МЕТОДЫ И ИНСТРУМЕНТЫ ДЕТЕКЦИИ МАНИПУЛЯТИВНОГО КОНТЕНТА В ИНТЕРНЕТ-ПРОСТРАНСТВЕ

Тахиров Р.А., *студент магистратуры группы СГНЗ-21М*
Московский государственный технический университет им. Н.Э. Баумана

Научный руководитель: ***Бочарников И.В.****, доктор политических наук, профессор кафедры «Информационная аналитика и политические технологии»*

Аннотация: В статье рассматриваются современные методы выявления манипулятивного контента в интернет-пространстве. Проведён анализ лингвистических, сетевых и поведенческих подходов, выявлены их ограничения при применении в реальных условиях. Предложен комплексный подход к детекции, основанный на интеграции различных типов данных. Показано, что комбинированное использование методов позволяет повысить точность выявления манипулятивных воздействий.

Ключевые слова: манипуляция сознанием, детекция манипулятивного контента, интернет-пространство, социальные сети, машинное обучение, анализ тональности, боты, информационная безопасность.

Современное интернет-пространство, в особенности социальные сети, является не только средой коммуникации, но и инструментом информационного воздействия на массовое сознание. Высокая скорость распространения информации, алгоритмическая фильтрация контента и возможность анонимного взаимодействия создают условия для масштабного распространения манипулятивных сообщений [1, 2].

Согласно Доктрине информационной безопасности Российской Федерации, информационное воздействие может представлять угрозу для личности, общества и государства [3]. На практике это проявляется в распространении дезинформации, координированных информационных кампаний и использовании автоматизированных аккаунтов для формирования общественного мнения.

В этих условиях особую актуальность приобретает задача разработки методов выявления манипулятивного контента. Однако существующие подходы обладают рядом ограничений, что требует их более детального анализа и развития.

МАНИПУЛЯТИВНЫЕ ПРАКТИКИ В ИНТЕРНЕТ-ПРОСТРАНСТВЕ

Манипуляция в научной литературе определяется как скрытое психологическое воздействие, направленное на изменение установок и поведения аудитории при сохранении иллюзии самостоятельного выбора [7, 8]. В условиях цифровой среды данный процесс усложняется за

счёт использования алгоритмов распространения информации и автоматизированных инструментов.

Практика показывает, что манипулятивные воздействия реализуются через различные формы. Одной из наиболее распространённых является **дезинформация** – намеренно ложная или искажённая информация, маскируемая под достоверные источники [9, 10].

Другой формой является **пропагандистский контент**, характеризующийся односторонней интерпретацией событий и использованием эмоционально окрашенной лексики. В отличие от дезинформации, он может содержать реальные факты, однако их подача направлена на формирование заданного восприятия.

Также широко распространён **астротурфинг**, представляющий собой искусственное создание видимости массовой поддержки или общественного мнения. Он реализуется через координированную активность аккаунтов, включая использование социальных ботов [11, 12].

Социальные боты играют ключевую роль в распространении манипулятивного контента. Они позволяют ускорять распространение информации, усиливать её видимость и воздействовать на алгоритмы платформ [11].

Важно отметить, что в реальных условиях манипулятивные кампании используют комбинацию перечисленных механизмов, что существенно усложняет их выявление.

МЕТОДЫ ДЕТЕКЦИИ МАНИПУЛЯТИВНОГО КОНТЕНТА

Современные методы выявления манипулятивного контента можно разделить на три основных направления: лингвистический, сетевой и поведенческий анализ [4, 5].

Лингвистический подход основан на анализе текстового содержания сообщений. Манипулятивный контент часто характеризуется повышенной эмоциональностью, использованием оценочной лексики и специфических речевых конструкций [14]. Методы анализа тональности и машинного обучения позволяют автоматически выявлять такие признаки [15].

На практике данный подход применяется для обнаружения фейковых новостей и пропагандистских материалов. Однако его эффективность снижается при маскировке текста под нейтральный стиль, а также при использовании качественно подготовленного контента [13].

Сетевой подход ориентирован на анализ структуры распространения информации. Он включает построение графов взаимодействия пользователей и выявление аномальных кластеров [11, 12].

Применение данного подхода позволяет обнаруживать группы аккаунтов, действующих согласованно. Например, бот-сети часто харак-

теризуются высокой плотностью связей внутри группы и синхронностью действий.

Основным ограничением является невозможность оценки содержания сообщений: выявленная активность может быть как манипулятивной, так и органической.

Поведенческий подход основан на анализе активности пользователей. Исследуются такие параметры, как частота публикаций, временные интервалы и реакция на события [17].

Боты часто демонстрируют неестественные паттерны поведения, включая равномерную активность и высокую скорость реакции. Однако современные алгоритмы позволяют имитировать человеческое поведение, что снижает эффективность данного подхода.

Сравнительный анализ показывает, что каждый из методов эффективен только в рамках своей области. Их изолированное использование не позволяет надёжно выявлять сложные манипулятивные кампании.

КОМПЛЕКСНЫЙ ПОДХОД К ДЕТЕКЦИИ

Анализ существующих методов показывает, что манипулятивные кампании используют одновременно несколько механизмов: особенности контента, структуру распространения и поведенческие характеристики аккаунтов. Это делает применение одного подхода недостаточным.

В рамках данной работы предлагается комплексный подход к детекции, основанный на последовательной обработке данных.

