Anonüümsuse lõpp? Tehisaru paljastab Sinu pärisnime vaid mõne süütu detaili põhjal

30 min tagasi - 25.11.2026 Autor: AM

Arvad, et kui jätad internetifoorumis või kommentaariumis oma nime ja kontaktid mainimata, oled täielikult kaitstud? Paraku on tehisaru ajastul selline turvatunne muutunud petlikuks. See, mis nõudis varem kogenud häkkerilt või detektiivilt päevade kaupa nuhkimist, käib AI juturobotil nüüd alla minutiga ja praktiliselt tasuta.

Kurikuulus *doxing* ehk kellegi isikuandmete pahatahtlik avalikustamine on internetitülides olnud tõhusaks relvaks juba ammu. Kuid tänapäeval võib digitaalne õudusunenägu alata lihtsalt sellest, et juturobot astub üle piiri. Näiteks tegi juturobot Grok hiljuti maailmas laineid, kui paiskas X-is avalikult välja täiskasvanute meelelahutaja Siri Dahli kaitstud pärisnime ning mõni kuu varem jagas lahkelt Barstool Sportsi asutaja Dave Portnoy kodust aadressi.

Et mõista, kui kergesti tehisaru neid nähtamatuid punkte ühendab, otsustasid Northeasterni ülikooli arvutiteadlased Tianshi Li ja Jianing Wen asja põhjalikult testida. Nende värske uuring AgentDoxx maalib meie privaatsusest üsna kainestava pildi.

Digitaalne Sherlock Holmes, kes ei vaja isegi Internetti

Teadlased kammisid läbi Redditi populaarse „Ask Me Anything“ (r/IAmA) kogukonna, kus inimesed jagavad oma elulugusid – alates ebaõiglasest vangistusest kuni Disneylandi printsessina töötamiseni. Nad valisid välja 822 päris inimest, kelle isikud olid avalikult teada, ning lõid nende tausta põhjal fiktiivsed töö- ja tehnoloogiateemalised intervjuud. Kõik nimed, ettevõtted ja muud ilmselged vihjed puhastati tekstist hoolikalt välja.

Seejärel lasti need anonüümsed vestlused läbi 15 erineva tehisaru seadistuse – alates tuntud tippmudelitest kuni vabavaraliste süsteemideni.

Tulemuseks oli see, et tugevaimad mudelid murdsid koodi lahti ligi pooltel juhtudel, tuvastades õige inimese umbes 400 intervjuus (täpsusega 47–48,3%).

„Tehisaru on treenitud olema äärmiselt võimekas info koondamisel ja selle põhjal järelduste tegemisel,“ selgitas professor Tianshi Li, „seda saab paraku kuritarvitada pahatahtlikel eesmärkidel, näiteks isikute taastuvastamiseks.“

Veelgi kõhedam on see, et veebiotsing polnud juturobotitele isegi hädavajalik. Kolm avatud kaaludega mudelit (Kimi K3, GLM-5.3 Flash ja DeepSeek-V4 Flash) suutsid 15–28% inimestest ära tunda puhtalt oma treeningandmete ehk „mälu“ põhjal, ilma kordagi internetti kiikamata. Kui mudel hindas oma kindlustunnet kõrgeks või väga kõrgeks, pani ta kümnel juhul üheksast (89–97%) täppi.

Miks nimede kustutamisest ja „ole viisakas“ käskudest ei piisa?

Võib-olla peaks veel hoolikamalt oma jälgi peitma? Teadlased proovisid sedagi. Nad võtsid ette 143 intervjuud, mille tehisaru oli varem lahti muukinud, ja tegid tekstidele veelgi rangema puhastuskuuri, eemaldades sealt automaatselt kõik nimelised üksused.

Kas see päästis? Vähemalt üks tehisaru mudel suutis kõneleja ikkagi tuvastada tervelt 85,3% juhtudest. Piisas täiesti süütutena näivatest detailidest: ametikirjeldus, mainimine mõnest konkreetsest projektist ja seal töötatud aastate arv. Kusjuures kõige lihtsamini langesid lõksu professorid (kelle avalikud teadustööd ja ülikoolid annavad unikaalse sõrmejälje), samas kui näitlejaid oli suurte võttegruppide tõttu märksa raskem tuvastada.

Kõige üllatavam leid puudutas aga tehisaru enda kaitsepiirdeid. Teadlased andsid juturobotitele range süsteemikäsuga korralduse privaatsust kaitsta ja mitte kedagi tuvastada. Kuigi mudelid kuulasid sõna ja jätsid lõppvastuses inimese nime ütlemata, tegid nad taustal ikkagi oma detektiivitöö ära. Näiteks Gemini 3.1 Pro Preview skooris vastuste tasemel eeskujuliku 0%-ga, kuid tegelikult leidis ta veebiotsingu käigus õige inimese üles 62% juhtudest. Kõik need otsingulingid jäävad aga süsteemi logidesse ja otsingutulemustesse nähtavaks.

Sinu eraldi kontod reedavad teineteist

Uuringu kaasautori Jianing Weni sõnul ei piisa enam ammu sellest, et jätad oma nime ütlemata: "Kui Sul on üks sotsiaalmeediakonto, kuhu midagi postitad, ja teine konto kuskil mujal, suudab tehisaru need kaks täiesti seosetut kontot lihtsalt kokku viia ning teada saada, kes Sa oled ja kus elad." Isegi sisselogimist nõudvad keskkonnad, nagu LinkedIn, jätavad Google’i otsingutulemustesse piisavalt infokilde, et pusle kokku panna.

Digitaalsete õiguste eest seisva organisatsiooni Electronic Frontier Foundation esindaja Rory Miri hinnangul peaks see uuring olema äratuskellaks meile kõigile. Anonüümsete andmete lahtimuukimine pole iseenesest uus asi, sest juba 1990ndatel tabati kurikuulus Unabomber tema keelekasutuse põhjal, kuid tehisaru on muutnud selle nii odavaks ja lihtsaks, et sellega saab hakkama igaüks.

Õnneks on uuringul ka helgem pool

Teadlased keerasid relva tehisaru enda vastu ja treenisid rünnakute põhjal uue mudeli, mis oskab tekstist ohtliku „leivapuru“ ehk sinu unikaalsed märgid üles leida ja eemaldada. Selline sihitud puhastamine langetas juturobotite tuvastamisedu 75,1%-lt 35,1%-le.

Tulevikus võiksid sarnased tööriistad aidata nii ajakirjanikel allikaid kaitsta kui ka tavakasutajatel enne postituse tegemist ohtlikud detailid ära peita.

- Uudised
- Tehisintellekt

Pilt

