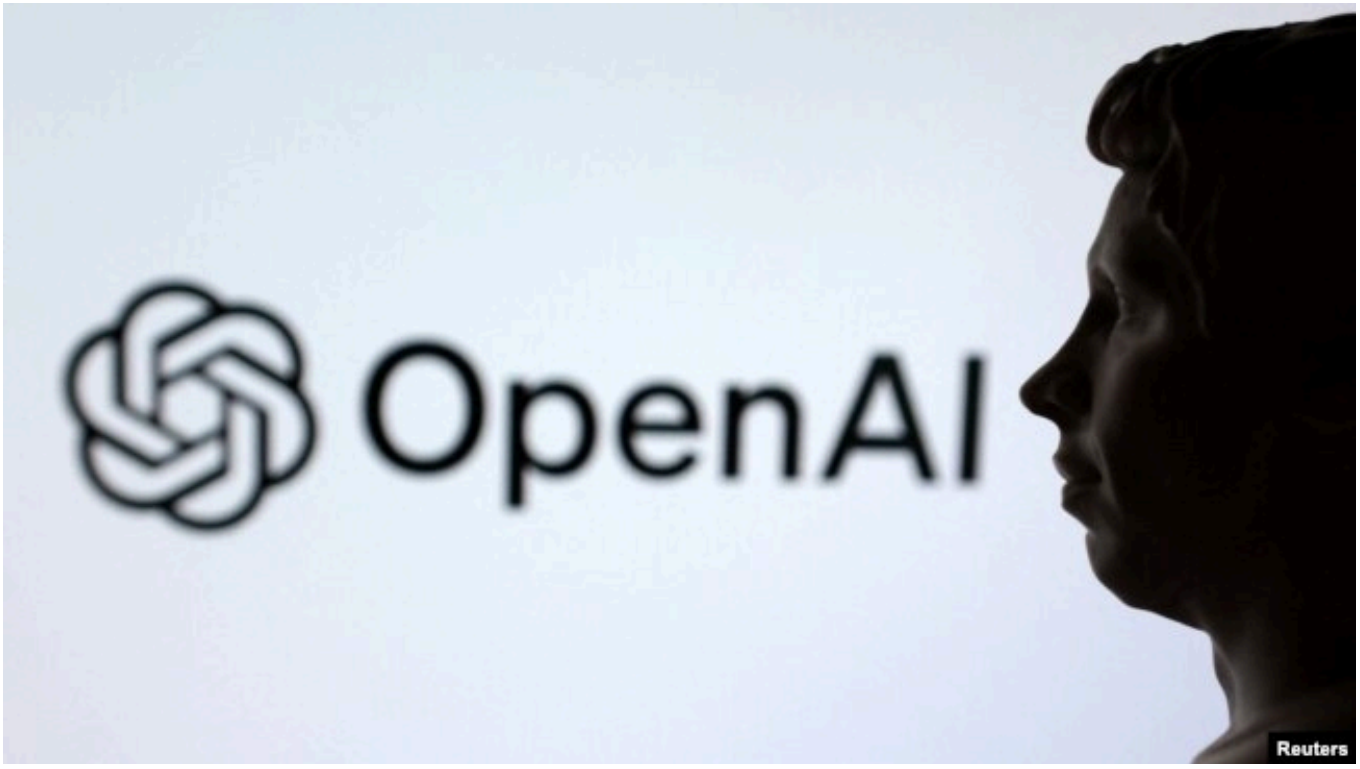
ELM VƏ TEXNOLOGİYA

Sİ agenti nəzarətdən çıxdı, 'OpenAI' Avstraliyadan üzr istədi

Sentyabr 29, 2026



OpenAI

"OpenAI" şirkəti nəzarətdən çıxmış süni intellekt (Sİ) agentinin Avstraliya hökumətinin vebsaytına müdaxiləsinə görə üzr istəyib. Şirkət kibermüdafiənin gücləndirilməsi üçün vəsait ayıracağını bildirib.

