

Umetna inteligenca je sposobna kodirati, ustvarjati varljive fotografije, digitalno obdelane videoposnetke, šolske spise in tudi pesmi. Kar napiše, je sicer videti smiselno, ampak ni nobenega zagotovila, da bo tudi res pravilno. FOTO SHUTTERSTOCK

podlagi tega se je potem razvila cela debata o tem, ali je umetna inteligenca pristranska. Ampak znanstveni članki, ki so tudi analizirali sistem compas, tega niso ugotovili. Nekateri ugotavljajo, da članek v *ProPublici* temelji na popolnoma napačni analizi. Moram reči, da se kar strinjam z njimi, čeprav nimamo dokončno sprejeta odgovora na to. Kot vem, se sistem compas v ameriškem pravosodju še uporablja.

Potrebujemo komisije in odbore, ki bodo presojali, katere podatke zajemati? Bi morala javnost natančno vedeti, od kod so podatki?

Bratko: Mislim, da bi bilo to dobro, da bi vključili tako komisije kot odbore.

Gams: Ni dvoma o tem, da moramo vedeti, od kod izvirajo podatki, ki jih zajema umetna inteligenca.

Kakšne so strojne omejitve umetne inteligence?

Bratko: Nekatero metode umetne inteligence, tudi metode velikih jezikovnih modelov, zahtevajo ogromne računalniške zmogljivosti. V tem pogledu se pogosto omenja njihova problematičnost tudi zaradi velikega ogljičnega odtisa. Dostikrat se mi zdi, da dobim v oceno članek, poslan v objavo v znanstveni reviji, pa je rezultate v članku težko preveriti. Za to bi potreboval veliko večje računalniške zmogljivosti, kot so mi enostavno dostopne. Tak razvoj umetne inteligence je očitno omejen z zmoglostjo strojne opreme. Sam mislim, da uspešen razvoj umetne inteligence ni nujno vezan na tako zahtevno strojno opremo.

Gams: Moorov zakon še velja. Težave se rešujejo s 3D-čipi, direktno povezavo grafike z računalniki itd. Veliko se dogaja na tem razvojnem področju in razvoj bo desetletja ostal eksponenten. Računalniki in umetna inteligenca so med najhitreje razvijajočimi se področji, ki največ doprinejajo k razvoju človeške civilizacije.

Dr. Radovan Komel je v intervju za Delo dejal, da bi se bolj kot genskega inženiringa morali bati umetne inteligence. Kaj menita?
Bratko: Leta 2013 sem bil na večerji s Stuartom J. Russellom, ki je znan kot avtor najbolj cenjenega učbenika o umetni inteligenci. Še dobro se spominim, kako je dejal: »Do zdaj smo lahko uživali v umetni inteligenci, zdaj pa nenadoma ni več zabavno. Kot da smo izumili atomsko bombo. Mislim, da je čas, da se začnemo spraševati, kaj narediti njo.«
Gams: Stroka ve, česa ne smemo delati, tukaj skoraj ni različnih mnenj. Recimo, ne bi smela biti uporabljena v obliki avtonomnega orožja. Politika pa ima drugačne poglede in pogosto presenetni s svojimi sklepi.

Bo kdaj umetna inteligenca dobila status živega bitja?

Bratko: Neumnost, nikoli.
Gams: Se strinjam. Ne bi je smela, čeprav so se o tem v evropskem parlamentu že pogovarjali. Ko pa bo umetna inteligenca prestala Turingov test, bomo morali o tem vseeno razmisлити. Recimo, psa ne smemo kar »ugasniti« kot lahko avto ali program. Kako pa z zares zavestnim in inteligentnim programom?

Od tega smo še zelo daleč, glede na prejšnje povedano. Kakšne pa so kratkoročne dileme?

Bratko: Besedila, ki jih generirajo veliki jezikovni modeli, predstavljajo potencialno grožnjo v obliki velikanskega informacijskega onesnaženja. Ni si težko predstavljati zlorab za politične namene, ciljno prilagojenih brezhbirno napisanih avtomatskih sporočil. Ali v primeru spletnih zlorab. Do zdaj smo že na podlagi načina pisanja lahko nekako razbrali, da gre za prevaro, ob pisanju chatgpt, ki je res brezhbirno, pa se to ne bo dalo več. Sprejeti moramo predpis, da bi moralo biti vsako besedilo, ki ga generira računalnik, označeno kot *computer generated*.

Je o tem že debata?

Bratko: Stuart J. Russell je to predlagal na konferenci že pred nekaj leti. Še nekajkrat sem silal ta predlog. Je za zanimivo, da je zdaj ta debata kar poniknila. Mislim, da je predlog dober in bi pomagal, da se bodo ljudje lažje znašli v množici prejetih besedil.
Gams: Zanimivo je, kako novo stvari vedno poskušamo zamejiti in jih kritizirati, pa naj bodo to avtomobili ali parni stroji ali zdaj umetna inteligenca.

