

LLM ЯК ІНСТРУМЕНТ ДЛЯ ЕМОЦІЙНОЇ НЕЙТРАЛІЗАЦІЇ ТЕКСТУ

Білоконь Б. О.

Науковий керівник – д.т.н., професор Терзіян В. Я.

Харківський національний університет радіоелектроніки, каф. ШІ

м. Харків, Україна

e-mail: bohdan.bilokon@nure.ua

Emotionality in communication can have both positive and negative effects, depending on how it is used and controlled. The aim of this paper is to investigate the effectiveness and accuracy of emotional neutralization of texts using LLM, while preserving semantics and main meaning. The study uses prompt engineering methods to build queries whose answers will have the best results.

Емоційність у спілкуванні може мати як позитивний, так і негативний вплив, залежно від того, як вона використовується та контролюється. Контроль над своїми емоціями допомагає уникнути конфліктів та покращити якість спілкування. Особливо це помітно в певних видах роботи зі взаємодією людей, як-от:

- журналістика та звітність: новини та журналістське письмо спрямовані на об'єктивність і безпристрасність, вимагаючи емоційно-нейтральної комунікації для представлення інформації без упередження;
- наука: наукові статті, статті та академічна комунікація повинні бути емоційно-нейтральними для збереження вірогідності та об'єктивності;
- юридичні та судові процеси: юридичні документи, судові засідання та офіційна комунікація часто потребують емоційної нейтральності для забезпечення справедливості та безпристрасності;
- модерація соціальних мереж: модерація контенту на соціальних мережах потребує емоційно-нейтральної комунікації для обробки повідомлень користувачів та виконання правил спільноти безпристрасно;
- військова справа: у військових операціях важливо зберігати секретність і передавати інформацію об'єктивно, без емоційного упередження. Застосування емоційно-нейтральної комунікації у плануванні операцій, звітності та командуванні може сприяти точному розумінню ситуації та прийняттю об'єктивних рішень. У випадках взаємодії з громадськістю або дипломатичними інстанціями, важливо подавати інформацію нейтрально та без емоційного впливу. Це може допомагати зберігати стабільність та міжнародний спокій.

Штучний інтелект розвивається дуже стрімко. Відносно нещодавно з'явилися великі мовні моделі (LLM), такі як ChatGPT. Такі моделі показали ефективність у багатьох сферах людського життя. ChatGPT демонструє здатність проводити аналіз настроїв, не вимагаючи тривалого

попереднього навчання, при цьому виконуючи його на рівні, порівнянному з людським емоційним судженням [1]. LLM може аналізувати текст для виявлення та оцінки емоційного забарвлення. Це може бути зроблено шляхом ідентифікації ключових слів, виразів та фраз, які мають емоційний тон.

Це свідчить про те, що ChatGPT можна використовувати і для корекції даних. LLM можуть стати потужним інструментом для тих сфер діяльності людини, де емоційна нейтральність та безпристрасність займає важливе місце. LLM можуть замінити емоційно заряджені слова на більш нейтральні або об'єктивні варіанти, зберігаючи сенс та значення тексту. Також використовуючи знання про мову та семантику, LLM можуть пропонувати альтернативні варіанти висловлювань, які допомагають зберегти інформацію, але менше піддаються емоційному спотворенню.

Проте варто зазначити, що LLM не завжди генерує бажаний результат. Це залежить від запиту, що подається моделі на вхід. Процес створення ефективних запитів для LLM називається *prompt engineering*. Ефективні запити грають ключову роль у взаємодії з моделлю, оскільки вони направляють її генераційний процес і впливають на зміст та якість згенерованого тексту. Результати різних досліджень доводять ефективність *prompt engineering* для мовних моделей, як-от ChatGPT [2].

Існують різні методи для побудови промптів для досягнення кращих результатів, починаючи від явних обмежень для виходу, наприклад, розмір відповіді, та поданням більш детального контексту на вхід, і до більш складних технік, таких як поєднання промптів в ланцюг та адаптація промптів до специфічного домену.

Зважаючи на все викладене вище, мета роботи полягає у дослідженні ефективності та точності емоційної нейтралізації текстів за допомогою LLM, зберігаючи при цьому семантику та основний зміст.

У доповіді запропоновано підхід, який використовує *prompt engineering* для досягнення найкращої якості нейтралізації емоцій у тексті для декількох LLM, таких як ChatGPT і BERT.

Для вимірювання рівня емоцій у тексті перед та після його нейтралізації можна використовувати наступні метрики:

- емоційна інтенсивність: метрика, що вимірює силу виражених емоцій у тексті. Перед та після нейтралізації порівнюються значення емоційної інтенсивності, забезпечуючи кількісну оцінку емоційного висловлювання;

- емоційний тон: метрика, що оцінює, наскільки текст виражає позитивні, негативні та нейтральні емоції. Порівнюючи тон тексту до та після нейтралізації, можна визначити зміни у спрямованості емоцій;

- різниця емоційного спектру: визначення змін у розподілі конкретних емоцій у тексті. Деякі емоції можуть бути активно нейтралізовані, що відобразатиметься у зміні частоти їх виявлення;

– семантична консистентність: метрика, що оцінює збереження семантики тексту. Вона може вимірювати те, наскільки точно нейтралізація зберігає основні значення та смислові зв'язки у тексті;

– об'єктивність: Оцінюється ступінь відсутності емоційного упередження у тексті. Чим вищий рівень об'єктивності, тим менше відчуття суб'єктивності та емоційного впливу.

Перед проведенням експериментів із нейтралізацією тексту важливо обрати або розробити відповідні метрики, які будуть враховувати особливості емоційного контексту та цілей дослідження.

Розрахування цих метрик можна проводити також за допомогою існуючих моделей штучного інтелекту, які зарекомендували себе і показують високі результати у таких задачах як визначення емоційної складової у тексті. До таких моделей відносяться BERT, EmoInt та інші популярні види трансформерів.

Список використаних джерел:

1. Ran C. Emotion analysis of dialogue text based on ChatGPT: a research study. Third International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2023), м. Yinchuan, China, 18–19 серп. 2023 р. / ред.: S. Saxena, C. Zhao. 2023. URL: <https://doi.org/10.1117/12.3011507> (дата звернення: 04.03.2024).

2. Wu Y., Hu G. Exploring Prompt Engineering with GPT Language Models for Document-Level Machine Translation: Insights and Findings. Proceedings of the Eighth Conference on Machine Translation, м. Singapore. Stroudsburg, PA, USA, 2023. URL: <https://doi.org/10.18653/v1/2023.wmt-1.15> (дата звернення: 04.03.2024).