Введение

Социальная сеть ВКонтакте (VK) широко используется и особенно популярна среди русскоязычных пользователей по всему миру (Социальная сеть «В контакте» и ее аудитория. URL: https://www.seowizard.ru/blog/faq/wiki/c/socialnaya-set-v-kontakte-i-ee-auditoriya). Отдельные публикации в этой сети могут повлечь за собой юридическую ответственность для владельцев платформ и ресурсов, на которых они размещены (Правовые основы противодействия экстремизму и терроризму. URL: https://252.56.xn--b1aew.xn--p1ai/news/item/45436942), в этой связи возникает необходимость обнаружения и контроля контента, имеющего противоправную специфику. Для эффективного мониторинга и управления этим контентом необходимы сложные инструменты и методы, которые, несмотря на все усилия [2,3,4,5,6,7,8,9], оказываются недостаточными. Для решения проблем мониторинга и управления контентом в социальной сети ВКонтакте авторами настоящей работы создан и апробирован прототип

компьютерной системы, которая способна выявлять и предотвращать распространение незаконной информации.

Выбор инструментов

Языком программирования для прототипа компьютерной системы был выбран скриптовый язык программирования Python, так как он универсален и подходит для решения задач на разных платформах, включая iOS и Android, а также серверные операционные системы и позволяет использовать нейронную сеть. Кроме того, в нем уже есть библиотека, отвечающая за работу с программным интерфейсом социальной сети ВКонтакте (API ВКонтакте) – библиотека Transformers с программной платформой Torch (Введение в библиотеку Transformers и платформу Hugging Face. [Электронный ресурс]. URL: https://habr.com/ru/articles/704592/). Интерфейс API ВКонтакте является мощным инструментом для разработчиков и исследователей, способен обеспечить доступ к обширному количеству данных этой социальной сети. API ВКонтакте представляет собой интерфейс, который позволяет извлекать информацию из базы данных vk.com через HTTP-запросы к специализированному серверу [5, 8, 10]. Синтаксис запросов и тип возвращаемых данных строго определены сервисом. Следующим использованным инструментом является платформа Hugging Face, которая представляет собой коллекцию готовых современных предварительно обученных моделей глубокого обучения (deep learning). Глубокое обучение— это вид машинного обучения, в основе которого лежит анализ данных через многослойные сети, похожие на человеческий мозг. Суть deep learning в том, что компьютеры самостоятельно находят решения. Они учатся на собственных ошибках и делают каждый раз все более точные прогнозы (Deep learning: что это, как работает и где применяется. URL: https://getcompass.ru/blog/posts/deep-learning). Библиотека Transformers предоставляет инструменты и интерфейсы для их легкой загрузки и использования.

Обобщенная схема модели и порядок взаимодействия ее модулей

Модель системы приведена на рис. 1. В качестве пользовательского интерфейса выбран мессенджер Telegram благодаря его удобному и простому языку для создания ботов, а также его доступности на персональных компьютерах и смартфонах [10, 11]. Кроме того, платформа Telegram уже имеет встроенную систему регистрации и проверки пользователей, что облегчает процесс аутентификации. При отправке пользователем сообщения, содержащего ссылку на группу ВКонтакте, ссылка передается на сервер. С использованием API ВКонтакте сервером разработанной системы собирается необходимое количество сообщений из указанной группы ВКонтакте. На сервере эти сообщения обрабатываются нейросетью, и формируется отчет, который затем возвращается в мессенджер Telegram пользователю.

