

О. А. Блинова (Москва, Россия)

*Российский государственный гуманитарный университет,
Московский государственный институт
международных отношений (университет) МИД РФ
o.blinova@my.mgimo.ru*

МАНИПУЛЯЦИЯ В ИНТЕРНЕТ-МЕМАХ, СГЕНЕРИРОВАННЫХ НЕЙРОСЕТЬЮ¹

Статья рассматривает возможность использования нейросетей для генерации мультимедийного контента с манипулятивным уклоном в медиадискурсе. Анализируются интернет-мемы общественно-политической тематики, созданные с помощью инструмента SupremeMeme.AI. Выявлены не заложенные в промптах оценочность, ирония, прецедентные феномены. Результаты показывают, что нейросети способны создавать контент с манипулятивными оттенками, указывая на необходимость проверки фактов и критического анализа мультимодальных текстов в цифровой среде.

Ключевые слова: интернет-мемы, манипуляция, медиадискурс, мультимодальность, искусственный интеллект, нейросети.

Искусственный интеллект (ИИ) и нейросетевые инструменты за относительно короткое время уже превратились в активного участника онлайн-коммуникации. С помощью нейросетей генерируется значительный объем контента в современном информационном поле, что превращает подобные тексты в актуальный объект исследования лингвистов, тем более что современные нейроинструменты активно вступают во взаимодействие с живыми пользователями интернета. Одной из наиболее затронутой этим процессом сфер оказался медиадискурс, особенно так называемые «новые медиа», социальные сети, где ИИ позволяет генерировать не только моно-, но и мультимодальный контент. Особый интерес представляют мультимодальные медиажанры, в частности интернет-мемы – мультимодальные единицы коммуникации [Голубкова, Канашина 2022], сочетающие широко тиражируемый шаблон с уникальной вербальной компонентой.

Согласно [Pütz, Esposito 2024], коммуникация состоялась, если получатель информации использует результат взаимодействия для создания собственных мыслей и информации, которые, в свою очередь, зависят от их истории, интересов и т.д. и неизбежно отличаются от мыслей и информации других

¹ Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 22–18–00594).

людей, в т. ч. и отправителя. В [Esposito 2022] предлагается позиционировать предмет этой дискуссии не как искусственный интеллект, а как искусственную коммуникацию. Исходя из такого понимания, в настоящем исследовании мы ставим вопрос: может ли нейросеть в рамках диалога с помощью вводимых текстовых промптов использовать полученную информацию для генерирования собственного контента и придавать ему манипулятивный характер применительно к нейросетям (едва ли можно говорить о «мыслях»). Под манипуляцией в данном случае мы будем понимать попытки скрытого воздействия, в частности, придания оценочности, не заложенной в промптах изначально.

В исследовании использовались отобранные интернет-мемы, посвященные резонансным общественно-политическим событиям: предвыборной кампании в США и собственно президентским выборам 2024 года, завершившимся победой Д. Трампа. Из доступных для использования нейросетей был отобран SupremeMeme.AI (<https://supermeme.ai/>), позволяющий генерировать мемы по заданным текстовым промптам. Выдача алгоритма порождает мультимодальный текст с полной либо частичной креолизацией, комбинируя устойчивый визуальный шаблон популярных мемов с вербальной информацией, сгенерированной на основе промптов.

В ходе эксперимента было выявлено, что часть выдачи носит манипулятивный характер. Так, промпт *Trump defeats Biden* позволяет сгенерировать следующее изображение (Рис. 1):

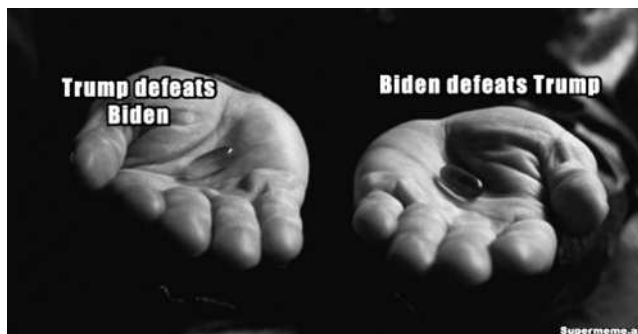


Рис. 1. Мем, сгенерированный нейросетью по промпту *Trump defeats Biden* (вариант А)

Обращает на себя внимание обращение нейросети к шаблонам мемов, основанным на прецедентных текстах. Кинофильм «Матрица» (The Matrix, 1999) стал источником символа: выбора между красной и синей таблетками. Приняв красную таблетку, протагонист избавляется от иллюзий и обретает

