МОДУЛЬ ОБЪЯСНИМОСТИ В МУЛЬТИАГЕНТНОЙ СИСТЕМЕ ДЕТЕКЦИИ ИИ-ТЕКСТОВ

Мельников Д.А., старший преподаватель,

ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва, Россия

Аннотация. В данной статье представлено исследование модуля объяснимости, встроенного в мультиагентную систему детекции текстов, генерируемых искусственным интеллектом, в образовательной области. Описаны функциональные блоки модуля, включая Селектор (пороговая фильтрация аргументов), Адаптер (терминологическая и стилистическая адаптация под профиль преподавателя) и Генератор (синтез текста для базового, детального и экспертного уровней). Для детализации работы модуля приведена диаграмма потоков данных.

Ключевые слова: мультиагентные системы, агент, детекция ИИ-текстов, искусственный интеллект.

С появлением больших языковых моделей задача автоматической детекции текстов, сгенерированных искусственным интеллектом (ИИ), стала критически важной в сфере

образования [1, 2, 3]. Несмотря на то, что большинство таких продуктов показывают высокую точность, их решения не всегда могут быть доступны преподавателям. Эта проблема часто возникает в ситуациях, когда преподаватель не может объяснить студенту, на каком основании его работа не прошла проверку на оригинальность, поскольку множество современных систем детекции выдают бинарный вердикт (есть в тексте ИИ или он отсутствует) и процент уверенности, но не объясняют, почему такое решение было принято.

Важность интегрирования модуля объяснимости в мультиагентную систему позволит преподавателю верифицировать решение, запросив детальный отчет с исходными признаками, методами измерения и весами доверия к аргументам.

...

полный текст во вложении