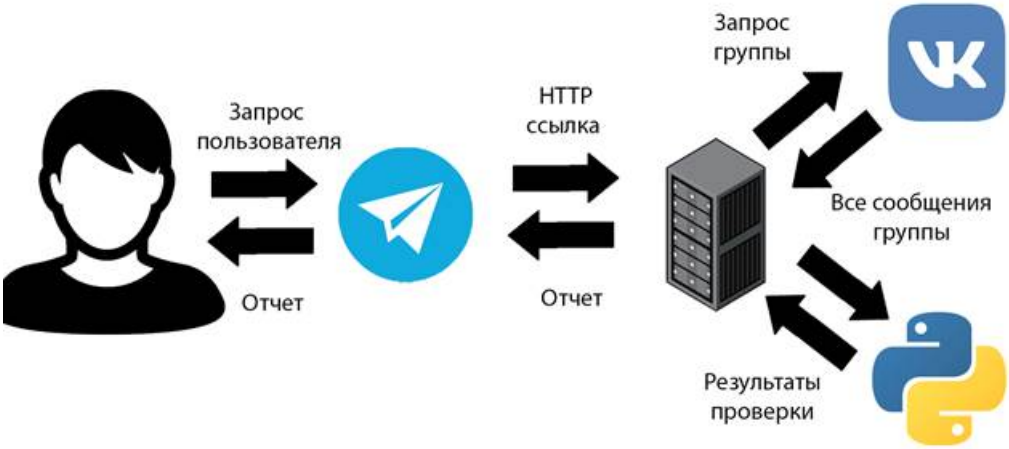

Рис. 1. Обобщенная схема модели компьютерной системы.

Всю структуру процесса обработки и получения данных можно разделить на отдельные функциональные модули, схема взаимодействия которых приведена на рис. 2 и каждый из которых выполняет свою задачу. Модули имеют названия: bot, creating_network, messagedumper, processing, reportgen.

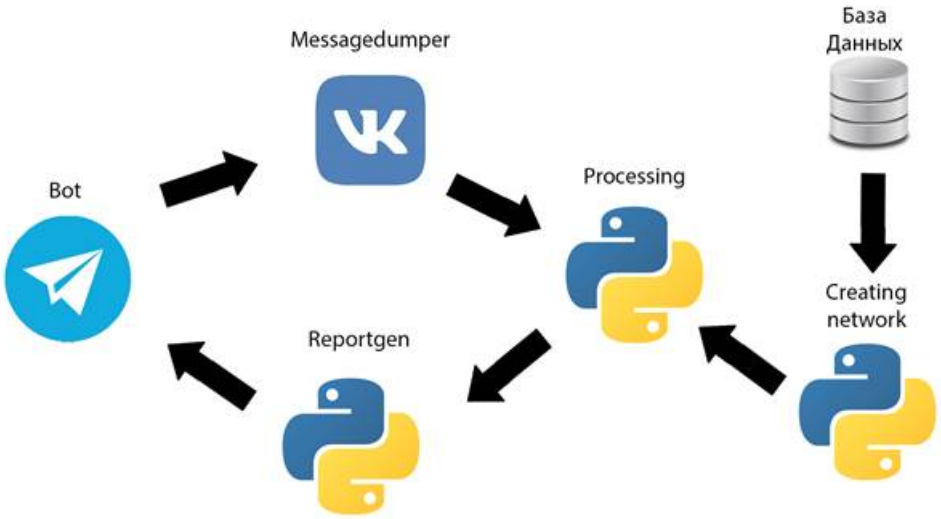

Рис. 2. Схема взаимодействия модулей системы.

На начальном этапе требуется создать исходную базу данных (опорный файл). Для этого был разработан модуль creating_network, который принимает от пользователя на вход файл в формате csv. В этом опорном файле, созданном разработчиками, находятся сообщения (фразы, слова), разделенные на две группы, с соответствующими метками: 1 — информация содержит противоправную информацию и 0 — не содержит. Данные опорного файла позволяют нейросети выполнить бинарную классификациию текста [13, 14, 15, 16, 17, 18, 19].

В работе применяется предварительно обученная модель DeepPavlov/ rubert-base-cased-sentence, предназначенная для обработки русского языка [15, 18]. Данные опорного файла разделяются пользователем на три группы в соотношении 60/20/20:

train — тренировочные данные;

val — валидационные данные;

test — тестовые данные.

Тренировочные данные используются для обучения модели. Однако, чтобы избежать переобучения, создаются валидационные данные. Эти данные поэтапно включаются в тренировочный набор во время обучения модели, пока не будет достигнуто оптимальное соотношение, при котором модель демонстрирует высокую точность распознавания неправомерного контента без признаков переобучения. Тестовые данные играют ключевую роль в оценке эффективности и надежности модели на новых, ранее не введённых данных [19, 20].

