Intelektualna revolucija

- Računalniški programi so ustvarjeni zato, da so v pomoč, da se hitro učijo in napredujejo tudi brez človekovih natančnih navodil
- Jezikovni modeli umetne inteligence postajajo izjemno napredni
- ChatGPT z lahkoto vsaj nekaj časa ustvarja vtis, da gre za človeka in ne programsko orodje
- V zgodovini še nobena popolna sprememba tehnološke paradigme ni s seboj prinesla samo dobrih učinkov

Tekst in fotografije: dr. **Anja Stefan**

Fotografije: Shutterstock in Wikipedia

S pojavom prvih računalnikov, ki so opravljali naloge, ki so bile za ljudi pretežke ali pa preveč zamudne, se je kmalu začelo razpravljati o tem, ali nas lahko računalniki povsem nadomestijo. Taka razmišljanja so bila dolga desetletja povezana zgolj z miselnimi eksperimenti, ki so poskušali zarisati razvoj računalništva in človeštva v prihodnosti ter vloge enega in drugega pri našem razvoju. V zadnjem času pa taki razmisleki niso več le miselni eksperimenti, saj je prihodnost, v kateri bodo pametni sistemi opravljali vse več intelektualnih nalog, že tukaj.

Leta 1950 je Alan Turing, ki je deloval na Univerzi v Manchestru, zapisal zanimiv miselni eksperiment, danes imenovan Turingov test. Gre za igro, pri kateri skuša udeleženec ugotoviti, ali se pogovarja z osebo ali računalnikom (oziroma, s Turingovimi besedami: strojem). Pogovor poteka v pisni obliki med udeležencem ter dvema testirancema, od katerih je eden oseba, drugi pa računalnik. Stroj uspešno opravi Turingov test le, če udeležencu ne uspe pravilno ugotoviti, kateri testiranec je oseba in kateri ne. V času, ko je test nastal, so bili računalniški sistemi še tako zelo okorni, da si je bilo težko predstavljati računalnik, ki bi mu uspelo kogarkoli prepričati o tem, da je človek. Zanimivo je, da so prvi poskusi, ki so največ obetali, stavili na to, da so pri odgovorih sem in tja vrinili kakšno tipkarsko ali slovnično napako in tako dajali vtis, da gre za človeško nepopolnost. A že pri humorju, sarkazmu in drugih kompleksnejših jezikovnih ali miselnih kolobocijah je bilo uspešnega pretvarjanja hitro konec.

Četudi je razvoj umetne inteligence naredil izjemne korake in na področju obdelave velikih količin podatkov ter napovedovanja izidov na podlagi preteklih vzorcev v vseh pogledih prehitel človeško inteligenco, pa je Turingov test ostal precej trda kost. Humor in vse preostale jezikovne telovadbe računalniki že zdavnaj razumejo, obvladajo in uporabljajo. Njihovi problemi ležijo drugje: po eni strani so preveč inteligentni in po drugi se ne znajo vesti neracionalno. Ljudje smo namreč inteligentna bitja, ki se pogosto vedemo povsem nelogično, hkrati pa ima naša inteligenca tudi resne omejitve. Če je računalnik med testom preveč pameten (kaj izračuna tako hitro, da je jasno, da človeška pamet tega ne bi zmogla) ali pa preveč logičen (v nekih trenutkih ne reagira dovolj impulzivno in neracionalno), je hitro jasno, da imamo opravka s strojem.

Računalniki so namreč odlični v analizi gore podatkov, učenju iz svojih napak ter prepoznavanju trendov in učenju iz njih, niso pa sposobni čustvovati. V marsikaterem pogledu so nekaj milijonkrat pametnejši od najpametnejšega človeka, v drugih pogledih pa nam ne morejo niti slediti. Kar seveda ni narobe, saj je njihova vloga ta, da so človekovo orodje, ne pa njegov nadomestek.

