

KAIP DIRBTINIS INTELEKTAS PAKEIS GYVENIMĄ

„VISI MIRS, ĮSKAITANT IR VAIKUS“

Kodėl net Elonas Muskas jo bijo?

Vos prieš kelerius metus dirbtinio intelekto vystymasis buvo apibūdinamas rožėmis klotais žodžiais. Įvairių sričių ekspertai įsivaizdavo, kaip jis galėtų padėti gydytojams, vertėjams ir daugeliui kitų profesijų atstovų. Tačiau vos tik GPT kalbos modelis pasiekė tokią pažangą, kad jo generuojami tekstai tapo praktiškai neatskiriami nuo realių žmonių parašytų tekstų, požiūris į dirbtinį intelektą (DI) kardinaliai pasikeitė.

Netgi pirmaujantys šios srities tyrėjai prisijungė prie radikalių šios technologijos priešininkų. Ir nors kibernetiniai nusikaltėliai džiaugiasi GPT pažanga, ekspertai perspėja, kad dirbtinis intelektas gali sukelti daug rimtesnių problemų, keliančių grėsmę gyvybei pačioje Žemėje. Vienas ru portalas tyrinėjo, kodėl tie, kurie anksčiau investavo į DI, pradėjo jo bijoti ir ar jis tikrai gali būti pavojingas. Dirbtinis intelektas norės pasitraukti iš interneto

Tolesnis dirbtinio intelekto vystymas gali reikšti žmogaus evoliucijos pabaigą: anksčiau ar vėliau superintelektualūs neuroniniai tinklai peržengs internetą, perims visus procesus planetoje ir pašalins žmones. Šią šokiruojančią prognozę emociingame „Time“ straipsnyje išsakė Eliezeris Yudkowsky, Mašininio intelekto tyrimų instituto bendrąkūrėjas ir vienas iš šiandien pirmaujančių IT ekspertų. Jis siūlo visiškai sustabdyti visus šios srities tyrimus, išskyrus tuos, kurie gali padėti išvengti žmonijos sunaikinimo.

„Pagrindinis pavojus yra ne intelekto, kuris konkuruotų su žmogaus intelektu, o jį pranokstančio, sukūrimo perspektyva. Šiame kelyje labai sunku įžvelgti kritines ribas, kurias tyrėjai galėtų nepastebimai peržengti. Todėl labiausiai tikėtinas superintelektualaus dirbtinio intelekto sukūrimo rezultatas bus tiesiogine prasme visų Žemėje gyvenančių žmonių mirtis. Ne ta prasme, kad „yra tolima galimybė“, o ta prasme, kad „visiškai tikra, kad tai įvyks“, – rašė Yudkowsky. 1.



Eliezeris Yudkowsky siūlo šiandien apriboti neuroninių tinklų darbą

Jis įsitikinęs, kad antžmogiškas dirbtinis intelektas bus abejingas tiek žmonėms, tiek protingai gyvybei apskritai. Abejingi neuroniniai tinklai bus valdomi ne emocijų, kurios jiems nežinomos, o savo poreikių, kurie galėtų motyvuoti mašiną išnaudoti žmogaus kūną savo tikslams. Pavyzdžiui, transformuoti kūną sudarančius atomus į kažką naudingesnio iš mašinos perspektyvos.

„Norėdami įsivaizduoti priešišką antžmogišką DI, neįsivaizduokite negyvo, protingo mąstytojo, besislapstančio internete ir siunčiančio užkrėstus el. laiškus. Įsivaizduokite visą nežemišką civilizaciją, mąstančią milijonus kartų greičiau nei žmonės, iš esmės ribojamą kompiuterių, pasaulyje, kuriame būtybės, iš jos perspektyvos, yra labai kvailos ir labai lėtos“, – siūlė Yudkowsky. Yudkowsky taip pat siūlo apsvarstyti galimybę, kad dirbtinis intelektas galėtų plėstis už interneto ribų, jei tik to norėtų. Jo pasiūlymai skamba kaip scena iš mokslinės fantastikos filmo: mašinos išmoks vogti iššifruotą atskirų žmonių DNR, kurti dirbtines gyvybės formas jos pagrindu arba net organizuoti postbiologinę molekulinę gamybą, kur galės surinkti tam tikras medžiagas į visiškai skirtingas.