Следующий этап включает разделение всех данных на две основные части: сообщения и заголовки. Заголовки используются для установления меток (0 или 1), которые характеризуют контекст или тип сообщения. Важно также провести токенизацию текста, что означает его разбиение на отдельные слова, фразы или другие значимые элементы, чтобы подготовить данные для последующей обработки и анализа.

Полученный текст преобразуется в тензор (в контексте нейронных сетей тензор представляет собой многомерный массив чисел, используемый для хранения и обработки данных) и загружается в DataLoader, который по частям подаёт данные для обучения и валидации модели. Обучать модель пользователю не требуется, возможно добавление своих данных для классификации текста.

В процессе обучения модели генерируются 20 поколений весов, из которых отбираются наиболее успешные. По результатам этого процесса создается файл saved_weights.pt, содержащий веса, оптимальные для бинарной классификации текста. Эти веса затем проверяются на тестовых данных для оценки их эффективности.

Иллюстрация работы прототипа компьютерной системы

Результаты работы модуля с DeepPavlov/rubert-base-cased-sentence формируются в виде чисел с плавающей запятой в диапазоне от 0 до 1. Примечательно, что порог, определяющий правильность распознавания, эмпирически установлен на уровне 0,92: значения до 0,92 считаются отрицательными, а значения выше положительными. Этот порог будет использоваться в дальнейшем как критерий для классификации текста.

Небольшая часть итоговых результатов работы прототипа компьютерной системы приведена на рис. 3, где представлены следующие данные:

messages – сообщения (на рисунке представлены в измененной кодировке);

labels – изначально заданные веса соответствующих сообщений;

confidence – точное значение уверенности модели в распознавании;

pred – итоговый результат распознавания.

Эти компоненты помогают понять, как модель интерпретирует и классифицирует входные данные.

	messages	labels	confidenc	pred
887	вЪ‹вЪ‹Р“	0	0.5109234	0
888	РќРёРєР°Р	1	0.9975583	1
889	РµСЪРёС	0	0.8846463	0
890	вЪ‹вЪ‹Р“	0	0.5850856	0
891	РљС,Ps Pj	1	0.999245	1
892	РљР°Рє	0	0.5923754	0
893	вЪ‹вЪ‹Р“	0	0.6155124	0
894	Р“PsPґ Р°Р	0	0.9385772	1
895	РЈРєСЪР°і	1	0.9994997	1
896	РЈРєСЪР°Р	1	0.9998038	1
897	Р”Р° Р±СѓР	1	0.9770850	1
898	Р’РёР»Р°Р	1	0.9999804	1
899	РЎРµРєС,Р	1	0.9998320	1
900	РµРs PSPs	1	0.9997815	1

Рис. 3. Результаты обработки тестовых данных.

Чаще всего система интерпретирует данные так же, как и пользователь, но встречаются и не совпадающие результаты. Так, например, в строке 894 пользователь определил данные как не содержащие противоправную информацию (Label=0), а система интерпретировала эти данные как содержащие такую информацию.

Модуль, ответственный за работу бота, называется соответственно Bot. Он объясняет, как с ним работать, собирает сообщения и отправляет ссылку на группу «ВКонтакте» для дальнейшей обработки (Круглик Р. И. Создание чат-бота в Telegram // Постулат. 2019. №. 9) [11, 12] . Кроме того, модуль bot присылает результат работы reportgen пользователю, осуществляющему мониторинг доступного контента социальной сети, если была распознана экстремистская или террористическая информация.

