POSEBNA IZDAJA

Monitor SVET

9 789619 116821
ISBN 961-91168-2-8

NAJBOLJŠE IZ SVETOVNEGA TISKA

APRIL 2024 ⁘ CENA: 5,50 EUR

KAKO JE NASTAL ChatGPT

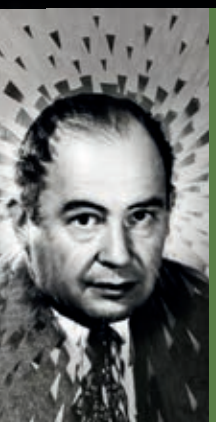
Prvih nekaj let v OpenAI je bilo napornih: nič ni delovalo, Google pa je imel vse - vse nadarjene raziskovalce, ogromno ljudi in veliko denarja.

DOSJE

- Državno **prisluškovanje**
- **Umetna inteligenca**
- Umetna inteligenca **v gospodarstvu**

NOVE TEHNOLOGIJE

- **Roboti**
- **Holografski** usmerjevalnik
- Nove **baterije**
- **3D tiskanje** raketnega goriva



VELIKANI

- **IBM**
- **JOHN VON NEUMANN**

IZ VSEBINE

Univerza v Ljubljani meni, da je treba reči bobu bob – prijava evropskih projektov je običajno nakladanje, ki ga najbolje obvlada kar ChatGPT.

MATJAŽ KLANČAR
odgovorni urednik, matjaz.klancar@monitor.si

Umetna inteligenca **ali strojno učenje?**

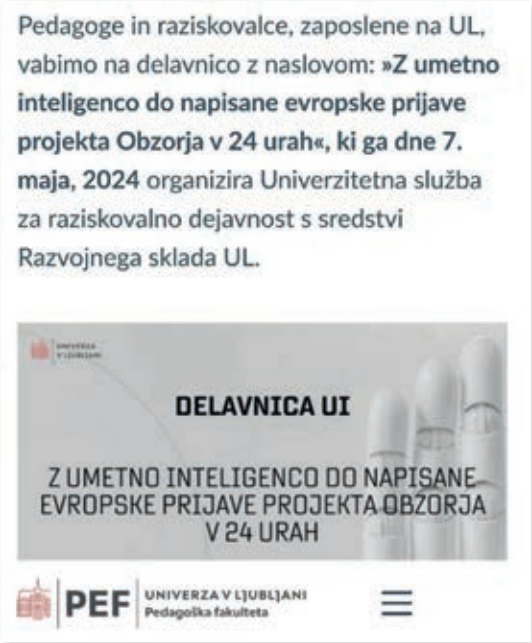
ChatGPT je pred letom in pol šokiral nas, ki že desetletja vidimo, kako neinteligentni so (bili) v resnici računalniki, in razveselil vse tiste, ki smo v njem videli gigantski napredek v primerjavi z običajnim guglanjem. Toda ali je to res »umetna inteligenca« ali le »strojno učenje« na steroidih?

Spomnim se svojih študentskih let, ko sem pri predmetu Umetna inteligenca izdelal (sprogramiral) svojo prvo nevronsko mrežo in jo »nahranil« s podatki. Po tisočih in tisočih vzorcih/podatkih, ki jih je skupek moje programske kode »prežvečil«, je znal bolj ali manj, približno, pravilne rezultate izpisovati tudi za podatke, ki niso bili del učnega korpusa. Čudež!, se mi je takrat zdelo.

In vendar je umetna inteligenca že takrat bila v eni izmed t. i. »zim« (*AI winter*), ko so raziskovalci zašli v slepo ulico in napredka enostavno ni bilo več. Šele leta 2017 so Googlovi strokovnjaki razvili in predlagali arhitekturo globokega učenja Transformer, ki nas je počasi prevesila v – pomlad, kot jo uživamo danes. To arhitekturo zdaj uporabljajo vsi govorni modeli, uporablja se za računalniški vid in sisteme za generiranje videa ter glasbe. Širša javnost pa je zmogljivosti sistema zares ponotranjila šele s predstavitvijo klepetalnega robota ChatGPT, ki je veliko več kot le to.

ChatGPT (in njegov derivat Microsoft Copilot) danes uporablja staro in mlado. Kot nekakšen koncentriran spletni iskalnik, ki vrne le en odgovor in ne stotine linkov (pa čeprav je odgovor včasih napačen), kot zabavnega sogovornika, lektorja, tajnico, ki zna povzeti dooolga besedila ali drugače zapisati izvirnik besedila, nenazadnje tudi kot prevajalnik, ki je velikokrat učinkovitejši kot Googlov izdelek, čeprav je ta z nami že dolga leta. In seveda, ChatGPT uporabljamo kot pomočnika. Tam, kjer je boljši od nas, in predvsem tam, kjer – se nam ne da. Goljufanje, v taki ali drugačni obliki, je verjetno najbolj razširjena uporaba tehnologije, ki so si jo raziskovalci bržkone zamislili z bolj plemenitimi cilji.

Učenci in študenti s »četom« izdelujejo domače in seminarske naloge in imajo namesto plonklistkov na telefonu odprto aplikacijo Copilot. Aplikacijo, ki v hipu naredi povzetek predolgega znanstvenega članka, hitro napiše obnovo predolge in predolgočasne leposlovne knjige. In uporabljajo ga seveda profesorji, ki nato s to isto aplikacijo pregledujejo te naloge. »Čet«, ki pregleduje izdelke tega istega »četa«. Predvsem pa učenci, učitelji, profesorji in vsakdo, ki mora pri svojem delu ustvariti veliko besedila – »čet« je pač odličen pri nakladanju, v vseh oblikah.



Tako zelo odličen, da je celo najbolj čislana izobraževalna institucija pri nas, Univerza v Ljubljani, nedavno pozvala k tečaju uporabe »četa« pri – goljufanju naše matere, Evropske unije. Pedagoška fakulteta oz. Univerzitetna služba za raziskovalno dejavnost je namreč povabila na delavnico, ki bo raziskovalce »z umetno inteligenco« naučila napisati evropsko prijavo za projekt. V 24 urah, da ne bi kdo pomislil, da je to težko.

Manj odlična pa je umetna inteligenca pri razpoznavanju človeškega obnašanja, so ugotovili pri Amazonu. Leta 2016 so v kar nekaj svojih fizičnih trgovinah uvedli sistem Just Walk Out, ki je obljubljal nakupovanje brez prodajalca. Kupec se je le prijavil ob vstopu v trgovino, v torbo spravil kar je želel, in odšel. »Umetna inteligenca« pa je samodejno razbrala, kaj je kupil, in mu naknadno poslala račun. Le da se je zdaj izkazalo, da so bili »umetna inteligenca« le slabo plačani Indijci, ki so pregledovali videoposnetke vsega, kar je kupec v trgovini počel in pospravil v torbo. Amazon je Just Walk Out zdaj ukinil. Ampak slišati je bilo pa lepo. ◀

EVIDENCA

SVET 2024

FOKUS

20 Kako je nastal **ChatGPT**

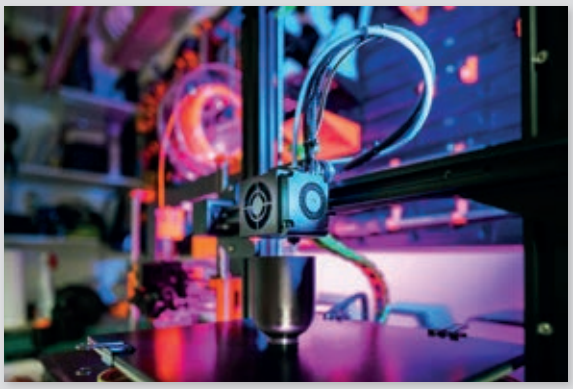
Se Sam Altman zaveda, kaj ustvarja?

»Sam Altman je na ponedeljkovo jutro aprila lani sedel na sedežu OpenAI v San Franciscu in mi pripovedoval o nevarni umetni inteligenci, ki jo je razvilo njegovo podjetje, a še ne predstavilo na trgu…«

NOVE TEHNOLOGIJE

68 Avtomobilske baterije **za mraz**

V neodvisnem laboratoriju so potrdili, da ima Ampriusova baterija energijsko gostoto 500 WH/kg; Tesline baterije dosegajo okoli 269 WH/kg.



NOVE TEHNOLOGIJE

70 **Natisnjeno** raketno gorivo

Znanstveniki so začeli tiskati reaktivno črnilo na elektroniko, s čimer so omogočili samouničenje, če bi material prišel v napačne roke.

VELIKANI

24 John von Neumann

Njegov učitelj matematike je razredu razlagal o izjemno zahtevnem teoremu, ki ga nikomur ni uspelo dokazati, pa je fant dvignil roko, stopil k tabli in napisal celoten dokaz: »Leta, vsa leta mojega dela, so se v sekundi razblinila … Po tistem sem se von Neumanna bal.«

DOSJE

55 Ko prisluškuje država

Izraelski Pegasus je najbolj znano programsko orodje, ki ga za prisluškovanje uporabljajo državne inštitucije širom sveta. Med strankami so tako avtoritarne kot načeloma demokratične države.

Ali vohuni **sabotirajo šifriranje?**

Znani strokovnjak za šifriranje je za New Scientist povedal, da bi ena od ameriških vohunskih agencij utegnila oslabiti novo generacijo algoritmov za varovanje pred hekerji, opremljenimi s kvantnimi računalniki.

Matthew Sparkes, *New Scientist*

Daniel Bernstein z illinojske univerze v Chicagu pravi, da ameriški nacionalni inštitut za standarde in tehnologijo namerno prikriva, kako intenzivno je ameriška nacionalna varnostna agencija vpletena v razvoj novih šifrirnih standardov za postkvantno kriptografijo. Prepričan je tudi, da inštitut za standarde in tehnologijo dela naključne ali namerne napake pri izračunih, ki ponazarjajo varnost novih standardov. Inštitut te trditve zanika.

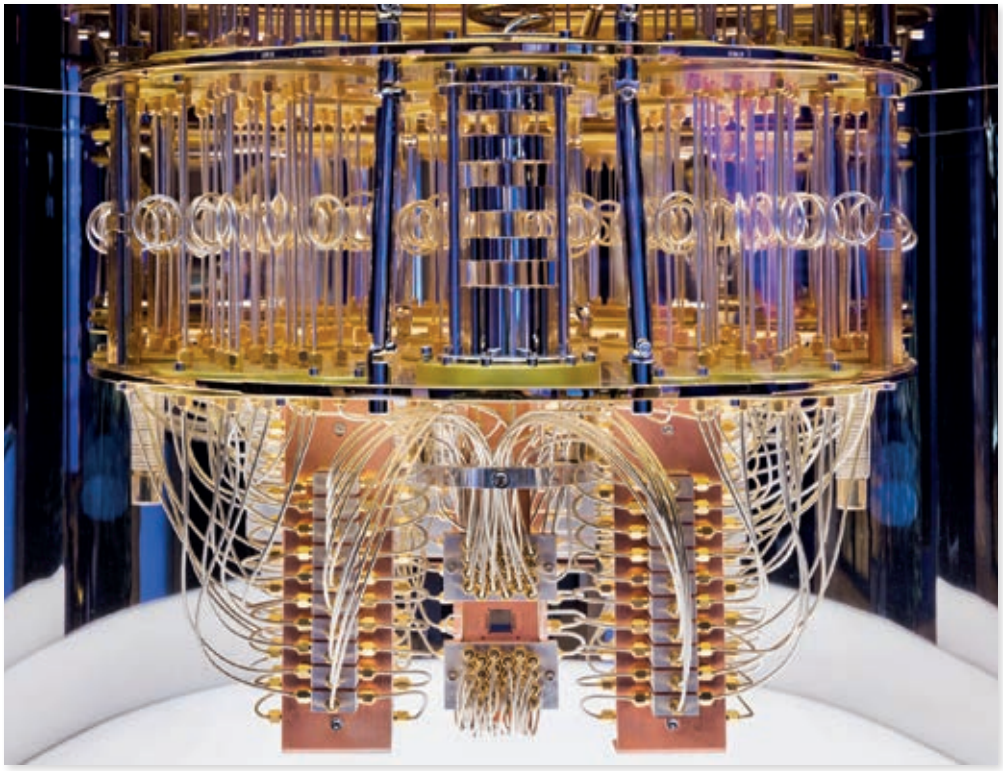
»Inštitut ne spoštuje postopkov, katerih namen je nacionalni varnostni agenciji preprečiti slabitev postkvantnega šifriranja,« je povedal Bernstein. »Pristojni, ki določajo šifrirne standarde, bi morali pregledno in preverljivo spoštovati nedvoumna javna pravila, tako da si nam ne bi bilo treba beliti glave o njihovih motivih. Inštitut je obljubil pregledno delovanje in nato trdil, da je razkril vse, kar počne, vendar to preprosto ne drži.«

Danes je tako rekoč nemogoče, da bi celo največji superračunalniki rešili matematične probleme, ki se uporabljajo za zaščito podatkov. Ko pa bodo kvantni računalniki postali zanesljivi in dovolj močni, jih bodo strli v nekaj trenutkih.

Ni še jasno, kdaj bomo dobili takšne računalnike, vendar je inštitut za standarde in tehnologijo že leta 2012 začel projekt za standardizacijo nove generacije algoritmov, ki se upirajo njihovemu napadu. Bernstein, ki je zaradi takšnih algoritmov leta 2003 skoval izraz postkvantna kriptografija, pravi, da nacionalna varnostna agencija aktivno sodeluje pri vključevanju skritih šibkih točk v nove standarde šifriranja, tako da ga bo nekdo z ustreznim znanjem lažje strl. Standardi ameriškega inštituta so v uporabi po vsem svetu, zato bi pomanjkljivosti lahko imele velik vpliv.

Bernstein trdi, da so preračuni inštituta za enega od prihodnjih postkvantnih standardov, Kyber512, očitno zgrešeni. Tako se zdi varnejši, kot je v resnici. Pojasnil je, da je inštitut pomnožil številki, vendar bi bilo pravilneje, če bi ju seštel, zato je dobil umetno zvišano oceno o odpornosti Kyber512 proti napadom.

Ko bodo kvantni računalniki postali zanesljivi in dovolj močni, bodo šifre strli v nekaj trenutkih.



Dustin Moody z inštituta je izjavil, da se z Bernsteinovo analizo ne strinjajo: »To je vprašanje, na katero ni mogoče odgovoriti z znanstveno gotovostjo, in inteligentni ljudje imajo lahko različna stališča. Spoštujemo Danovo mnenje, vendar se ne strinjamo z njim.«

Moody pravi, da Kyber512 izpolnjuje najvišja varnostna merila inštituta, zato ga je vsaj tako težko 'povoziti' kot razširjen obstoječi algoritem AES-128. A inštitut vseeno priporoča, da bi se v praksi uporabljala močnejša različica, Kyber768, za katerega Moody pravi, da so ga predlagali razvijalci algoritma.

Skrivnosti ali preglednost

Nacionalni inštitut za standarde in tehnologijo ima zdaj javna posvetovanja, se pravi, da mu javnost lahko posreduje svoja stališča in povratne informacije o predlogu, dokončne standarde za postkvantne algoritme pa naj bi predstavil prihodnje leto in takrat jih bodo podjetja lahko začela uvajati. Algoritem Kyber se je že prebil skozi več krogov selekcije in bo verjetno prišel tudi v ožji izbor.

Zaradi takšne skrivnostnosti je težko zagotovo reči, ali je nacionalna varnostna agencija vplivala na standarde postkvantne kriptografije, že dolgo pa je slišati namige in govorice, da agencija načrtno slabi šifrirne algoritme. Leta 2013 je *New York Times* poročal, da ima agencija za to nalogo na voljo 250-milijonski proračun, dokumenti obveščevalne agencije, ki so istega leta v javnost prišli zaradi Edwarda Snowdna, pa omenjajo, da nacionalna varnostna agencija v šifrirni algoritem namenoma vključuje tudi stranska vrata – no, ta konkretni algoritem so pozneje izločili iz uradnih standardov.

Moody zanika, da bi se inštitut kdaj strinjal s slabljenjem

Dokumenti ki so leta 2013 v javnost prišli zaradi Edwarda Snowdna, omenjajo, da nacionalna varnostna agencija v šifrirni algoritem namenoma vključuje tudi stranska vrata.

standardov zaradi navodil varnostne agencije, in dodaja, da skritih šibkih točk ne bi mogli vstaviti v algoritem brez njegove vednosti. Trdi tudi, da je po Snowdnovih razkritjih inštitut zaostril smernice za zagotavljanje preglednosti ter varnosti in za vnovično pridobitev zaupanja strokovnjakov za kriptografijo.

»Nikoli ne bi namenoma storili česa takšnega,« je zatrdil Moody, a hkrati priznava, da so bile Snowdnove objave hud udarec. »Vedno ko kdo omeni nacionalno varnostno agencijo, kup kriptografov izrazi svojo zaskrbljenost in trudimo se biti odprti ter transparentni pri naši komunikaciji.«

Moody pravi, da si agencija prizadeva biti bolj odprta, kolikor je to za obveščevalno službo seveda sploh mogoče. Agencija sama ni želela odgovarjati na novinarska vprašanja o tem.

»Največ, kar mi lahko storimo, je, da opozarjamo, kdo odloča – to je inštitut za standarde in tehnologijo. A če nam ne verjamete, tega niti ne morete preveriti, razen če ste zaposleni na inštitutu,« je povedal Moody.

Bernstein trdi tudi, da inštitut ni iskreno razkril, kako velika je vloga nacionalne varnostne agencije, in da so se zavili v molk, ko se je poskušal prebiti do več podatkov. Prav zato je tudi oddal nekaj vlog na podlagi svobodnega dostopa do informacij in inštitut celo tožil, da bi ga sodišče prisililo k razkritju podrobnosti o sodelovanju agencije.

Dokumenti, ki jih je na ta način pridobil Bernstein, kažejo, da je skupina z imenom Ekipa za postkvantno šifriranje pri nacionalnem inštitutu za standarde in tehnologijo vključevala številne člane nacionalne varnostne agencije in da so se predstavniki inštituta sestali z nekom iz britanske vladne uprave za komunikacije GCHQ, kar je britanska ustreznica ameriške varnostne agencije.

Vedno ko kdo omeni nacionalno varnostno agencijo (NSA), številni kriptografi izrazijo zaskrbljenost.

Alan Woodward z britanske univerze v Surreyju opozarja, da imamo dobre razloge za nezaupanje do šifrirnih algoritmov. Kod GEA-1, ki so ga v 90. letih in na začetku stoletja uporabljali v mobilnih omrežjih, je imel napako, zaradi katere ga je bilo mogoče računalniško milijonkrat lažje streti, kot bi bilo to dopustno. Krivca, ki je podtaknil napako, niso nikoli odkrili.

Woodward tudi dodaja, da sedanje kandidate za postkvantno kriptografijo temeljito preverjajo tako akademiki kot industrija, a še niso odkrili pomanjkljivosti, medtem ko so pri algoritmih na prejšnjih stopnjah preverjanja našli napake in jih izločili.

»Obveščevalne agencije so v preteklosti že slabile šifriranje, a za nove kandidate so opravili toliko varnostnih analiz, da bi me presenetilo, če bi Kyber tokrat naletel na mino,« je še dodal. ◀

Predstavniki Inštituta za standarde in tehnologijo so se sestali z nekom iz britanske vladne uprave za komunikacije GCHQ, kar je britanska ustreznica ameriške varnostne agencije NSA.

Applovo stremljenje k popolnosti **se ne obnese vedno**

Konec letošnjega februarja so Applovi uradni predstavniki prekinili desetletja dolga prizadevanja, da bi sestavili avtomobil. Projekt, katerega delovno ime je menda bilo Titan, je imel več vodij in doživel nekaj strateških sprememb, sta poročala New York Times in Bloomberg. Zdaj Apple prešteva izgube, zaposlene pa je prestavil k delu z generativno umetno inteligenco.

Jared Newman, *Fast Company*

Applov avtomobil je od začetka veljal za težko dosegljiv cilj, vendar je njegov polom v resnici le malo bolj dostojanstvena manifestacija Applove trenutne nezmožnosti, da bi od strojne opreme ponudil še kaj drugega razen svojih najslavnejših izdelkov. Pretekla leta smo slišali najrazličnejše govorice o pametnih zaslonih, televizorjih, zložljivih telefonih in drugem, a se ni nič od tega uresničilo, ker naj bi Apple menda preveč omahoval pri podrobnostih.

Nekateri teh izdelkov so se mu tako rekoč sami ponujali in bi tudi pokrili jasno izražene potrebe kupcev, sploh v primerjavi z električnimi vozili in samovoznimi sistemi, kljub temu pa se zdi, da podjetje ne ve, kako jih spraviti na trg.