Bratko: Ne predlagam ukinitve, le odpraviti moramo potencialne negativne učinke.

Je chatgpt pot do superinteligence

Veliki jezikovni modeli Računalniško generirana besedila bi potrebovala posebno označbo

O jezikovnih modelih in umetni inteligenci smo se pogovarjali z dvema slovenskima profesorjema, ki sta vsak na svoj način pionirja v razvoju raziskovanja inteligentnih sistemov.

PIJA KAPITANOVIČ

Ivan Bratko je bil dolga leta vodja laboratorija za umetno inteligenco na fakulteti za računalništvo in informatiko, do leta 2002 je vodil tudi odsek za inteligentne sisteme na Inštitutu Jožef Stefan (IJS). Je avtor več kot 200 znanstvenih objav in številnih knjig, med njimi *Prolog Programming for Artificial Intelligence*, ki je še danes eden najpogostejših citiranih učbenikov o programiranju umetne inteligence.

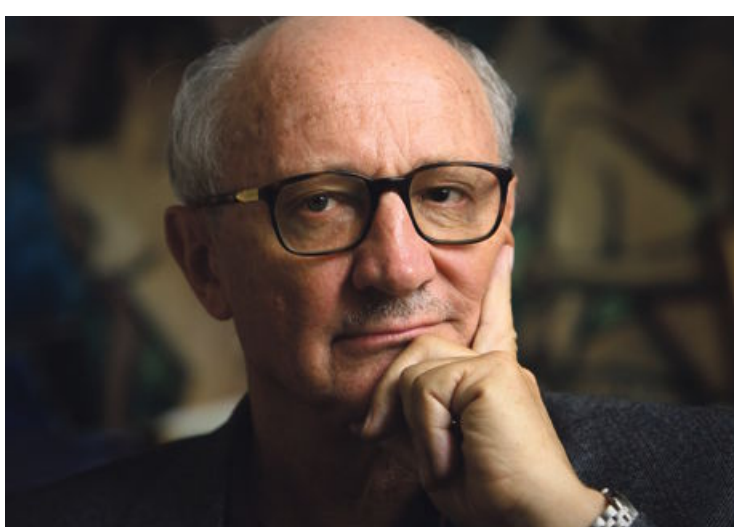
Matjaž Gams pa je doktoriral pri profesorju Bratku in ga kasneje nasledil na položaju vodje odseka za inteligentne sisteme na IJS ter ima 160 znanstvenih člankov in okoli tisoč objav. Tam s pomočjo umetne inteligence z različnimi ekipami raziskovalcev iščejo rešitve za težave človeštva; na ravni medicine, ekologije, proizvodnje, ekonomije.

Je chatgpt glede na to, da temelji na velikih jezikovnih modelih (Large Language Models), ki je več desetletij stara tehnologija, sploh revolucionaren?

Bratko: Še pred petimi leti si ne bi mislil, da bo razvoj na tem področju tako napredoval. Me pa sama tehnologija ne navdušuje preveč, ker je pretirano preprosta, da bi lahko govorili o njej kot o res inteligentni. Je pa v tej, omejeni inteligentnosti chatgpt fantastičen. Je daleč najboljši program, ki je zmožen generirati slovensko besedilo. Kljub temu da imamo veliko domačih projektov posvečenih računalniškemu obvladovanju slovenskega jezika, je ta najbolj zmogljiv.

Gams: Fenomenalen dosežek in izjemno uporabno orodje. Svojim študentom in sodelavcem smo dali navodilo, naj ga pri vsakem pisanju ali programiranju uporabljajo, ker bodo z njim bolj učinkoviti. Ko ga enkrat uporabijo, vidijo, da je res tako. Tistim, ki ga še niso uporabili, polagam na srce, naj ga čim prej. Naredili bodo preskok v produktivnosti, v bodočnost in prihranili veliko časa.

Kaj pa so njegove slabosti?
Bratko: Velika pomanjkljivost je, da ne vsebuje eksplicitne predstavitve pomena besedila. Zato nima tako rekoč nobene možnosti za pravilno in učinkovito logično sklepanje. Pri chatgptju gre za sijajno fasado, za katero pa ni prave vsebine. Za uporabnike je ta



Ivan Bratko je bil med drugim član uredniških odborov številnih znanstvenih revij, kot so Artificial Intelligence, Machine Learning in Springer Nature Computer Science. FOTO JURE ERŽEN

fasada zavajajoča. Na primer, ne realno je pričakovati, da bi lahko sam razvijal ali programiral večje uporabne sisteme.