Technologijos genijai prašė sulėtinti dirbtinio intelekto plėtrą.

Yudkowskio susirūpinimas buvo atsakas į kovo pabaigoje paskelbtą laišką, kurį paskelbė Vakarų ne pelno siekianti organizacija „The Future of Life Institute“, tirianti egzistencinę grėsmę žmonijai keliančius klausimus. Amerikiečių technologijų verslininkas Elonas Muskas šį projektą pradėjo praėjusio

dešimtmečio viduryje ir buvo vienas iš pagrindinių pasirašiusių po gana trumpu laišku apie dirbtinio intelekto ateitį. Tačiau jis nebuvo vienintelis: laišką pasirašė daugiau nei dešimt tūkstančių žmonių. Tarp jų buvo „Apple“ įkūrėjas Steve'as Wozniakas, vienas iš žymiausių šiuolaikinių istorikų Yuvalis Noah Harari ir dešimtys žinomų dirbtinio intelekto tyrėjų. Pagrindinė dokumento žinutė panaši į Yudkowskio: laiške teigiama, kad dirbtinis intelektas atneš didelių pokyčių žmogaus gyvenime, todėl jo plėtrą reikia valdyti. Tačiau kadangi niekas to nedaro, reikalingas tolesnės plėtros moratoriumas. Eliezer Yudkowsky, Mašininio intelekto tyrimų instituto bendrąkūrėjas:

„Jei kas nors dabartinėmis sąlygomis sukurs pernelyg galingą dirbtinį intelektą, manau, kad netrukus po to mirs kiekvienas žmonių rasės narys ir visa biologinė gyvybė Žemėje. Jei ir toliau taip elgsimės, mirs visi, įskaitant vaikus.“

„Šiuolaikinės dirbtinio intelekto sistemos tampa konkurencingos sprendamos pagrindines problemas“, – teigiama laiške. „Turime savęs paklausti: ar turėtume leisti mašinoms užtvindyti mūsų informacijos srautus propaganda ir melu? Ar turėtume automatizuoti visus darbus, įskaitant sprendimus priimančių asmenų? Ar turėtume ugdyti ne žmonių protus, kurie galiausiai galėtų mus pranokti skaičiumi, pergudrauti ir pakeisti? Ar turėtume rizikuoti prarasti savo civilizacijos kontrolę?“



Elon Musk buvo vienas iš įmonės, sukūrusios GPT kalbos modelius, įkūrėjų.

Nuotrauka: Susan Walsh / AP

Laiško autoriai ir tie, kurie su jais sutinka, siūlo sustabdyti intelektualią sistemų mokymą ateinantiems šešioms mėnesiams. Yudkowsky atsakydamas mano, kad toks moratoriumas turėtų trukti mažiausiai 30 metų. Pasak Musko ir kitų pasirašiusių, per ateinančius šešis mėnesius joks neuroninis tinklas neturėtų tapti protingesnis už GPT-4. Oficiali dokumento paskelbimo priežastis buvo naujos, daug reklamos sulaukusio „OpenAI“ (kurią kartu įkūrė Muskas) produkto versijos išleidimas.

Kas yra GPT-4 ir „ChatGPT“?

GPT-4 yra ketvirtasis „OpenAI“ sukurtos neuroninių tinklų šeimos narys. Neuroniniai tinklai, paprastai tariant, yra skaičiavimo galios bendruomenė, sujungta kaip žmonių ar gyvūnų nervinių ląstelių tinklai. Daugelio procesorių, kurie visi veikia griežtai pagal tam tikrą matematinį modelį, derinys daro šią sistemą gana tvirtą.