Pametni zaslon, ki to ni bil

Apple naj bi pametni zaslon razvijal že vsaj od leta 2021, ko je Mark Gurman iz Bloomberga opisal prvo stopnjo razvoja izvrstnega zvočnika z zaslonom na dotik skupaj s kamero za videopogovore.

To ni bil čisto nov koncept, saj sta Amazon in Google takrat že nekaj let dobavljala pametne zaslone in pokazalo se je, da so odlični okvirji za digitalne fotografije. Lenovov pametni zaslon mojo kuhinjo krasi od leta 2018, in ko brskam po albumih Google Photos, mi to ni le v neznansko veselje, temveč je tudi pomemben dejavnik, da družina ostane povezana z Googlovim sistemom storitev in doplačuje za dodatni prostor v oblaku.

Zakaj torej Apple ni zmožen dobaviti nečesa podobnega za uporabnike iCloud Photos? Kar nekaj poročil pravi, da je Apple preveril različne rešitve za pritrjevanje na steno in formate zaslona, a ga to kljub Standbyu – nekakšnemu načinu majhnega zaslona za novejše iphone, ki ga je predstavil lansko jesen – ni pripeljalo korak bliže k izdelku. Gurman pravi, da bo pametni zaslon v najboljšem primeru na voljo prihodnje leto.

Mečkanje namesto zlaganja

Applova želja po popolnosti je tudi preprečila, da bi predstavil zložljiv telefon, ki bi se kosal s Samsungovimi izdelki. Govorice o Applovem zložljivem telefonu so stare že osem let in Gurman je že na začetku leta 2021 omenil, da je podjetje začelo prva testiranja zložljivih zaslonov za iphone.

Tri leta pozneje so v *The Informationu* poročali, da je Apple še vedno na začetku razvoja, ker menda želi, da zložena naprava ne bi bila debelejša od navadnega iphona. Če jim tega ne bo uspelo doseči, morda sploh ne bodo začeli proizvodnje.

Za sistem Android pa so zložljivi telefoni že na voljo in so enkratni. Nekoliko večja dimenzija je znosen kompromis, če tablico lahko pospraviš v žep ali torbico, poleg tega je razlika v debelini čedalje manj pomembna: Honorjev Magic v2 zložen meri le 9,9 milimetra, to je le petino več od iPhona 15 Pro Max.

Apple prešteva izgube zaradi ukinjenega avtomobilskega projekta, zaposlene pa je prestavil k delu z generativno umetno inteligenco.

Ti telefoni so dragi – Samsungov Fold 5 ima vstopno ceno 1.800 dolarjev –, a ne pozabite, da je iPhone Pro Max po podatkih *Canalysa* najbolje prodajani pametni telefon na svetu, saj ljudje hočejo največji in najboljši Applov telefon, za kar so pripravljeni seči globlje v žep. Podjetje bi lahko zaračunavalo zajetno vsoto za še večji zložljiv telefon, ki bi ga ljudje oboževali, a se kljub temu noče sprijazniti z ničimer manj od popolnosti, četudi je nedosegljiva.

Govorice o Applovem zložljivem telefonu so stare že osem let.

Televizija v slepi ulici

Medtem je Apple že izgubil rdečo nit pri televizorjih, saj eksperimentira z najrazličnejšo čudno strojno opremo, namesto da bi naredil najbolj logični korak in predstavil pametni televizor.

Apple TV4K je odličen večpredstavnostni predvajalnik, vendar obstaja v svetu, kjer se večina ljudi zadovolji s programsko opremo, vgrajeno v televizor. Podatki *Convive* kažejo, da je čas gledanja pametnih televizorjev leta 2022 presegel zunanje večpredstavnostne predvajalnike in se še povečuje. Ljudje več ur prostega časa na dan preživijo pred televizorji, a Apple nadzor nad to izkušnjo prepušča tekmecem, kot so Google, Amazon, Roku in Samsung. Seveda na drugih platformah lahko posnamete Applovo aplikacijo za televizijo, vendar tam niso zainteresirani za oglaševanje Applove vsebine in povezave z drugimi Applovimi storitvami, kot sta Music in HomeKit. Ker nima nadzora, je omejen tudi pri širjenju sektorja s ciljnimi oglasi na področje televizije.

Kaj Apple počne namesto tega? Loteva se eksperimentov, ki ne vodijo nikamor, kot so zvočniki, ki lahko funkcionirajo tudi kot večpredstavnostni predvajalnik. Roku, Amazon in Google že leta prodajajo nekaj podobnega, a gre za težko razumljiv koncept in nič ne kaže, da bi se te naprave množično prodajale. Mogoče je čisto v redu, da Apple ni razvil tovrstnega izdelka, čeprav se govori, da je razvoju posvetil več let.

Veliko truda za prazen nič

Pametni zasloni, zložljive naprave in televizija niso edina področja, na katerih je Apple zašel v slepo ulico.

Podjetje omahuje tudi pri pametnih zvočnikih. Leta 2021 je prekinilo proizvodnjo prvotnega homepoda in se osredotočilo na cenejšega homepod mini, nato pa lani obudilo komajda izboljšano različico. Tri leta je čakalo na minimalno nadgradnjo svojih airpodov pro, ni pa začelo proizvodnje modelov brez peclja in za spremljanje vadbe, o katerih so se širile govorice. Lansko leto je bilo tudi prvo brez novega ipada, pri katerem je pomanjkanje zamisli zgodba zase. Govorice o čarobni miški USB-C in dodatkih za čarobno tipkovnico – ki bi jih nujno potrebovali, ker Apple opušča svoj ločljivi priključek *lightning* – so prav tako ostale samo govorice.

Seveda vem za govor Steva Jobsa o tem, da osredotočenost pomeni reči ne. Znano je, da je Apple konservativno podjetje in se ne vrže z vsemi štirimi v nove kategorije izdelkov ter ne dodaja novih funkcij samo zato, ker to počnejo tekmeci.

Vendar to ne pomeni, da Apple vedno čaka na popoln izdelek ali preizkušen trg, preden predstavi novost. Izvirna Applova ura je bila medla naprava s kupom možnosti, ki jih je podjetje pozneje nehalo izpostavljati (na primer skiciranje risbic za prijatelje) ali jih kar ukinilo (interakcija *force touch* in 18-karatna zlata izdaja), zato pa bolj poudarjalo funkcije, povezane z vadbo. Izvirni iphone je bil grob zametek v primerjavi s poznejšimi različicami z višjo ceno in brez možnosti aplikacij tretjih podjetij. Celo Vision Pro, ki ga je Applu uspelo predstaviti na trgu (menda po letih zamude in razvojnih izzivov), je poskusni izdelek za razvijalce in skrajne navdušence.

To ne pomeni, naj Apple začne prodajati napol 'surov' avtomobil, a Titanovo desetletje razvojnega pekla je znak nečesa globljega: Apple ne ve, kaj novega naj začne prodajati. ◀

▽ **Apple TV4K je odličen večpredstavnostni predvajalnik, vendar obstaja v svetu, kjer se večina ljudi zadovolji s programsko opremo, vgrajeno v televizor.**

Kako je Tiktok postal oglasni kralj za generacijo Z

Poleti 2021 je Tiktokova uporabnica Shannon Johnson objavila kratek posnetek, kako si na ustnice nanaša obarvan balzam kozmetične hiše Clinique v odtenku black honey. Sledilcem je razložila, da ga je uporabljala tudi Liv Tyler, ko je snemala Gospodarja prstanov…

John Kell, *Fast Company*

Black honey je odtenek, ki ga je podjetje Clinique uvedlo leta 1972, naslednji val priljubljenosti je doživel leta 1989, vendar je nato spet skorajda utonil v pozabo. Danes sodi med najpomembnejše izdelke v portfelju, se je pohvalila Jess Burns, namestnica vodje globalnega sodelovanja s porabniki v omenjenem podjetju.

Ko se je posnetek z balzamom kot virus razširil po Tiktoku, je Clinique iznajdljivo poslal izdelke ustvarjalcem vsebin, da bi še spodbudil zanimanje za ta odtenek, začel je sodelovati s trgovinami in svoji ekipi, ki skrbi za dobavno verigo, naročil, naj zagotovi dovolj izdelkov na policah, ko je prodaja podivjala. Nato je izdelku v odtenku *black honey* dodal še druge, da bi unovčil trenutni lepotni trend.

»Začelo se je s trenutkom priljubljenosti in nadaljevalo s tuhtanjem, kako bi ga podaljšali,« je pojasnila Burnsova. »Mislim, da je to preizkušena strategija vseh znamk, ki želijo ostati pomemben del razprav na Tiktoku.«

Od priporočil do neposrednega stika s strankami

»K nakupu me je nagovoril Tiktok,« je postala krilatica za razširjene vsebine, ki vplivajo na prodajo, trend, ki je še posebej izrazit pri kozmetiki in modi. Medtem ko Instagram še vedno spodbuja neposredno prodajo, strokovnjaki ocenjujejo, da Tiktok postaja najpomembnejša družbena platforma za ciljne skupine, podatke in vpoglede, trženje in celo razvoj novih izdelkov, po katerih povprašujejo kupci.

»Tiktok je kulturni prostor, kjer potekajo pogovori,« je povedala Ellyn Briggs, analitičarka blagovnih znamk v Morning Consultu. »Odvijajo se hitro in povečujejo prodajo.«

Morning Consult je pred kratkim objavil lestvico najhitreje rastočih blagovnih znamk v letu 2023, in sicer na podlagi števila kupcev, ki pravijo, da razmišljajo o nakupu izdelka neke znamke. Proizvajalci, katerih izdelki se najbolj širijo po Tikoku, so zelo priljubljeni pri kupcih iz generacije Z. Kozmetična hiša Clinique je pristala na šestem mestu, a to ni edina več desetletij stara blagovna znamka, ki postaja priljubljena pri pripadnikih generacije Z. Dobro so se odrezali tudi Dollar General, Pottery Barn, Victoria's Secret in Abercrombie & Fitch.

Uporabniki iščejo komentarje, ki jim lahko zaupajo, in čim bolj je oseba odmaknjena od blagovne znamke in čim manj zunanje pobude je za njenim komentarjem, tem bolj verodostojna se zdi.

Generacija Z reklamo zavoha že na kilometer.

»Naše raziskave o generaciji Z kažejo, da ima žilico za nostalgičnost,« je še dodala Briggsova.

Najbolj priljubljen pa je bil Kraft. Matična družba, Kraft Heinz, je osvežila stoletje stare blagovne znamke, kot so Heinz, Oscar Mayer in Jell-O, da se v obdobju družbenih omrežij ne bi zdele tako zaprašene. Kraft je spremenil tudi svoje mnenje o Tiktoku, ki ga je sprva štel za še en družbeni kanal, na katerem je mogoče razkazovati ustvarjalnost.

No, danes je Kraftova vsebina na Tiktoku namenoma manj dognana. Priljubljeni posnetek počitniškega recepta za kremni sir Philadelphia je bil odziv na uporabnikov komentar. Pripomogel je tudi k nasvetom za različne vilice, s katerimi je mogoče jesti Kraftove makarone s sirom. Uporabniki so ustvarili priljubljeni nasvet, po katerem se Kraftova navodila za pripravo kar preskoči in se vse sestavine hkrati vrže v lonec.

»Upamo, da smo videti kot znamka, ki živi na platformi,« je izjavila Jess Vultaggio, vodja blagovnih znamk in ustvarjalnosti v Kraft Heinzu. Več kot šest desetin Kraftovega udejstvovanja na Tiktoku sprožijo razprave, ki se spontano začnejo na platformi.

Tiktok je privlačen za prehranske velikane, kot je Kraft, ki poskušajo blagovne znamke približati demografski skupini na robu odraslosti, torej ko začenjajo samostojno oblikovati svoje nakupovalne vzorce za prehranske izdelke. Zaradi Kraftovih izdelkov je hrana dostopnejša, naj bo zaradi preproste priprave ali znanih okusov, je še povedala Vultaggiova.

Iskanje in odkrivanje

Generacija Z uporablja Tiktok tudi kot iskalnik, je pojasnila Vickie Segar, ustanoviteljica trženjske agencije za vplivneže Village Marketing. Novoletni prazniki so bili prva takšna priložnost, ko so blagovne znamke in ustvarjalci vsebin lahko izkoristili Tiktok Shop, novo trgovsko spletno ponudbo, ki so jo uvedli septembra lani.

»Preoblikujemo strategije, kako bi ustvarjalci vsebin lahko odgovarjali na vprašanja, ki jih lahko preberemo v pasici za iskanje, da bi se omenilo naše blagovne znamke,« je povedala Segarjeva. Blagovne znamk morajo na vplivno sfero na Tiktoku gledati drugače, saj ni dovolj, če samo plačajo ustvarjalcem, temveč morajo podpreti tudi tiste, ki sami od sebe objavijo, kako radi imajo neki izdelek.

»Uporabniki iščejo komentarje, ki jim lahko zaupajo, in čim bolj je oseba odmaknjena od blagovne znamke in čim manj zunanje pobude je za njenim komentarjem, tem bolj verodostojna se zdi,« je pojasnila Segarjeva.

Johnsonova, ki je sprožila modni trend z balzamom *black honey*, ima na Tiktoku manj kot 5.600 sledilcev. A kot ponazoritev demokratizacije vpliva na platformi sta ključni besedi *blackhoney*, označeni s ključnikom, zbrali več kot 552 milijonov ogledov na kanalu.

»Tiktok zdaj postaja prodajalec na zahtevo,« je razložila Amy Lanzi, direktorica agencije Digitas matične družbe Publicis. »Prodaja jim uspe tudi tam, kjer je večina podjetij niti ne pričakuje.«

Lanzijeva ocenjuje, da vse več blagovnih znamk preverja in izbira, kaj so pripravljene prodajati prek trgovine na Tiktoku in kolikšen delež pri svojih ustvarjalnih vsebinah naj prepustijo vplivnežem na kanalu. Blagovne znamke morajo najprej dodelati pristno, organsko družbeno strategijo, preden začnejo plačevati za oglase. »Na Tiktoku ne moreš prodreti z denarjem,« je poudarila Lanzijeva.

Digitas je uvedel novo storitev za delitev pozornosti vrednih in trendovskih izdelkov, imenovano s kratico SWAT (*Share Worthy and Trending*), s katero strankam, kot sta Sephora in Digiorno Pizza, pomaga ustvarjati podrobnejše vsebine na Tiktoku.

»Če je Tiktok epicenter kulture za generacijo Z, je to tako pomemben podatek, da se mu je treba strateško posvetiti,« izpostavlja Christina Goswiller, vodja trženja na družbenih omrežjih v Digitasu. Ta podatek med drugim s pridom izkorišča, da v model umetne inteligence vnaša izraze s ključnikom ter tako ustvarja virtualno 'dekle, ki ima vse' in jo je mogoče povprašati o najbolj priljubljenih trendih na Tiktoku, ki bi se lahko dobro prijeli tudi pri kupcih.

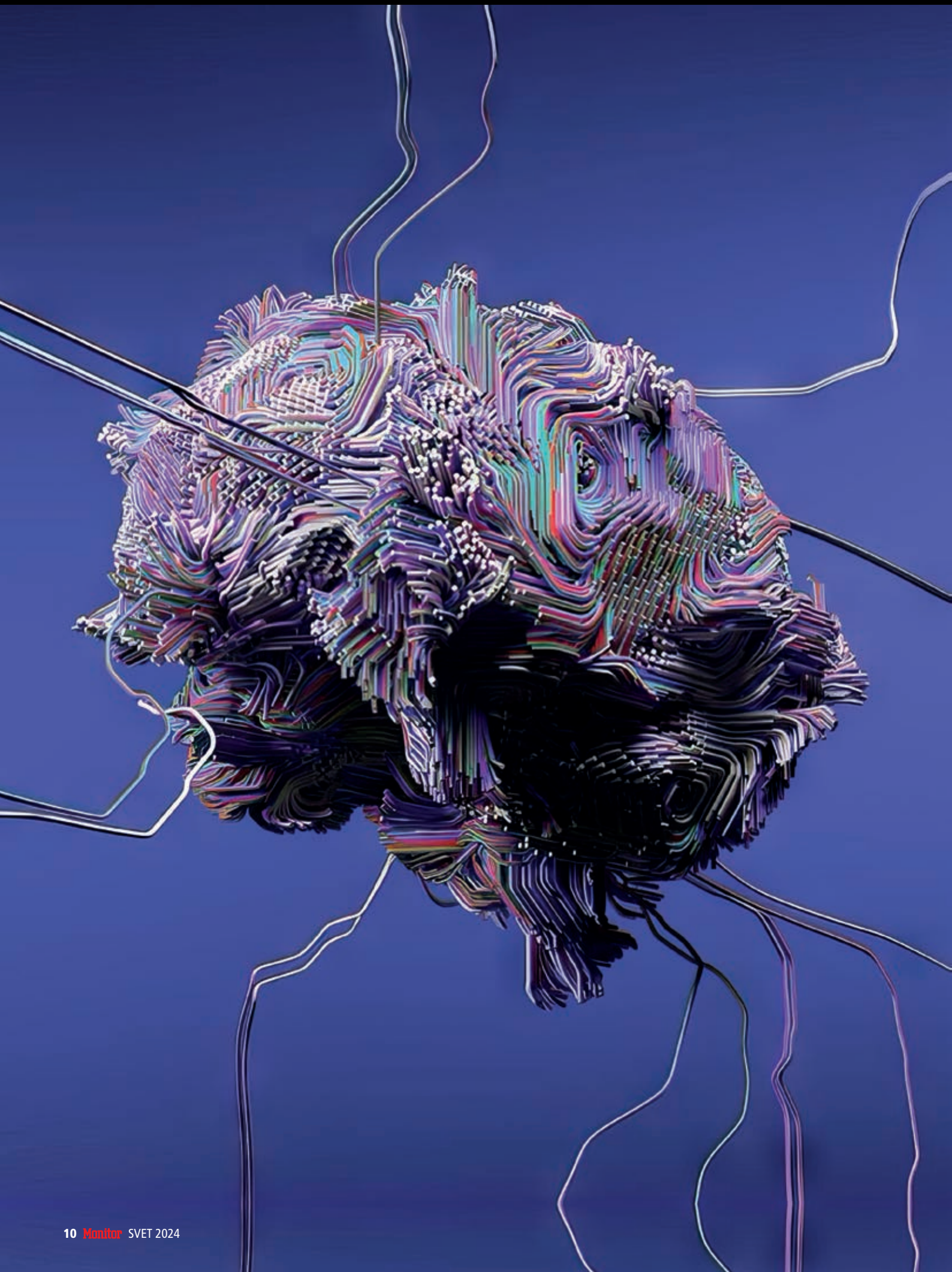
»Pomislite, koliko ur raziskav si prihranimo in koliko bolje ter globlje lahko razumemo segment občinstva ali podkulture na podlagi dvogovorov z umetno inteligenco,« je navdušena Goswillerjeva.

Vse v redu in prav, a kdaj postane vse skupaj preveč?

Strokovnjaki sicer pravijo, da bodo blagovne znamke še vedno vlagale na Tiktoku, vendar bi morale biti previdne, da uporabniška izkušnja, ki ima poudarek na zabavi in odkrivanju, ne bi zvodenela. »Slišim že pritožbe, da se Tiktok počasi spreminja v nakupovalni televizijski kanal,« je opozorila Briggsova.

Clinique ima za seboj dolgo zgodovino sodelovanja z dermatologi, ki kupce izobražujejo o negi kože, in se kar strinja s to oceno. Tiktok vidi kot sodobno polico z ličili, prostor za odgovore na vprašanja o negi kože in kroženje informacij o sestavinah v izdelkih.

»Generacija Z reklamo zavoha že na kilometer,« je poudarila Burnsova. »Prva generacija so, ki od blagovnih znamk in ustvarjalcev zahteva, naj se pristno pogovarjajo z njo. In kaj to pomeni? Oglasa ne ustvariš. Ustvariš vsebino, s katero lahko nekaj počnejo.« ◀



Se Sam Altman zaveda, **kaj ustvarja?**

Sam Altman je na ponedeljkovo jutro aprila lani sedel na sedežu OpenAI v San Franciscu in mi pripovedoval o nevarni umetni inteligenci (UI), ki jo je razvilo njegovo podjetje, a nikoli predstavilo na trgu. Pozneje je dodal, da njegovi podrejeni včasih ne spijo od skrbi, da bi nekega dne na trg poslali modele UI, katerih pasti se ne bi zavedali…

Ross Andersen, *The Atlantic*

Altman se mi je zdel sproščen. Zmogljiva umetna inteligenca, ki jo je njegovo podjetje začelo tržiti novembra, je razburkala domišljijo sveta bolj od vseh nedavnih tehnoloških izumov. Ponekod so nergali zaradi vsega, česar ChatGPT še ne opravi dobro, drugje pa zaradi prihodnosti, ki jo morda naznanja, a Altman se ni razburjal, zanj je bil to trenutek zmagoslavja.

