Магистранты ИТМО совместно с компанией Raft разработали первую отечественную систему защиты искусственного интеллекта (ИИ) HiveTrace. Решение способно отражать кибератаки на веб-приложения с генеративным ИИ.

Система защищает от семи наиболее опасных уязвимостей, включая промт-инъекции, утечки конфиденциальной информации и некорректную обработку данных. В будущем функционал планируют расширить для борьбы с новыми видами угроз.

HiveTrace работает с любыми популярными ИИ-моделями — как открытыми, так и закрытыми. Это позволяет компаниям гибко настраивать правила безопасности под свои потребности.

Разработчики отмечают, что зарубежные аналоги не адаптированы под российский рынок и русский язык. При этом существующие корпоративные решения крупных IT-компаний недоступны для широкого использования.

Система создана в лаборатории AI Security ИТМО при поддержке онлайн-магистратуры AI Talent Hub. По мере роста внедрения ИИ в бизнес-процессы потребность в таких защитных решениях будет только увеличиваться.