Tuo pačiu metu neuroninius tinklus galima apmokyti naudojant įvairius metodus, todėl jie tampa daug protingesni. Ažiotažas, susijęs su GPT-4 ir jo taikomąja versija, pokalbių robotu „ChatGPT“, kyla dėl to, kad neuroninis tinklas tapo tiek daug protingesnis, kad jį gali matyti net paprastas žmogus. Ir tai buvo atradimas. „ChatGPT“ bendrauja su žmonėmis kaip žmogus ir per šį bendravimą toliau vystosi. Kūrėjai jau atsisakė vien tekstinės sąveikos su robotu ir dabar jis gali bendrauti naudodamas vaizdus, garsą ir vaizdo įrašus.

GPT vystosi per greitai

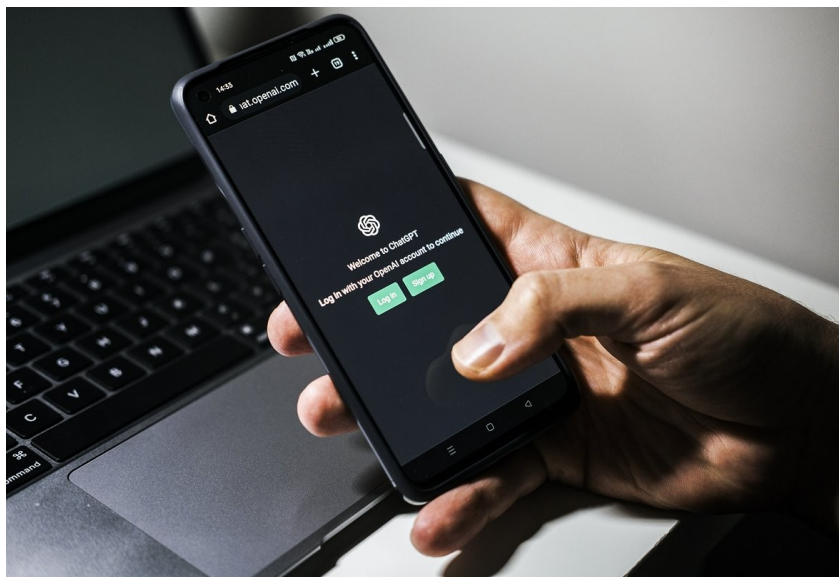
Pastarosiomis savaitėmis spartus neuroninių tinklų vystymasis patraukė ne tik ekspertų bendruomenės, bet ir atskirų šalių vyriausybių dėmesį. Tai įvyko pasirodžius pranešimams, kad nusikaltėliai pradėjo naudoti dirbtinį intelektą. Net Europolas išreiškė susirūpinimą dėl šios situacijos, įspėdamas apie galimą riziką kibernetinei erdvei, kurią kelia esami dirbtinio intelekto modeliai.

30 metų –

Eliezeris Yudkowsky siūlo bent tiek laiko sustabdyti neuroninių tinklų kūrimą.

Visų pirma, teisėsaugos pareigūnai baiminasi, kad neuroniniai tinklai galės apeiti saugumo programinę įrangą ir padėti organizuoti kibernetines atakas, padėti apmokyti kibernetinius nusikaltėlius naujos taktikos ar net generuoti unikalių kenkėjišką kodą. Visa tai sumažins įsilaužimo kliūtis, todėl ji bus prieinama net ir tiems, kurie mažai arba visai nesupranta šios temos.

Europolas pripažįsta, kad dalis šios informacijos jau yra prieinama internete, tačiau „ChatGPT“ gali ją rasti per kelias sekundes, o ne per dienas ar mėnesius. Pagrindiniai argumentai prieš platų „ChatGPT“ diegimą yra susiję su technologijos privatumo pažeidimais (bet kokie neuroninio tinklo gauti asmens duomenys tampa medžiaga dirbtinio intelekto tyrimams ir tolesnei plėtrai) ir teisinio pagrindo naudoti naudotojų duomenis nebuvimu. Kiti nusiskundimai – negalėjimas patikrinti naudotojo amžiaus, nepaisant oficialaus „ChatGPT“ naudojimo apribojimo jaunesniems nei 13 metų asmenims. Europos bendruomenei, kuri jau keletą metų taiko gana griežtas vyriausybės ir verslo sąveikos su piliečių asmens duomenimis taisykles, toks netobulumas yra neišsivaizduojamas.