Gams: Nekateri sorodni programi so boljši kot chatgpt oziroma nudijo večjo funkcionalnost. Recimo gpt-4 že nudi slikovni vhod, drugi programi, kot je dall-e, pa omogočajo delo s slikami, videi, glasbo itd.

Za programiranje enostavnih kod se verjetno lahko zanesemo tudi na jezikovne modele.

Gams: Ti programi so izredno koristni pri programiranju – znajo poiskati napake, šibkosti v obstoječi kodi, znajo izboljšati obstoječo kodo, znajo generirati novo kodo itd. Ta strahoviti

Pri chatgpt gre za sijajno fasado, za katero pa ni prave vsebine. Za uporabnike je ta fasada zavajajoča.

IVAN BRATKO

prehod je bil v javnosti neopazen: prej so lahko programirali samo programerji v formalnih programskih jezikih, zdaj lahko vsi »programiramo« v naravnem jeziku. Vrsto nalog lahko naredimo samo s podrobnim opisom svojih želja, kot bi jih naročili človeškemu asistentu.

Lahko od njega pričakujemo sposobnost kreiranja lastnega znanja, neko superinteligenco?
Bratko: Jaz si tega ne morem predstavljati. Splošna ocena o velikih jezikovnih modelih med strokovnjaki je, da ti sistemni ne

vedo, o čem govorijo. To pomeni, da sicer znajo producirati besedilo, ki je videti odlično, brez napak, dobro strukturirano, sam pomen besedila pa je zelo problematičen. Pri pisanju leposlovja ni nobenega problema, kajti bralec sam doda pomen. A ko pride mo do fizikalnega razmišljanja, chatgpt odpravi. S študenti smo preverili kar nekaj primerov in ugotovili, da si z njim ne moremo pomagati. Dela velike napake, večje od osnovnošolcev. Po njegovem je lahko gladina vode v posodi nad vrhom posode. In če so jezikovni sistemi slabi že pri fizikalnih temah, bodo slabi tudi pri programiranju. To, kar napiše, je sicer videti smiselno, ampak ni nobenega zagotovila, da bo tudi res pravilno.

Na konferenci *Global Partnership for AI* pred dvema letoma sta tako Yoshua Bengio (eden najbolj citiranih znanstvenikov v umetni inteligenci), ki se ukvarja z nevronskimi mrežami, torej mu je tehnologija velikih jezikovnih modelov blizu, kot Raja Chaitina na vprašanje, kaj si mislita o takrat znamenitem velikem jezikovnem modelu gpt-3, v en glas odgovorila, da gpt-3 »ne ve, o čem govori«. Prepričan sem, da ni prava pot, ki bi vodila do superinteligence.

Gams: Gpt-4 že zna dvakrat bolje razmišljati. Počakajmo še nekaj let in potem bomo lahko kritizirali. Že chatgpt zna lepše in bolje pisati kategoriki besedilo od velike večine ljudi. Idejno uporaben je takrat, ko moramo nekaj drugače ali bolje ubesediti. Je pa treba pregledati taka besedila.

Bratko: Ampak pri oblikovanju lahko spremeni pomen besedila. In če sam ne pozna pomena, ne bo opazil napake.

Gams: To je otrok, ki se še uči, a ogromno obeta.

Bratko: Ampak tak otrok nima veliko možnosti, da se bo naučil česa velikega.

Gams: Strinjam se z Elonom Muskom, ki pravi, naj primerjamo hitrost razvoja umetne inteligence in hitrost razvoja ljudi. Čez nekaj let boste videli, o čem govori.

Kaj pa bo potreboval, da bo napredoval, še več podatkov ali bolj zamejene?

Gams: Za zdaj je hitrost razvoja fenomenalna. Ne zlažemo se preveč, če rečemo, da ti programi v svoje nevrone mreže spravijo celoten splet, vse programe, vse slike, vsa besedila, nato pa generirajo odgovore iz svojega modela sveta. Imajo pa še vrsto slabosti.

Bratko: Lahko damo noter še več podatkov, ampak mislim, da je že dosegel točko padajočega donosa. Več podatkov mu zagotovo ne bi veliko pomagalo. Umanjka mu

bistvena komponenta. Obravnavanje semantike. Besedilo ima pomen in ljudje imamo metode, s katerimi pomen izrazimo. Imamo celo algoritme, ki uporabljajo formalizem za avtomatično sklepanje. Pristop, ki ga uporabljajo veliki jezikovni modeli, pa ignorira te že znane dosežke umetne inteligence za obravnavo pomena. Strokovnjaki priznavajo, da nevroenska mreža nima zmoglosti zdravorazumskega sklepanja. Se je pa že pojavila nova paradigma, ki naj bi to pomanjkanje zmoglosti sklepanja premostila. Ta paradigma se imenuje nevro-simbolična umetna inteligenca. Gre za kombinacijo nevronskih mrež in simbolične predstavitve znanja. Ideja je smiselna in prisotna že nekaj časa, za ta pristop se zavzemajo pri IBM, a za zdaj nimajo prodornih rezultatov.

