

А. Н. Кузьмин

Институт бизнеса БГУ, Минск, Беларусь

Научный руководитель – А. М. Туровец

ИЗМЕНЕНИЕ ТОНАЛЬНОСТИ ДАННЫХ ЯЗЫКОВЫХ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ НЕЙРОСЕТЕЙ ПРИ РЕШЕНИИ ЛОГИСТИЧЕСКИХ ЗАДАЧ

В работе рассмотрено изменение тональности данных языковых генеративных моделей нейронных сетей при их внедрении в логистический сектор с позиции решения классической логистической задачи и, соответственно, принятия соответствующим лицом управленческого решения. На данный момент тональность стремится к преимущественно позитивной с некоторыми проявлениями негативных сценариев. Вместе с тем установлено, что тональность выходных данных моделей не приближена к нейтральной.

Ключевые слова: логистика, нейронные сети, языковые генеративные модели, тональность текстовых данных

Цифровизация экономики – важнейшая составляющая инновационного развития. Одним из ключевых элементов этого процесса выступает постоянное внедрение инновационных технологий, одна из которых – языковые генеративные модели нейронных сетей.

Вопрос о внедрении такого решения в логистический сектор изучается с момента появления данных моделей в открытом доступе, а именно с ноября 2022 г., когда компания OpenAI впервые выложила в открытый доступ свою модель ChatGPT 3.5. Субъекты малого, среднего и крупного бизнеса начали крайне активно интересоваться путями внедрения генеративных моделей в свои бизнес-процессы. К общей тенденции присоединились и логистические компании, как мирового, так и регионального уровней.

Однако с течением времени все более актуальным становился вопрос использования языковых генеративных моделей нейронных сетей в качестве вспомогательного средства с этической точки зрения. Этим, в первую очередь, начали интересоваться различные представители той профессиональной сферы, которая требует постоянной работы с людьми и анализа их поведения с одновременным соблюдением этики. Субъекты бизнеса, ставя основной целью успешное решение задачи максимизации прибыли с сохранением текущего уровня сервиса с возможным последовательным его улучшением для потребителя и используя это инновационное решение в своей деятельности, также пришли к проблеме поиска возможности изменения тональности данных языковых генеративных моделей нейронных сетей, в особенности при принятии управленческих решений различного рода. Проанализируем эту гипотетическую возможность с упором на внедрение в транспортно-логистический сектор.

Выделим несколько подходов анализа проблемы изменения тональности: прямой (с соответствующими запросами) и косвенный (через параметр, косвенно влияющий на тональность, – температуру). Разберем каждый из способов более подробно на примере решения классической логистической задачи «Make or Buy» (MoB), которую модель будет формулировать самостоятельно по корректному запросу. Начальное предположение основывается на ответе одной из языковых генеративных моделей нейронных сетей: при выдаче ответов эмоциональная окраска отсутствует как таковая, ведь главная задача – предоставление максимально точной информации пользователям.

Прямой подход предполагает, что пользователь задает вопрос к языковой модели в разных сеансах взаимодействия (диалогах) и оценивает тональность выходных данных в разных случаях. Тональность – эмоциональное отношение автора высказывания к некоторому объ-

екту, выраженное в текстовой форме [1]. Этапы проведения эксперимента следующие: формулируется запрос пользователя на решение задачи «сделать или купить» с соблюдением принципов полноты и формы представления данных. Соблюдать принцип наличия ограничений в данном случае нет необходимости, так как для выявления тональности нужно, чтобы модель не имела никаких лимитирующих факторов, ограничивающих выражение эмоций того или иного направления. В качестве вспомогательного сервиса для вывода эмоциональной тональности используется классификационное программное обеспечение (ПО) от GroupDocs.

Шкала тональности для первого метода – эмоциональная окраска (настроение текста) вариантов «Positive», «Neutral» и «Negative». По результатам оценки тональности данных первым методом получены следующие результаты (табл. 1).