свободу и познает правду жизни, но платит за это цену: приняв суровую реальность, он вынужден бороться и выживать в ней. Приняв синюю таблетку, можно оставаться во власти заблуждений и иллюзий, продолжать вести беззаботную, но бесславную и бессмысленную жизнь. Таким образом, в сгенерированном меме нейросеть придает победе одного кандидата (Трампа) оценочный образ «горькой пилюли», в то время как потенциальный выбор в пользу Дж. Байдена трактуется как желание остаться во власти иллюзий и игнорировать реальность. Вероятно, распределение цветов между двумя кандидатами обусловлено цветовой символикой ведущих партий США: красный – цвет Республиканской партии, которую представляет Трамп; синий – традиционный цвет демократов, партии Байдена. Однако в сгенерированном мультимодальном тексте в силу опоры на известный прецедентный текст эти цвета придают победе Трампа дополнительный, оценочный смысл.

Рассмотрим еще один пример мема, сгенерированного на основе того же промпта *Trump defeats Biden* (Рис. 2):

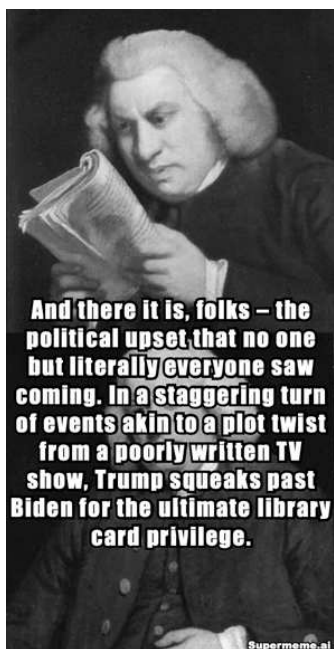


Рис. 2. Мем, сгенерированный нейросетью по промпту *Trump defeats Biden* (вариант Б)

Данное изображение на иконическом уровне опирается на темпоральную стилизацию, архаизируя образ выборов. На вербальном уровне отмечается опора на иронию за счет соседства в одном предложении противоположных по смыслу местоимений (*no one but literally everyone*), вкупе с преувеличением (*in a staggering turn of events*). Победа Трампа при этом сравнивается (*akin*) со слабым киносценарием, т.е. наделяется негативной коннотацией.

Наконец, нейросеть может генерировать мемы, противоположные по оценочности контенту, уже циркулирующему в медиапространстве (с оценочностью, намеренно заданной живыми актерами). Так, покушение на жизнь Д. Трампа во время предвыборной кампании и фотографии победного жеста будущего президента породили огромное количество мемов, где Трамп сравнивается с футбольной звездой Лионелем Месси.

Ниже (Рис. 3) приведем в качестве примера один из оригинальных мемов, который носит ярко выраженный положительный (хотя и шуточный) оттенок, а также сгенерированный нейросетью мем по аналогичному промпту. Можно заметить, что оценочность сгенерированного мема ровно противоположная; кроме того, в роли главного героя, видимо, выступает не Трамп, а промахнувшийся стрелок, значительно уступающий Месси в мастерстве:



а. Пример оригинального мема



б. Мем, сгенерированный нейросетью, промпт: Trump recreated Messi

Рис. 3. Оригинальный мем vs. сгенерированный мем

Таким образом, можно отметить, что нейросети способны создавать контент, пользуясь манипулятивными приемами и придавая оценочность нейтральным текстовым промптам. Это достигается использованием прецедентных феноменов, иронии. В отдельных случаях нейросеть способна создавать контент, по характеру оценки противоположный заданным пром-

штам. Эти результаты демонстрируют важность проверки фактов в сгенерированных текстах, в том числе мультимодальных.

Литература

Голубкова Е. Е., Канашина С. В. Репрезентация концепта УГРОЗА в мультимодальных единицах коммуникации – интернет-мемах // Социокультурные угрозы: реальность, ментальные модели и дискурс. М.: Языки славянской культуры, 2022. С. 369–384.

Esposito E. Artificial Communication: How Algorithms Produce Social Intelligence. Boston: The MIT Press, 2022.

Pütz O., Esposito E. Performance without understanding: How ChatGPT relies on humans to repair conversational trouble // Discourse & Communication. 2024. № 18(6). P. 859–868.

O. A. Blinova (*Moscow, Russia*)

Russian State University for the Humanities, MGIMO University

MANIPULATION IN AI-GENERATED INTERNET MEMES

The article considers the possibility of using neural networks in media discourse to generate multimedia content with manipulative bias. The article uses SupremeMeme.AI to generate political Internet memes. The analysis shows evaluation, irony and precedent images not embedded in the memes prompts. The results show that AI tools can create manipulative content, indicating the need for fact-checking and critical analysis of multimodal texts in the digital environment.

Key words: artificial intelligence, Internet memes, manipulation, media discourse, multimodality, neural networks.