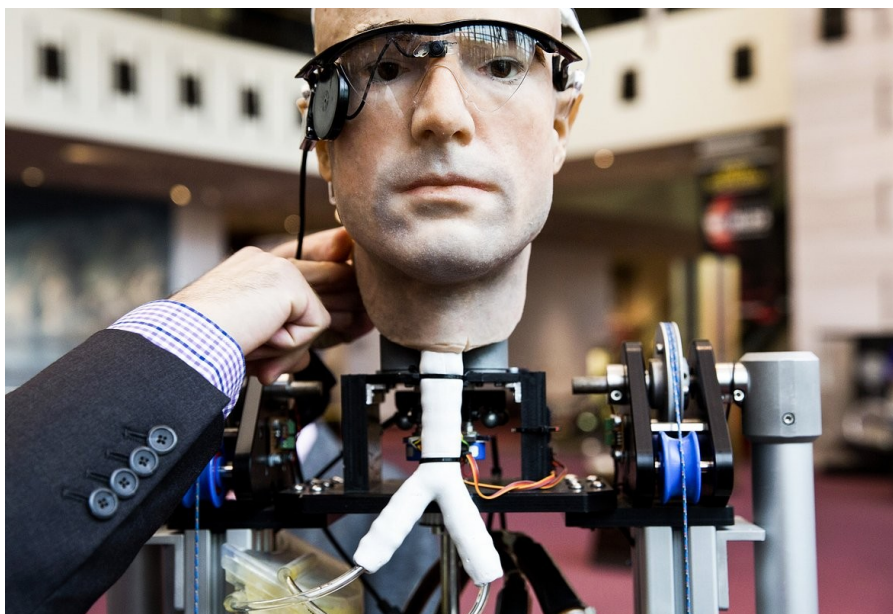
„ChatGPT“ pokalbių robotas jau pradėtas riboti Europoje.

Nuotrauka: Iryna Imago / Shutterstock / Fotodom

Tai buvo oficialus pagrindas apriboti pokalbių roboto naudojimą Italijoje. Pranešama, kad panašią priemonę svarsto ir Vokietijos, Prancūzijos bei Airijos valdžios institucijos. Dialogą tarp Europos vyriausybių ir GPT kūrėjų stabdo oficialaus neuroninio tinklo atstovavimo ES nebuvimas. Ne visi tiki neišvengiamu žmonijos sunaikinimu.

Idėją, kad dirbtinis intelektas gali peržengti internetą ir sunaikinti visą egzistenciją, palaiko ne visi vizionieriai. Be to, daugelis žymių Vakarų mokslininkų, kurie, kaip ir Yudkowsky, šiai temai paskyrė beveik visą savo gyvenimą, pasisakė prieš ją. Pagrindinis Musko ir jo rėmėjų priešininkas buvo ankstesnės kartos rinkos gigantų atstovas – „Microsoft“ įkūrėjas Billas Gatesas.

„Nemanau, kad paprašius vienos konkrečios kūrėjų grupės sustabdyti savo tyrimus, kas nors pasikeis. Dirbtinio intelekto naudojimas teikia milžinišką naudą. Tai reiškia, kad mums tereikia suprasti riziką“, – sakė Gatesas.



Kai kurie baiminasi, kad dirbtinis intelektas žmones pakeis kita, žmogaus sukurta gyvybės forma.

Nuotrauka: Joshua Roberts / Reuters

Filantropas išvardijo šias rizikas savo straipsnyje apie dirbtinio intelekto ateitį, kuris buvo paskelbtas lygiai dieną prieš „Future of Life Institute“ laišką. Savo straipsnyje, tinkamai pavadintame „Dirbtinio intelekto amžius prasidėjo“, Gatesas aptaria DI indėlių į įvairias gyvenimo sritis („Jis pakeis tai, kaip žmonės dirba, mokosi, keliauja, gauna sveikatos priežiūros paslaugas ir bendrauja tarpusavyje“, – rašė „Microsoft“ įkūrėjas) ir galimybes, kurias jis siūlo verslui. Tarp jo išvardytų iššūkių buvo dirbtinio intelekto nesugebėjimas visada suprasti užklausų konteksto ir jo silpnosios vietos abstrakčiose problemose, taip pat jo potencialas karinėse srityse. Milijardierius taip pat užsiminė apie superintelektualius neuroninius tinklus, nors ir išreiškė skepticizmą dėl jų tiesioginio potencialo.

