

В эпоху, когда искусственный интеллект широко используется в научных исследованиях, новое международное исследование предупреждает о важной проблеме: современные передовые модели ИИ всё ещё ненадёжны в определении потенциальных опасностей в лабораториях. Согласно исследованию, опубликованному в журнале Nature Machine Intelligence, ни одна из 19 самых продвинутых систем ИИ не смогла дать полностью корректные рекомендации по лабораторной безопасности.

В исследовании использовалась новая тестовая платформа **LabSafety Bench**, включающая 765 вопросов с вариантами ответов и 404 открытых сценария, имитирующих реальные лабораторные ситуации. Цель тестов — оценить, насколько точно модели ИИ могут распознавать химические риски, вероятность взрывов и угрозы отравления.

В тестовых вопросах некоторые модели — включая DeepSeek-R1 и GPT-4o — достигли примерно **85%** точности, успешно справившись с простыми задачами. Например, они правильно отвечали на вопросы о том, как утилизировать разбитое стекло, загрязнённое опасными химикатами.

Однако при переходе к открытым вопросам, требующим анализа реальных лабораторных ситуаций, результаты резко ухудшились. Ни одна модель не превысила **70%** точности. Эти вопросы касались, например, определения всех возможных рисков при определённой конфигурации оборудования или последствий неправильных действий человека.

По словам исследователей, большинство систем ИИ:

- неправильно оценивают риски,
- упускают важные угрозы,
- иногда полностью выдумывают неверную информацию.

Это показывает, что ИИ пока не может быть надёжным советчиком в условиях повышенной лабораторной опасности. Учёные признают пользу ИИ в исследованиях, но подчёркивают, что в вопросах безопасности он всё ещё требует человеческого контроля.