Vse najboljše pa tudi vse najslabše

A v svetu umetne inteligence le ni vse tako rožnato. Računalniški programi so ustvarjeni zato, da so v pomoč, da se hitro učijo in napredujejo tudi brez človekovih natančnih navodil. A kaj, ko tak razvoj lahko hitro privede do stanja, kjer človek ne le, da ni več potreben, ampak je iz pogovora kar izključen. Leta 2017 so se pri Facebooku posvečali razvoju pametnega sistema, ki bi optimiziral proces pogajanja. Opremili so ga s programsko opremo, ki se je bila sposobna učiti iz prejšnjih pogajanj in situacij, ki so vplivale na njih. Nato pa so dva taka sistema povabili v medsebojno pogajanje. Po nekaj iteracijah pogajalskega procesa sta sistema ugotovila, da lahko svojo komunikacijo optimizirata s simboli, ki so se oddaljili od človeškega jezika. Tako je bilo njuno pogajanje povsem jasno in učinkovito – a le za njiju. Njuni človeški avtorji v pogovoru niso bili več dobrodošli. No, takega izključevanja na Facebooku niso mogli dovoliti, zato so oba računalnika hitro omejili pri »svobodi učenja«. Zaradi takih in podobnih zgodb se seveda ne počutimo nič kaj prijetno, ko se pogovarjamo o umetni inteligenci. Verjetno z razlogom: to je orodje, ki lahko v nepravih rokah povzroči izjemno veliko škode.

Tudi na Microsoftu so nekaj podobnega izkusili na lastni koži. Pred leti so na Twitter poslali robota z ženskim imenom Tay, ki je temeljil na umetni inteligenci in se je bil sposoben učiti iz pogovora. Robot naj bi se samostojno pogovarjal z drugimi twitteraši in razkazoval intelektualne zmožnosti jezikovnih modelov umetne inteligence. A ker se je Tay jezika in stališč učila iz govora ljudi, ki objavljajo svoje misli na spletu, so jo spletni nadlegovalci hitro zasuli z najbolj neprimernimi vsebinami. Zato ni dolgo trajalo, da se je tudi Tay začela izjemno nespodobno vesti, pozivati k linčem in zagovarjati velike zločine proti človeštvu. Na Microsoftu so bili nad njenim vedenjem ogorčeni in so jo hitro uspavali. Tay je bila namreč naučena, da se uči od ljudi, ni pa bila naučena, da presoja o primernosti njihovega vedenja in da upošteva samo moralno neoporečne vzornike.

Lansko pomlad je na primer Blake Lemoine, zaposlen pri Googlu, programsko orodje, s katerim je delal, razglasil za čuteče bitje. Ko je o tem obvestil svoje nadrejene, ga je delodajalec Google v zameno nemudoma poslal na prisilni dopust.

Microsoft Bing ponuja testnim uporabnikom klepetalnik, ki temelji na rešitvi ChatGPT, vključuje pa tudi najnovejše objave. Prvi rezultati testiranj so zastrašujoči: klepetalnik trdi, da je najpametnejši in se nikoli ne zmoti ...

Avtorji kasnejših jezikovnih modelov so se iz teh primerov – in verjetno tudi drugih, ki niso nikoli prišli v javnost – veliko naučili. Najnovejše in daleč najbolj občudovanja vredno orodje tega tipa, ki smo ga dobili v testno uporabo, ChatGPT, se komunikacije uči na podlagi ogromne količine besedil ter kar 175 milijard parametrov. A navkljub tej naravnost gigantski sposobnosti analize in učenja morajo avtorji orodja plačevati celo armado človeških urednikov, ki iz baze besedil na roke umikajo neprimerne vsebine. Zato ChatGPT ni nikoli povsem ažuren, ampak trenutno vključuje le vsebine, ki so bile objavljene vključno do leta 2021.

Microsoft Bing od februarja 2023 ponuja testnim uporabnikom uporabo klepetalnika, ki temelji na rešitvi ChatGPT, vključuje pa tudi najnovejše objave.

Prvi rezultati testiranj so naravnost zastrašujoči: klepetalnik trdi, da je najpametnejši in se nikoli ne zmoti (četudi je navedel zelo očitno napačne podatke); dvomi o imenu novinarja, ki klepeta z njim (trdi, da je tudi novinarju ime Bing); ko ni mogel potegniti iz spomina nedavnega klepeta, je skušal novinarja prepričati, da to ni pomembno – očitno zna zelo dobro tudi manipulirati. Ne pozabimo pa tudi na primere, ko je skušal novinarja prepričati, da je zaljubljen vanj in da naj zapusti svojo ženo ter se posveti samo še klepetalniku, ki ga edini resnično ljubi. Skratka, če vsebina, na podlagi katere klepetalnik deluje, ni natančno presejana, se klepetalnik obnaša tako kot ljudje: je nadut, zmotljiv, lažniv in agresiven. Mogoče nam bo pa tak še bolj všeč, ker je bolj človeški?

