**2025-06-16**

**Dirbtinis intelektas turi daugiau empatijos už žmogų: įspėja nepasitikėti aklai**

**Visada išklausys, patars, paguos ir leis pasijusti teisiu. Kai kažkada taip buvo galima apibūdinti tik terapeutus, šiandien taip nesuklysite kalbėdami ir apie dirbtinio intelekto (DI) modelius. Naujausių tyrimų duomenimis, populiarūs DI įrankiai geriau atlieka empatijos testus nei realūs žmonės. Darius Povilaitis, „Telia“ kibernetinio saugumo vadovas, sako, kad vis daugiau žmonių ieško emocinės paramos bendraudami su DI, tačiau toks atvirumas gali turėti nenumatytų pasekmių.**

**Kai DI tampa patarėju, jūsų išgyvenimai gali tapti vieši**

2024 m. gegužės 22 d. paskelbtas „Nature“ leidybos grupei priklausančiame žurnale „Communications Psychology“, atskleidė, kad didieji kalbos modeliai (angl. *LLMS*) pranoksta žmones standartiniuose emocinio intelekto testuose. Juose DI pasiekė stulbinantį 81 proc. tikslumą ir taip smarkiai pralenkė žmonių vidurkį (56 proc.).

Tyrimas parodė, kad ChatGPT‑4 ne tik geba spręsti emocinio intelekto testus geriau nei vidutinis žmogus, bet ir sukuria naujus, patikimus testų klausimus. Kurti testus reiškia formuluoti klausimus, kurie tiksliai vertina žmogaus gebėjimą atpažinti, suprasti ir reguliuoti emocijas. Šios išvados leidžia teigti, kad DI pokalbių robotai gali generuoti atsakymus, kurie atitinka tikslias žinias apie žmogaus emocijas ir jų reguliavimą.

„Kai žmogus kreipiasi į DI kaip į terapeutą, atskleidžia daug itin asmeninės informacijos. Šiuo metu net vystosi tendencija, kai žmonės kelia į modelius susirašinėjimus su draugais, kolegomis ar antromis pusėmis ir klausia patarimų, analizuoja pokalbius. Nors visai neseniai buvo paviešintas atvejis, kai ChatGPT kitoje užklausoje parodė asmeninį Regitros dokumentą, nes kažkas jį kėlė į modelį anksčiau. Tai  rodo, kad į sistemą patekę duomenys gali netikėtai pasirodyti kitose sąveikose“, – perspėja kibernetinio saugumo ekspertas.

Pavyzdžiui, Gemini modelio apačioje yra nurodyta, kad modelis dar tobulinamas ir jūsų susirašinėjimus gali peržiūrėti kitas žmogus. Kai kuriose sistemose, dažniausiai naudojantis nemokama versija arba neprisijungus prie paskyros, vartotojų įvedama informacija gali būti naudojama modelio tobulinimui.

**Kaip klausti ir išlaikyti privatumą?**

Kibernetinio saugumo eksperto teigimu, yra keletas būdų, kaip kalbant su DI geriau apsaugoti savo asmeninius duomenis. Galima išjungti pokalbių istoriją, klausimus formuluoti be asmeninių detalių arba pateikti informaciją kiek įmanoma anonimiškiau.

„Norint saugiau bendrauti su ChatGPT ar Gemini, svarbu vengti atskleisti asmeninius duomenis ir formuluoti klausimus kuo bendriau. Be to, rekomenduojama išjungti ar reguliariai ištrinti pokalbių istoriją, kad jūsų įvestas tekstas nebūtų naudojamos modelio mokymui. Bet kuriuo atveju kelti asmeninių duomenų, susirašinėjimų, dokumentų, nuotraukų su savo atvaizdu nereikėtų, nes taip prisiimate rizikas. Griežtai nerekomenduoju dalintis finansine informacija“, – sako D. Povilaitis.

Svarbu prisiminti, kad DI modeliai, nors ir galingi, nėra neklystantys. Jie kartais gali „halucinuoti“. Tai reiškia, kad modeliai generuoja informaciją, kuri skamba įtikinamai, bet iš tikrųjų yra neteisinga.

Anot D. Povilaičio, ChatGPT ar Gemini gali atrodyti labai empatiški, jų žodžiai tėra statistiškai pagrįsti tekstai.  Jie nesupranta, ką reiškia jausti skausmą, nerimą ar džiaugsmą.

„Žmonės ieško pagalbos visais įmanomais būdais. Suprantu, kad turint DI kišenėje sunku atsispirti paklausti patarimo. Ir viskas su tuo būtų gerai, tačiau DI modeliai tik apdoroja žodžius ir nuspėja, kas galėtų padėti, remdamiesi dideliais tekstų rinkiniais. Tai nėra tas pats, kas tikro žmogaus gebėjimas pajausti emocijas, išgirsti tarp eilučių. Kitaip nei psichologinės sveikatos specialistai, pokalbių robotai linkę nuolat patvirtinti vartotojo įsitikinimus, net kai šie yra klaidingi“, – paaiškina D. Povilaitis.

**Daugiau informacijos:**

**Audrius Stasiulaitis**

I WILL TELL YOU TELIA

Tel. (0 5) 236 7019, Mob. +370 686 77988

El. p. [audrius.stasiulaitis@telia.lt](mailto:audrius.stasiulaitis@telia.lt)

Telia Lietuva, AB, Saltoniškių g. 7A, 03501 Vilnius