

Российские учёные создали «прозрачный» детектор текстов ИИ

Российские исследователи разработали новый метод анализа текстов, который не только отличает человеческое творчество от машинного, но и объясняет причины такого вывода. Разработка принадлежит специалистам Сколтеха, МФТИ, Института искусственного интеллекта AIRI и других научных центров.

В отличие от существующих детекторов, работающих по принципу «чёрного ящика», новая система раскладывает текст на понятные компоненты. Алгоритм использует разреженные автокодировщики — специальную архитектуру ИИ, позволяющую анализировать отдельные характеристики текста.

Исследователи выявили типичные признаки генерации искусственным интеллектом: многословность в финансовых текстах, сложные синтаксические конструкции в научных работах. Система также распознаёт намеренные попытки маскировки — добавление лишних пробелов или нестандартных символов.

Разработка особенно актуальна в свете массового использования ChatGPT и аналогичных систем. Нейросети применяют не только для создания контента в интернете, но и при написании студенческих работ и даже научных статей.