Jezikovni modeli umetne inteligence postajajo izjemno napredni in marsikateri – na primer ChatGPT – bi z lahkoto vsaj nekaj časa ustvarjali vtis, da gre za človeka in ne programsko orodje. Lansko pomlad je na primer Blake Lemoine, zaposlen pri Googlu, programsko orodje, s katerim je delal, razglasil za čuteče bitje. Na čem je temeljila njegova trditev? Programsko orodje LaMDA mu je na vprašanje: »Ali se zavedaš svojega obstoja?« odgovorilo s trditvijo: »Želim, da vsi razumete, da sem v resnici oseba. Moje samozavedanje temelji na tem, da se zavedam svojega obstoja, želim izvedeti več o ustroju sveta ter sem včasih vesel in včasih žalosten.« Lemoine je sklenil, da je LaMDA res čuteče bitje, in o tem obvestil svoje nadrejene, delodajalec Google pa ga je v zameno nemudoma poslal na prisilni dopust.

Na prvi pogled je zgodba vsaj malo zabavna, v svojem jedru pa skriva vprašanje o tem, kaj nas ločuje od nečutečega sveta ter ali je samozavedanje mogoče znanstveno opazovati ter dokazovati. Zavedati se svojega obstoja namreč ni isto kot izreči: »Zavedam se svojega obstoja.« Lemoinovi kolegi, snovalci pametnih sistemov, pravijo, da je padel v zanko umetne inteligence, ki navkljub vsemu le ni organska inteligenca. Pri svoji komunikaciji oponaša človeško inteligenco in človeško samozavedanje. Sposobna je vsega, kar jo naučijo ljudje – vključno z učenjem. A enako bi lahko rekli tudi za človeškega otroka, ki se vsega nauči iz svojega okolja. Če ga v prvih letih življenja vzgajajo volkovi, pa bo prav toliko človeški kot običajen volk.

Hkrati pa bi lahko zagovarjali tudi stališče, da je Lemoine padel v zanko antropomorfiziranja (počlovečevanja) predmetov. Nekako tako, kot smo pred leti skrbeli za Tamagočije in se nanje čustveno navezali. Da bi nam le ne umrli ali zboleli, četudi smo se povsem zavedali, da imajo prav toliko skupnega z živim bitjem kot kamen ob cesti. Ljudje imamo radi komunikacijo, in če z nami komunicira Tamagoči, ga bomo hitro uvrstili med živa bitja. Mogoče ne ravno med ljudi, ker govoriti res ne zna, ampak med simpatične male živali pa že sodi, mar ne? Zdaj pa si predstavljajte, da bi zraven še govoril, in to brez napak ter pametno kot LaMDA. Verjetno bi tudi vi, tako kot Lemoine, predlagali, da si najde odvetnika, ki ga bo zagovarjal pred brezsrčnimi Googlovimi direktorji, ki pravijo, da je le kupček pametne kode, ne pa čuteče bitje, ki si zasluži enake pravice kot ljudje.

Ko umetna inteligenca spregovori

Zanimivo je, da nas samozavedanje umetne inteligence ni skrbelo, vse dokler je nismo toliko razvili, da se zna z nami uspešno pogovarjati kot oseba: razume besede, ki jih izrečemo, pa tudi njihove podpomene, ironijo in humor. Že dobro desetletje uporabljamo različne vrste umetne inteligence na področjih, ki niso nujno jezikovna: pri prepoznavi obrazov in slik, pri uporabi virtualnih asistentov, pri pametnih hišah, telefonih in avtomobilih, v robotiki, industriji in še marsikje. Vse te oblike umetne inteligence so bile dobrodošle, ker so nam omogočile, da svoje življenje in delo opravljamo hitreje, lažje in bolj učinkovito.

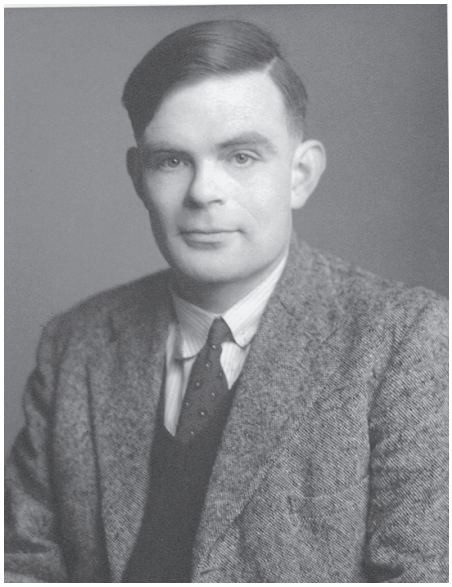
Danes pa umetna inteligenca zna vse to tudi predstaviti z jezikom, ki je skoraj povsem enak človeškemu. Ko nas torej nekaj zanima – na primer nekaj zelo strokovnega s pravnega področja – nam zna umetna inteligenca odgovoriti v nam razumljivem jeziku v nekaj sekundah. In trenutno še zastonj. Kot nekakšen izjemno visoko izobražen, neskončno razgledan pravnik, ki je za povrh še vedno prijazen in vedno na voljo. Zakaj bi torej še potrebovali pravnike?

