

## Распознавание фейков в социальных медиа с помощью технологий управления данными

Научный руководитель – Копытина Евгения

*Копытина Евгения Александровна*

*Выпускник (магистр)*

Финансовый университет, Факультет социальных наук и массовых коммуникаций,  
Москва, Россия

*E-mail: evgenie.kopytina@gmail.com*

В последние десятилетия с развитием интернет-технологий и социальных сетей фейковые новости стали серьёзной угрозой для информационного общества. Они влияют на общественное мнение, политические процессы, экономику и даже национальную безопасность. Фейки распространяются вирусно, часто без возможности быстрого опровержения, что затрудняет принятие решений на государственном уровне, провоцирует беспорядки и создаёт риски для безопасности.[2] Например, ложные новости о террористических угрозах могут вызвать панику и излишнюю мобилизацию сил правопорядка. По данным ВЦИОМ, 74% россиян регулярно сталкиваются с фейковой информацией в СМИ и соцсетях, что подчёркивает актуальность проблемы.[1]

В прошлой статье автором были выявлены маркеры идентификации фейковой информации в условиях информационного противоборства [3]. В данной работе исследование было продолжено.

Для борьбы с фейками предложен алгоритм, основанный на технологиях искусственного интеллекта, в частности, моделях GPT (Generative Pre-trained Transformer). Он анализирует текст, метаданные, мультимедийный контент и поведение пользователей, что делает его эффективным инструментом против дезинформации.[4]

Алгоритм включает несколько этапов. Сначала проводится предобработка данных: очистка текста от лишних символов и токенизация. Если новость содержит изображения или видео, текст извлекается с помощью OCR и транскрибации. Затем GPT анализирует текст на наличие эмоционально заряженных слов, противоречий с известными фактами и манипулятивных техник. Оценивается тональность текста — нейтральный, негативный или манипулятивный.

На этапе контекстного анализа текст сравнивается с данными из проверенных источников, таких как FactCheck.org или PolitiFact. Проверяется соответствие реальным событиям и выявляются манипуляции, например, создание иллюзии общепризнанной правды. Далее анализируются метаданные: репутация источника, время публикации, географическая локализация и активность пользователей. Если новость распространяется через ботов или подозрительные аккаунты, это становится индикатором манипуляции.

Для мультимедийного контента используется компьютерное зрение для выявления подделок, например, deepfakes. На этапе оценки достоверности результаты анализа текста, контекста, метаданных и мультимедиа объединяются. Новость помечается как достоверная (вероятность фейка < 20%), подозрительная (20-70%) или фейковая (> 70%). Формируется объяснение, почему новость может быть фейковой.

Преимущества алгоритма — комплексный подход, использование современных технологий, адаптивность и масштабируемость. Он может быть интегрирован с социальными платформами для мониторинга в реальном времени. Пример применения: новость о том, что "все врачи рекомендуют новый препарат X для лечения COVID-19", анализируется на соответствие данным из проверенных источников. Если выявляются противоречия или подозрительные метаданные, новость помечается как фейковая.

Перспективы развития включают интеграцию с соцплатформами, использование краудсорсинга для повышения точности, регулярное дообучение моделей на новых данных и создание системы рангов для оценки доверия к источникам. Этот подход позволяет не только выявлять фейки, но и минимизировать их влияние на общественное мнение, создавая устойчивую информационную экосистему.

### Источники и литература

- 1) Аналитический обзор «Фейк-ньюс»: масштаб проблемы” // ВЦИОМ. 21.10. 2022 URL - <https://wciom.ru/analytical-reviews/analiticheskii-obzor/fejk-nyus-masshtab-problemy>
- 2) Кошкин П. Сделаем факты снова великими : Можно ли сегодня противостоять фальшивым новостям? // Российский совет по международным делам. 2017. .<https://russiancouncil.ru/analytics-and-comments/analytics/sdelajem-fakty-snova-velikimi-mozhno-li-segodnya-protivostoyat-falshivym-novostyam>
- 3) Сушкевич М.А., Копытина Е.А. Маркеры идентификации фейковой информации в условиях информационного противоборства // современные тенденции развития социальных наук.- Москва: МГУ имени М.В. Ломоносова, 2023. - С. 86-90.
- 4) Devlin, J., Chang, M., Lee, K., Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019. URL: <https://arxiv.org/abs/1810.04805>