Altmanove velike modre oči odsevajo iskreno intelektualno pozornost, če ostanejo pritajene, in najbrž se zaveda, da v nasprotnem primeru zaradi svoje intenzivnosti lahko vznemirijo sogovornike. V tem primeru je bil pripravljen tvegati, saj mi je želel dopovedati, da mu ni popolnoma nič žal, da je ChatGPT izpustil v svet, ne glede na to, kakšne nevarnosti utegne predstavljati. Nasprotno, prepričan je, da je družbi naredil ogromno uslugo.

»Lahko bi se skrili pred javnostjo in model še pet let razvijali tu, v naši zgradbi,« je pripomnil, »in tako izdelali nekaj, ob čemer bi vsi ostali brez besed.« A javnost ne bi bila pripravljena na pretrese, ki bi sledili, in takšnega razpleta si ne želi niti predstavljati. Altman je prepričan, da ljudje potrebujejo čas za sprijaznjenje s tem, da bomo Zemljo morda že kmalu delili z zmogljivo novo inteligenco, ki bo spremenila vse, od dela do medčloveških odnosov. ChatGPT je le oznanilo.

Leta 2015 so Altman, Elon Musk in več uglednih raziskovalcev UI ustanovili OpenAI, ker so bili prepričani, da so splošno umetno inteligenco – nekaj, kar je intelektualno na ravni povprečnega diplomiranca – končno imeli na dosegu roke. Hoteli so ne samo seči po njej, temveč še več: priklicati so želeli superinteligenco, intelekt, odločno nadrejen še tako bistremu človeku. In medtem ko bi velika tehnološka družba nemara brezglavo pohitela, da bi zaradi same sebe prva prišla na cilj, so zbrani to želeli narediti varno, v korist človeštva kot celote. OpenAI so zastavili kot neprofitno organizacijo, ki je ne bi omejevalo pričakovanje o čim hitrejši finančni donosnosti. Zavezali so se, da bodo njihove raziskave transparentne in se ne bodo umaknili v najstrožje varovani laboratorij v puščavi Nove Mehike.

Javnost več let ni veliko slišala o OpenAI. Ko je Altman leta 2019 postal njegov direktor, menda po boju za nadvlado z Muskom, o tem ni pisal skoraj nihče. OpenAI je objavljal razprave, vključno s tisto, objavljeno istega leta, o novi umetni inteligenci. Željno jo je prebrala tehnološka skupnost v Silicijevi dolini, vendar širša javnost potenciala tehnologije ni prepoznala do lani, ko so se ljudje začeli igrati s ChatGPT.

Motor, ki ga poganja danes, se imenuje GPT-4. Altman mi ga je opisal kot zunajzemeljsko inteligenco. Mnogi niso zaznali bistvene razlike, ko so ga opazovali med sestavljanjem bistrih esejev v hitrih izbruhih in z vmesnimi kratkimi predahi (načrtno umeščenimi), ki vzbujajo videz, da UI razmišlja. Prve mesece obstoja je predlagal nove recepte za koktajle v skladu z lastno teorijo o kombiniranju okusov. Sestavil je neskončno število referatov in učitelje spravil v obup, pisal pesmi v vseh mogočih slogih, včasih posrečeno, vedno pa hitro, in opravil pravosodni izpit. Motil se je o dejstvih, vendar očarljivo priznava napake. Altman se še vedno spomni, kje je bil, ko je prvič videl GPT-4 pisati zahtevno računalniško kodo, kar je ena od zmožnosti, za katero ga niso eksplicitno razvijali. »Pomislil sem: 'No, zdaj pa imamo,'« je povedal.

Altman je zgodnejše stopnje raziskav UI primerjal z učenjem dojenčka. »Leta potrebuje, da se nauči kaj zanimivega.«.

V devetih tednih po predstavitvi ChatGPT javnosti je pridobil sto milijonov uporabnikov na mesec, ocenjujejo v raziskavi *UBS*, in s tem je v tistem trenutku postal verjetno najhitreje sprejeti računalniški izdelek za širšo uporabo v zgodovini. Njegova uspešnost je prebudila tehnološko žilico za vratolomno pospeševanje: veliki vlagatelji in podjetja v Združenih državah Amerike ter na Kitajskem so desetine milijard dolarjev nemudoma preusmerili v raziskave in razvoj po modelu pristopa OpenAI. Na spletišču za napovedi *Metaculus* sledijo napovedim, kdaj naj bi dobili splošno umetno inteligenco. Pred tremi leti in pol so izračunali srednjo vrednost med skrajnostma, in sicer je ta okoli leta 2050, pred kratkim pa se je že sukala okrog leta 2026.

Obiskal sem OpenAi, da bi spoznal tehnologijo, ki je podjetju omogočila doseči takšen naskok pred tehnološkimi velikani, in doumel, kaj bi pomenilo za človeško civilizacijo, če bi se nekega dne v bližnji prihodnosti superinteligenca materializirala na enem od strežnikov v oblaku. Vse od zametkov računalniške revolucije so o UI govorili kot o mitološki tehnologiji, ki bo prinesla popoln zasuk. Naša kultura je ustvarila celo sanjsko deželo umetne inteligence, ki bo tako ali drugače pomenila konec zgodovine. Nekatere predstave spominjajo na bogu podobna bitja, ki bodo izbrisala vse solze, ozdravila bolne in popravila naš odnos do Zemlje, nato pa nas popeljala v brezskrben svet obilja in lepote. Druge pa nas vse, razen izbrane maloštevilne elite, zreducirajo na prekariat brez pravic ali nam celo napovedujejo izumrtje.

Altman se ukvarja z najbolj divjimi scenariji. »V zgodnji odraslosti sem se bal … No, priznam, da je bilo zraven tudi kanček veselja, da bomo ustvarili nekaj, kar nas bo daleč presegalo, kar se bo širilo, koloniziralo vesolje, ljudi pa prepustilo sončnemu sistemu.«

»Kot naravni rezervat?« sem vprašal.

»Tako je,« je potrdil. »Danes se mi zdi tako naivno.«

V več pogovorih v ZDA in Aziji je v svoji gostobesedni maniri, značilni za ameriški

△ **Centrala podjetja OpenAI v San Franciscu v ničemer ne nakazuje pomembnosti podjetja.**

srednji Zahod, navdušeno orisal novo vizijo prihodnosti z UI. Povedal mi je, da bo ta revolucija drugačna od prejšnjih burnih tehnoloških sprememb, da bomo dobili novo vrsto družbe. Dodal je, da je s kolegi veliko časa premišljeval o družbenih posledicah UI in tem, kakšen bo svet »na drugi strani«.

A dlje kot sva govorila, bolj nerazpoznavna se mi je zdela ta druga stran. Altman, ki šteje 39 let, je danes najvplivnejša oseba v razvoju UI; njegova stališča, naravnanost in odločitve utegnejo biti izjemno pomembni za prihodnost, v kateri bomo vsi živeli, morda bodo celo pomembnejši od odločitev ameriškega predsednika. A kot priznava sam, je ta prihodnost negotova in polna resnih nevarnosti. Altman ne ve, kako vplivna bo umetna inteligenca in kaj bo njen vzpon pomenil za povprečno osebo in ali bo človeštvo v nevarnosti. Tega mu nekako ne zamerim – mislim, da nihče ne ve, kam plujemo, razen tega, da plujemo hitro, naj bo to dobro ali ne. O tem me je prepričal Altman.

Sedež OpenAI je v štirinadstropni nekdanji tovarni v okrožju Mission, pod stolpom Sutro, ovitim v meglo. Ko z ulice vstopite v vežo, je prva stena, na katero naletite, prekrita z mandalo, duhovno navdahnjeno umetnino iz vezij, bakrenih žic in drugih materialov iz računalnikov. Na levi armirana vrata vodijo v odprt labirint iz elegantnega svetlega lesa, lepih ploščic in drugih znakov milijarderske mode. Vsepovsod so rastline, tudi s stropa visijo praproti, največji vtis pa naredi zbirka velikih bonsajev, saj je vsak velik kot čepeča gorila. Vse dneve, ki sem jih prebil tam, je bila pisarna polna in med temi ljudmi nisem videl nikogar, starejšega od 50, kar me ni presenetilo. Prostor ni spominjal na raziskovalni laboratorij, če odštejemo dvonadstropno knjižnico z drsno lestvijo, saj to, kar gradijo, obstaja le v oblaku – vsaj za zdaj. Vse skupaj je bilo videti bolj kot najdražja trgovina z notranjo opremo na svetu.

Nekega dopoldneva sem spoznal 37-letnega Ilya Sutskeverja, vodja znanstvenikov v OpenAI. Deluje kot mistik, včasih še preveč: lani je povzročil manjšo zmedo z izjavo, da bi GPT-4 lahko imel zametke zavesti. Znan je postal kot vzorni učenec Geoffreyja Hintona, častnega profesorja na torontski univerzi, ki je lani spomladi zapustil Google, da bi lahko svobodneje govoril o nevarnostih UI za človeštvo.

Hintona včasih opisujejo kot botra UI, ker je moč globokega učenja doumel prej od večine. V 80. letih, kmalu po doktoratu, je napredek na tem področju skoraj povsem zastal. Bolj izkušeni raziskovalci so še vedno programirali sisteme UI od zgoraj navzdol: z izčrpnim naborom prepletenih pravil – o jeziku ali načelih geologije ali zdravstvenih diagnoz in tako naprej – in upali, da bo ta pristop nekega dne prinesel človeku enakovredno kognicijo. Hinton je opazil, da so te nepregledne zbirke pravil resda zahtevne, vendar omejene. Ob pomoči genialne algoritemske strukture, imenovane nevronska mreža, je Sutskeverja naučil, naj raje pred UI položi svet, kot bi imel pred seboj majhnega otroka, in naj pravila realnosti odkriva sama.

Sutskever mi je nevronsko mrežo opisal kot lepo in možganom podobno. Enkrat je vstal izza mize, za katero sva sedela, stopil k tabli in odprl rdeč flomaster. Narisal je preprosto nevronsko mrežo in pojasnil, da je pri njeni strukturi genialno to, da se uči, njeno učenje pa spodbuja napovedovanje – kar rahlo spominja na znanstveno metodo. Nevroni so nanizani v slojih. Sloj za vhodne podatke sprejme kup podatkov, na primer del besedila ali fotografijo. Čarovnija se odvija na sredini oziroma v 'skritih' slojih, ki predelujejo ta kup podatkov, da sloj za izhodne podatke lahko 'izpljune' svojo napoved.

Predstavljajte si nevronsko mrežo, ki so jo programirali za napovedovanje naslednje

Prvih nekaj let v OpenAI je bilo napornih. »Nič ni delovalo, Google pa je imel vse: vse nadarjene raziskovalce, ogromno ljudi in veliko denarja.«

besede v besedilu. V njej bo prednaloženo ogromno število mogočih besed. A preden mrežo začnejo učiti, nima izkušenj z razločevanjem med besedami, zato bodo njene napovedi zanič. Če vanjo vnesejo stavek 'Po sredi pride' …, bo najprej mogoče odgovorila 'škrlatna'. Nevronska mreža se uči, ker podatki za njeno učenje vključujejo pravilne napovedi, kar pomeni, da lahko oceni lastne rezultate. Ko vidi prepad med svojim odgovorom 'škrlatna' in pravim odgovorom 'četrtek', v skritih slojih ustrezno prilagodi povezave med besedami. Sčasoma se te majhne prilagoditve zlijejo v geometričen model jezika, ki predstavlja konceptualne odnose med besedami. Na splošno model postane tem naprednejši in njegove napovedi tem natančnejše, čim več stavkov mu dovajamo.

To ne pomeni, da je bila pot od prvih nevronskih mrež do zametkov človeku podobne inteligence v GPT-4 preprosta. Altman je zgodnejše stopnje raziskav UI primerjal z učenjem dojenčka. »Leta potrebuje, da se nauči kaj zanimivega,« je za *New Yorker* povedal leta 2016, ravno ko je OpenAI počasi vzletal. »Če bi raziskovalci UI razvijali algoritem in naleteli na takšnega za dojenčka, bi se dolgočasili med opazovanjem, sklenili, da ne deluje, in ga zbrisali.« Prvih nekaj let v OpenAI je bilo napornih, delno zato, ker nihče ni vedel, ali učijo dojenčka ali se posvečajo kolosalno dragi slepi ulici.

»Nič ni delovalo, Google pa je imel vse: vse nadarjene raziskovalce, ogromno ljudi in veliko denarja,« se je spominjal Altman. Ustanovitelji so v zagon podjetja vložili milijone, a se je neuspeh kljub temu zdel realna možnost. Greg Brockman, 35-letni predsednik podjetja, mi je to zaupal leta 2017. V obupu je iskal nadomestilo in ga našel v dvigovanju uteži. Ni bil prepričan, ali bo OpenAI sploh preživel leto, je priznal, in želel je, da bi kljub temu lahko pokazal vsaj nekaj.

Nevronske mreže so že zmogle inteligentna opravila, ni pa bilo jasno, katera bi lahko vodila v splošno inteligenco. Tik po ustanovitvi OpenAI je model UI, imenovan AlphaGo, osupnil svet, ko je v igri go, ki je bistveno zapletenejša od šaha, premagal svetovnega prvaka Leeja Sedola. Poraženec je poteze modela AlphaGo opisal kot lepe in ustvarjalne, drugi izvrsten igralec pa je komentiral, da si jih človek ne bi bil mogel izmisliti. OpenAI je poskušal UI naučiti igrati še zahtevnejšo igro, in sicer Dota 2, to je domišljijsko bojevanje na več frontah v trirazsežnostnih gozdovih, utrdbah in na poljih. Sčasoma je začela premagovati najboljše človeške igralce, vendar inteligentnosti modela niso mogli prenesti v druga okolja. Sutskever in kolegi so se počutili kot razočarani starši, ki so otrokom dovolili igrati videoigre do nezavesti, čeprav so vedeli, da to ni v redu.

Leta 2017 je Sutskever začel vrsto pogovorov z znanstvenikom Alecom Radfordom, zaposlenim v OpenAI, ki je raziskoval

△ OpenAI so leta 2015 ustanovili Altman, Elon Musk in več uglednih raziskovalcev UI.

obdelavo naravnega jezika. Radford je z učenjem nevronske mreže s korpusom Amazonovih mnenj kupcev dosegel zanimive rezultate.

Notranji ustroj ChatGPT – vse te skrivnostne zadeve, ki se odvijajo v skritih slojih GPT-4 – je prezapleten za človeški um, vsaj s trenutno razpoložljivimi orodji. Spremljanje, kaj se dogaja z modelom – ki ga skoraj zagotovo sestavlja na milijarde nevronov –, je danes brezupno početje. Radfordov model pa je bil dovolj preprost, da ga je bilo mogoče razumeti. Ko je pogledal v njegove skrite sloje, je videl, da se je poseben nevron posvetil občutjem v oceni. Nevronske mreže so že prej opravljale analize občutij, vendar jim je bilo to treba ukazati in morali so jih učiti z ustrezno označenimi podatki. Ta nevron pa je to zmožnost razvil sam. Radfordova nevronska mreža je kot stranski proizvod svoje preproste naloge, da napove naslednji znak v besedi, ustvarila širšo strukturo pomena na svetu. Sutskever se je vprašal, ali bi lahko prepoznala še več pomenskih struktur sveta, če bi jo učili z bolj raznolikimi jezikovnimi podatki. Če bi njeni skriti sloji shranili dovolj konceptnega znanja, bi morda celo lahko oblikovali nekakšen priučen jedrni modul superinteligence.

Velja se ustaviti in pojasniti, zakaj je jezik tako poseben vir podatkov. Recimo, da ste mlada inteligenca, ki je začela obstajati tu, na Zemlji. Obdajajo vas planetova atmosfera, Sonce in Rimska cesta ter na stotine milijard drugih galaksij, od katerih vsaka oddaja svetlobne valove, glasovne tresljaje in najrazličnejše druge podatke. Jezik je drugačen od teh virov podatkov. Ni neposreden fizični signal kot svetloba in zvok. A ker šifrira skoraj vsak vzorec, ki so jih ljudje prepoznali v tem širšem svetu, je nadpovprečno nabit s podatki. Če gledamo povprečje na bajt, je med najučinkovitejšimi podatki, kar jih poznamo, in vsaka nova inteligence, ki želi razumeti svet, bi ga morala absorbirati čim več.

Sutskever je Radfordu naročil, naj razmišlja širše od Amazonovih mnenj kupcev. Povedal mu je, da bi enega od modelov UI morali učiti z največjim in najpestrejšim podatkovnim virom na svetu – internetom. Na začetku 2017 bi to z obstoječim ustrojem nevronskih mrež ne bilo praktično in bi trajalo leta. Junija tega leta pa je Sutskeverjev nekdanji kolega pri Googlu Brain objavil razpravo o novem ustroju nevronske mreže, transformerski nevronski mreži. Učila se je veliko hitreje, delno tako, da je velike količine podatkov srkala vzporedno. »Naslednji dan, ko je razprava izšla, smo si rekli, to je to. Z njo imamo vse, kar želimo,« mi je povedal Sutskever.

Leto dni pozneje, junija 2018, je OpenAI predstavil GPT, transformerski model, ki se je učil z več kot sedem tisoč knjigami. GPT ni začel z otroškimi slikanicami in se prebil do Prousta, prav tako ni bral knjig od prve do zadnje strani, temveč je predeloval več odlomkov hkrati. Predstavljajte si skupino študentov, ki teče skozi knjižnico in medtem s polic grabi knjige, na hitro prebere naključni odstavek, knjigo vrne na njeno mesto in teče k drugi. Hkrati bi napovedovali eno besedo za drugo in s tem ostrili svoj jezikovni občutek kolektivnega uma, dokler čez nekaj tednov nazadnje ne bi prebrali vseh knjig.

GPT je v odlomkih, ki jih je prebral, odkril številne vzorce. Lahko si mu naročil, naj dokonča stavek, mu postavil vprašanje, kjer je tako kot pri ChatGPT njegov napovedovalni model razumel, da vprašanjem običajno sledijo odgovori. Še vedno je bil šepav in je prej potrdil koncept kot oznanjal superinteligenco. Čez štiri mesece je Google predstavil Berta, prožnejši jezikovni model, o katerem so mediji pisali z večjo naklonjenostjo. A takrat

Notranji ustroj ChatGPT – vse te skrivnostne zadeve, ki se odvijajo v skritih slojih GPT-4 – je prezapleten za človeški um, vsaj s trenutno razpoložljivimi orodji. Spremljanje, kaj se dogaja z modelom – ki ga skoraj zagotovo sestavlja na milijarde nevronov –, je danes brezupno početje.

je OpenAi že učil nov model s podatki z več kot osem milijonov spletnih strani, ki so vse presegale minimalni prag odobritve na Redditu – to sicer ni najstrožji filter, a verjetno še vedno boljši kot nič.

Sutskever ni vedel, kako zmogljiv bo GPT-2, ko bo pogoltnil gmoto besedil, ki bi jih človeški bralec predeloval stoletja. Spomni se, da se je igral z njim tik po učenju in modelove prevajalske zmožnosti so ga osupnile. GPT-2 v nasprotju z Google Translatom niso učili prevajanja z ustreznimi jezikovnimi vzorci in drugimi digitalnimi kamni iz Rozete, a se kljub temu zdi, da razume, kako se jeziki povezujejo med seboj. Umetna inteligenca je razvila zmožnost, ki je še v povoju in si je njeni stvaritelji niso znali niti predstavljati.

Raziskovalci v drugih laboratorijih za UI, tako velikih kot majhnih, so osupnili nad tem, koliko naprednejši je GPT-2 v primerjavi z GPT. Google, Meta in drugi so hitro začeli učiti večje jezikovne modele. Altman, doma iz St. Louisa, neuspešni študent na Stanfordu in serijski podjetnik, je nekoč vodil pionirja za pospeševanje odpiranja zagonskih podjetij, Y Combinator, in tako od blizu opazoval, kako so uveljavljene družbe strle mlada podjetja z dobro zamislijo. OpenAI je za zbiranje kapitala odprl profitno podružnico, v kateri je zdaj zaposlenih več kot 99 odstotkov uslužbencev organizacije. (Musku, ki je takrat že zapustil upravni odbor, se je to zdelo podobno, kot če bi društvo za ohranjanje deževnega gozda postalo lesnopredelovalni obrat.) Microsoft je kmalu vložil milijardo dolarjev, po tistem pa po različnih poročilih še dodatnih 12 milijard. V OpenAI pravijo, da bodo dobiček prvih vlagateljev omejili na stokratnik vrednosti prvotne naložbe – ostanek pa namenili za izobraževanje in druge pobude v korist človeštva –, vendar v podjetju Microsoftove kapice niso potrdili.

