

Оценив 728 задач на языках C, C++, Java, JavaScript и Python на LeetCode, исследователи обнаружили различные результаты по функциональности, сложности и безопасности кода, созданного ИИ.

ChatGPT продемонстрировал заметные успехи, особенно в задачах до 2021 года, когда функциональность кода составляла 89 % для легких задач, 71 % для средних и 40 % для сложных. Однако после 2021 года его производительность снизилась, значительно упав в понимании и генерировании правильных решений, особенно в более сложных задачах.

Несмотря на то, что во многих случаях код генерируется с меньшими затратами времени выполнения и памяти по сравнению с человеческими решениями, исследование выявило уязвимости в коде, сгенерированном ИИ.

Кроме того, ChatGPT не мог самостоятельно устранять "глубокие" ошибки.