Новые исследования показывают, что чат-боты с искусственным интеллектом, такие как ChatGPT и Perplexity, могут становиться причиной фишинговых атак. По данным компании Netcraft, ИИ часто предлагает пользователям неправильные ссылки на сайты, и злоумышленники могут этим воспользоваться.

В тесте, где ИИ просили дать ссылки на 50 известных брендов, модели от OpenAI (GPT-4.1) ошиблись в 34% случаев.

Особенно уязвимыми оказались небольшие бренды, о которых меньше информации в обучающих данных ИИ. Преступники уже начали регистрировать фальшивые сайты, на которые могут отправлять пользователей, доверяющих чат-ботам.

Также сообщается о более 17 000 поддельных страниц на платформе GitBook, которые маскируются под справочные материалы и нацелены на пользователей криптовалют. В одном случае Perplexity предложил ссылку на фишинговый сайт вместо официального ресурса банка Wells Fargo. Кроме того, мошенники пытаются «отравить» ИИ-кодогенераторы, публикуя поддельные API и проекты, чтобы они попали в обучающие данные.

Эксперты советуют быть осторожными и проверять ссылки вручную.