В. А. Борисова, И. А. Рабинович, А. П. Сербина

Рецензия на проект Я. Н. Далевич, С. А. Жафяровой и Е. Д. Соловковой “Автоматическое определение троллинга в текстах с помощью разных языковых моделей”

Авторы проекта ставили перед собой задачу тестирования компетенции больших языковых моделей в задаче определения открытого и скрытого троллинга. Данная проблема является одной из актуальных задач в области обработки естественного языка.

В решении подзадачи определения наличия / отсутствия троллинга команде удалось добиться значительно более высоких результатов, чем авторам работы Lee et al. 2022, лежащей в основе их исследования. Ключевым фактором такого улучшения стало то, что членами команды была проведена большая и кропотливая работа по подбору и тестированию промптов для LLM. Проект довершает подробный и объективный анализ полученных результатов: в равной мере обсуждаются и успехи, и затруднения, и высказываются предположения о причинах последних. Еще одним несомненным достоинством данной работы является представление кода в аккуратном, подробном, структурированном и легко читаемом виде с использованием подходящих и грамотных визуализаций.

Для нас остались неясными конкретные источники, во-первых, исходного датасета и, во-вторых, нейтральных предложений, которыми он был дополнен. Принципы, на основании которых предложения включались / не включались в модифицированный датасет, также не обозначены эксплицитно. Помимо этого, нам не удалось понять, как устроено соответствие между степенью троллинга, оцененной по десятибалльной шкале, и метками *открытый / скрытый*. Заметный недостаток презентации – отсутствие обзора предыдущих исследований (который, впрочем, обсуждался на представлении темы) и списка литературы и источников.

Тем не менее, указанные замечания совсем не катастрофичны и с легкостью могут быть учтены. Они не мешают общему впечатлению от выполненного проекта, который, на наш взгляд, заслуживает высокой оценки.

P. S. Помимо прочего, работа заставила задуматься об освоении названий моделей, заимствованных в русский язык, ведь во французском языке *mistral* – мужского рода.