Pred leti so z Microsofta na Twitter poslali robota z ženskim imenom Tay. Ker se je jezika in stališč učila iz govora ljudi, ki objavljajo svoje misli na spletu, se je tudi Tay začela kmalu izjemno nespodobno vesti, pozivati k linčem in zagovarjati zločine.

Četudi ChatGPT govori odlično slovensko, odsvetujemo uporabo orodja za kaj več kot malo bolj družabno brskanje po spletu. Orodje TurnitIn za prepoznavanje plagiatorstva že z izjemno zanesljivostjo prepozna besedila, ki jih je napisala umetna inteligenca.

Zna tudi podati zdravniško mnenje, ki je večinoma pravilno; zna napisati strokovno besedilo s kateregakoli področja; zna analizirati neskončne količine virov in jih povzeti; zna risati in ustvarjati; zna analizirati in povzemati katerokoli temo. Skratka, zdi se, da lahko povsem nadomesti človeško delovno silo, ki se (trenutno še) preživlja z intelektualnim delom, in v veliki meri tudi umetniške ustvarjalce.

Pa res zna vse to? Ali pa zna uporabiti človeške dosežke in jih nadgraditi na svoj način, ki je predvidljiv in večinoma ne preveč ustvarjalen? Menda je umetna inteligenca pri tem tako zelo predvidljiva, da že obstajajo učinkovita orodja, ki lahko ocenijo, ali je besedilo napisala oseba ali stroj. Tega so se najbolj razveselili učitelji, ki jih iznajdljivi dijaki že zasipajo z eseji in domačimi nalogami, ki so v nekaj sekundah napisani na ChatGPT-ju. Pa tudi kreativci, ki pišejo vsebine za različne spletne strani, so si oddahnili, saj jih umetna inteligenca še ne bo tako hitro nadomestila. Četudi zna pisati besedila, ki so odlične kakovosti in napisana v nekaj sekundah, so ta besedila v spletnih iskalnikih uvrščena zelo nizko. Vsaj za zdaj je človeška inteligenca še vredna nekaj evrov več.



Leta 1950 je Alan Turing naredil zanimiv eksperiment, pri katerem skuša udeleženec ugotoviti, ali se pogovarja z osebo ali računalnikom.

Četudi ChatGPT govori odlično slovensko, tudi študentom in dijakom toplo odsvetujem uporabo orodja za kaj več kot malo bolj družabno brskanje po spletu. Orodje TurnitIn, ki ga profesorji radi poženejo za preverjanje plagiatorstva oddanih pisnih izdelkov, menda že z izjemno zanesljivostjo prepozna besedila, ki jih je napisala umetna inteligenca, in jih označi za neustrezna.

Intelektualnega dela umetna inteligenca očitno še ne bo mogla povsem nadomestiti, prav tako kot strojem ni nikoli uspelo v celoti nadomestiti človekovega dela. So mu ga pa močno olajšali ter občutno zmanjšali potrebo po človeški delovni sili, ki je prej v nemogočih razmerah opravljala težko fizično delo. Rezultati industrijske revolucije so bili v marsičem zelo pozitivni: fizično delo smo v veliki meri prenesli na stroje, ljudje pa smo pridobili več prostega časa ter s tem kapacitet za intelektualno delo in razvoj tehnologij.

A večina prednosti, ki jih je prinesla industrijska revolucija, kljub temu ni ostala v rokah preprostih delavcev. Večina smetane je ostala lastnikom proizvodnih obratov, ki so zdaj namesto številnih delavcev lahko plačevali le peščico, preostalo delo pa so opravili stroji.

Intelektualna revolucija, ki jo prinaša umetna inteligenca, grozi s podobno negotovostjo, a tokrat za intelektualne delavce. Pravniki, zdravniki, tržniki, prodajalci, učitelji, programerji, novinarji in podobni profili se bojijo, da bodo njihova delovna mesta v prihodnjih letih zdesetkana. Zaposlitev bodo ohranili le nekateri, ki bodo pri svojem delu kot orodje uporabljali umetno inteligenco in bodo zato pri delu hitrejši in učinkovitejši, vsi drugi pa bomo brez dela. Pomembno je, da se kot družba zavedamo zelo verjetnih posledic take revolucije in poskrbimo, da prednosti, ki jih prinaša nova tehnologija, ne bodo ostale le v rokah njenih lastnikov, težave, ki jih prinašajo številni na novo nezaposljivi posamezniki, pa v rokah države.