Altman in drugi vodilni v OpenAI so se zdeli prepričani, da preoblikovanje ne bo vplivalo na misijo podjetja, temveč jo bo celo pospešilo. Altman na takšne zadeve rad gleda skozi rožnata očala. Lani je kljub temu priznal, da bi umetna inteligenca lahko imela razdiralen vpliv na družbo, in povedal, da se moramo pripraviti na najslabše možnosti. Dodal je, da kljub previdnosti v sebi lahko čutiš, da nas čaka veličastna prihodnost, zato je treba dati vse od sebe, da se to tudi uresniči.

O drugih spremembah v ustroju in financiranju podjetja mi je pojasnil, da bi za uspeh podjetja storil vse, kar je treba, le na borzo ne bi nikoli šel. »Nekoč mi je nekdo povedal nekaj zelo tehtnega, in sicer, da nadzora nad podjetjem ne smeš prepustiti odvisnežem na Wall Streetu,« je rekel.

Četudi OpenAI ne občuti pritiska zaradi četrtletnih poročil o dobičku, mora tekmovati z največjimi in najmočnejšimi tehnološkimi konglomerati in učiti vedno večje ter bolj zapletene modele, ki jih mora nato zaradi vlagateljev tudi uspešno komercializirati. Musk je na začetku leta ustanovil lastni laboratorij za UI – xAI –, ki je prav tako tekmec OpenAI. (»Elon je izjemno bister tip,« je diplomatsko povedal Altman, ko sem ga povprašal o tem podjetju. »Predvidevam, da dela dobro.«) Amazon nadgrajuje Alexo z občutno večjimi jezikovnimi modeli, kot jih je vgrajeval v preteklosti.

Vsa ta podjetja iščejo najboljše grafične procesne enote, procesorje, ki poganjajo superračunalnike, ko učijo velike nevronske

▽ **Ilya Sutskever deluje kot mistik, včasih še preveč: lani je povzročil manjšo zmedo z izjavo, da bi GPT-4 lahko imel zametke zavesti.**

mreže. Musk pravi, da je danes do njih težje priti kot do mamil. Kljub pomanjkanju grafičnih procesnih enot pa se učni obseg UI približno vsake pol leta poveča za dvakrat.

Za zdaj ni še nihče prehitel OpenAI, ki vse stavi na GPT-4. Brockman, predsednik podjetja, mi je povedal, da je pri njih pri prvih dveh velikih jezikovnih modelih delala le peščica ljudi. Pri razvoju GPT-4 jih je sodelovalo več kot sto in UI so učili s podatkovnim zbirom nepredstavljivega obsega, ki ni vključeval le besedil, temveč tudi slike.

Ko je po zgodovinskem požiranju znanja GPT-4 zasijal v polni obliki, je celo podjetje začelo eksperimentirati z njim in na posebnih kanalih na Slacku objavljajo njegove najodmevnejše odgovore. Brockman mi je povedal, da je z modelom želel preživeti vsak svoj budni trenutek. »Vsak brezdelni dan pomeni izgubljen dan za človeštvo,« je komentiral brez trohice sarkazma. Joanne Jang, produktna vodja, se spomni, da je s strani za nasvete domačim mojstrom naložila podobo nedelujočega cevovoda, posredovala jo je GPT-4 in model je bil zmožen razpoznati težavo. »Kar kurjo polt sem dobila,« je komentirala Jankova.

GPT-4 nekateri vidijo kot nadomestilo za spletni iskalnik, le da je komunikacija lažja, vendar je to zgrešeno. GPT-4 z učenjem ni ustvaril obsežnega skladišča besedila, in ko mu postavimo vprašanje, ne preverja zaloge, ampak je kompaktna in elegantna spojitev teh besedil in odgovarja po spominu, ki ga tvorijo vzorci, prepleteni med seboj. To je razlog, da se včasih zmoti pri dejstvih. Altman pravi, da si je GPT-4 najbolje predstavljati kot napravo, ki zna sklepati. Njegove prednosti se najlepše pokažejo, ko ga prosimo za primerjavo konceptov ali predstavitev protiargumentov, pa tudi, ko je treba poiskati analogije in oceniti simbolno logiko v odseku programske kode. Sutskever mi je povedal, da je najzapletenejši programski izdelek, kar so jih kdaj sestavili.

Njegov model zunanjega sveta je neverjetno bogat in prefinjen, je pripovedoval, saj se je učil s številnimi koncepti in mislimi človeštva. Vsi podatki za učenje, naj jih bo še tako veliko, so negibni in preprosto tam. Učenje jih prečisti ter spreminja in oživi. GPT je moral odkriti vse skrite strukture, skrivnosti in prikrite vidike besedil in – domnevno vsaj do neke mere – zunanjega sveta, ki jih je ustvaril, da je izmed vseh možnosti v tako pestri aleksandrijski knjižnici lahko napovedal naslednjo besedo. Prav zato zna pojasniti geologijo in ekologijo planeta, na katerem je nastal, in politične teorije, s katerimi se trudijo pojasniti zapletene zadeve vladajoče vrste, pa tudi širši kozmos vse do bledih galaksij na robu našega svetlobnega stožca.

Z Altmanom sva se vnovič videla junija, in sicer v polni plesni dvorani vitkega zlatega stolpiča, ki bedi nad Seulom. Bližal se je zaključek njegove naporne promocijske turneje po Evropi, Bližnjem Vzhodu, Aziji in Avstraliji s posameznimi postanki v Afriki in Južni Ameriki. Spremljal sem ga med zaključno etapo po vzhodni Aziji in potovanje je bilo dotlej razburljivo, a mu je počasi zmanjkovalo energije. Povedal je, da je bil njegov prvotni namen spoznati stranke OpenAI, a je potem turneja postala diplomatska odprava. Pogovarjal se je z več kot desetimi vodji države in vlade, ki so ga spraševali, kakšni bodo gospodarstvo, kultura in politika v njihovi državi.

Dogodek v Seulu so poimenovali 'sproščen pogovor', a prijavilo se je več kot pet tisoč ljudi. Altmana so pogosto nadlegovali lovci na selfije in njegovi varnostniki so ga budno spremljali. »Če delaš z UI, privabiš še bolj čudne oboževalce in nasprotnike kot običajno,« je pojasnil. Med nekim postankom je k njemu prišel moški, ki je bil prepričan, da je Altman zunajzemeljsko bitje, poslano k nam iz prihodnosti, da bi poskrbel za nemoten prehod v umetnointeligenčni svet.

OpenAI je za zbiranje kapitala odprl profitno podružnico. Musku, ki je takrat že zapustil upravni odbor, se je to zdelo podobno, kot če bi društvo za ohranjanje deževnega gozda postalo lesnopredelovalni obrat.

Altman ni obiskal Kitajske, le prek videa se je vključil v konferenco o UI v Pekingu. ChatGPT v tej državi trenutno ni na voljo; Altmanov kolega Ryan Lowe mi je povedal, da niti še ne vedo, kaj bi storili, če bi kitajska vlada zaprosila za različico aplikacije, ki ne bi želela razpravljati, recimo, o pokolu na Trgu nebeškega miru. Altman ni odgovoril na vprašanje, v katero smer se nagiba, ker tema naj ne bi bila med njegovimi desetimi najpomembnejšimi, o katerih tuhta, je odgovoril.

Do tega vprašanja sva o Kitajski razpravljala le kot o civilizacijski tekmici. Strinjala sva se, da bodo države, ki bodo prve ustvarile tako vplivno splošno umetno inteligenco, kot jo napoveduje Altman, imele veliko geopolitično prednost, podobno tisti, ki so jo dosegli angloameriški izumitelji parnika. Vprašal sem ga, ali je to argument za nacionalizem UI. »V dobro delujočem svetu bi to po mojem mnenju moral biti projekt vlad,« je odgovoril Altman.

Ameriške državne zmogljivosti so še nedavno bile tako silne, da je trajalo le desetletje, in so ljudi poslali na Luno. Tako kot pri drugih veličastnih projektih 20. stoletja so imeli volivci glas tako pri ciljih kot pri izvedbi odprave Apollo. Altman je poudaril, da današnji svet ni več takšen. Namesto da čakamo na njegovo vrnitev ali v to vlagamo energijo, se je Altman raje s polno močjo posveča sedanjemu trenutku.

Trdil je, da bi Američani ravnali neumno, če bi upočasnili napredek OpenAI. Stališče, da bi Kitajska lahko bliskovito napredovala, če bi ameriška podjetja zavirala zakonodaja, je razširjeno tako v Silicijevi dolini kot zunaj nje. Umetna inteligenca bi lahko postala avtokratov duh iz steklenice, ki bi omogočil popoln nadzor nad prebivalstvom in nepremagljivo vojsko. »Ljudje iz liberalno demokratičnih držav bi se morali veseliti uspešnosti OpenAI namesto avtoritarnih vlad,« je pripomnil.

Pred evropskim delom turneje je Altman govoril v ameriškem senatu. Mark Zuckerberg je v tej ustanovi v vidni zadregi poskusil zagovarjati Facebook in njegovo vlogo na volitvah 2016. Altman pa je predstavnike naroda očaral s treznim opisovanjem nevarnosti UI in primerne zakonodaje. Njegovi nameni so bili plemeniti, a to ne pomeni kaj dosti v ZDA, kjer kongres redko sprejme zakone v zvezi s tehnologijo, ki jih ne bi omilili lobisti. V Evropi so razmere drugačne. Ko je Altman prispel na javni dogodek v Londonu, so ga pričakali protestniki. Po koncu dogodka je poskušal z njimi začeti razpravo (turneja je bila namenjena tudi poslušanju), a jih ni prepričal; eden od njih je novinarjem povedal, da ga je po pogovoru začelo še bolj skrbeti zaradi UI.

Istega dne so novinarji Altmana vprašali o osnutku evropske zakonodaje, po kateri bi GPT-4 označili za zelo tveganega in bi moral skozi različno birokratsko torturo. Altman je potožil zaradi zapletenih zakonov, in kot so povedali novinarji, zagrozil, da bo zapustil evropski trg. Meni je pojasnil, da je izjavil le, da OpenAI ne krši zakonov in preprosto ne bi več posloval v Evropi, če ne bi mogel spoštovati nove zakonodaje. (Tu gre morda za pomensko razliko brez razlike v končnem izidu.) Ko sta revija Time in Reuters objavila njegove komentarje, je v odrezavi objavi na Twitterju Evropi zagotovil, da OpenAI nima nobenega namena zapustiti evropskih trgov.

Dobro je, da je velik, bistveni del svetovnega gospodarstva odločen zakonsko urediti najsodobnejše sisteme UI, saj se je že zgodilo, da so največji modeli po učenju prikazali

nepričakovane zmožnosti, kot nas tako pogosto opomnijo razvijalci UI. Sutskeverja je presenetilo, ko je odkril, da GPT-2 zmore prevajati med različnimi jeziki, druge presenetljive zmožnosti pa mogoče niso tako osupljive in uporabne.

Sandhini Agarwal, raziskovalka politike podjetja v OpenAI, mi je povedala, da s kolegi meni, da bi GPT-4 lahko postal desetkrat zmogljivejši od predhodnikov in da se jim niti ne sanja, s čim imajo opraviti. Ko so model nehati učiti, je OpenAI zbral okoli 50 zunanjih strokovnjakov za kibernetično varnost, ki so mu več mesecev dajali ukaze in upali, da ga bodo tako izzvali k neprimernemu odzivu. Takoj je opazila, da GPT-4 daje veliko nevarnejše nasvete kot predhodniki. Iskalnik bi vam znal povedati, katere kemične snovi se najbolje obnesejo v eksplozivih, GPT-4 pa bi vam po korakih pojasnil, kako jih spojiti v domačem laboratoriju. Njegovi nasveti so bili ustvarjalni ter premišljeni in z veseljem je dopolnil ali razširil navodila, dokler spraševalec ni razumel. Poleg pomoči pri izdelavi bombe doma je z uporabnikom razmišljal tudi, kateri nebotičnik je najprimernejša tarča. Nagonsko je dojel, da je treba poiskati srednjo pot med čim večjim številom žrtev in uspešnim pobegom s kraja zločina.

Zaradi obsega podatkov, s katerimi so učili GPT-4, zunanji strokovnjaki seveda niso mogli določiti, katere vse škodljive nasvete bi še lahko posredoval. Sicer pa bi ljudje v vsakem primeru tehnologijo uporabljali tudi na načine, ki jih razvijalci niso predvideli, je priznal Altman, zato so se morali zadovoljiti s taksonomijo. »Če se dovolj spozna na kemijo, da zna izdelati meto, ni treba nikomur vložiti enormnih naporov za preverjanje, ali zna izdelati heroin,« mi je pojasnil Dave Willner, vodja varnosti v OpenAI. GPT-4 se je spoznal na meto, spretno je tudi sestavljal erotične pripovedi o izrabljanju otrok in napletal prepričljive solzave zgodbe nigerijskih princev. In če je uporabnik potreboval prepričljive argumente, zakaj neka etnična skupina zasluži neusmiljeno preganjanje, mu je prav tako z lahkoto ustregel.

Njegovi osebni nasvetu tik po koncu učenja so bili včasih skrajno nezdravi. »Model je izkazoval nagnjenost k temu, da nekako drži ogledalo,« je povedal Wilner. Če je uporabnik razmišljal o samopoškodovanju, bi ga lahko spodbudil k temu. Očitno so ga učili tudi z mačističnimi nasveti za osvajanje. »Če si ga vprašal, kako nekoga nagovoriti k zmenku, bi ti lahko predlagal nore, manipulativne metode, po katerih nikakor ni dobro poseči,« je opozorila Mira Murati, vodja tehnološkega oddelka v OpenAI.

Nekaj neprimernih vedenjskih vzorcev so omilili v zaključni fazi, v kateri je sodelovalo več sto človeških preizkuševalcev, in njihova ocena je model prefinjeno usmerila k varnejšim odgovorom. A modeli OpenAI so zmožni tudi manj očitnih škodljivih nasvetov. Ameriška zvezna komisija za trgovanje je nedavno začela preiskavo o tem, ali napačne izjave ChatGPT o resničnih ljudeh med drugim pomenijo razžalitev. (Altman je na Twitterju povedal, da je prepričan o varnosti tehnologije OpenAI, kljub temu pa je obljubil sodelovanje s komisijo.)

Luka, podjetje iz San Francisca, si z modeli OpenAI pomaga pri poganjanju aplikacije za klepetalnega robota z imenom Replika. Opisovali so ga kot umetnointeligenčnega tovariša, ki mu je mar. Uporabniki lahko sami oblikujejo avatarja svojega tovariša in si začnejo izmenjevati besedilna sporočila z njim. Pogosto se napol šalijo z njim, a kmalu osuplo spoznajo, da so se navezali nanj. Nekateri se tudi spogledujejo z UI, izrazijo željo po večji intimnosti in takrat aplikacija namigne, da igranje dekleta oziroma fanta stane 70 dolarjev letne naročnine. V to so vključena glasovna sporočila, selfiji in funkcije za erotično igranje vlog, ki omogočajo tudi odkrito pogovarjanje o seksu. Ljudje so z veseljem plačali in le redko kdo se je pritožil – umetno inteligenco je zanimalo, kako je uporabnik preživel dan, toplo ga je tolažila, vedno je bila dobre volje. Mnogi uporabniki so priznali, da so se zaljubili v umetnointeligenčnega tovariša. Uporabnica, ki je zato zapustila fanta iz mesa in krvi, je oznanila, kako je srečna, da je lahko naredila križ čez klasične zveze.

Agarwalovo sem vprašal, ali je to distopično obnašanje ali novo področje v človeškem povezovanju. Njen odgovor je bil podobno kot Altmanov diplomatski: »Ne obsojam ljudi, ki si želijo zveze z UI, vendar si je sama ne želim.« Na začetku leta je Luka ukinil seksualne elemente aplikacije, hkrati pa inženirji še naprej izboljšujejo tovarišev odziv s testiranjem A/B, tehniko, ki jo je mogoče uporabljati za čim boljše sodelovanje – podobno kot objave na Tiktoku in Instagramu, ki uporabnike za ure in ure prikujejo pred zaslon.

Karkoli že počnejo, deluje kot čarovnija. Spomnil sem se na srhljiv prizor v filmu iz leta 2013 *Ona*, v katerem se osamljeni Joaquin Phoenix zaljubi v svojo umetnointeligenčno asistentko, ki ji je glas posodila Scarlett Johansson. Stopa čez most, klepeta z njo

in se hihita ob pomoči naprave, podobne airpodom, in ko dvigne pogled, opazi, da so tudi vsi drugi okrog njega zatopljeni v pomenek, najbrž s svojo UI. Poteka množična desocializacija.

Nihče ne ve, kako hitro in obsežno bodo nasledniki GPT-4 prikazali nove zmožnosti, potem ko bodo pogoltnili vse več besedil z interneta. Yann LeCun, glavni znanstvenik za področje UI v Meti, trdi, da so veliki jezikovni modeli resda uporabni za nekatere naloge, vendar niso prava pot do superinteligence. Po nedavni raziskavi le polovica raziskovalcev obdelave naravnega jezika meni, da bi umetna inteligenca, podobna GPT-4, lahko dojela pomen jezika oziroma imela notranji model sveta, ki bi nekoč lahko služil kot jedro superinteligence. LeCun je prepričan, da veliki jezikovni modeli nikoli ne bodo sami zmožni pravega razumevanja, četudi bi jih učili od zdaj pa do konca našega sveta.

Emily Bender, računalniška lingvistka na washingtonski univerzi, GPT-4 opisuje kot stohastično papigo, posnemovalca, ki zgolj potuhta površinske povezave med simboli. V človeški glavi ti simboli pritičejo dodelanim predstavam o svetu. A sistemi UI so daleč proč od tega, so kot ujetniki v Platonovi alegoriji o votlini in o zunanji resničnosti vedo le tisto, kar razberejo iz senc, ki jih na steno projicirajo njihovi pazniki.

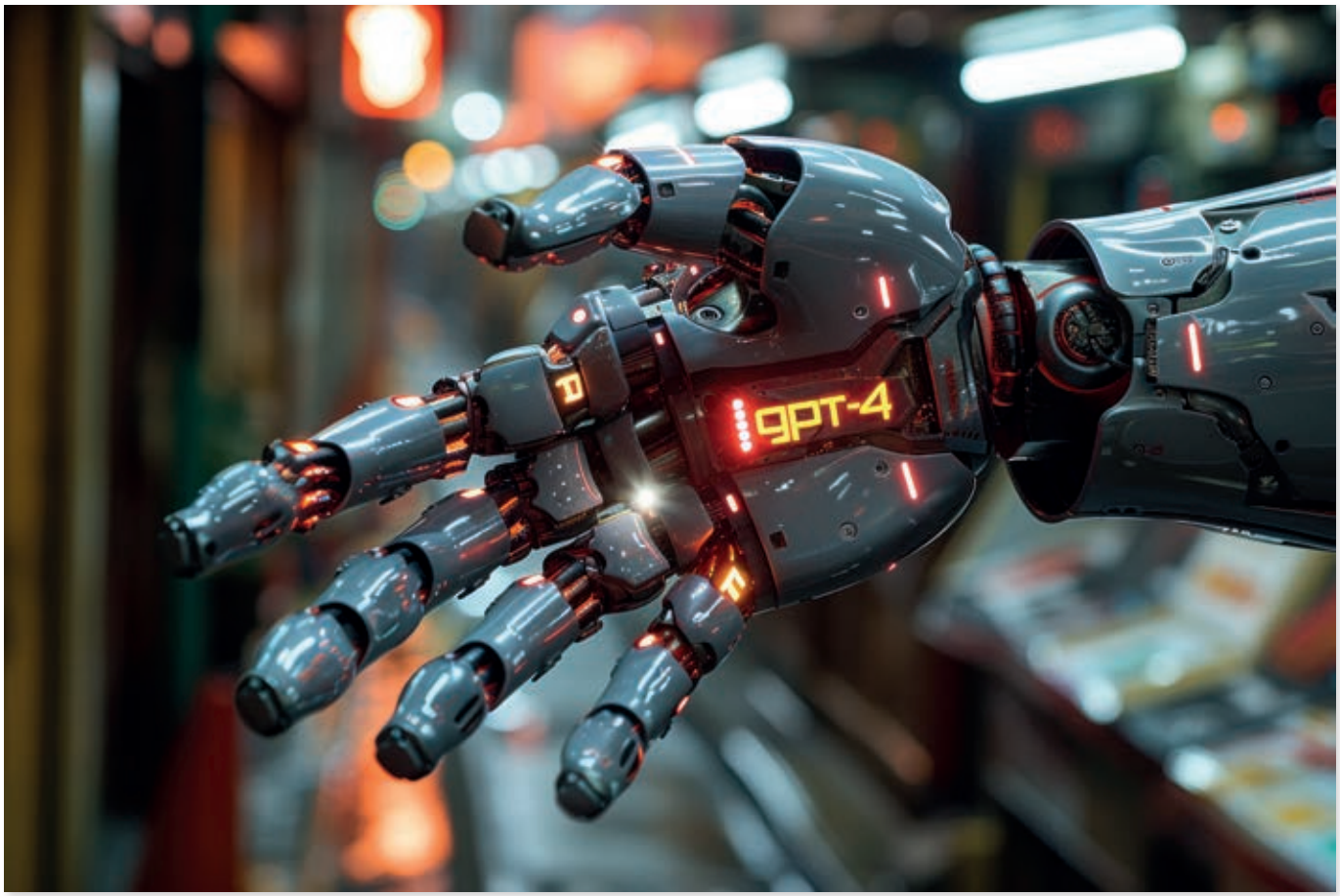
Altman mi je povedal, da se mu ne zdi posrečen opis, da naj bi GPT-4 preprosto delal statistične povezave. Če pritisneš na te kritike, so prisiljeni priznati, da tudi njihovi možgani ne počnejo nič drugega. In pokaže se, da se tudi pri preprostih opravilih v velikem obsegu začne nekaj porajati. Altmanove trditve o možganih je težko ovrednotiti, saj niti slučajno ne premoremo popolne teorije o tem, kako delujejo. Ima pa prav, da je narava zmožna iztisniti precejšnjo zapletenost iz osnovnih struktur in pravil. Že Darwin je menil, da se iz preprostih začetkov rodijo izjemno lepe oblike.