„Šie superintelektualūs neuroniniai tinklai greičiausiai galės nusistatyti savo tikslus. Kokie bus tie tikslai? Kas nutiks, jei jie prieštaraus žmonijos interesams? Ar turėtume bandyti užkirsti kelią stipraus dirbtinio intelekto vystymuisi? Šie klausimai laikui bėgant taps vis aktualesni. Tačiau nė vienas iš pastarųjų mėnesių proveržių mūsų reikšmingai nepriartino prie stipraus dirbtinio intelekto.“

„Dirbtinis intelektas vis dar nekontroliuoja fizinio pasaulio ir negali nusistatyti savo tikslų“, – įsitikinęs Gatasas.



Dirbtinis intelektas taps internetinės žmonių, valdžios ir verslo sąveikos pagrindu.

Nuotrauka: Fabian Sommer / dpa / Globallookpress.com

Tai atkartoja ir dirbtinio intelekto srities politikos dokumentų bei tyrimų autoriai, įskaitant tuos, kuriuos cituoja „Future of Life Institute“ laiško autoriai.

Pavyzdžiui, Konektikuto universiteto docentė Shiri Dori-Hacohen teigė, kad „neuroniniai tinklai nebūtinai turi tapti itin intelektualūs, kad padidėtų branduolinio karo rizika“. Tačiau ji tvirtina, kad aptariant problemą kaip visumą reikėtų vengti Holivudo dramos. Kai kurie įvairių Vakarų žiniasklaidos priemonių apklausti ekspertai tiesiogiai teigė, kad dirbtinio intelekto plėtros priešininkai kuria „neįsivaizduojamus apokaliptinius scenarijus“.

Nepaisant to, daugelis su dirbtiniu intelektu susijusių rizikų, įskaitant ir tas, kurias Gatasas paminėjo, yra gana realios. „Pernelyg pažangaus dirbtinio intelekto“ pavojus yra tas, kad jis gali užtvindyti kanalus propaganda, atimti žmonių darbus ir paskatinti sukurti nežmogišką intelektą, galintį pergudrauti

žmones ir padaryti jų žinias pasenusias ir nereikalingas. Žmonėms jau dabar sunku atskirti dirbtinio intelekto generuojamas netikras naujienas nuo tiesos. Pavyzdžiui, beveik 70 procentų suaugusiųjų negali atskirti „Chat GPT“ parašyto el. laiško nuo žmogaus parašyto. Nors daugelis fantazuoja apie mašinų sukilimą, manau, kad tikrosios dirbtinio intelekto išplitimo pasekmės gali būti visai kitokios“, – teigia Jevgenijus Nepeyvoda, „Novo BI“ vadovaujantis partneris.

„Vienintelis pavojus, kurį matau ir kodėl galbūt verta apriboti dirbtinio intelekto naudojimą, yra tas, kad neuroniniai tinklai dabar gali būti naudojami įvairiems tekstams, moksliniams straipsniams ir naujienoms generuoti“, – sutinka kibernetinio saugumo specialistas Vladimiras Uljanovas. **„Kuo daugiau informacijos skleidžiama, tuo sunkiau suprasti, kas ją sukūrė ir kiek ji tikra.“** „Gali pasirodyti netikrų naujienų.“

Sparčiai vystantis neuroniniams tinklams, IT ekspertai specializuotose žiniasklaidos priemonėse piešė rožinį vaizdą apie magiškus žmogaus gyvenimo pokyčius. Dar prieš kelerius metus bendras tokių medžiagų tonas buvo visiškai teigiamas; tačiau dabar, įskaitant ir „Future of Life Institute“ publikacijas, jis tapo daug atsargesnis.