Автономная некоммерческая организация
«Институт Искусственного интеллекта»

121170, Москва, Кутузовский пр-т, 32, к. 1, 4.В.08
E-mail: info@airi.net, Web: airi.net
ОГРН 1207700493978 ИНН 7730261209 КПП 773001001

ОТЗЫВ РЕЦЕНЗЕНТА

на выпускную квалификационную работу Белякина Евгения Олеговича

на тему: Методы доменной адаптации больших языковых моделей для решения комплексных задач

В работе проведено всестороннее исследование и реализация различных подходов к доменной адаптации больших языковых моделей (LLM) для задач олимпиадного программирования. Рассмотрены методы Prompt Engineering, Supervised Fine-Tuning (SFT) с применением LoRA, а также Retrieval-Augmented Generation (RAG). Наиболее эффективным оказался подход RAG, SFT и Prompt Engineering также продемонстрировали определённые преимущества, однако их эффективность ограничивалась качеством разметки данных и соответствующих инструкций. Было отмечено, что при использовании RAG как интегратора сложных внешних знаний модель сильнее повышает производительность на более сложных задачах.

Сильной стороной работы является высокая актуальность поставленной задачи, а также тщательно выполненный обзор существующих подходов к доменной адаптации больших языковых моделей. Автор рассмотрел широкий спектр современных методов, что демонстрирует хорошее понимание области. Для решения задачи были применены разнообразные техники соответствующие современным тенденциям в исследовании LLM. Для метода Supervised Fine-Tuning был создан синтетический датасет с разумной системой верификации объектов.

К числу недостатков можно отнести недостаточную проработку отдельных элементов методологии. Так, в разделе 4.2 указаны параметры сегментации(длина и перекрытие), однако не пояснен выбор именно этих значений. В разделе 4.3 используется понятие сложности задач без его явного определения или формализации. Более существенным недостатком является отсутствие анализа качества синтетических данных: не обосновано, почему они могут использоваться в качестве замены реальных. Не упомянуты известные подходы к оценке такого качества, например, Сходство Распределений, Human Evaluation или Train on Synthetic, Test on Real (TSTR). Несмотря на предположение о применении TSTR, наблюдается просадка качества модели при SFT на синтетических данных по сравнению с бейзлайном на ряде доменов (alloc, geometry, search, string), что требует дополнительного анализа и обсуждения.

Учитывая объем проделанной работы описанные недостатки, оценка выпускной работы 8 (*«отлично»*).

Рекомендую присвоение студенту квалификации магистра.

Рецензент,
PhD, Младший научный сотрудник Чекалина В.А.

АНО «Институт искусственного интеллекта»

Подпись рецензента, сотрудника АНО «Институт искусственного интеллекта»,

удостоверяю
HR Директор – начальник отдела
АНО «Институт искусственного интеллекта» Звонарева М. И.