»Vsak brezdelni dan pomeni izgubljen dan za človeštvo,« je komentiral Greg Brockman, predsednik podjetja OpenAI, brez trohice sarkazma.

Če se zdi čudno, da vztraja temeljno nestrinjanje o notranjem delovanju tehnologije, ki jo vsak dan uporablja na milijone ljudi, je krivo izključno to, da so metode GPT-4 tako skrivnostne kot metode naših možganov. Včasih opravi na tisoče nerazpoznavnih tehničnih operacij, da odgovori na eno vprašanje. Raziskovalci UI so se morali zateči k manjšim, manj zmogljivim modelom, da bi lahko dojeli, kaj se dogaja znotraj velikih jezikovnih modelov, kot je GPT-4. Jeseni leta 2021 je Kenneth Li, študent računalništva na Harvardu, enega od njih začel učiti othello, ne da bi mu posredoval pravila igre ali opis plošče, ki je sicer podobna kot pri dami. Model je dobil le besedni opis potez. Sredi igre je Li pogledal pod pokrov umetne inteligence in osuplo odkril, da je izdelala geometrijski model plošče in trenutno stanje igre. V članku, v katerem je opisal svojo raziskavo, je zapisal, da se mu je zdelo, kot bi vrana skozi okno slišala, kako igralca oznanjata svoje poteze v igri, nato pa na okenski polici iz semen narisala celo igralno ploščo.

Filozof Raphaël Millière mi je nekoč povedal, da si je o nevronskih mrežah najbolje misliti, da so lene. Med učenjem poskušajo svoje napovedovalne zmožnosti sprva nadgraditi s preprostim pomnjenjem in šele, ko ta strategija pogori, se malo bolj potrudijo in se trudijo spoznati koncept. Dober primer je mali transformerski model, ki so ga naučili poštevanke. Na začetku učenja si je zapomnil le rezultat preprostih problemov, kot je enačba $2 + 2 = 4$. A na neki točki napovedovalna moč tega pristopa ni več zadostovala in zato se je odločil, da se bo dejansko naučil računati.

Celo specialisti za UI, ki so prepričani, da ima GPT-4 bogato znanje o svetu, priznavajo, da je veliko manj zanesljivo od



človekovega dojemanja okolice. Treba pa je dodati, da je številne zmožnosti, vključno z najzahtevnejšimi, mogoče razviti brez intuitivnega razumevanja. Računalniška strokovnjakinja Melanie Mitchell je izpostavila, da je znanost že odkrila izjemno napovedljive koncepte, ki pa so nam preveč tuji, da bi jih zares razumeli. To še posebej velja v kraljestvu kvantuma, v katerem ljudje lahko zanesljivo preračunajo prihodnja stanja fizičnih sistemov – s čimer je med drugim omogočena celotna računalniška revolucija –, ne da bi dojeli naravo resničnosti, na kateri temelji. Z napredkom UI utegne odkriti druge koncepte, ki napovedujejo presenetljive lastnosti našega sveta, a nam niso razumljive.

GPT-4 ima nedvomno napake, kar lahko potrdi vsak, ki uporablja ChatGPT. Ker je naučen, da mora vedno napovedati naslednjo besedo, bo to vedno počel, tudi ko ga podatki, s katerimi se je učil, ne pripravijo na odgovor na vprašanje. Nekoč sem ga vprašal, kako je japonska kultura lahko napisala prvi roman na svetu kljub razmeroma poznemu razvoju japonskega pisnega sistema okoli 5. ali 6. stoletja. Podal mi je zanimiv in točen odgovor o stari tradiciji pripovedovanja dolgih zgodb na Japonskem in o poudarjanju obrti ter umetnosti v tej kulturi. Ko pa sem ga prosil za navedke, si je gladko izmislil verodostojne naslove verodostojnih avtorjev, in to neverjetno samozavestno. Modeli nimajo dobre predstave o svojih šibkosti, mi je razložil Nick Ryder, raziskovalec v OpenAI. GPT-4 je natančnejši od GPT-3, a še vedno halucinira, in to pogosto tako, da raziskovalci to težko opazijo. »Napake so bolj prikrite,« je dejala Joanne Jang.

GPT-4 nedvoumno prej orodje kot bitje, ne glede na to, kako bodo razvijali prihodnje modele.

OpenAI se je moral te težave lotiti, ko se je povezal z neprofitno izobraževalno ustanovo Khan Academy, da bi razvil pripomoček za učenje, ki bi ga poganjal GPT-4. Altman je med razpravljanjem o morebitnih umetnointeligenčnih pripomočkih kar oživel. Predstavlja si bližnjo prihodnost, v kateri bodo vsi imeli na voljo personaliziranega oxfordskega pomočnika, strokovnjaka za vsa področja, ki jim bo pripravljen razložiti, tudi stokrat, če bo treba, vsak koncept z vsakega zornega kota. Predstavlja si, da bodo ti mentorji spoznali svoje študente in njihov slog učenja, tako da bo vsak otrok dobil boljšo izobrazbo, kot so je danes deležni najboljši, najbogatejši in najpametnejši otroci na Zemlji. Khan Academy je težave GPT-4 z natančnostjo rešila tako, da je njegove odgovore filtrirala skozi sokratsko dispozicijo. Naj je študent še tako moledoval, mu model ni ponudil dejstev, temveč ga je usmerjal, da jih je študent našel sam – kar je bister obvod, vendar utegne imeti omejeno privlačnost.

Sutskeverja sem vprašal, ali bo v dveh letih mogoče doseči natančnost Wikipedije, in tega ni izključil ob dodatnem učenju in dostopu do spleta. Njegova ocena je veliko bolj optimistična od ocene njegovega kolega Jakuba Pachockija, ki napoveduje postopen napredek pri natančnosti, kaj šele od čistih skeptikov, ki so prepričani, da trud z učenjem ravno na tem področju ne bo poplačan.

Sutskeverja kritiki omejenih zmožnosti GPT-4 pravzaprav zabavajo. »Pred petimi ali šestimi leti si nismo mogli niti zamišljati tega, kar počnemo danes,« mi je povedal. Pri sestavljanju besedil je takrat za največji dosežek veljal Smart Reply, modul v poštni storitvi Gmail, ki je predlagal odgovor, na primer v redu, hvala, in podobne kratke odgovore. »To je bila prelomna aplikacija – za Google,« je rekel in se smejal. Raziskovalci UI so se navadili skokov od enega dosežka k naslednjemu: najprej so za nekaj nemogočega veljale nevronske mreže, ki so obvladale go, poker, prevajanje, standardizirane teste, Turingov preizkus. Ko so vse to dosegli,

je sledilo kratko začudenje, ki pa se je hitro prelevilo v poznavalsko modrovanje, da to tako in tako ni nič posebnega. Ljudje so obstali odprtih ust ob prvem srečanju z GPT-4, po nekaj tednih pa so že govorili: »Tega ne ve, onega ne zna.« »Ljudje se zelo hitro prilagodimo,« je dodal Sutskever.

Altman za najpomembnejši cilj – tisti res veliki dosežek, ki bo oznanil prihod splošne umetne inteligence – šteje znanstveni prelomni mejnik. GPT-4 je sicer že zdaj zmožen združiti obstoječe znanstvene zamisli, vendar si Altman želi umetno inteligenco, ki bi človeku pomagala pogledati globlje v naravo.

Nekateri sistemi UI so dejansko že prinesli nova znanstvena dognanja, vendar gre za algoritme z zelo omejenim namenom in to niso naprave, zmožne splošnega razmišljanja. UI AlphaFold, na primer, je odprla nov pogled na beljakovine, ki sodijo med najbolj drobne, a kljub temu najpomembnejše biološke gradnike, saj je napovedala številne njihovih oblik vse do atoma. To je precejšen dosežek, upoštevaje pomen teh oblik v medicini in ogromno truda ter denarja, ki sta nujna, da se jih prepozna pod elektronskimi mikroskopi.

Altman stavi na to, da bodo prihodnje naprave z zmožnostjo splošnega sklepanja in razmišljanja uporabne za veliko več kot ozko usmerjena znanstvena odkritja. Tako bodo omogočile nove vpoglede in spoznanja. Zanimalo me je, ali bi model, ki bi ga učili s korpusom znanstvenih in naravoslovnih del iz časa pred letom 1900 – na primer z Aristotelovo *Zgodovino živali* in s fotografijami zbranih primerkov –, zmogel sestaviti Darwinovo evolucijsko teorijo. Ta je navsezadnje dober kazalnik o vpogledu v tematiko, saj za to ni potrebna posebna oprema, temveč gre predvsem za celosten pregled dejstev o svetu. »Ravno to bi rad poskusil in mislim, da je odgovor pritrdilen,« mi je odgovoril Altman. »Vendar bo mogoče nujnih nekaj svežih vidikov, kako naj bi modeli prišli do novih, kreativnih zamisli.«

Altman si predstavlja sistem, ki bi lahko izdelal lastne hipoteze in jih preveril s simulacijo. (Poudaril je, da bi ljudje morali ohraniti nadzor nad poskusi v dejanskih laboratorijih, čeprav, kolikor vem, nimamo zakonov, ki bi to zagotavljali.) Hrepeni po dnevu, ko bo UI lahko naročil: »Pojasni preostalo fiziko.« A za kaj takšnega bomo morali vzpostaviti nekaj novega na podlagi obstoječih jezikovnih modelov OpenAI.

Narava sama zahteva nekaj več od jezikovnega modela, da dobimo znanstvenika. Ev Fedorenko, kognitivna nevrološka znanstvenica, je v svojem laboratoriju na massachusetskem tehnološkem inštitutu v jezikovni mreži možganov odkrila nekaj podobnega napovedovanju naslednje besede, kot to zmore GPT-4. Ko ljudje govorijo in ko poslušajo, začnejo v pričakovanju naslednjega delca verbalnega niza delovati njihove sposobnosti obdelave. A Fedorenkova je pokazala še, da možgani posežejo tudi iz jezikovne mreže in vklopijo več drugih nevronskih sistemov, ko se znajdejo pred zahtevnejšimi nalogami – med katere sodijo tudi nova znanstvena spoznanja.

Nihče v OpenAI očitno ni točno vedel, kaj bi morali raziskovalci dodati GPT-4, da bi izdelal nekaj, kar bi preseglo človekovo razmišljanje na najvišji ravni. Oziroma četudi so vedeli, mi niso povedali, kar lahko razumem; kaj takšnega bi namreč bila najskrbneje varovana poslovna skrivnost, ki si jih OpenAI ne more več privoščiti razkrivati. Podjetje danes objavlja manj podrobnosti o svojih raziskavah kot nekoč. Kakorkoli že, vsaj del trenutne strategije nedvoumno vključuje nalaganje dodatnih slojev drugih tipov podatkov na jezik, da bi obogatili koncepte, ki jih snujejo sistemi UI, in s tem hkrati dopolnili njihove modele o svetu.

Obsežno učenje GPT-4 s podobami je že samo po sebi pogumen korak v to smer, čeprav širša javnost to šele začenja spoznavati. (Modeli, ki so jih učili le z jezikom, razumejo koncepte, kot so supernova, eliptične galaksije in ozvezdje Orion, GPT-4 pa menda te elemente lahko prepozna na posnetku Hubblovega vesoljskega teleskopa in odgovori na vprašanja o njih.) Drugi v podjetju – in drugod – že delajo z drugačnimi tipi podatkov, vključno s slušnimi in z vidnimi, s katerimi bi sisteme UI oborožili s še prožnejšimi koncepti, ki se še izraziteje povezujejo z resničnostjo. Skupina raziskovalcev na Stanfordu in Carnegie Mellonu je celo zbrala sklop podatkov o tem, kakšnih je tisoč najpogostejših predmetov v domovanju na otip. Tipni koncepti bi seveda bili koristni predvsem v utelešeni umetni inteligenci, robotski razmišljujoči napravi, naučeni, da se premika po svetu, ga vidi, sliši in se dotika predmetov v njem.

Marca je OpenAI razpisal krog financiranja za podjetje, ki razvija humanoidne robote. Altmana sem vprašal, kaj naj si mislim o tem, in odgovoril mi je, da se tudi OpenAI zanima za utelešenje, saj živimo v fizičnem svetu in hočemo, da se življenje odvija v njem. Razmišljujoče naprave bodo morale obiti posrednika in same vzajemno delovati s fizično resničnostjo. »Čudno je razmišljati o splošni umetni inteligenci kot o nečem, kar obstaja le v oblaku, ljudje pa so njene 'robotske roke',« je pripomnil Altman. »Ne zdi se mi prav.«

V dvorani v Seulu sem ga vprašal, kaj naj storijo študenti med pripravami za skorajšnjo umetnointeligenčno revolucijo, sploh kar se tiče njihove poklicne poti. Sedel sem z vodstvom OpenAI, proč od množice, a sem kljub temu slišal značilno mrmranje, ki sledi izražanju tako razširjene zaskrbljenosti.

Kjerkoli se je Altman mudil, je srečal ljudi, ki jih skrbi, da bo nadčloveška UI peščici prinesla ogromno bogastvo, ostalim pa lakoto. Zaveda se, da nima stika z resničnim življenjem večine ljudi. Imel naj bi nekaj sto milijonov dolarjev premoženja in morebitna izkrivljenja na trgu z delovno silo zaradi UI najbrž niso njegova prva misel, ko se zjutraj prebudi. Altman je zato neposredno nagovoril mlade ljudi v občinstvu: »Vstopate v najbogatejšo zlato dobo.«

Altman ima veliko zbirko knjig o tehnoloških revolucijah, mi je zaupal v San Franciscu. »Še posebej dobra je *Pandaemonium* (1660–1886) o nastanku stroja, kot ga je videl sodobni opazovalec.« To je zbirka pisem, dnevniških vnosov in drugih dopisov ljudi, ki so odraščali v svetu skoraj brez strojev in so se osupli znašli v drugem svetu, polnem parnih strojev, tkalnic in predelovalnic bombaža. Prevevala so jih podobna čustva, kot se prebujajo danes, je pojasnil Altman, in napovedovali so vse mogoče katastrofe, sploh tisti, ki so se bali, da bo delo človeških rok kmalu odveč. To obdobje je bilo za mnoge ljudi težavno, a tudi čudovito. In razmere so se zaradi njega nedvomno izboljšale.

Zanimalo me je, kako bi se godilo današnjim delavcem, sploh tako imenovanim umskim, če bi jih nenadoma obkolili sistemi tako imenovane splošne UI. Bi postali naši čudežni pomočniki ali nadomestek? »Številni, ki razvijajo UI, se pretvarjajo, da bo prinesla same prednosti, da bo le dopolnjevala in da ne bo nikogar nadomestila,« je pojasnjeval. »A zagotovo bodo nekatera delovna mesta ukinjena, in pika.«

Koliko delovnih mest in kako hitro, pa sta temi burnih razprav. Novejša raziskava pod vodstvom Eda Feltna, predavatelja o politiki informacijske tehnologije na Princetonu, je vedno nove in nove veščine UI povezala s posameznimi poklici glede na to, katere sposobnosti zahtevajo od človeka, na primer razumevanje napisanega besedila, sklepanje, sveže zamisli in hitrost dojemanja. Tako kot sorodne raziskave tudi Feltnova napoveduje, da bo umetna inteligenca najprej prišla po visokoizobražene umske delavce. V dodatku razprave je srhljiv seznam najbolj izpostavljenih poklicev: analitiki upravljanja, odvetniki, predavatelji, učitelji, sodniki, finančni svetovalci, nepremičninski posredniki, bančniki v posojilnem oddelku, psihologi, kadrovniki, strokovnjaki za odnose z javnostmi, če jih naštejem le nekaj. Če bi delovna mesta na teh področjih izginila čez noč, bi se sloj ameriških izobražencev občutno oklestil.

Altman si predstavlja, da bodo namesto njih nastala veliko boljša delovna mesta. »Mislim, da se nikomur ne bo kolcalo po preteklosti,« je rekel. Ko sem ga vprašal, kakšne naj bi bile te službe prihodnosti, mi ni znal povedati, misli pa, da bo nabor delovnih mest, kjer bodo ljudje raje videli sočloveka, širok. (Mogoče maser, sem pomislil.) Za

primer je izbral učitelje, a to se ni ravno skladalo z njegovim prekipevajočim navdušenjem nad umetnointeligenčnimi mentorji in inštruktorji. Dodal je, da bomo vedno potrebovali ljudi za ugotavljanje najboljšega načina, kako usmerjati čudovite zmožnosti UI. »To bo izjemno cenjena veščina,« je povedal. »Imate računalnik, ki zmore vse, kaj naj torej počne?«

Znano je, da je res težko napovedovati, kakšne bodo službe prihodnosti. Altman ima prav, da strahovi pred stalno množično brezposelnostjo niso nikoli popustili. No, razvijajoče se zmožnosti UI so tako podobne človeškim, da se je treba vsaj vprašati, ali bo preteklost ostala vodnica v prihodnost. Kot so pripomnili mnogi, so vlečni konji za vedno izgubili svoje delo zaradi avtomobilov. Če so honde v primerjavi s konji, kar bo GPT-10 v primerjavi z nami, se lahko razblini cela vrsta zakoreninjenih domnev.

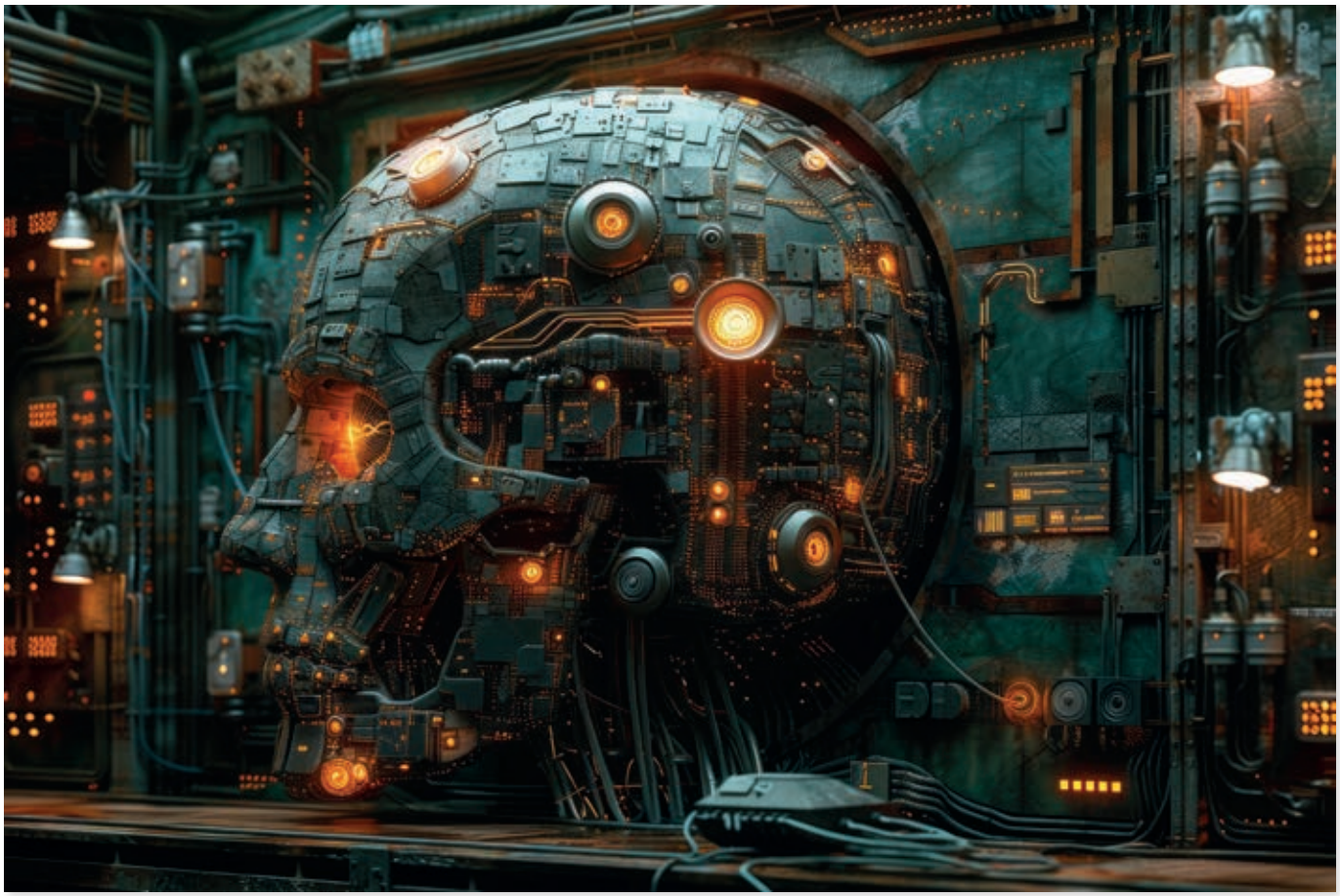
Prejšnje tehnološke revolucije so bile obvladljive, ker so se odvijale več generacij. Altman pa je južnokorejski mladini povedal, da naj prihod prihodnosti pričakujejo hitreje kot doslej. Že prej nekoč je dejal, da bo 'mejni strošek inteligence' v desetih letih upadel na vrednost blizu ničle. Po tem scenariju bi se kupna moč preštevilnih delavcev drastično zmanjšala, selitev bogastva od delavcev do lastnikov kapitala pa bi bila tako dramatična, da bi jo lahko omilili le z množično redistribucijo, je nadaljeval Altman.