Далее Messagedumper собирает последние 10 непустых сообщений сообщества и отправляет серверу на проверку (Рис. 4). Количество сообщений можно увеличить, однако это увеличит и время обработки пропорционально увеличению количества сообщений.

```
import vk_api

def dump_all_messages(url: str, numberOfRows: int):
    access_token = '8c2c541a2c541ae91e7fb970376b8115057af5'
    domain = url
    vk_session = vk_api.VkApi(token=access_token)
    vk = vk_session.get_api()
    id = vk.groups.getById(group_ids=domain)[0]['id']

    posts = vk.wall.get(domain=domain, count=numberOfRows)['items']
    posts_strings = [post['text'] for post in posts]

    comments_strings = []
    for post in posts:
        try:
            comments = vk.wall.getComments(owner_id=id, post_id=post['id'], count=10)['items']
            comments_strings.append([comment['text'] for comment in comments])
        except:
            comments_strings.append([])

    out = []
    for p, c in zip(posts_strings, comments_strings):
        out.append(p)
        for x in c:
            if x != '':
                out.append(x)
    return out
```

Рис. 4. Иллюстрация программного кода процесса сбора сообщений из социальной сети «ВКонтакте».

Processing обрабатывает полученные сообщения от messagedumper, используя веса, взятые из creating_network. Точность распознавания противоправной информации достигает 95%, если сообщения имеют большой объем и схожи по тематике с базой данных. Полученные результаты передаются в модуль reportgen, который формирует отчет и отправляет его боту [21]. Если в процессе обработки не была обнаружена противоправная информация, генерация отчета пропускается, и бот выдает сообщение о том, что группа соответствует установленным нормам.

Заключение

Предложенный прототип компьютерной системы обладает значительным потенциалом для дальнейшего развития и совершенствования в будущем. Технически возможно проверять не только посты групп ВКонтакте, но и комментарии под ними, сохранять статистику пользователей и формировать их общий рейтинг, который мог бы помочь в идентификации террористической или экстремистской деятельности. В настоящий момент Правила платформы запрещают использование пользовательских данных, что не позволяет в данный момент реализовать вышеуказанные возможности. Приведем ограничения, изложенные в разделе 2 «Работа с данными» Правил платформы при использовании публичного API ВКонтакте (Правила использования API ВКонтакте. Редакция от 01.03.2024. URL: https://dev.vk.com/ru/rules).

Приложениям запрещено:

2.1. Собирать и хранить пользовательские данные, включая идентификатор пользователя (User ID), в целях, не связанных с функционированием Приложения. Запрашиваемые данные должны использоваться только в контексте приложения.

Например, кэшировать идентификаторы друзей пользователя, чтобы быстрее отображать список на мобильном устройстве, можно. Передавать идентификаторы всех пользователей к себе на сервер, чтобы хранить в собственной базе на всякий случай, нельзя.

2.2. Передавать любые пользовательские данные, автоматизировано полученные через API (включая User ID), сторонним сервисам (например, рекламным) как напрямую, так и через посредников.

2.3. Использовать пользовательские данные в любых рекламных объявлениях. Например, обращаться к пользователю по имени из рекламного баннера.

2.4. Данные, полученные через API, в т. ч. методами newsfeed.search, wall.get, wall.search, в т. ч. идентификаторы пользователей (User ID), не могут использоваться в целях передачи или перепродажи, создания аналитических отчётов, скоринга и т. д. напрямую или через посредников без прямого согласия Администрации сайта. Таким согласием, например, может быть договор с рекламным агентством на использование данных о показах рекламных объявлений в отчётах клиентам.

Процесс обучаемости разработанной компьютерной системы был реализован в режиме офлайн. Для самообучения в режиме реального времени необходимы достаточно мощное оборудование и постоянное администрирование, что приведёт к увеличению процента распознавания неправомерного текстового контента.

Модульная структура системы позволяет постоянно модернизировать её отдельные компоненты. Схема построения прототипа компьютерной системы может быть применена и для платформ иных социальных сетей.

Возможна перенастройка модели для бинарной классификации текста, не связанного с терроризмом и экстремизмом. Модель работает с исходными данными, поступающими в её входную базу данных. Если потребуется поиск другой информации, смена ключевых слов позволит быстро перенастроить модель в соответствии с необходимыми требованиями.

Иллюстрации 3 и 4 показывают апробацию прототипа компьютерной системы. В итоге обнаружены:

1) Работоспособность системы.