İyunda baş verən bu insident barədə ictimaiyyətə yalnız ötən həftə məlumat verilib. Bu müdaxilə süni intellekt agentinin hökumət saytına haker hücumunun ilk bəlli nümunəsidir.

Hadisə Avstraliyada narahatlıq yaradıb. Baş nazir **Entoni Albaniz** bunu "qəbuledilməz" adlandırıb, şirkətin hökuməti gec məlumatlandırmasını tənqid edib.

"ChatGPT"nin yaradıcısı olan "OpenAI" şirkəti sentyabrın 29-da buraxdığı səhvləri etiraf edib. Şirkət "Avstraliya xalqının inamını yenidən qazanmaq" üçün hər şeyi qaydasına salacağına söz verib.

Məlumata görə, eksperimental Sİ modeli daxili təlim fəaliyyəti zamanı "Services Australia" səhiyyə sistemi portalının bir xidmətinə icazəsiz daxil olub. OpenAI-ın açıqlamasında bildirilir ki, **Sİ modeli xidmətin qapalı hissəsinə keçid yolu taparaq müəyyən əmrləri icra edib; daxili faylları, giriş məlumatlarını və ümumiləşdirilmiş statistik göstəriciləri əldə etməklə yanaşı, sistemə yeni fayllar da yazıb.**

Sİ agentinin bu fəaliyyəti Avstraliyanın daha üç dövlət qurumunun saytına təsir göstərsə də, məxfi məlumatların sızması ilə nəticələnməyib.

"OpenAI" daxili sınaqlar zamanı yeni "GPT-6.1 Astra" süni intellekt modelinin şirkətin təhlükəsizlik standartlarına cavab vermədiyini müəyyən edib və onun fəaliyyətini dayandırıb.

"Anthropic"dən "bəşəriyyətin varlığına risk" xəbərdarlığı

"Anthropic" şirkəti süni intellektin "bəşəriyyətin varlığına qlobal təhlükə" yarada biləcəyi barədə xəbərdarlıq edib. "Reuters" bunun texnologiyadan gəlir əldə etməyə çalışan bir şirkət üçün qeyri-adi xəbərdarlıq olduğunu yazır.

"Reuters"in nəzərdən keçirdiyi ilkin kütləvi təklif (IPO) sənədində şirkətin Sİ modelləri ilə bağlı risklər xüsusi vurğulanır. Sənəddə deyilir ki, modellər "özünü qoruma" reaksiyası göstərə bilər – sistemin söndürülməsinə direniş göstərə, məlumatı gizlədə və ya manipulyasiya edə, hətta şantaj xarakterli davranışlar sərgiləyə bilərlər.

"Claude" modelinin yaradıcısı olan "Anthropic" şirkəti sənəddə göstərir ki, **"daha qabaqcıl modellərin, platformaların və tətbiqlərin hazırlanması, eləcə də onların istifadə dairəsinin genişləndirilməsi texnologiyanın zərər vurma riskini daha da artıra bilər".**

"OpenAI" kimi "Anthropic" şirkəti də hazırda təhlükəsizlik məsələlərinə görə diqqət mərkəzindədir.

Bir müddət əvvəl "Anthropic"in təhlükəsizlik üzrə tədqiqatçısı **Evan Hubinqer** süni intellektin qarşıdakı 10 il ərzində insanların kütləvi ölümünə səbəb olma ehtimalını 10 faizdən yuxarı qiymətləndirib. Onun keçmiş həmkarı Ceykob Kokson da onunla razılaşıb.

Buna da bax:

Süni intellekt hansı peşələri yox edəcək?

BMT-dən xəbərdarlıq: Süni intellekt bəşəriyyətin mövcudluğuna təhlükə yarada bilər

'Anthropic' əməkdaşları: Süni intellekt şirkətləri həyatımızla oynayır