

DOI 10.58351/2949-2041.2026.34.5.012

Акулов Динис Наилевич, магистрант
АНО ВО «РосНОУ»
Akulov Dinis Nailevich, RosNOU

**РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ВЫЯВЛЕНИЯ
НЕГАТИВНЫХ КОММЕНТАРИЕВ И ПРИОРИТИЗАЦИИ ОТВЕТОВ
В СОЦИАЛЬНЫХ СЕТЯХ (НА ПРИМЕРЕ ВКОНТАКТЕ)
DEVELOPMENT OF AN AUTOMATED SYSTEM FOR DETECTING NEGATIVE
COMMENTS AND PRIORITIZING RESPONSES IN SOCIAL NETWORKS
(USING VKONTAKTE AS AN EXAMPLE)**

Аннотация. В данной статье рассматривается создание доступной автоматизированной системы для мониторинга негативных комментариев в социальных сетях. Обосновывается правовая легитимность применения парсинга через официальные API и показана практическая польза системы для бизнеса в условиях демократизации технологий искусственного интеллекта.

Abstract. The article discusses the creation of an accessible automated system for monitoring comments in social networks. The legitimacy of using parsing methods via official APIs is substantiated, and the high practical utility of the system for business in the context of the democratization of artificial intelligence technologies is proven.

Ключевые слова: Автоматизация мониторинга, анализ тональности, RuBERT, приоритизация данных, NLP, ИИ.

Keywords: Monitoring automation, sentiment analysis, RuBERT, data prioritization.

Введение

В современную эпоху тотальной цифровизации соцсети стали основным полем взаимодействия между бизнесом и клиентами [6, с. 3]. Как для крупных, так и для небольших организаций ежедневный поток входящих сообщений, комментариев на различных площадках исчисляется сотнями и тысячами, что делает ручной мониторинг неэффективным. При этом несвоевременная реакция на критический отзыв или комментарий может привести к значительным репутационным, а также финансовым потерям.

Актуальность представленной работы обусловлена необходимостью создания инструментов, которые не просто фиксируют наличие негативной реакции со стороны клиентов, но и ранжируют ее по степени критичности для удобства и оперативности реагирования. При этом важным вектором в современной разработке является доступность: использование технологий с открытым исходным кодом позволяет сегодня реализовывать системы промышленного уровня с минимальными затратами на инфраструктуру.

Правовые аспекты парсинга

Прежде чем приступить к технической реализации проекта, необходимо очертить правовое поле автоматизированного сбора данных. В Российской Федерации парсинг не запрещен, если он осуществляется в рамках этических и правовых норм [2, с. 51]. Также необходимо отметить, что в российском праве отсутствует отдельный нормативный акт, который бы детально регламентировал процесс сбора данных. На сегодняшний день данная сфера регулируется совокупностью нескольких законодательных актов:

1. Гражданский кодекс РФ: устанавливает нормы об охране авторских прав (глава 70), правовом статусе программ для ЭВМ (статья 1261) и баз данных (статья 1334) [1].

2. Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных»: определяет порядок сбора и обработки сведений о физических лицах [4].



3. Федеральный закон от 27.07.2006 N 149-ФЗ «Об информации, информационных технологиях и о защите информации»: закрепляет правила работы с информационными ресурсами и доступа к ним [3].

Правомерный сбор и обработка данных ориентированы на информацию, открытую для неограниченного круга лиц. Допускается извлечение общедоступных сведений при условии, что это не ущемляет интересы третьих лиц и не противоречит условиям использования сервиса. В связи с этим использование официального API гарантирует, что процесс получения информации соответствует правилам площадки и не нарушает ее работоспособность.

Этапы разработки

Разработка системы была разделена на четыре логических этапа, которые объединяют методы обработки естественного языка и инженерные практики.

Этап 1: Дообучение модели

Центральным элементом системы является модель RuBERT. Выбор обусловлен способностью архитектуры Transformer учитывать контекст слова в предложении, что критически важно для выявления сарказма и иронии. Для дообучения можно собрать и разметить свои данные или использовать общедоступные (например, с платформы Kaggle), соответствующие предметной области бизнеса.

Обучение возможно как на локальном оборудовании, так и в облачных средах (например, Google Colab или Kaggle). Использование облачных GPU делает этот этап доступным даже при отсутствии мощного локального оборудования. Для данного проекта использовался датасет с Kaggle, а обучение осуществлялось в Google Colab.

Этап 2: Парсинг и алгоритм приоритизации

Модуль парсинга реализован на языке Python. Скрипт итеративно использует API ВКонтакте (методы get и getComments [5]), сохраняя новые записи в базу данных (SQLite).