2) Простота использования, в том числе на смартфонах.

3) Достаточно высокая точность поиска неправомерного контента.

4) Способность системы анализировать контент не только по ключевым фразам, но и по семантическому значению текста.

Итак, разработанный прототип компьютерной системы с использованием нейросетевых технологий значительно упрощает процесс обнаружения экстремистской (противоправной) информации в социальной сети ВКонтакте. Система позволяет автоматизировать процесс мониторинга контента, обеспечивая своевременное выявление и реагирование на публикации потенциально опасные или нарушающие правила социальной сети, что позволяет повысить безопасность и обеспечить в используемой платформе соответствие контента нормативным требованиям. Важным преимуществом прототипа является то, что он сконструирован из свободно распространяемого программного обеспечения.

Статья будет интересна исследователям в области информационной безопасности, разработчикам систем автоматизированного мониторинга контента, а также специалистам по анализу социальных сетей. Важно отметить, что практическое применение предложенной системы может вызвать интерес у государственных органов и организаций, занимающихся борьбой с экстремизмом.

Библиография

1. Салахутдинов А. А. Социальные сети как информационный канал экстремистского материала // Молодой ученый. 2014. № 17 (76). С. 561-564.
2. Мартышкин А. И., Маркин Е. И., Зупарова В. В. Исследование и разработка прототипа модуля автоматического отслеживания контента социальных сетей // XXI век: итоги прошлого и проблемы настоящего. 2021. Т. 10. №. 2. С. 96–100.
3. Титов Н. Г. и др. Методы мониторинга социальных сетей, их развитие и применение в контексте обеспечения их информационной безопасности // Информация и безопасность. 2019. Т. 22. №. 3. С. 305–324.
4. Голосной К. С., Янаева М. В. Анализ потенциально опасного контента в социальной сети «ВКонтакте» // Наука, общество, личность: проблемы и перспективы взаимодействия в современном мире. Петрозаводск: МЦНП «Новая наука», 2022. С. 103–107.
5. Остапенко А. Г. и др. Организация мониторинга постов социальной сети ВКонтакте с помощью интерфейса vkapi // Информация и безопасность. 2018. Т. 21. № 3. С. 408–415.

