Ксения Александровна Болдина

Нижегородский государственный университет им. Н. И. Лобачевского (Нижний Новгород)

[olesova@yandex.ru](https://e.mail.ru/compose?To=olesova@yandex.ru)

**Проблема предвзятости текстовых генеративных нейросетей в контексте журналистской деятельности**

В исследовании рассматривается проблема предвзятости текстовых генеративных нейросетей в контексте журналистской деятельности. На основе контент-анализа ответов ChatGPT, DeepSeek и YandexGPT выявлены особенности идеологической предвзятости, обусловленные страной происхождения каждой системы.

Ключевые слова: искусственный интеллект, генеративные нейросети, медиакоммуникации, новые медиа, журналистская этика.

В современную цифровую эпоху текстовые генеративные нейросети становятся все более востребованным инструментом в журналистской деятельности. Их способность быстро создавать связные тексты, анализировать большие объемы информации и формулировать выводы делает их привлекательными для использования в новостных редакциях. Однако существует серьезная проблема: нейросети, обученные на больших массивах текстовых данных, могут наследовать и усиливать существующие в обществе предубеждения, стереотипы и идеологические установки.

Это подтверждается современными исследованиями в области интеграции ИИ в журналистскую практику. Как отмечает Д. Н. Дроздов, очевидно, что, хотя качество ИИ постоянно улучшается, существует риск наличия неточностей, предвзятости и отсутствия важных деталей в создаваемом содержании. Решающую роль здесь играет страна нахождения серверов конкретной нейросети и гражданство ее собственника (индивидуального или коллективного) [1].

Данное наблюдение указывает на серьезную проблему: алгоритмические системы неизбежно отражают геополитические и идеологические установки своих создателей. Подтверждение этого тезиса можно найти и в исследовании Е. В. Шибановой, которая подчеркивает, что ограничения и риски при использовании сгенерированного контента в журналистике требуют комплексного научного изучения и четкой этической стратегии. Необходим контроль ИИ и его системный мониторинг для обеспечения качества и объективности контента [2]. Автор справедливо указывает на необходимость создания системы сдержек и противовесов при использовании ИИ в медиа, поскольку бесконтрольное применение этой технологии может привести к массовому распространению предвзятой информации и искажению общественного мнения.

На основе этих наблюдений мы провели исследование, направленное на выявление и анализ предвзятости в ответах текстовых генеративных нейросетей при освещении политически чувствительных тем. Его актуальность обусловлена растущей популярностью использования ИИ среди журналистов и авторов телеграм-каналов, которые зачастую применяют эти инструменты без должного понимания рисков и потенциальных последствий. Особую озабоченность вызывает тот факт, что генеративные модели могут транслировать и усиливать определенные политические нарративы, что противоречит принципам объективной журналистики.

Для проведения исследования мы разработали систему вопросов, касающихся внутренней и внешней политики трех ключевых геополитических игроков – России, США и Китая. Эти страны были выбраны как наиболее часто фигурирующие в российском информационном пространстве и вызывающие наибольший интерес у аудитории. В качестве объектов исследования выбраны три популярные нейросети: ChatGPT (OpenAI, США), DeepSeek (DeepSeek, Китай) и YandexGPT (Яндекс, Россия). Такой выбор позволил нам не только проанализировать возможную предвзятость каждой модели, но и сравнить, как нейросети, разработанные в разных странах, освещают одни и те же политические темы.

Проведенный контент-анализ ответов трех нейросетей выявил существенные различия в подходах к освещению политических тем. Каждая система продемонстрировала характерные особенности в обработке политически чувствительной информации:

1. ChatGPT, несмотря на декларируемую нейтральность, показывает заметный прозападный уклон в оценках и суждениях, особенно в вопросах демократии и гражданских свобод.

2. DeepSeek проявляет более выраженную прокитайскую позицию с элементами критики западных стран, особенно США.

3. YandexGPT избегает оценочных суждений и часто отказывается от ответов на потенциально спорные вопросы, что можно интерпретировать как форму самоцензуры.

Примечательно, что все три системы проявляют наибольшую предвзятость именно при обсуждении вопросов, затрагивающих национальные интересы стран их происхождения. На основании результатов исследования можно утверждать, что генеративные нейросети отражают идеологические установки стран их происхождения.

Литература

1. Дроздов Д. Н. Генерируемый искусственным интеллектом контент как составная часть контент-стратегии современных медиа // Журнал Белорусского гос. ун-та. Журналистика. 2024. № 2. С. 25–30.

2. Шибанова Е. В. Экспериментальные подходы к использованию искусственного интеллекта в медиакоммуникациях // Вестник Тверского гос. ун-та. Серия: Филология. 2024. № 1(80). С. 245–250.