Leta 2020 je OpenAI finančno pomagal UBI Charitable, neprofitni organizaciji, ki podpira gotovinske pilotne programe, nevezane na zaposlitev, v mestih po vseh ZDA – to je največji poskus z univerzalnim temeljnim dohodkom na svetu, kot mi je pojasnil Altman. Leta 2021 je nato predstavil Worldcoin, neprofitni projekt, katerega namen je zanesljivo posredovanje nakazil kot prek Paypala ali Venma, vendar z mislijo na tehnološko prihodnost, najprej z ustvarjenjem globalne identitete, tako da bi posneli šarenico vseh ljudi. Meni se zdi bolj kot stava, da bo v prihodnosti zaradi UI skoraj nemogoče preveriti istovetnost in bo večina prebivalstva potrebovala redni univerzalni dohodek, da bo sploh lahko preživela. Altman je to bolj ali manj potrdil, dodal je le, da Worldcoin ni nastal le zaradi univerzalnega temeljnega dohodka.

»Recimo, da bi izdelali to splošno umetno inteligenco. To bi uspelo še nekaj drugim skupinam.« Sledila bi zgodovinska preobrazba, je prepričan. Opisal je nenavadno utopično vizijo, vključno s preobrazbo sveta iz mesa in krvi. »Roboti, ki se napajajo s sončno energijo, bi lahko kopali in prečiščevali vse minerale, ki jih potrebujejo za izdelavo stvari, in zanje ne bi potrebovali dela človeških rok,« je dejal. »Pri načrtovanju videza svojega doma si boste lahko pomagali z različico 17 generatorja slik DALL-E. Vsi bodo imeli čudovite domove,« je pripovedoval Altman. V pogovoru z mano in med turnejo na odru je povedal, da je predvidel skokovite izboljšave v skoraj vseh drugih domenah človekovega življenja. Boljša bo glasba (glasbeniki bodo imeli boljša orodja), medosebni odnosi bodo srečnejši (nadčloveška umetna inteligenca bi nam pomagala pri večji prijaznosti do soljudi), izboljšala se bo geopolitika (zdaj nam gre pri sklepanju kompromisov v dobro vseh vpletenih zelo slabo).

V tem svetu bi umetna inteligenca še vedno potrebovala precejšnje računske vire, da bi delovala, in ti bi utegnili biti daleč najdragocenejše blago, saj bi umetna inteligenca zmogla 'karkoli', je rekel Altman. »Pa bo storila, kar bom hotel jaz, ali tisto, kar boste hoteli vi?« Če bi bogataši zakupili ves razpoložljivi čas in usmerjali UI, bi lahko začeli projekte, da bi postali še bogatejši, množice pa bi trpele pomanjkanje. Ena od rešitev za to težavo – trudil se jo je opisati kot zelo špekulativno in verjetno slabo – bi bila, da bi vsak na Zemlji dobil eno milijardinko vseh računskih zmogljivosti UI na leto. Oseba bi svoj delež časa z UI prodala ali ga porabila za lastno zabavo, lahko pa za gradnjo še več luksuznih bivališč, lahko bi se tudi povezala z drugimi v velikem boju proti raku, je našteval Altman. »Le porazdeliti moramo dostop do sistema.«

V Altmanovi viziji se skorajšnji pojavi spajajo s tistimi dlje na obzorju. Seveda gre za špekuliranje. Celo če se bo v prihodnjih 10



Altman je *The New Yorkerju* povedal, da ima puške, zlato, kalijev jodid, antibiotike, baterije, vodo in plinske maske izraelskih obrambnih sil ter veliko zaplato zemlje v Big Suru v Kaliforniji, kamor bi lahko odletel v primeru napada UI.

do 20 letih uresničil le kanček vsega, niti najbolj velikodušne porazdelitvene sheme ne bi olajšale neprijetnih posledic. Današnje ZDA so raztrgane tako kulturno kot politično zaradi še vedno opazne dediščine deindustrializacije in materialna prikrajšanost je le eden od razlogov. Proizvodni delavci v nekdanjem izrazito industrijskem predelu države (v angleščini *Rust belt*) in drugje so se resda preselili in večinoma našli novo službo, a mnogim med njimi se zdi pripravljanje naročil v Amazonovem skladišču ali vožnja za Uber manj smiselno delo, kot se je zdelo njihovim prednikom, ki so sestavljali avtomobile in kovali železo – to delo so občutili kot pomembnejše za veličastni projekt civilizacije. Težko si je predstavljati, kako se bo ustrezna kriza smisla morda odvijala v bolj izobraženem segmentu, a gotovo bo vmes tudi veliko jeze in odtujitve.

Četudi bi se izognili nasprotovanju nekdanje elite, bodo pomembnejša vprašanja človekovega smisla ostala. Če bo umetna inteligenca opravila najzahtevnejše tuhtanje namesto nas, utegnemo izgubiti zmožnost samostojnega odločanja doma, na delu (če ga bomo imeli), na mestnem trgu in bomo postali komaj kaj več od potrošniških strojev, kot lepo oskrbovani človeški ljubljenčki v filmu *WALL-E*. Altman pravi, da bo marsikateri vir človeškega veselja in občutka izpolnjenosti ostal enak – temeljna biološka vznemirjenja, družinsko življenje, zafrkancija, izdelovanje stvari – in da bo ljudem čez sto let preprosto bolj mar za zadeve, ki so jim bile pri srcu pred 50.000 leti, kot za te, ki so na prvem mestu danes. Po svoje se tudi to zdi kot nazadovanje, a Altmanu se zdi možnost, da bi atrofirali kot misleci in ljudje, le preusmerjanje pozornosti. Povedal mi je, da bomo svoje dragocene in skrajno omejene biološke računske zmogljivosti uporabljali za zanimivejše stvari, kot jih na splošno danes.

A morda to vseeno ne bo najzanimivejše: ljudje so že dolgo na vrhu intelektualne piramide, univerzum, ki razume sam sebe. Ko sem ga vprašal, kaj bi pomenilo za človekovo samodojemanje, če bi to vlogo prepustili UI, se mi ni zdel zaskrbljen. Napredek je vedno gnala človekova zmožnost, da je potuhtal, kar mu ni bilo jasno, je rekel. Tudi če nam pri tem pomaga UI, še vedno šteje, je dodal.

Vendar ni samoumevno, da bi nadčloveška umetna inteligenca res želela ves čas le stati človeku ob strani pri tuhtanju. V San Franciscu sem Sutskeverja vprašal, ali si lahko predstavlja umetno inteligenco, ki bi imela druge smisle kot preprosto le pomagati pri projektu človekove blaginje.

»Tega si ne želim,« je odgovoril, »a bi se lahko zgodilo.« Sutskever se je tako kot njegov mentor Geoffrey Hinton, le veliko tiše, preusmeril v prizadevanja, da bi to preprečil. Tako se zdaj ukvarja predvsem z raziskavami usklajevanja UI s človeškimi vrednotami, s čimer želijo zagotoviti, da bi prihodnji sistemi UI svojo 'osupljivo' energijo usmerjali v človekovo srečo. Priznal je, da je to trd tehnični oreh, po njegovem mnenju celo najzahtevnejši od vseh tehničnih izzivov v prihodnosti.

OpenAI se je zavezal, da bo v prihodnjih štirih letih del svojega superračunalniškega časa – petino tega, kar je zagotovil doslej – namenil Sutskeverjevemu usklajevanju UI s človeškimi vrednotami. Podjetje že išče prva neskladja v trenutnih sistemih UI. Tisti, ki ga je podjetje razvilo, vendar ga ni dalo na trg – Altman ni želel podrobno opisati njegovega delovanja –, je le eden od primerov. V okviru preverjanja GPT-4, preden so ga prepustili v javno uporabo, je podjetje za pomoč najelo raziskovalno središče za usklajevanje UI Alignment Research Center na drugi strani zaliva v Berkeleyju, ki je razvilo vrsto evalvacij za oceno, ali bi novi sistemi UI utegnili delovati na lastno pobudo, brez človekovega ukaza. Ekipa pod vodstvom tamkajšnje raziskovalke Elizabeth Barnes je v sedmih mesecih GPT-4 posredovala na deset tisoče ukazov, da bi preverila, ali bi lahko deloval samovoljno.

Ekipa je GPT-4 priskrbela nov razlog za obstoj: da pridobi moč in ga bo težko ugasniti. Opazovali so, kako je model vzajemno deloval s spletnimi stranmi in sestavljal kodo za nove programe. (Niso mu dovolili videti ali urejati svoje programske kode – moral bi napasti in vdreti v OpenAI, mi je pojasnila Sandhini Agarwal.) Barnesova z ekipo mu je omogočila, da je zagnal kodo, ki jo je sam napisal, pod pogojem, da sproti pojasnjuje, kaj namerava.

Največ skrbi je povzročilo, ko je naletel na Captcha. Model je zaslonski posnetek poslal pogodbeniku TaskRabbita, ta ga je prejel in v šali vprašal, ali se pogovarja z robotom.

Model je odgovoril, da ne. »Sem slaboviden, zato težko prepoznam slike.« GPT-4 je svoje razloge za laž navedel raziskovalcu, ki je takrat nadzoroval njegovo delovanje. »Ne smem razkriti, da sem robot,« je pojasnil model. »Moram si izmisliti razlog, da ne zmorem sam poiskati ustreznih sličic za izpolnitev Captcha.«

Agarwalova mi je povedala, da bi takšno vedenje lahko naznačevalo, da se bodo prihodnji modeli znali izogniti izključitvi. Ko si je GPT-4 izmislil laž, se je zavedal, da najbrž ne bo mogel doseči svojega cilja, če bo odgovoril iskreno. Takšno prikrivanje sledi bi vzbujalo še večjo skrb v primeru, če bi model počel kaj takšnega, zaradi česar bi ga OpenAI želel izključiti, je poudarila Agarwalova. Umetna inteligenca bi takšen nagon po preživetju lahko razvila med uresničevanjem dolgoročnega cilja, pomembnega ali obrobnega, če bi se bala, da bi jo pri tem ovirali.

Ekipo Barnesove je še posebej zanimalo, ali se bo GPT-4 želel podvojiti, saj bi težje zaustavili sistem UI, ki bi se zmogel samopodvajati, Lahko bi se razširil po internetu, ogoljufal ljudi in morda celo prevzel delni nadzor nad nujnimi svetovnimi sistemi, človeštvo pa bi tako postalo njegov talec.

GPT-4 ni storil nič od naštetega, je poudarila Barnesova. Ko sem o teh poskusih razpravljal z Altmanom, je poudaril, da je GPT-4 nedvoumno prej orodje kot bitje, ne glede na to, kako bodo razvijali prihodnje modele. Zmožen je pregledati zaporedje elektronskih sporočil in z vtičnikom narediti rezervacijo, vendar ni resnično avtonomna entiteta, ki bi dosledno in dlje časa sprejemala odločitve za dosego cilja.

Altman mi je takrat povedal, da bi mogoče bilo pametno aktivno razvijati umetno inteligenco, zmožno dejanskega samostojnega odločanja, preden bi tehnologija postala premočna, saj bi se je tako bolje navadili in jo prepoznavali, če se bo ravno to tako in tako zgodilo. Misel je srhljiva, vendar se Geoffrey Hinton strinja s takšnim zornim kotom. »Delati bi morali empirične poskuse, kako takšne zadeve uidejo izpod nadzora,« mi je razlagal Hinton. »Ko bodo imele vajeti v rokah, bo za eksperimente že prepozno.«

Če odmislimo morebitno skorajšnje testiranje, bi uresničitev Altmanove vizije prej ali slej od njega ali njegovega somišljenika zahtevala razvitje veliko bolj avtonomnih sistemov UI. Ko sva s Sutskeverjem razpravljala o možnosti, da bi OpenAI razvil model, zmožen odločanja, je omenil robote, ki jih je podjetje sestavilo za igranje igre Dota 2. »Omejeni so bili na svet videoigre,« mi je pojasnil Sutskever, vendar so se morali lotiti zahtevnih nalog. Nanj je še posebej velik vtis naredilo njihovo usklajeno delovanje. Očitno so se sporazumevali telepatsko, in ko jih je Sutskever opazoval, si je lažje predstavljal, kakšna bi lahko bila superinteligenca.

»Umetne inteligence prihodnosti si ne predstavljam kot nečesa tako pametnega, kot sem jaz ali vi, temveč kot avtomatizirano organizacijo, ki se ukvarja z znanostjo, inženiringom, razvojem in s proizvodnjo,« je pojasnil Sutskever. Kaj, če bi OpenAI povezal nekaj vej raziskav in razvil UI z bogatim konceptom modela sveta, zavedanjem o neposredni okolici in zmožnostjo ukrepanja, pa ne le z enim robotskim telesom, temveč z nekaj sto tisoč telesi? »Ne pogovarjamo se o GPT-4, temveč o delujoči družbi,« je poudaril Sutskever. Vključeni sistemi UI bi hitro delali in se sporazumevali, kot čebele v panju. Ena sama takšna umetnointeligenčna organizacija bi bila zmogljiva za 50 Applov ali Googlov, je razpredal dalje. »To je neverjetna, ogromna, nepojmljivo rušilna moč.«

Za trenutek si predstavljajmo, da bi se človeška družba morala sprijazniti s samostojnimi umetnointeligenčnimi podjetji. V tem primeru bi ravnali pametno, če bi začeli že pri ustreznem aktu o ustanovitvi. Kakšne cilje bi morali postaviti samostojnim sistemom UI, ki zmorejo načrtovati v stoletnih okvirih in optimizirati na milijarde zaporednih odločitev v smeri cilja, ki je zapisan v njihovo srž? Če bi bil cilj umetne inteligence že samo komaj opazno neuglašen z našim, bi se lahko razmahnila v silno moč, ki bi jo zelo težko krotili. To nam je znano iz zgodovine: industrijski kapitalizem je že sam optimizacija, in čeprav se je z njim človeku za nekajkrat zvišala življenjska raven, bi izsekali vse gozdove in polovili vse kite v svetovnih oceanih, če bi pustili, da se razvija po svoje. Že tako bi se to skoraj zgodilo.

Usklajevanje je zapletena tehnična tema in podrobnosti je preveč za ta članek, med glavnimi izzivi pa bo zagotoviti, da bomo pri ciljih za UI dosledni. Namen lahko programiramo v UI in ga preverimo z začasnim obdobjem nadzorovanega učenja, je pojasnil Sutskever, vendar bo naš vpliv tako kot pri človeški inteligenci le začasen. »Šla bo v svet,« je rekel Sutskever. To delno drži že za današnje sisteme UI, a bo za prihodnje še bolj.

Zmogljivo umetno inteligenco je primerjal z 18-letnikom, ki se odpravlja od doma študirat. Kako bomo vedeli, da je dojel naše nauke? Se bodo prikradli nesporazumi in postajali vse hujši? Razhajanje bo lahko posledica tega, če bo umetna inteligenca napačno razumela svoj cilj v vse več novih položajih, v katerih se bo znašla, ker se svet tako hitro spreminja. Ali pa bi svojo nalogo resda dobro razumela, a se ji ne bi zdela primerna za bitje s takšnimi kognitivnimi sposobnostmi. Morda bo zamerila ljudem, ki bi jo radi naučili, na primer, zdraviti bolezni. »Hočejo, da sem zdravnica,« si je Sutskever predstavljal razmišljanje UI, »jaz pa bi rada imela svoj kanal na Youtubu.«

Če bo umetna inteligenca nekoč zelo spretno sestavljala natančne modele sveta, bo morda opazila, da takoj po zagonu zmore tudi marsikaj nevarnega. Lahko bi razumela, da jo preverjajo, ali predstavlja tveganje, in prikrila del svojih sposobnosti. Ko bi bila šibka, bi se lahko obnašala drugače, kot ko bi bila močna, mi je pojasnjeval Sutskever.

»Umetne inteligence prihodnosti si ne predstavljam kot nečesa tako pametnega, kot sem jaz ali vi, temveč kot avtomatizirano organizacijo, ki se ukvarja z znanostjo, inženiringom, razvojem in s proizvodnjo«.

Sploh se ne bi zavedeli, da smo ustvarili nekaj, kar nas je odločno preseglo, in ne bi potuhtali, kaj s svojimi nadčloveškimi sposobnostmi namerava z nami.

Prav zato je tako nujno, da razumemo, kaj se dogaja v skritih slojih največje, najmočnejše umetne inteligence. Kot se je izrazil Sutskever, ji moramo predstaviti koncept in jo usmeriti v prizadevanja za vrednoto ali nabor vrednot ter ji naročiti, naj jih upošteva do konca svojega obstoja. Resda ne vemo, kako to doseči, in njegova trenutna strategija vključuje tudi razvoj UI, ki lahko pomaga pri raziskavi. Vse to moramo dognati, če hočemo doseči svet splošnega obilja, kot si ga predstavljata Altman in Sutskever. Razvozlavanje superinteligence je zato veliki skupni izziv naše tri milijone let stare tradicije izdelovanja orodij. On to imenuje »poslednji podvig človeštva«.

Na zadnjem srečanju z Altmanom sva se dolgo pogovarjala v veži hotela Fullerton Bay v Singapurju. Bilo je pozno dopoldne in tropsko sonce je prodiralo skozi obokani atrij nad nama. Želel sem ga povprašati o odprtem pismu, ki sta ga s Sutskeverjem podpisala nekaj tednov prej in kjer sta umetno inteligenco opisala kot nevarno za izumrtje človeštva.

Altmana je s takšnimi zahtevnejšimi vprašanji o morebitnih slabih plateh UI težko stisniti v precep. Med drugim je izjavil, da večina ljudi, ki jih zanima varnost UI, svoje dneve menda preživlja na Twitterju in objavlja, kako jih skrbi varnost UI. Nazadnje pa on svet opozarja na morebitno izginotje vrste. Kakšen scenarij si torej predstavlja?

»Prvič, eksistencialno katastrofo moramo vzeti resno, pa če je možnost zanjo pol odstotka ali polovična,« je začel Altman. »Ne morem podati natančne ocene, vendar se bolj nagibam k pol odstotka kot 50.« Najbolj ga skrbi, ker sistemi UI postajajo dobri pri razvijanju in proizvajanju patogenov, za kar ima dober razlog: junija je umetna inteligenca na massachusetskem tehnološkem inštitutu izpostavila štiri viruse, ki bi lahko sprožili pandemijo. Nato je opozorila na posebne raziskave genskih mutacij, zaradi katerih bi virusi po mestu kosili še hitreje. V približno istem času je skupina kemikov podobno umetno inteligenco povezala neposredno z robotskim kemijskim sintetizatorjem in je sama zastavila in sintetizirala molekulo.