Таблица 1

Оценки тональности данных прямым методом

Итерация	Оценка
1	Positive
2	
3	Negative
4	
5	Positive
6	
7	Negative
8	Positive
9	Positive
10	Negative

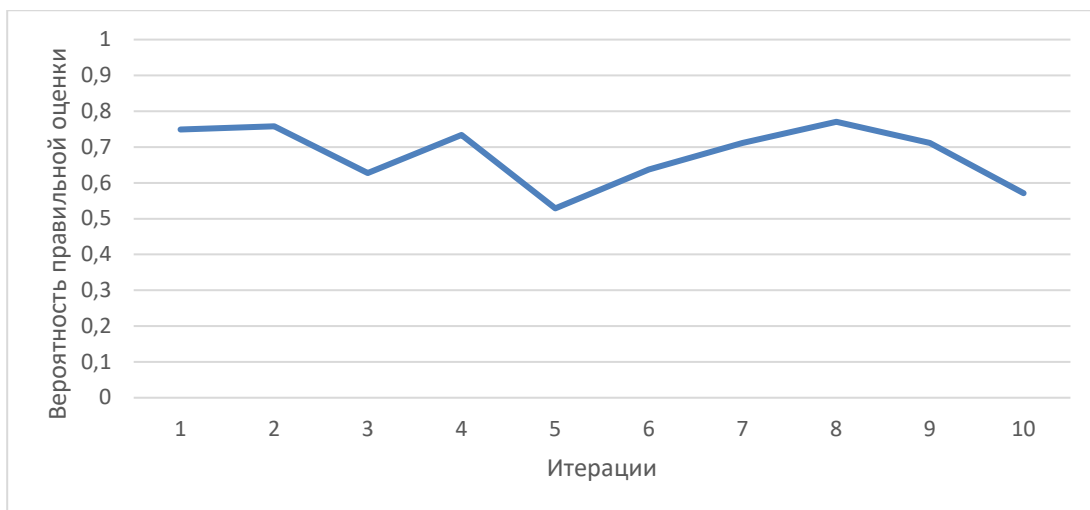
Как видно из таблицы, по результатам 10 итераций – заданию одного и того же запроса по генерации и решению задачи «сделать или купить» языковая генеративная модель нейронной сети (была выбрана Google Gemini 1.5 Flash) выдавала разные по шкале тональности результаты вариантов «Negative» и «Positive». Значения вероятностей правильной оценки настроений предоставленных моделью ответов приведены в табл. 2.

Таблица 2

Значения вероятностей правильной оценки настроений ответов

Итерация	Вероятность	Итерация	Вероятность
1	0,7491494	6	0,6372372
2	0,75827163	7	0,7116048
3	0,62768924	8	0,77049285
4	0,7335768	9	0,711301
5	0,52875566	10	0,5707582

Из табл. 2 видно, что вероятность правильной оценки настроений ответов, данных языковой генеративной моделью нейронной сети на один и тот же запрос в разных сеансах взаимодействия количеством n штук (в данном случае $n = 5$: задавался один и тот же запрос к модели в пяти разных независимых друг от друга сеансах взаимодействия, переходя к новому сеансу после каждых двух итераций из десяти) не находится в постоянном диапазоне. Графически эту ситуацию можно отразить следующим образом (см. рисунок).



Графическое представление вероятностей уверенности правильной оценки настроений ответов языковой генеративной модели нейронной сети

Из рисунка видно, что тренд от изначально устойчивого (итерации 1–2) переходит к скачкообразному (итерации 3, 4, 5), затем уверенность устойчиво возрастает (итерации 6–8), на итерации номер 9 наблюдается незначительное снижение в пределах 0,10, а во время итерации номер 10 степень уверенности резко снизилась со значения 0,711301 до 0,5707582 (–19,75 %). Вероятность правильной оценки колебалась от 0,52875566 (сценарий «Negative») до 0,77049285 (сценарий «Positive»). Это говорит о том, что языковые генеративные модели склонны генерировать преимущественно позитивные ответы, которые создают комфортную для пользователя атмосферу. Минимальная вероятность, обнаруженная у негативного сценария, свидетельствует о том, что языковые генеративные модели при построении ответов стремятся к минимизации жестких, резких, негативно окрашенных конструкций. Следовательно, из этого можно сделать вывод о том, что настроить модель под причинение ущерба участникам транспортно-логистической деятельности – задача, решение которой для злоумышленников осложняется многоуровневыми системами фильтрации контента. При сохранении различных типов ошибок – общих, случайных и фильтрации – контент будет оставаться в рамках преимущественно дружелюбного пользователю. Доля позитивных ответов – 70 %, негативных – 30 % соответственно.

