

На конференции RSAC 2025 исполняющий обязанности директора DARPA Роб МакГенри поделился смелым видением: объединение искусственного интеллекта с проверенными программными методами может в один прекрасный день устранить большинство программных ошибок, положив конец постоянной потребности в исправлениях.

Говоря об инициативе DARPA AI Cyber Challenge, МакГенри сказал, что цель состоит в том, чтобы разорвать цикл поиска и исправления уязвимостей программного обеспечения после того, как они были использованы. Вместо этого системы искусственного интеллекта, особенно большие языковые модели (LLM), могли бы автоматически находить и устранять недостатки до того, как у хакеров появится шанс.

В конкурсе участвуют такие крупные технологические партнеры, как Google, Microsoft, OpenAI и Anthropic. В ходе недавнего теста команды, использующие ИИ, успешно выявили и устранили поддельные недостатки в ключевых программах с открытым исходным кодом, таких как ядро Linux и SQLite.

Один из прорывов – объединение ИИ с формальными методами – математическим способом доказательства правильности работы программного обеспечения. Этот подход долгое время был слишком медленным и дорогим для большинства применений, но теперь ИИ может сделать его быстрее и дешевле.