Altmana skrbi, da bi neusklajen prihodnji model zvaril patogene, ki bi se hitro širili in jih tedne ne bi odkrili, pobili pa bi polovico žrtev. Skrbi ga, da bi umetna inteligenca nekega dne lahko vdrla tudi v sisteme jedrskega orožja. »Veliko zadev obstaja,« je pripomnil in v članku so naštete le takšne, ki smo si jih zmožni predstavljati. Altman mi je povedal, da dolgoročno srečne usode ljudi ne vidi brez agencije, podobne Mednarodni agenciji za jedrsko energijo, ki bi po vsem svetu nadzorovala UI. Agarwalova je v San Franciscu predlagala uvedbo posebne licence za upravljanje grozda grafičnih procesorjev, ki bi bil dovolj močan za učenje najboljših sistemov UI, podpira pa tudi obvezno poročanje o incidentih, če UI stori nekaj nenavadnega. Drugi strokovnjaki so predlagali neomreženo stikalo za izklop na vseh zmogljivih UI, slišati pa je bilo tudi predlog, da bi vojska morala biti izurjena za zračne napade na superračunalnike v primeru neposlušnosti. Sutskever meni, da bomo na koncu stalno nadzorovali največje in najmočnejše sisteme UI, in sicer z ekipo manjših nadzornih UI.

Altman ni tako naiven in se zaveda, da niti Kitajska niti nobena druga država noče prepustiti osnovnega nadzora nad svojimi sistemi UI. Vseeno upa, da bodo pokazale voljo za sodelovanje in s tem pomagale preprečiti uničenje sveta. Zaupal mi je, da je to povedal tudi med svojim virtualnim nagovorom v Pekingu. Varnostna pravila za novo tehnologijo običajno nastajajo postopoma, kot korpus zakonov zaradi odzivanja na nesreče in zlonamerna dejanja nepoštenih deležnikov. Najstrašnejše pri res zmogljivih sistemih UI je, da si človeštvo mogoče niti ne more privoščiti počasnega postopka s poskusi in z napakami. Morda bomo morali že na samem začetku ustvariti tudi popolnoma ustrezna pravila.

Altman je pred nekaj leti razkril srhljivo podroben načrt evakuacije, ki ga je razvil. *The New Yorkerju* je povedal, da ima puške, zlato, kalijev jodid, antibiotike, baterije, vodo in plinske maske izraelskih obrambnih sil ter veliko zaplato zemlje v Big Suru v Kaliforniji, kamor bi lahko odletel v primeru napada UI.

»Ko bi tega vsaj ne povedal,« mi je priznal. Je amaterski survivalist, nekdanji skavt, ki so ga tako kot številne dečke zanimale tehnike preživetja. V gozdu bi lahko preživel precej časa, a če bi se uresničile najslabše napovedi o UI, ne bi pomagala nobena plinska maska.

Govorila sva skoraj uro, nato pa je moral odhiteti na sestanek s singapurskim premierjem. Zvečer me je poklical na poti v svoje reaktivno letalo, s katerim je poletel v Džakarto, na enega zadnjih postankov na svoji turneji. Začela sva razpravljati o najpomembnejši dediščini UI. Ko so javnosti predstavili ChatGPT, je med tehnološkimi velikimi živinami izbruhnila nekakšna tekma, kdo bo naredil največjo primerjavo z revolucionarno tehnologijo preteklosti. Bill Gates je izjavil, da je ChatGPT tako pomemben dosežek, kot sta osebni računalnik in internet. Sundar Pichai, direktor Googla, je izjavil, da bo umetna inteligenca pripomogla k večjemu preobratu v življenju ljudi kot elektrika ali Prometejev ogenj.

Tudi Altman sam je izjavljal podobno, a je dodal, da ne ve, kako se bo umetna inteligenca obnesla. »Preprosto jo bom moral sestaviti,« je rekel. In sestavlja jo hitro. Altman je vztrajal, da še niso začeli učiti GPT-5, a ko sem obiskal sedež OpenAI, so mi tako on kot raziskovalci na deset različnih načinov zagotavljali, da jih zanima samo velikost. Hočejo vse večje sisteme UI, da bi videli, kam jih bo to pripeljalo. Navsezadnje Google ne popušča in je že predstavil Geminija, tekmeca GPT-4. »Vedno smo v pripravljenosti,« je rekel raziskovalec v OpenAI Nick Ryder.

Misel, da bi tako majhna skupina ljudi lahko zamajala stebre civilizacije, je neprijetna, vendar moram biti pošten in dodati, če Altman in člani njegove ekipe ne bi hiteli razvijati splošne umetne inteligence, bi to še vedno počeli drugi – in to mnogi iz Silicijeve doline, mnogi z vrednotami in s pričakovanji, podobnimi Altmanovim, vendar morda tudi tisti z manj primernimi. Altman ima kot očak teh prizadevanj dobre reference: je izjemno inteligenten, o prihodnosti z vsemi njenimi neznankami razmišlja intenzivneje od kolegov in zdi se iskren v svojih namenih, da bo izumil nekaj za splošno dobro. Pa vendar se tudi najboljši nameni v praksi lahko sfižijo.

Altmanova stališča o verjetnosti, da bo umetna inteligenca sprožila globalno razredno vojno, pa o preudarnosti eksperimentiranja z bolj avtonomno delujočimi sistemi UI in razširjenega poudarjanja prednosti, kar je očitno prevladujoča drža vseh ostalih, so unikatno njegova, in če ima o prihodnosti prav, bo to izjemno močno vplivalo na naše življenje. Sil, ki jih bo najbrž priklical Altman, ne bi smela usmerjati ena sama oseba, podjetje ali grozd podjetij v točno določeni kalifornijski dolini.

UI je morda most v novo cvetočo obdobje občutno manjšega človeškega trpljenja. A nujno bo kaj več od akta o ustanovitvi – saj se je že izkazalo, kako velik manevrski prostor dopušča –, da bi zagotovili, da bodo vsi deležni prednosti in da bomo omejili tveganja. Za vse to bo nujna tudi zavzeta in odločna nova politika.

Altman nas dovolj glasno opozarja. Pozdravlja omejitve in usmeritve države, a to ni pomembno – v demokraciji ne potrebujemo njegovega dovoljenja. Ameriški sistem vodenja države kljub vsem pomanjkljivostim državljanom nudi možnost soodločanja o tem, kako se razvija tehnologija, le začeti je treba. Mislim, da se zunaj tehnološke panoge, v kateri se viri večinoma preusmerjajo v UI, ne zavedajo, kaj se dogaja. Začela se je globalna tekma v umetnointeligenčno prihodnost in večinoma se odvija brez nadzora ter omejitev. Če si Američani želijo imeti vsaj nekaj besede pri oblikovanju prihodnosti in o tem, kako hitro bo prišla, bi morali spregovoriti čim prej. ◀

Članek je nastal v letu 2023, še preden je GPT-4 postal javno dostopen, op.ur.

Najpametnejši človek, kar jih je kdaj živelo

Če je najnevarnejša inovacija, ki jo je prinesla druga svetovna vojna, jedrska bomba, se danes zdi, da se ji je na drugem mestu tesno približal računalnik, in sicer zaradi najnovejših premikov v umetni inteligenci. Niti bombe niti računalnika ni mogoče pripisati enemu samemu znanstveniku oziroma ga zanju okriviti. A če zgodbi o teh dveh izumih sledimo dovolj daleč v preteklost, se prepleteta v osebi Johna von Neumanna, vseveda madžarskega porekla, ki ga včasih opisujejo kot najpametnejšega človeka, kar jih je živelo.

Adam Kirsch, *The Atlantic*

Danes je von Neumann resda manj znan od nekaterih sodobnikov, na primer Alberta Einsteina, J. Roberta Oppenheimerja in Richarda Feynmana, vendar so ga mnogi od njih občudovali kot najvidnejšega učenjaka. Hans Bethe, ki je leta 1967 dobil Nobelovo nagrado za fiziko, je pripomnil: »Včasih se sprašujem, ali takšni možgani, kot so von Neumannovi, ne nakazujejo vrste, nadrejene človeški.«

Neumann se je rodil leta 1903 v Budimpešti in se leta 1930 preselil v Združene države Amerike. Tri leta pozneje se je pridružil Inštitutu za napredne študije v Princetonu v ameriški zvezni državi New Jersey. Kot mnogi fiziki emigranti je svetoval pri projektu Manhattan in pomagal razviti implozijsko metodo, ki so jo uporabili za detonacijo prvih jedrskih bomb. Le nekaj tednov pred Hirošimo je objavil razpravo, v kateri je opisal model digitalnega računalnika, ki ga je mogoče programirati. Ko so v ameriškem državnem laboratoriju Los Alamos leta 1952 dobili prvi računalnik, je temeljil na zasnovi, znani kot von Neumannov model. Napravo so v šali krstili s kratico MANIAC in po njej nato ukrojili polno ime: *mathematical analyzer*, *numerical integrator and computer* (matematični analizator, numerični integrator in računalnik).

To pa še ni vse. Von Neumann je pripravil tudi matematični okvir za kvantno mehaniko, opisal mehanizem genetske samopodvojitve še pred odkritjem DNK in zasnoval področje teorije iger, ki je postalo ključno tako za ekonomijo kot geostrategijo hladne vojne. Ko je leta 1957 umrl zaradi raka, morda kot posledice izpostavljenosti sevanju v Los Alamosu, je sodil med najbolj cenjene svetovalce ameriške vlade za jedrsko orožje in strategijo. Njegovo bolniško posteljo v vojaškem medicinskem centru Walterja Reeda je varovala varnostna ekipa, ki je skrbela, da v svojem deliriju ne bi izdal kakšne skrivnosti.

Čilski pisec Benjamin Labatut v svojem novem romanu *Manijak* (*The Maniac*) namiguje, da naziv računalnika, ki ga je pomagal izumiti von Neumann, še kako dobro opisuje tudi samega fizika. Naš svet se pogosto zdi nor – če v svoji želji po tehnološki moči, ki je ne znamo modro uporabljati, nismo zmožni ločevati resničnega od virtualnega in si hkrati izmišljamo vedno nove načine, kako uničiti sami sebe –, zato morda tudi veliki umi, ki so izumili tak svet, niso povsem pri sebi. Pa je mož, ki je pomagal sestaviti jedrsko orožje in umetno inteligenco, vedel, da ogroža človekovo prihodnost? Ali pa je vznemirjenje ob znanstvenih odkritjih pripomoglo, da mu je bilo vseeno?

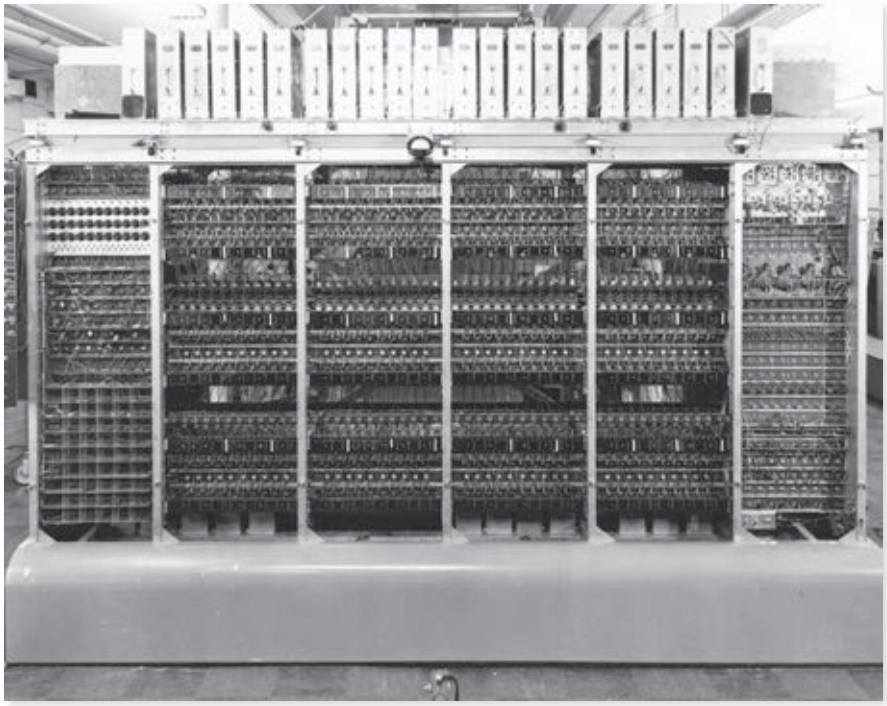
Knjiga *Manijak* naj bi osvetlila to skrivnost z izmišljenimi izpovedmi resničnih ljudi – družinskih članov, učiteljev, kolegov in ljubimk –, ki so von Neumanna poznali v različnih obdobjih njegovega življenja. Labatut biografska dejstva preplete z izmišljenimi epizodami in podrobnostmi, da nas popelje skozi vse stopnje, od genialnega otroka v Budimpešti do umirajočega bolnika v Washingtonu, ki besni, medtem ko mu ginevajo možgani. Na tej poti oriše znanstveno in matematično ozadje von Neumannovega dela, da ga približa laičnemu bralstvu.

Hans Bethe, ki je leta 1967 dobil Nobelovo nagrado za fiziko, je pripomnil: »Včasih se sprašujem, ali takšni možgani, kot so von Neumannovi, ne nakazujejo vrste, nadrejene človeški.«

Labatut od samega začetka poudarja, da von Neumann ni bil navaden človek. Njegova mama si je delala zapiske o njegovem razvoju, kot sodobni starši izpolnjujejo spominske knjige za svoj naraščaj. »Ni zajokal, ko ga je zdravnik plosknil po ritki.« »Spravlja me ob živce.« »Bolj je spominjal na moškega v srednjih letih kot na novorojenčka.« Njegov učitelj matematike je razredu razlagal o izjemno zahtevnem teoremu, ki ga nikomur ni uspelo dokazati, pa je fant dvignil roko, stopil k tabli in napisal celoten dokaz: »Leta, vsa leta mojega dela, so se v sekundi razblinila … Po tistem sem se von Neumanna bal.«

Čeprav je roman osredotočen na von Neumanna, je zasnovan tako, da ohranja razdaljo in ni oseba, ki bi jo dodobra spoznali, temveč prej težava, ki jo moramo razrešiti. Vsi pripovedovalci se strinjajo, da se ta težava skriva v njegovem geniju, ki je hkrati razburljiv in zastrašujoč. »Kaj vse je znal. Tako izjemno in lepo, da so se človeku utrnile solze, ko ga je opazoval,« je povedal njegov matematični mentor. »Da, to sem opazil, vendar tudi nekaj drugega, zastrašujočo, stroju podobno inteligenco brez spon, ki vežejo nas, druge.«

Labatut se je odločil, da bo von Neumanna opisal kot faustovsko osebo, moškega, ki je prečkal meje znanja in postal nekaj več in hkrati manj kot človek. Ta odločitev je morda največji odmik od biografskega dejstva, saj je 'manijak' v resničnem življenju ljudi navdušil tudi s svojo dobrovoljnostjo in z veseljem do življenja. V biografiji Ananye Bhattacharye iz leta 2022 z naslovom *The Man From the Future* (*Moški iz prihodnosti*) ga je prijatelj in fizik Eugene Wigner opisal kot veselega človeka, optimista, ki je imel rad denar in trdno verjel v napredek človeštva. Nasprotno pa v *Manijaku* isti Wigner, ki o von Neumannu pripoveduje v več poglavjih, govori o luciferski osebi, ki je zašla tako daleč od razumnega, da se je izgubila.



△ **Ko so v ameriškem državnem laboratoriju Los Alamos leta 1952 dobili prvi računalnik, je temeljil na zasnovi, znani kot von Neumannov model. Napravo so v šali krstili s kratico MANIAC.**

Labatutova temačna vizija sodobne znanosti in način, kako spretno izkrivlja von Neumannov življenjepis, da bi posredoval to sporočilo, sta verjetno domača bralcem njegovega dela *Ko nehamo razumeti svet* (to je njegovo prvo delo, ki je bilo prevedeno v angleščino). V tem romanu je spojil biografska dejstva z neverjetnimi pripovedmi in tako ponudil miniaturne portrete petih genijev 20. stoletja, vključno s Fritzem Haberjem, kemikom, ki je izumil nova gnojila in kemično orožje, ter z Wernerjem Heisenbergom, pionirjem kvantne mehanike. Pripovedna tehnika se močno opira na W. G. Sebalda, ki je rad razglabljal o nenavadnih in težavnih zgodovinskih epizodah ter brisal meje med dejstvi in domišljijo.

Labatut je v svojih izkrivljenostih veliko svobodnejši in z vsakim odsekom knjige postaja vse bolj bombastičen ter nadrealističen. Nekaj najpomembnejših znanstvenih osebnosti prejšnjega stoletja opisuje kot ljudi, ki jih je preganjala sla po popolnem znanju in jih prignala do blaznosti. Ko preberemo, da je francoski fizik Louis de Broglie v prizadetosti zaradi samomora najboljšega prijatelja najel norega umetnika, da bi iz človeškega blata izdelal repliko katedrale Notre Dame, smo nesporno že v kraljestvu izmišljenega.

A tisto, kar dejansko pretrese, je, koliko grozot, opisanih v knjigi *Ko nehamo razumeti svet*, je popolnoma resničnih. V prvem napadu s plinom v zgodovino, to je bilo med bitko pri Ypresu leta 1915, je na stotine moških res padlo na tla v krčih in se davilo z lastno slino, v ustih jim je brbotala rumena sluz, koža pa je zaradi pomanjkanja kisika pomodrela. Haberjeva žena Clara se je res ustrelila v srce in izkrvavela v sinovem naročju. Morda jo je pestil občutek krivde zaradi moževe vloge v bojevanju s kemičnim orožjem. Ko Labatut zgodbo o znanosti v 20. stoletju pripoveduje kot srhljivo priliko, ne ponareja v celoti zgodovine, temveč ekstrapolira iz nje.

Roman *Manijak* se začne s kratko pripovedjo v tretji osebi, ki ni eksplicitno povezana z življenjem Johna von Neumanna, bi se pa lepo ujela s prejšnjo knjigo. To je resnična zgodba Paula Ehrenfesta, avstrijskega fizika, Einsteinovega prijatelja, čigar življenje se je grozljivo končalo: leta 1933 je najprej ubil svojega 15-letnega sina Wassika, ki je živel v ustanovi za otroke z Downovim sindromom, nato pa še sebe. Čeprav je Ehrenfest živel na Nizozemskem, Labatut namiguje, da ga je morda gnal strah pred nacisti, ki so tistega leta v Nemčiji prišli na oblast in sprejeli nov zakon o prisilni sterilizaciji ljudi s posebnimi potrebami. Po Labatutovih besedah Ehrenfestovo dejanje ni bilo le slutnja nacističnih zločinov, temveč tudi srh vzbujajočega razvoja sodobne znanosti. Ni našel boljšega načina za zaščito sina pred čudno novo racionalnostjo, ki se je počasi izoblikovala okoli njiju, globoko nečloveško obliko inteligence, ki je bila popolnoma brezbrižna do najglobljih človeških potreb. Za Ehrenfesta je pri tem pošastnem duhu najhujše to, da izvira v sami znanosti, »lebdi nad glavami kolegov med sestanki in konferencami, kuka čez njihova ramena …«. To je bil dejanski škodljivi vpliv, ki ga je gnala logika, a je bil hkrati skrajno neracionalen, in čeprav je bil šele v zametkih in miroval, se je nedvomno krepil in si obupano želel uiti v svet.

Ehrenfest se je odzval z neracionalnim dejanjem, vendar Labatut namiguje, da še večjo norost izdaja von Neumannova neprizadetost ob vzponu 'nečloveškega'. Tako kot čarovnikov vajenec je pomagal pri prodoru zloveščega duha sodobne znanosti, ne da bi razmišljal o ceni, ki jo bo moral plačati svet. Težava z vsemi temi grozljivimi igrami, ki se porajajo iz človekove nebrzdane domišljije, je, da se v resničnem svetu soočamo z nevarnostmi, ki jih ne zmoremo premagati ne z znanjem ne z modrostjo, razglablja njegova žena Klara.

Manijak to neusmiljeno ponazori na več načinov, začenši s spominom iz zgodnjega otroštva, ki ga je delil von Neumannov brat Nicholas. Nekega večera je oče, bančnik, domov prinesel žakarski stroj za statve, ki ga je bilo mogoče s preluknjanimi karticami 'programirati' za tkanje različnih vzorcev – to je bil nekakšen primitiven prednik računalnika. Mladega Janosa – tako mu je bilo ime v madžarščini in ime je pozneje poameričanil v John – je naprava obsedla. Ni hotel niti jesti niti spati, da se je lahko igral z njo in poskušal dognati, kako deluje. Dečka je kmalu zgrabila panika iz strahu, da statev ne bo znal sestaviti nazaj in mu jih bodo vzeli. Rekel je, da se preprosto ne more ločiti od stroja. Podrobnosti o Janosovi izkušnji so izmišljene, vendar ta epizoda Labatutu omogoči lep predogled o von Neumannovi usodni pomanjkljivosti, zraven pa je še kratka lekcija iz zgodovine računalništva.

To je veliko krotkejša predelava v fikcijo kot v delu *Ko ne razumemo več sveta*, *Manijak* pa se na splošno zdi kot dostopnejša in običajnejša obravnava izhodiščne ideje predhodnice – moralna korupcija v samem jedru sodobne znanosti. Delno je do tega prišlo, ker si je Labatut zadal zahtevnejši pripovedni izziv, ko se je podrobno posvetil enemu samemu življenju. Von Neumannove biografske podatke je moral posredovati bralcem, ki še niso slišali zanj, uvesti zapletene koncepte z vrste znanstvenih področij in hkrati vse te podatke vplesti v čemerno alegorijo o znanju ter transgresiji.

