

УДК 004.8

Анализ диалоговых данных с использованием больших языковых моделей

Братусь Д.А (ИТМО)

Научный руководитель – кандидат технических наук, доцент Федоров Д.А. (ИТМО)

Введение. Автоматизация рецензирования научных выступлений является актуальной задачей в условиях роста числа конференций и образовательных программ.

Существующие массовые ИИ-ассистенты (ChatGPT и аналоги) ориентированы на общую генерацию ответов и не способны проводить структурированную оценку по заданным экспертным рубрикам. Главной особенностью предлагаемого подхода является мультимодальность: научный доклад рассматривается как синтез устной речи (транскрибация) и визуальной презентации (слайды). В данном исследовании разрабатывается специализированный агент-критик, который в реальном времени анализирует оба потока данных для формирования сбалансированной обратной связи. Подробное описание архитектуры системы и экспериментальных данных будет представлено в основной части.

Основная часть. При создании системы автоматической оценки возникает ряд сложностей, связанных с риском галлюцинаций модели, субъективностью экспертных оценок и необходимостью анализа визуального контента. Сформулированная гипотеза исследования заключается в том, что использование пула специализированных LLM-агентов (по одному на каждый критерий: новизна, обоснованность, валидность, качество подачи) с последующей эволюционной оптимизацией позволяет приблизить оценки системы к предпочтениям конкретного эксперта-человека. В отличие от существующих решений, система не просто выставляет баллы, но и генерирует развернутое обоснование каждого критерия, что критически важно для образовательных задач. Эксперименты, проведенные в ходе сессий, показали, что система стабильно занижает оценку по сравнению с человеком из-за более строгих критериев, заложенных в промпт, однако демонстрирует высокую согласованность при оценке текстовой части доклада. Визуальный анализ слайдов (структура, наличие графиков, читаемость) позволяет уточнить итоговый балл, особенно по критериям «качество подачи» и «практическая значимость». Текущие результаты подтверждают,

что ключевым фактором качества работы является точность транскрибации и четкость формулировки критериев в промптах. Дальнейшее исследование направлено на создание датасета из реальных докладов с оценками нескольких экспертов для обучения модели и автоматической настройки весов критериев под стиль конкретного научного руководителя или программного комитета.

Выводы. Проведен анализ проблематики автоматизированной оценки научных докладов. Определены ключевые критерии (актуальность, новизна, валидность, качество подачи, практическая значимость) и разработана архитектура мультимодального агента-критика. Выявлена зависимость качества оценки от точности распознавания речи и визуального контента. Планируется доработка системы до прототипа, способного выполнять предварительную экспертизу докладов на молодежных конференциях и выступать в роли тренажера для студентов, позволяя им сравнивать свою самооценку с эталонной.