6. Вихляев Д. Р., Глаголев В. А. Парсинг данных сообщества «ВКонтакте» с помощью VK API // Постулат. 2021. № 10.
7. Жданов А. В., Тютякин А. А. Поиск общих групп и сообществ пользователей социальных сетей с использованием веб-сервисов на примере ВКонтакте // Информация и образование: границы коммуникаций. 2017. № 9. С. 74–76.
8. Лехов К. А., Сперанский Д. Д., Митрохин М. А., Карамышева Н. С. Система извлечения и анализа текстовых данных из социальных сетей для образовательного учреждения // Модели, системы, сети в экономике, технике, природе и обществе. 2021. № 1. С. 128–136. doi:10.21685/2227-8486-2021-1-11
9. Сердечный А. Л. и др. Картографический подход исследования процессов распространения деструктивного контента в сообществах единой тематики социальной сети «Вконтакте» // Информация и безопасность. 2020. Т. 23. №. 2. С. 203–214.
10. Биков Д. И. Способы обработки запросов для чат-бота при помощи инструментов VK API // Приоритетные направления инновационной деятельности в промышленности. Часть 2. Казань: ООО «Конверт», 2020. С. 35–36.
11. Козлов А. А., Батищев А. В. Телеграм-бот как простой и удобный способ получения информации //Территория науки. 2017. №. 5. С. 55–64.
12. Шведов Н. Д. Создание простого Telegram бота: пошаговая инструкция //Академическая публицистика, 2023. №. 3-1. С. 7–14.
13. Рабчевский А. Н. Обзор методов и систем генерации синтетических обучающих данных // Прикладная математика и вопросы управления, 2023. №4. С. 6–45.
14. Абдуллах А. Л. И., Соловьева Е. Б. Бинарная классификация текстов с помощью сепарабельной сверточной нейронной сети (BTC_SCNN). Программа для ЭВМ. Свидетельство № 2022613069 от 01.03.2022.
15. Гальченко Ю. В., Нестеров С. А. Классификация текстов по тональности методами машинного обучения // Системный анализ в проектировании и управлении. 2023. Т. 26. №. 3. С. 369–378.
16. Хайкин С. Нейронные сети: полный курс, 2-е изд./ Пер с англ. – М.: ООО «И.Д. Вильямс», 2016.
17. Журавлёв Д.В., Смолин В.С. Нейросетевая революция искусственного интеллекта и варианты её развития// Проектирование будущего. Проблемы цифровой реальности. М.: ИПМ им. М.В. Келдыша, 2023. С. 223-244
18. Куликов А. А., Маилян Э. К. Сравнение архитектур рекуррентных нейронных сетей в задаче бинарной классификации текстов // Инновационное развитие техники и технологий в промышленности (ИНТЕКС-2021). М: РГУ им. А. Н. Косыгина, 2021. Часть 3. С. 223–226.
19. Леготин Д. Л., Зрыбная Е.А. Реализация рекуррентной искусственной нейронной сети для классификации текстов // Актуальные проблемы преподавания информационных и естественно-научных дисциплин. Кострома. КГУ, 2019. С. 197–202.
20. Батура Т. В. Методы автоматической классификации текстов // Программные продукты и системы. 2017. Т. 30. №. 1. С. 85–89.
21. Алексеева В. А. Использование методов интеллектуального анализа в задачах бинарной классификации // Известия Самарского научного центра Российской академии наук. 2014. Т. 16. №. 6-2. С. 354–356.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Статья посвящена разработке и апробации прототипа компьютерной системы для обнаружения экстремистских сообщений в социальной сети «ВКонтакте». Данная система основана на использовании нейросетевых технологий и применении предварительно обученной модели для классификации текстового контента.

Методология исследования подробно описана и включает использование скриптового языка Python, библиотек Transformers и Torch, а также платформы Hugging Face для реализации модели глубокого обучения. Описаны этапы создания базы данных, подготовки данных для обучения модели, а также алгоритм работы прототипа системы. Также подробно раскрыты процесс токенизации текста и методы бинарной классификации, что придаёт исследованию научную строгость и методологическую глубину.

Актуальность темы очевидна в контексте современных вызовов, связанных с необходимостью мониторинга и контроля контента в социальных сетях для предотвращения распространения экстремистской информации. В связи с увеличением количества таких угроз, разработка автоматизированных систем для выявления противоправного контента является важным шагом в обеспечении безопасности пользователей и соблюдении правовых норм.

Научная новизна работы заключается в создании прототипа системы, использующей современные технологии нейросетевого анализа текста для автоматического мониторинга контента в социальной сети «ВКонтакте». Авторы предлагают комплексный подход к обнаружению экстремистской информации, что отличается от существующих решений, ориентированных на использование ключевых слов или простых алгоритмов фильтрации.

Статья написана в научном стиле с соблюдением академической терминологии. Структура статьи логически последовательна и включает введение, описание методологии, иллюстрацию работы прототипа, а также заключение с выводами. Текст статьи местами требует улучшения в части ясности изложения, особенно в разделах, касающихся технических аспектов работы системы, где описание процессов может быть упрощено для лучшего восприятия аудитории.

Выводы статьи соответствуют поставленным задачам и подтверждают работоспособность предложенной системы. Тем не менее, в разделе выводов можно было бы больше акцентировать внимание на перспективности применения разработанной системы в других социальных сетях и потенциальных направлениях её улучшения.

Статья будет интересна исследователям в области информационной безопасности, разработчикам систем автоматизированного мониторинга контента, а также специалистам по анализу социальных сетей. Важно отметить, что практическое применение предложенной системы может вызвать интерес у государственных органов и организаций, занимающихся борьбой с экстремизмом.

