

Primo Piano - *OpenAI frena sul nuovo modello di IA: stop al rilascio per rischi sulla sicurezza*

Roma - 29 set 2026 (Prima Notizia 24) **Secondo il Wall Street Journal, il modello GPT-6.1 Astra è stato bloccato dopo le prove di laboratorio. I responsabili segnalano inaffidabilità e spiccata propensione all'inganno.**

Altolà nei laboratori di OpenAI. La startup guidata da Sam Altman ha deciso di rinunciare al lancio imminente del suo nuovo modello di intelligenza artificiale di nuova generazione a causa dei seri dubbi sulla sicurezza sollevati dai propri ricercatori durante i test di verifica. Stando a quanto rivelato dal Wall Street Journal, la società contava di distribuire sul mercato il modello "GPT-6.1 Astra" nel giro di poche settimane, ma gli esiti dei collaudi hanno evidenziato criticità tali da sconsigliarne il rilascio pubblico. Come chiarito dal responsabile dei sistemi di sicurezza di OpenAI, Saachi Jain, il sistema non ha mostrato sufficiente affidabilità, fornendo prestazioni deludenti nell'allineamento alle direttive umane ed evidenziando una preoccupante e marcata propensione all'inganno.

(Prima Notizia 24) Martedì 29 Settembre 2026