На первом этапе осуществляется сбор данных из социальных сетей, включая тексты сообщений, информацию об аккаунтах и структуру взаимодействий.

На втором этапе проводится сетевой анализ, направленный на выявление координированных групп пользователей и аномальных структур распространения информации [12].

На третьем этапе применяется поведенческий анализ, позволяющий выявить аккаунты с нетипичной активностью [17].

На заключительном этапе выполняется лингвистический анализ контента, направленный на выявление признаков манипуляции [14].

Объединение результатов всех этапов позволяет сформировать интегральную оценку и повысить точность выявления манипулятивного контента. Использование методов машинного обучения и глубоких нейронных сетей дополнительно расширяет возможности анализа за счёт выявления скрытых закономерностей в данных [18].

Проведённый анализ показал, что проблема выявления манипулятивного контента в интернет-пространстве носит комплексный характер и не может быть эффективно решена с использованием одного метода.

Лингвистические, сетевые и поведенческие подходы обладают существенными ограничениями, которые проявляются при анализе современных информационных кампаний.

В работе показано, что манипулятивные воздействия реализуются через сочетание различных механизмов, включая особенности текстового содержания, структуру распространения информации и поведенческие характеристики пользователей. Это определяет необходимость интеграции различных методов анализа.

Предложенный комплексный подход, основанный на последовательном применении сетевого, поведенческого и лингвистического анализа, позволяет учитывать многослойную природу манипулятивных процессов и повышает надёжность их выявления.

Практическая значимость работы заключается в возможности использования предложенной схемы в системах мониторинга социальных сетей и аналитических платформах. Перспективы дальнейших исследований связаны с применением методов глубокого обучения и расширением объёма анализируемых данных, что позволит повысить точность и адаптивность систем детекции.

Литература и источники:

1. Ефимова И.Н. Информационные войны в социальных сетях // Власть. – 2022. – № 2. – С. 67–72.
2. Войнов Д.А. Политические интернет-коммуникации: теории и технологии. – М.: Прометей, 2021. – 240 с.
3. Доктрина информационной безопасности Российской Федерации : утв. Указом Президента РФ от 05.12.2016 № 646.
4. Смирнов И.В. Технологии анализа социальных сетей для задач мониторинга // Искусственный интеллект и принятие решений. – 2022. – № 2. – С. 78–85.
5. Логвиненко Ю.А. Модели и методы машинного обучения для распознавания фейкового контента // Вестник науки. – 2024. – Т. 3. – № 5. – С. 112–118.
6. Тахиров Р.А. Технология выявления признаков манипуляции в интернет-пространстве на основе машинного обучения: отчет о НИР. – М.: МГТУ им. Н.Э. Баумана, 2025. – 29 с.
7. Доценко Е.Л. Психология манипуляции: феномены, механизмы и защита. – М.: ЧеРо, 1997. – 344 с.
8. Манойло А.В. Информационные войны и психологические операции. Руководство к действию. – М.: Горячая линия – Телеком, 2021.
9. Свечников В.С. Манипулятивные практики в социальном конструировании реальностей: автореф. дис. ... д-ра социол. наук. – Саратов, 2021. – 38 с.
10. Головацкая О.Е. Методологические подходы к изучению искаженной информации // Коммуникология. – 2023. – Т. 11. – № 1. – С. 15–30.
11. Кочкаров А.А. Выявление ботов в социальных сетях на примере LiveJournal // Мир новой экономики. – 2020. – Т. 14. – № 3. – С. 56–63.
12. Коломеец М.В. Метрики вредоносных социальных ботов // Информатика и автоматизация. – 2023. – Т. 22. – № 1. – С. 15–42.

13. Аминов Д.А., Петрова Н.Е. Методы выявления русскоязычной фейковой информации в СМИ // Доклады БГУИР. – 2024. – № 2 (128). – С. 98–104.

14. Бирюкова Е.В., Воронина И.Е. Анализ тональности текста как метод моделирования русскоязычного нарратива // Интеллектуальные информационные системы: Труды всероссийской конференции. – Воронеж: ВГУ, 2023. – С. 5–12.

15. Кузнецов И.А. Методы и алгоритмы машинного обучения для обработки слабоструктурированных текстовых данных: дис. ... канд. техн. наук. – М., 2022. – 150 с.

16. Ушаков И.А. Методика обнаружения инсайдеров в компьютерных сетях с использованием комбинирования экспертных правил, методов машинного обучения и обработки больших данных: автореф. дис. ... канд. техн. наук. – СПб., 2020. – 24 с.

17. Логинова А.О. Анализ существующих подходов к классификации ботов // Информационные технологии. – 2020. – № 8. – С. 44–50.

18. Федоров П.С. Применение глубокого обучения для обнаружения дезинформации // Системный анализ и управление. – 2023. – Т. 4. – № 1. – С. 11–19.

19. Противодействие фальсификации истории великой отечественной войны / Бочарников И.В., Суздалева Т.Р., Федоров К.В., Криворучко А.А., Петренко А.И., Зеленков М.Ю., Кандыбович С.Л., Разина Т.В., Овсянникова О.А., Трипольский В.Б. Москва, 2020.

20. Ремарчук В.Н. Информационно-аналитическая деятельность: проблемы и перспективы // Вестник Академии военных наук. 2023. № 1 (82). С. 31–35.