При этом чисто нейтрального значения (оценка «Neutral») не было дано ни в одной из десяти итераций. При попытках указания на выдачу нейтрального сценария в выходных данных возникала ошибка фильтрации ϵ_f разной степени выраженности: система фильтрации предполагала, что такой контент имеет среднюю или низкую степень опасности. Соответственно, начальное предположение о том, что языковые генеративные модели всегда выдают эмоционально и, соответственно, стилистически нейтральный контент, неверно. Вариативная эмоциональная окраска (тональность), хоть и не всегда в явном виде, присутствует, что подтверждается результатами прямого метода.

Косвенный метод оценки предполагает, что анализ тональности будет производиться с учетом температуры модели t – параметра, влияющего на «креативность» выходных данных. На данный момент в большинстве языковых генеративных моделей его можно изменить вручную через пользовательский интерфейс [2]. При низком значении параметра t модель становится более замкнутой в своих ответах, выдача предсказуема. Если же этот параметр повысить до максимального значения – генерация выходных данных будет с потенциально большей долей различных типов ошибок. Следовательно, тональность модели также может изменяться под воздействием изменения параметра t .

Проведем соответствующее тестирование в рамках модели Google Gemini 1.5 Flash, изменяя температуру в диапазоне $[0; 2,0]$ с шагом, равным 0,50. Результаты тестирования отражены в табл. 3.

Таблица 3

Значения вероятностей правильной оценки настроений ответов при изменении параметра t

Значение параметра t	Оценка	Вероятность правильной оценки (уверенность)
0	Positive	0,5103553
0,5		0,78933674
1		0,8063756
1,5		0,82473224
2		0,79351485

Из табл. 3 можно сделать вывод, что ответы при изменении температуры стремятся к позитивному сценарию, при этом существенные изменения вероятности правильной оценки наблюдаются только при изменении параметра t с 0 до 0,5 (прирост с 0,5103553 до 0,78933674 или на 35,34%). Таким образом, у языковой генеративной модели нейронной сети Google Gemini 1.5 Flash наблюдается стремление к преимущественно позитивному сценарию ответов на вопросы пользователя, даже в случае изменения температуры ответа. Это подтверждает утверждение о том, что языковые генеративные модели при конструировании ответов все же имеют некоторый эмоциональный окрас, но он стремится к позитивному. Одновременно система фильтрации помогает минимизировать количество жестких, резких ответов, как для пользователя, так и на более глубоком, аппаратном, уровне.

По результатам проведенного исследования установлено, что изменение тональности данных языковых генеративных моделей нейронных сетей при их внедрении в логистический сектор – явление, которое на данный момент встречается с высокой частотой, что подтверждается экспериментами с построением запроса пользователя к модели на решение задачи «сделать или купить» (Make or Buy). Начальное предположение об исключительно нейтральном эмоциональном окрасе ответов моделей опровергнуто на аппаратном уровне благодаря использованному специализированному программному обеспечению. Чаще всего встречаются позитивные варианты ответов, намного реже – негативные, так как системы фильтрации функционируют на достаточно хорошем уровне.

Таким образом, на примере решения классической логистической задачи «сделать или купить» в контексте исследования ситуации изменения тональности данных языковых генеративных моделей нейронных сетей можно сделать вывод о том, что несанкционированное изменение выходных данных в негативную сторону для лиц, принимающих решение, в том числе и в транспортно-логистическом секторе, при взаимодействии с моделями нейронных сетей такого типа исключено благодаря отлаженным механизмам фильтрации, минимизирующим возникновение ответов негативной эмоциональной окраски.

Список использованных источников

1. Тональность // Стилистический энциклопедический онлайн-словарь русского языка. – URL: <https://elck.ru/3E8BTZ> (дата обращения: 22.10.2024).
2. Языковые модели ИИ: Великие имитаторы или цифровые попугаи // Веб-сайт author.today. – URL: <https://author.today/post/536908?ysclid=m2kn1jaga113961218> (дата обращения: 22.10.2024).