

Prompt Engineering.

Проблемы работы с современными ИИ инструментами и способы их преодоления

Казанин М.А. (ИТМО)

Научный руководитель – Бутенко О. Р.

(ИТМО)

Введение. В последние несколько лет технологии, связанные с искусственным интеллектом (ИИ), шагнули вперед. Особое внимание привлекли к себе генеративные нейросети (ГН), так как позволяли пользователю на основе текстового описания генерировать изображения, текст или что-то другое. Прорывной стала технология ChatGPT – генеративная нейросеть, обученная на слепке информации, доступной в интернете, с которой можно переписываться. Так же популярными стали такие инструменты, как Midjourney и Dall-e, позволяющие генерировать изображения на основе текстовых запросов. Сейчас эти инструменты могут использоваться в огромном количестве отраслей: от дизайна до юриспруденции. Возникает вопрос: как увеличить свою профессиональную эффективность, используя инструменты ИИ? В качестве интерфейса взаимодействия они имеют текстовый ввод, следовательно, нужно научиться составлять запрос к ИИ таким образом, чтобы получить желаемый результат, ведь оказывается, что ГН может перепутать факты, проигнорировать часть запроса или, наоборот, добавить что-то, чего вы бы в результате видеть не хотели. Решением этих вопросов занимается Prompt Engineering (от англ. Prompt (“запрос”), Engineering (“инжиниринг”)) – направление по составлению запросов к ГН.

Основная часть. При генерации текста с помощью нейросетей выделяют следующие проблемы:

1. Логические ошибки: Модель может создавать нелогичные высказывания или противоречивые утверждения из-за отсутствия последовательного понимания контекста;
2. Галлюцинации: Модель может добавлять вымышленные данные или ложные факты, которые кажутся реальными;
3. Обобщение: Модель может предоставить общие или неопределенные ответы, оставив часть информации без внимания;
4. Нарушение правил: Модель может сгенерировать текст, который имеет этические проблемы, неприличен для определенного общества или включает в себя нецензурную лексику;
5. Непонимание контекста: Модель генерирует текст, основываясь на ограниченном количестве информации, которое ей предоставлено, она не может быть полностью актуальна;

Для решения этих проблем Prompt engineer’ы составляют подсказки, которые помогают нейросети генерировать контент, наиболее подходящий потребителю [1]. Для оценки качества необходимо тестировать запросы, а для этого следует составить метрики приемлемости сгенерированного контента на основе определенного запроса. Среди этих метрик могут быть проверка на включение задуманного в результат, проверка отсутствия задуманного в результате, оценка на основе нейросети и другие способы [2].

Выводы. Проведено исследование техник написания запросов, составлены метрики качества выполненной работы, получен ряд рекомендаций в работе с ГН

Список использованных источников:

1. Open AI - Лучшие практики использования ChatGPT. Режим доступа: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>
2. Йен Вебстер – Документация к библиотеке тестирования запросов. Режим доступа: <https://promptfoo.dev/docs/>