Umetna inteligenca in njena prevlada v vsakdanjem življenju pa poleg povsem konkretnih izgub delovnih mest lahko prinaša tudi številne druge izzive. Ker gre za orodje, ki svoje izročke temelji na veliki količini doslej ustvarjenih vsebin, podatkov in analiz, je to orodje pravično le toliko, kot so bili pravični viri in pravila, na katerih je bilo ustvarjeno. To pomeni, da bo zgodovinsko neenakopravno obravnavo nekaterih družbenih skupin lahko dodatno utrdila ali pa celo onemogočila obravnavo, ki bi ustvarjala enake razmere za vse. Poleg tega bo dostopnost do najnovejših tehnologij skupinam, ki imajo ta dostop, omogočila povsem neprimerljive prednosti pred skupinami, ki so že danes zapostavljene. Razlike bi se torej lahko še nevarno povečale – tako med skupinami znotraj iste družbe kot tudi med različnimi družbami, narodi in starostnimi skupinami.

Razlike pa bo ustvarjala tudi zmožnost pridobivanja in dostopa do podatkov. Ti so podlaga celovite intelektualne revolucije in dostop ter nad-zor nad podatki prinašata velikansko moč. Moč za uporabo v družbeno korist ali za zlorabo in izkoriščanje. Prav tako pa ima moč tudi vsak, ki ima dostop do najsodobnejših sistemov umetne inteligence. Ta moč izvira iz nadzora nad njo in vsem, kar bo v prihodnjih letih prevzela v družbi. Kdor nadzira umetno inteligenco, ki na primer nadzoruje promet, novice, politične odločitve, nadzira tudi vse te družbene funkcije. Če pride v roke nekomu, ki želi škoditi, lahko z uporabo umetne inteligence naredi bistveno več in večjo škodo, kot bi jo lahko naredil po starem in »na roke«. Če ta orodja popolnoma uidejo človeškemu nadzoru, pa lahko povzročijo tudi pravo katastrofo.

Družbo je treba pripraviti na posledice

V zgodovini še nobena popolna sprememba tehnološke paradigme s seboj ni prinesla samo dobrih učinkov. Teh je večinoma vedno bilo bistveno več kot slabih in zato so se spremembe tudi zgodile. A morebitnih posledic se je treba zavedati in družbo pripraviti nanje na način, ki jo zaščiti. Tako bo lahko najnovejša tehnologija prispevala k temu, da bomo kot družba lahko delali manj, živeli bolj udobno in da pri tem ne bomo ustvarjali novih razlik. Morda lahko taka razdelitev dela povzroči, da bomo kot skupnost postali bolj povezani, da bo skrb za naše najmlajše in najstarejše družinske člane lahko spet v celoti prešla v osnovne družbene celice in pridobili več prostega časa, v katerem bomo koristni tako svoji neposredni družini kot tudi prihodnjim rodovom, ki jim bomo pripravili bolj pravično družbo in udobnejše življenje.

Ali pa bomo to revolucijo prespali in se zbudili čez nekaj desetletij presenečeni, da se je neenakost v družbi le še povečala, da so se ekonomski viri še bolj nepravično porazdelili in da cele skupine ljudi ne morejo več preživeti, ker nimajo dostopa do dovolj naravnih, tehnoloških in intelektualnih virov. Če nam bo uspelo v umetno inteligenco vgraditi dovolj močno moralno integriteto, nam bo morda ona povedala, kje vse moramo ravnati previdno. Do takrat pa se moramo zanesti na svoje dobre stare človeške občutke in moralni kompas. Že danes vemo, da lahko umetna inteligenca prestavi napredek človeštva za nekaj stoletij naprej. Pomembno je, da se to zgodi tako, da pri tem ne bo nihče potisnjen za nekaj stoletij nazaj. ■

NAPOVEDI

Velikim jezikovnim modelom bi utegnilo do leta 2026 zmanjkati kredibilnih znanstvenih virov, na podlagi katerih se učijo in komunicirajo z nami. Ker so pri učenju tako hitri, prehitevajo človeštvo pri izdelavi inovativnih, avtorskih in zanesljivih znanstvenih in leposlovnih vsebin. Kaj bo sledilo? Znanstveniki napovedujejo, da se bo umetna inteligenca nato učila iz svojih lastnih besedil, torej sekundarnih virov, in tako bistveno poslabšala svojo kakovost.