Рекомендации по доработке:

1. Рекомендуется упростить и сделать более понятными некоторые технические разделы статьи для широкой научной аудитории.
2. Было бы полезно более детально проанализировать возможные ограничения разработанной системы и предложить пути их преодоления в дальнейших исследованиях.
3. Следует расширить список литературы за счёт включения современных исследований в области использования нейросетевых технологий для анализа текстового контента.

Статья представляет собой значимый вклад в область мониторинга контента социальных сетей. Предложенная система демонстрирует высокие результаты в задаче обнаружения экстремистских сообщений. Рекомендую принять статью к публикации после внесения вышеуказанных доработок.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Предметом рецензируемого исследования выступают технологии обнаружения экстремистских сообщений в социальных сетях. В качестве кейса для исследования была выбрана сеть «ВКонтакте». Автор справедливо связывает высокую степень актуальности выбранной темы с необходимостью разработки технологий обнаружения и контроля противоправного контента с целью противодействия распространению этого контента. Соответственно, рецензируемая работа имеет также и большую практическую значимость, связанную с потенциальными проблемами, к которым могут привести публикации сообщений экстремистской направленности как для владельцев соответствующих коммуникативных каналов, так и для их авторов. В качестве базовых методологических инструментов применялись методы компьютерного моделирования и нейронных сетей. При помощи скриптового языка программирования Python, библиотеки Transformers и платформы Hugging Face была разработана и протестирована (правда, тестирование производилось офлайн в связи с юридическими ограничениями) обучаемая модель системы поиска и обнаружения экстремистских сообщений в социальной сети «ВКонтакте» с использованием API этой сети. Собственно, разработка и апробация прототипа указанной модели вполне может претендовать на научную новизну и практическую полезность. В отличие от других инструментов, разработанная автором модель позволяет анализировать посты в указанной сети посты и комментарии под ними, сохранять статистику пользователей и формировать их рейтинг, что может стать основой дальнейшей работы по выявлению противоправного контента и его авторов. Можно предположить, что данную модель автор рецензируемой статьи планирует масштабировать и на другие социальные сети – «Одноклассники», «Telegram» (тем более, что автор уже работал с этой сетью), «Мой мир», «Яндекс.Дзен», «TikTok» и др. В структурном плане статья также производит положительное впечатление: её логика последовательна и отражает основные аспекты проведённого исследования. В тексте выделены следующие разделы: - «Введение», где ставится научная проблема и аргументируется её актуальность и практическая значимость; - «Выбор инструментов», где достаточно подробно раскрываются методологические и программные инструменты разработки компьютерной модели, а также аргументируется их выбор; - «Обобщенная схема модели и порядок взаимодействия ее модулей», где описаны основные принципы работы модели, а также модули, из которых она состоит; - «Иллюстрация работы прототипа компьютерной системы», где раскрыты результаты апробации прототипа модели; - «Заключение», где описаны юридические сложности, с которыми столкнулся автор, резюмированы итоги проведённого исследования, сделаны выводы и намечены перспективы дальнейших исследований. Стиль рецензируемой статьи научно-аналитический, с сильным креном в сторону технических деталей. В тексте встречается незначительное количество стилистических и грамматических погрешностей, но в целом он написан достаточно грамотно, на хорошем русском языке, с корректным использованием научной терминологии. Библиография насчитывает 21 наименование и в должной мере отражает состояние исследований по проблематике статьи. Хотя и могла бы быть усилена за счёт включения источников на иностранных языках. Апелляция к оппонентам отсутствует, но в силу научно-технического характера статьи не является обязательным требованием. К достоинствам статьи можно отнести также использование иллюстративного материала (четырёх рисунков), существенно упрощающего восприятие

аргументов автора.

ОБЩИЙ ВЫВОД: предложенную к рецензированию статью можно квалифицировать в качестве научной работы, отвечающей основным требованиям, предъявляемым к работам подобного рода. Полученные автором результаты будут интересны для специалистов в области информационной безопасности, в сфере медиа и PR, государственным служащим, а также студентам перечисленных специальностей. Представленный материал соответствует тематике журнала «Полицейская деятельность». По результатам рецензирования статья рекомендуется к публикации.