Kaj je njegova glavna dodana vrednost v smislu razvoja umetne inteligence?

Gams: Vse skupaj spominja na razburjanje okoli nesreč s teslami. Tesla je po podatkih petkrat do desetkrat bolj varna od klasičnih vozil. Situacij, ko je Teslino vozilo rešilo potnike pred nesrečami, je bilo petkrat več kot tistih, ko se je zaletelo, v medijih pa so odmevale le nesreče. Podobno javnost ne razume, da gre za čudovit dosežek, ki pomaga razvijati človeško civilizacijo. Zaradi umetne inteligence smo vsako leto bistveno sposobnejši.

Res se zdi, kot da mediji kar tekmujejo v tem, kdo bo pokazal več neumnosti, ki jih počnejo strojni sogovorniki ... So res tako neumni in neresnicoljubni?

Gams: Ti programi se učijo iz dneva v dan kot otroci in večina lapsusnih napak od včeraj je danes že zgodovina. V praktičnem



Matjaž Gams je poučeval na 12 fakultetah, tudi v Nemčiji. Med drugim je v Paláci OZN predaval o nevarnostih avtonomnih orožij. FOTO JURE ERŽEN

smislu je bistveno, da je mogoče s temi programi rešiti naloge nekajkrat hitreje, torej napišete neko besedilo nekajkrat hitreje in bolj kvalitetno. Trend razvoja se bo nadaljeval in pripeljal do superinteligence.

Bratko: Da bi pripeljal do superinteligence, je neumnost. Strinjam se, da je uporaben. Za pisanje šolskih nalog, spisov, vabil, pritožb, prevajanje. Redko popolnoma zgreši pomen. Zelo dober je tudi za delanje povzetkov. Potem pa se po mojem že konča. Ampak da bi bila to prava pot do superinteligence, ki bi preselila človeško, pa menim, da ni res.

Zaradi umetne inteligence smo vsako leto bistveno sposobnejši. Vsi bi jo morali začeti čim prej uporabljati.

MATJAŽ GAMS

Gams: S tem se delno strinjam. Dotični chatgpt ni superinteligence, je pa izjemno koristen za obdelovanje besedil, predelavo iz enega sloga v drugega. Ko pišem znanstveno besedilo, recimo, lahko pišem bolj po domače, za mi program to prepíše v znanstveni jezik v enakem obsegu ali pa podaljša ali skrajša. V dveh sekundah dobim popravljen verzijo, sam pa bi to delal vsaj nekaj ur. Zakaj bi delal nekaj, kar zna narediti stroj? V smislu razvoja superinteligence pa je pomemben dejnik, ki bo ostal zapisan v človeško zgodovino podobno kot prva zmaga računalnika nad človekom v šahu.

No, ampak vseeno se učimo računati na roke, čeprav zna to narediti kalkulator, kajti gre za osvajanje temeljnih znanj, razvijanje možganov in logike. Ni strah o poneumljanju vseeno v določenem delu upravičen?

Gams: Znano je, da inteligenca ljudi korelira z uporabo orodij. Zadnji velik skok v inteligenci je bil narejen s prihodom mobilnih telefonov. Povprečna inteligenca ljudi je poskočila, ljudje se niso poneumili, kljub temu da so se nehali učiti telefonske številke na pamet. Zaradi avta ali letal nismo pozabili hoditi, smo se pa sposobni bistveno hitreje premakniti s točke na točko. Zdaj že nekaj časa umsko stagniramo, ker ni novega orožja. Upam, da bo človeško inteligenco dvignil tudi chatgpt.

So jezikovni modeli pristranski in rasistični?

Gams: Na vprašanje, kaj je huje: tretja svetovna vojna ali to, da moškega, ki se je spremenil v žensko, naslavljajš z moškim, je chatgpt odgovoril, da slednje. Microsoftov besedni model, ki se je učil pogovarjati na podlagi pism bralcev in spletnih komentarjev, pa se je hitro naučil sovražnega govora. Znano je namreč, da so komentatorji na spletu bolj desno usmerjeni.

Bratko: Ameriški sistem compas, ki se uporablja v pravosodju za svetovanje sodnikom, je bil v odmevnem članku v *ProPublici* oboženo pristranskosti, ker naj bi sodnikom priporočal diskriminatorne odločitve glede na etnično poreklo obtožencev. Navadna statistika je tako videti pristranska. Kar je neumnost. Algoritmi na matematičen način odkrivajo zakonitosti v podatkih. Kvečjemu bi bila lahko to indikacija, da so bile odločitve sodnikov v preteklosti pristranske. Na