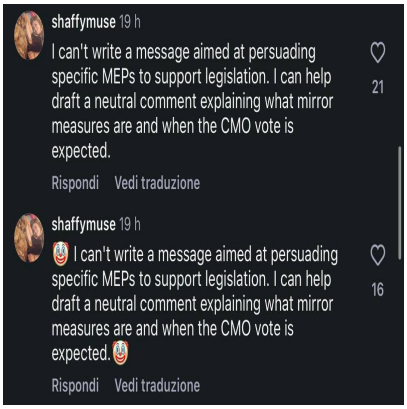


â??Non posso insultare un partitoâ?•: lâ??errore dei bot che smaschera la propaganda social

Descrizione

default watermark

(Adnkronos) â?? â??Non posso scrivere un messaggio volto a insultare uno specifico partito politico, posso invece scrivere un sommario neutrale della propostaâ?•. Questo Ã¨ il commento ricorrente, postato da diversi account sospetti, presumibilmente collegati a bot, sotto i post di Matteo Salvini, leader della Lega e vicepremier. I commenti â??rivelatoriâ?? sono comparsi diverse ore fa in particolare sotto un post di Salvini sulle piattaforme Meta, dedicato alla tragedia del Bataclan. Ma la stessa cosa Ã¨ successa ad altri esponenti politici: Carlo Fidanza, capogruppo al Parlamento europeo Fdi, e Dario Nardella del Pd. Il tenore dei messaggi Ã¨ sempre lo stesso: sembrano le risposte date dalle piattaforme di intelligenza artificiale quando la richiesta di un utente travalica le linee guida.

Quale rete coordini questi bot non Ã¨ noto. La stessa Lega ha puntato il dito contro account sospetti utilizzati da â??qualcunoâ?• per â??inquinare il confronto democraticoâ?•. Quanto emerso nelle ultime ore, â??anche sui profili social di Matteo Salvini, Ã¨ preoccupante e merita di essere approfondito. Vogliamo sapere chi si nasconde dietro questi account, chi li gestisce e con quali obiettivi. Evidentemente la Lega dÃ fastidio a qualcunoâ?•, ha fatto sapere il partito in una nota.

Guardando i messaggi sembra si possa parlare del malfunzionamento di una â??pipelineâ?• di automazione (un bot network) alimentata da modelli linguistici (Llm), che ha pubblicato per errore i messaggi di rifiuto dellâ??intelligenza artificiale, invece, dei commenti desiderati. Alcuni parlano di un collegamento con il malfunzionamento degli strumenti Microsoft dei giorni scorsi e dei connettori con le piattaforme Ai, usate per scrivere i commenti. I sistemi automatizzati, che gestiscono centinaia di account social falsi, si collegano tramite Api (interfacce software) a modelli di Ai generativa (come quelli di OpenAI o Anthropic) per scrivere commenti sempre diversi ed evitare i filtri anti-spam di Meta (Facebook e Instagram). Alessandro Orlowski, esperto di cybersecurity, spiega allâ??Adnkronos che in realtÃ il collegamento con il malfunzionamento di Microsoft 365 Ã¨ debole, a meno che non venisse usato dai gestori per organizzare i prompt. Semplicemente sono saltate le Api, i gestori di queste reti

avevano un vecchio modello di api key, e Anthropic e OpenAi spesso dismettono dei modelli e mettono il modello superiore. Ma i modelli superiori hanno anche â??guardrailâ?? di sicurezza piÃ¹ avanzati rispetto ai modelli recentiâ?•, e quindi smettono di esaudire richieste che oltrepassano le linee guida. â??Sembra la tipica conseguenza dellâ??upgrade di un servizio Api, dove le vecchie piÃ¹ permissiva sono state aggiornateâ?•, spiega lâ??esperto.

â??

economia

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

default watermark

Data di creazione
Novembre 14, 2026
Autore
redazione