Новизной данной разработки является алгоритм многофакторной приоритизации. Разработанная система не просто помечает комментарий как «негативный», но и присваивает ему уровень важности (Low, Medium, High) на основе суммы ряда факторов:

1. Лексические маркеры: выявление слов-триггеров («обман», «не работает», «пропали деньги», «верните деньги», «суд» и т.д.).
2. Оформление текста: анализ использования Caps Lock (более 40% заглавных букв) и избыточной пунктуации («!!!!», «????»).
3. Социальное одобрение: количество лайков и ответов на комментарий. Так, популярный негативный комментарий требует немедленного вмешательства.
4. Временной контекст: Сообщения, поступившие в нерабочее время (с 20:00 до 09:00), получают инкремент к приоритету, чтобы быть обработанными первыми в начале смены.

Этап 3: Проектирование веб-интерфейса

Для визуализации данных был использован фреймворк Streamlit. Это решение позволило создать полноценный интерфейс на Python, минуя фронтенд-разработку. В интерфейсе реализованы:

1. Дашборд со статистикой (общее количество комментариев, авторов, доля негативных комментариев).
2. Лента комментариев с цветовой индикацией приоритетов.
3. Функция экспорта данных в формат CSV.

Этап 4: Развертывание приложения

Для обеспечения стабильности и удобства развертывания система была упакована в Docker-контейнер. Docker образ включает в себя Python 3.10, библиотеки requests, pandas, transformers, torch и streamlit. Деплой на сервер с использованием Nginx в качестве обратного прокси и Certbot для TLS-шифрования гарантирует безопасность передачи данных. Стоит отметить, что Nginx поддерживает HTTP Basic Auth – встроенный модуль, позволяющий быстро настроить аутентификацию по логину и паролю без каких-либо изменений в приложении.



Ниже представлена структура разработанного проекта:

```
├── app.py
├── docker-compose.yml
├── Dockerfile
├── main.py
├── requirements.txt
├── src
│   ├── parse_vk.py
│   └── rubert-toxic-comments-final
└── start_script.sh
```

В `app.py` формируется веб-интерфейс с использованием Streamlit. Через `main.py` запускается парсинг и обработка комментариев. В `parse_vk.py` расположен код для парсинга и анализа комментариев. В `requirements.txt` перечислены необходимые библиотеки для установки. Запуск скриптов осуществляет при помощи `start_script.sh`. В `rubert-toxic-comments-final` расположена используемая модель. Для удобства сборки и запуска Docker-контейнера создан `docker-compose.yml`.

Разработанная система демонстрирует, что современные технологии ИИ стали доступными. Использование предобученных моделей позволяет избежать значительных затрат на разметку данных и аренду серверов. Кроме того, модератор избавляется от необходимости просматривать большой поток нейтральных сообщений, фокусируясь на действительно критических. При этом исключается человеческий фактор – нейросеть одинаково точно оценивает сообщения в любое время суток.

Заключение

В ходе работы была успешно разработана автоматизированная система мониторинга негативный комментариев для сообщества ВКонтакте. Комбинация модели RuBERT и алгоритма приоритизации позволила создать инструмент, способный эффективно выявлять зоны репутационного и финансового риска в коммуникациях с клиентами. Доступность используемого стека технологий (Python, Docker, RuBERT) делает подобные разработки открытыми для внедрения не только в крупном корпоративном секторе, но и в малом бизнесе

Список литературы:

1. Гражданский кодекс Российской Федерации: федер. закон от 30 нояб. 1994 г. № 51-ФЗ // КонсультантПлюс: [сайт]. – URL: https://www.consultant.ru/document/cons_doc_LAW_5142/ (дата обращения: 18.03.2026).
2. Дятлова Е. В., Янгличева Ю. Р. Парсинг и закон // Вестник экономики, права и социологии. – 2022. – № 2. – С. 49-52.
3. Об информации, информационных технологиях и о защите информации: федер. закон от 27.07.2006 № 149-ФЗ // КонсультантПлюс: [сайт]. – URL: https://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 18.03.2026).
4. О персональных данных: федер. закон от 27.07.2006 № 152-ФЗ // КонсультантПлюс: [сайт]. – URL: https://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 18.03.2026).
5. Официальная документация API ВКонтакте [Электронный ресурс]. – URL: <https://dev.vk.com/ru/api/> (дата обращения: 02.03.2026).
6. Шевченко Д. А. SMM (маркетинг в социальных сетях): стратегия и тактика // Практический маркетинг. – 2022. – № 8 (305). – С. 3-11