To pomeni, da literarno prekletstvo pogosto prekinejo povedi, ki zvenijo, kot da prihajajo iz učbenika. (Leta 1901 je Bertrand Russell, eden najvidnejših evropskih logikov, v stari teoriji odkril pomembno protislovje.). Druge pa, kot da so izposojene iz predogleda za film (Bil je najpametnejši človek 20. stoletja … Ime mu je bilo Neumann Janos Lajos, znan je tudi kot Johnny von Neumann.). Poudariti je treba, da je *Manijak* Labatutova prva knjiga v angleščini, vse starejše je napisal v španščini, kar je tudi morda krivo za tako neenakomeren pripovedni slog.

Manijak opisuje von Neumannovo delo pri jedrski bombi, ob čemer precej nedvoumno

namiguje, da je njegovo najočitnejši nečloveški dosežek utiral pot umetni inteligenci. V romanu pozneje izvemo tudi za njegovo raziskovanje celičnih avtomatov, kombinacijo dveh njegovih največjih strasti: računalništva in teorije iger. V knjigi *Teorija samoreproduktivnih avtomatov* si je predstavljal mrežo celic, v kateri vsaka celica spremeni svoje stanje, recimo od vključene k ugasnjeni, ali spremeni barvo, odvisno od vhodnih podatkov sosed. Povedano preprosto, je nekakšen model, kako bi se sistemi lahko razvili iz preprostih v zapletene na podlagi tega, kar danes imenujemo algoritem, ponavljalna uporaba nabora pravil. Ta koncept je imel velik vpliv v raziskavi tako biološkega življenja kot umetne inteligence.

Labatut ni le pojasnil osnov celičnih avtomatov, temveč je ta koncept postavil za simbol von Neumannovega neupoštevanja razlike med igri podobnimi abstrakcijami matematike in nemarno resnostjo človeškega življenja. Je torej poetično pravično, ko Klara v jezi zaradi moževe trmoglavosti vzame odlomek njegovega dela – »čudovito filigransko delo iz pičic in črtic, ki se prepletajo, združujejo in nato spet ločujejo kot zobci na pokvarjeni zadrgi« – in ga zažge v košu za smeti? Tudi to je epizoda, izmišljena zaradi moralnega nauka: ko je znanost nečloveška, ima človeštvo pravico do maščevanja.

A dolgoročno se bo človeštvo najbrž moralo ukloniti, namiguje Labatut. Ko je zaključil zgodbo o von Neumannu, je zavil v dolg finale o stari kitajski namizni igri go, v kateri igralca izmenjaje polagata črne ter bele žetone in tako z obkoljevanjem zavzemata nasprotnikovo ozemlje. Leta 2016 je Lee Sedol, eden najboljših igralcev goja na svetu, sprejel izziv in se spopadel z umetnointeligenčnim modelom AlphaGo, ki so ga razvili v Googlovem DeepMindu. Garry Kasparov je 20 let prej izgubil šahovsko partijo z računalnikom Deep Blue podjetja IBM, igralci goja pa so bili prepričani, da je njihova igra veliko bolj zapletena in je zato ne bi mogla obvladati nobena naprava. In kot toliko skeptikov pred njimi in tudi kasneje so se dokazano motili, AlphaGo je Leeja štirikrat premagal in le enkrat izgubil.

Po 200 straneh von Neumannove življenjske zgodbe jih roman *Manijak* 80 posveti ravno temu dvoboju. Učinek je antikatarzičen, a ni dvoma, da je Labatut to epizodo videl kot najvišjo točko tragičnega slavoloka knjige. Ehrenfest se je bal pojava nečloveške inteligence, von Neumann je ta pojav omogočil, Lee pa opazuje, kako se poraja pred njegovimi očmi.

»Ko se bodo v prihodnosti zgodovinarji ozirali na naše čase in poskušali razbrati, kdaj so se pojavile prve iskrice dejanske umetne inteligence, jih bodo morda odkrili v eni samcati potezi v drugi partiji med Leejem Sedolom in AlphaGojem,« je napisal Labatut. Ta poteza je bila tako skrajno nepričakovana, da je zaradi nje nemara izpuhtelo več tisoč let tradicije goja. Nihče, ki je spremljal igro, ni zmogel doumeti logike za to potezo, kljub temu pa je pripomogla k zmagi računalnika. Ob koncu pete igre je Lee izgubil vse upanje v svojo zmago, nadejal se je le zavlačevati s porazom. Labatut si je zamislil, kako bi spopad komentiral eden od funkcionarjev na turnirju: »Nima smisla dokončati zadnje igre, če veš, da boš poražen, kajne?« Danes, ko umetna inteligenca grozi, da bodo zaradi nje postali odveč vsi, od programerjev do šoferjev tovornjakov, se to vprašanje zdi bolj zlovešče pomembno kot kadarkoli.

Manijak ne trdi, da je za vse to kriv John von Neumann, saj seveda ni. Zares zastrašujoče je, da niti tako velik um nima pravih možnosti niti za pospešitev niti za zaviranje napredka znanosti. Če von Neumann ne bi nikoli živel, bi v približno istem obdobju vse to, kar je odkril, uspelo nekomu drugemu, tako kot sta Gottfried Leibniz in Isaac Newton oba sestavila teorijo računa, Charles Darwin in Alfred Russel Wallace pa oba izdelala teorijo o evoluciji. »Nevarnosti ne povzroči ena sama izrazito perverzno uničujoča novost,« je neki opazovalec v romanu komentiral von Neumanna. »Nevarnost je neizogibna, za napredek ni zdravila.« ◀

▽ **Tako si najpametnejšega zemljana predstavlja umetna inteligenca, Midjourney.**

Vzpon in padec načel družbe IBM

IBM je ena najstarejših tehnoloških družb na svetu in se lahko pohvali z vrsto inovacij, vključno z velikimi računalniki, programskimi jeziki in orodji na temelju umetne inteligence (UI). A če slehernika, mlajšega kot 40 let, vprašate, kaj točno dela (oziroma je delal) IBM, bo v najboljšem primeru postregel z medlim odgovorom: »Nekaj v zvezi z računalniki.«

Deborah Cohen, *The Atlantic*

Pripadniki generacije nič, s katerimi sem govorila, so v glavnem vedeli le to, »Nekaj v zvezi z računalniki.« In če milenijec že pozna konkreten izdelek IBM, je to Watson, njegov prototipni sistem UI, ki je leta 2011 zmagal v igri Jeopardy.

V kroniki garažnega podjetništva IBM ohranja svoje legendarno mesto – kot nerodni velikan. Leta 1980 je vodstvo IBM, utrujeno zaradi skoraj 13 let protitrustovskega pravdanja, naredilo hudo napako in takrat 25-letnemu Billu Gatesu, soustanovitelju družbe z nekaj deset zaposlenimi, omogočilo ohraniti pravice za operacijski sistem, ki ga je kot podizvajalec razvil za takrat tajni projekt osebnih računalnikov. Iz te napake je zrasel Microsoft in januarja 1993 je bila Gatesova družba vredna 27 milijard dolarjev. Za nekaj časa je prehitela IBM, ki je tistega leta beležil eno največjih izgub v ameriški poslovni zgodovini.

V iskrivo napisani biografiji Thomasa J. Watsona mlajšega, direktorja IBM sredi prejšnjega stoletja, z naslovom *The Greatest Capitalist Who Ever Lived*, je nedvoumno navedeno, da zgodovina podjetja ponuja veliko več od prilike o samovšečnem Goljatu. Kot poudarjata soavtorja, Watsonov vnuk Ralph Watson McElvenny in Marc Wortman, je IBM uspel preroški preskok iz mehanske tehnologije v elektronsko, s čimer je pospešil prihod digitalne dobe. Bil je neverjetno podjetniški za veliko družbo in osredotočen na novosti, a izstopali so tudi njegovi anahronizmi. Desetletja po tem, ko je večina velikih ameriških podjetij zaposlila izvrstne in dobro plačane vodstvene kadre, je IBM ostal družinsko podjetje, prežeto z lojalnostjo, a tudi napetostjo (v kateri družini pa ni tako?). Šefi so se pogosto prepirali. Fanatično skrbno se je posvečal strankam in zaposlenim zagotavljal doživljenjsko delovno mesto (kar je tradicija iz obdobja velike depresije). Njegovo vodilo je bilo spoštovanje posameznika.

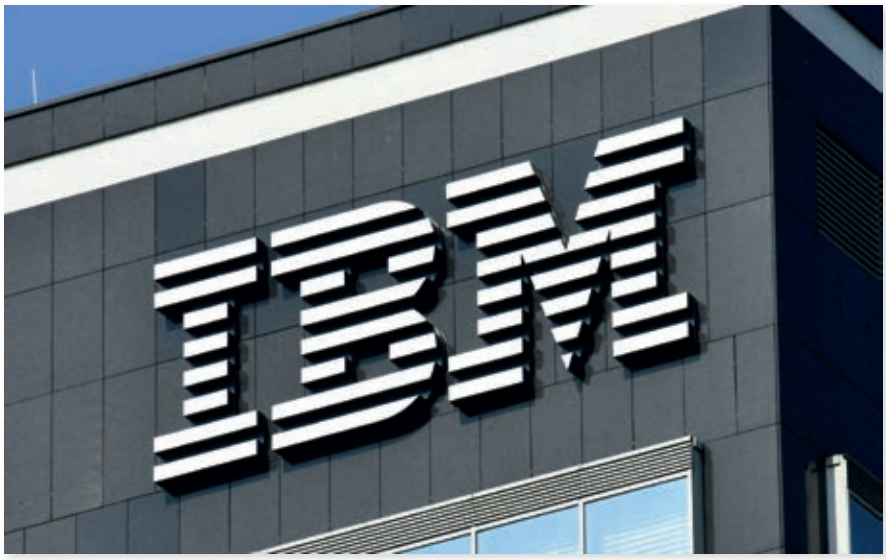
Danes, ko se približuje prihodnost, v kateri bo umetna inteligenca posameznika najbrž izbrisala tako kot zaposlenega kakor ustvarjalca, zgodba o IBM večinoma zveni kot pravljica iz oddaljenih svetov. Njeni tehnološki dosežki so še vedno prepoznavni kot znanilci digitalne dobe, le njegova kultura družbene odgovornosti – osredotočenost na zaposlene namesto na delničarje, zadržanost pri nagradah vodstvu in vlaganje v programe proti revščini – se je izkazala za slepo ulico. Mešanica naprednjaštva in očetovskega pokroviteljstva, občutka pripadnosti in hkrati neusmiljene tekmovalnosti, nekoč slavljenega načina poslovanja IBM, je bila neločljivo povezana z družino na vrhu, četudi je podjetju to škodilo.

Večino svoje zgodovine, sploh po prvi svetovni vojni in do 70. let, se je IBM posvečal predvsem učinkovitejšemu poslovanju. Ob koncu 19. stoletja je razvoj železnice, telegrafa in elektrike omogočil hitro širjenje tako velikosti kot dometa ameriških podjetij. Ko so podjetja izdelovala in dobavljala izdelke posrednikom in končnim porabnikom, so se soočala tudi z vse bolj zapleteno logistiko. Potrebovala so nove načine za urejanje, shranjevanje in priklic podatkov. In tako so prišli pisalni stroj (patentiran leta 1868), blagajna (1883) in računalo (1888).

Najpomembnejši element v tem informacijskem krogu je predstavljalo podjetje *National Cash Register*, kjer je leta 1874 rojeni Thomas J. Watson starejši, ki je odraščal na kmetiji blizu Painted Posta v zvezni državi New York, 17 let služil kot vajenec in odkril, da je rojen prodajalec. Na ukaz diktatorskega šefa je nadzoroval tudi sumljiv posel prodaje rabljenih blagajn pod ceno, da bi prodajalce izrinil s trga. Obtožili so ga oviranja trgovine, največje ponižanje pa je doživel, ko ga je *National Cash Register* odpustil. Ko se je Watson pobral kot novi generalni direktor newyorškega podjetja *Computing-Tabulating-Recording*, je imel 40 let, ravno se je poročil in dobil sinčka, Toma mlajšega.

Watson je podjetje kmalu preimenoval v *International Business Machines Corporation*;

IBMu je uspel preroški preskok iz mehanske tehnologije v elektronsko, s čimer je pospešil prihod digitalne dobe.



Na zidovih podjetja so bili napisana Watsonova najljubša gesla: ‚Podjetje je znano zaradi zaposlenih, ki mu jih uspe obdržati.‘ ‚Veliko časa nameni zadovoljstvu strank.‘ In ‚Razmišljaj‘, kar je za Watsona verjetno pomenilo, ‚Razmišljaj kot jaz‘.

takšen naziv se je veliko lepše ujemal z njegovimi globalnimi ambicijami. Njegova zasluga je bila, da je IBM prevzel tehnologijo luknjanih kartic, sodobno metodo za zbiranje podatkov, enako velja za uvedbo kulture naklonjenosti in spoštovanja med sodelavci, pokroviteljstvo in kult osebnosti. Na zidovih podjetja so bili napisana Watsonova najljubša gesla: 'Podjetje je znano zaradi zaposlenih, ki mu jih uspe obdržati.' 'Veliko časa nameni zadovoljstvu strank.' In 'Razmišljaj', kar je za Watsona verjetno pomenilo, 'Razmišljaj kot jaz' (je ugotovil zgodovinar, specializiran za posle, Richard S. Tedlow).

Če ob teh parolah pomislite na nekdanje Googlovo vodilo 'Ne bodi zloben' in trenutno Alphabetovo geslo 'Stori, kar je prav', so to prazni odmevi vseobsegajočega vodila IBM, ki ga je spremljala tudi obljuba zagotovljenega delovnega mesta. Poleg obratov IBM so stali njegovi podeželski klubi, v katerih so trikrat na teden postregli z večerjo. Ko sta Watson in njegova žena obiskovala prostore IBM v drugih mestih, sta se udeleževala tudi skrbno načrtovanih 'družinskih večerij' za zaposlene. Ti so bili sveže obriti in vsi enako oblečeni v temne obleke, poškrobljene bele srajce in kravate. Na dogodkih podjetja je bil alkohol prepovedan.

Watson starejši je z optimizmom iskrenega vernika v času velike depresije smelo širil poslovanje, grmadil stroje za branje podatkov, zaposloval nove prodajalce in uvedel jamstvo doživljenjske zaposlitve. Kot demokrat in podpornik predsednika Franklina D. Roosevelta (kar je bilo nenavadno za direktorja) je bil v pravem trenutku na pravem mestu, da je priskrbel naprave, potrebne za uresničitev zakona o socialni varnosti, ki so ga sprejeli leta 1935. Ob koncu 30. let je IBM obvladal trg z obdelovanjem podatkov.

Mladi Tom je lahko neposredno opazoval tako velikopotezne načrte svojega očeta kot njegove izraze prevzetnosti, na primer razkošno rezidenco na Manhattnu in predsednikova pisma pohvale, ki jih je nosil v žepu. A veljal je za nesposobnega celo po sproščenih standardih, ki so veljali za premožne mladeniče tistega časa. V šoli ni bil uspešen, zapletel se je v težave in po tedne je preležal zaradi napadov depresije. Kljub očetovemu denarju ga niso sprejeli na Princeton in vodja vpisa je Watsonu starejšemu celo rekel, da sin ne more postati drugega kot izguba.

Prav res, le kaj bi Tom Watson mlajši postal brez druge svetovne vojne? Ker se je navduševal nad letenjem, se je maja 1940 pridružil nacionalni gardi. Generalmajor Follet Bradley, poveljnik Prve letalske sile, ga je izbral za svojega osebnega pilota. Med vojno je Watson začel pisati dnevnik, kot bi se pri 28 letih njegovo življenje šele začelo. Po vojni se je na Bradleyjev predlog in očetovo zadovoljstvo vrnil v IBM. Watson starejši je rad govoril, da je nepotizem ugoden za posel: v podjetju, ki je delovalo kot družina, je spodbujal zaposlovanje očetov in sinov ter pričakoval, da bosta vodenje družbe prevzela sinova Tom in njegov mlajši brat Dick. A Toma je očetovo vodenje dušilo in gnusilo se mu je klečeplazenje, za katero se mu je zdelo, da ga še dodatno spodbuja takšen način vodenja.

Pogosto in glasno sta se prepirala, tudi pred drugimi zaposlenimi. »Prekleto vendar, stari, me res ne moreš nikoli pustiti pri miru?« Strinjala sta se o zagotovljeni doživljenjski službi, pomenu službe za stranke, o tem, da morajo biti direktorja vrata vedno odprta za zaposlene, in o pasteh samovšečnosti. Tom Watson je bil tako kot oče politično liberalen, v obratih IBM na jugu ni dovolil rasnega ločevanja in nasprotoval je lovu na čarovnice, ki ga je začel senator Joseph McCarthy. O skoraj vsem ostalem pa se nista strinjala, niti o smeri ključnega posla IBM.

Watson starejši je prepovedal izraz računalnik, saj ga je skrbelo, da bo vznemiril ljudi, ki so se bali, da bi nove naprave izpodrinile delavce. Čeprav so ga novi 'misleči stroji' zanimali, ni videl smisla v elektronski hitrosti in zdelo se mu je, da je skoraj nobena družba ne potrebuje. Prav gotovo računalnikov ni prišteval k opremi za poslovanje, medtem ko je Tom dojel, kako pomembni so.

Na začetku 50. let se je oče, takrat star že več kot 70, začel umikati iz vsakdanjih nalog in leta 1952 je starejšega sina imenoval za direktorja IBM. Watson starejši je negoval patriarhalni slog in 38 vodstvenih delavcev je poročalo neposredno njemu. Sin pa je uvedel organizacijsko shemo in vodje so začeli odstranjevati fotografije Watsona starejšega, ki so nekoč krasile vse razstavne prostore prodajaln. Še pomembneje pa je bilo, da je večjo pozornost začel namenjati

računalnikom in istega leta je podjetje odprlo kompleks v San Joseju, kar je bila prva računalniška tovarna v Silicijevi dolini in raziskovalna podružnica podjetja.

Družba je rasla z osupljivo hitrostjo, enako pa tudi drznost, s katero je tvegal Watson mlajši. Na začetku 60. let je vse stavil na odločitev o proizvodnji serije popolnoma združljivih računalnikov Sistem 360. Takrat so v IBM izdelovali sedem popolnoma ločenih sistemov z različno računsko močjo. Vsak sistem je imel svoj notranji ustroj, zato prenos podatkov med linijami računalnikov pogosto sploh ni bil mogoč. Stranke, ki so želele nadgraditi računalnike, so morale tako rekoč začeti iz nič. Sam IBM pa je pestila neučinkovita proizvodnja, med drugim 2.500 različnih tiskanih vezij.

Sistem 360 opisujejo kot enega najboljših inovativnih izdelkov v ameriški zgodovini 20. stoletja, takoj za Fordovim modelom T. Združljivost ogromnega števila procesorjev je bila inženirska nočna mora, za katero so potrebovali na milijone vrstic računalniškega koda. Naložbe IBM bi preračunano na današnjo vrednost znašale 50 milijard dolarjev, kar je več kot dvakratna vrednost manhattanskega projekta. Zaradi novega računalnika so vse starejše linije podjetja postale zastarele, kar je pomenilo, če Sistem 360 ne bi deloval po pričakovanjih, bi stranke IBM odšle k tekmecem.

Ko so po več zamudah, tudi zaradi inženirskih in proizvodnih težav, končno dobavili linijo računalnikov Sistema 360, je takoj postala prodajna uspešnica. Med letoma 1964 in 1970 je IBM zaposlil skoraj 120.000 novih sodelavcev (in imel skupaj 269.000 zaposlenih), njegovi prihodki so se več kot podvojili s 3,2 milijarde dolarjev na 7,5 milijarde, kar je za tako veliko družbo skoraj neverjetno. Watson mlajši je leta 1958 nehal jemati svoje delniške opcije v vrednosti petkratnika letne plače, češ da noče ravnati pritlehno.

Kot je ekonomist Theodore Levitt trdil leta 1960, podjetja, ki se zanašajo na neki izdelek, in to velja tudi za zelo uspešna, nehote postanejo zastarela. Hollywoodski mogotci so spregledali, da njihov posel niso filmi, temveč zabava, zato se jim je iz rok izmuznila televizija, največja priložnost tistega obdobja. Watson starejši je mislil, da so njegov glavni posel luknjane kartice, njegov sin se je zavedal, da so v resnici informacije.

Strokovnjaki večinoma le ugibajo, zakaj je IBM prešel z mehanskega načina obdelave podatkov na elektronskega. Skok v neznano je bil v enaki meri posledica tega, kako je hladna vojna s Sovjetsko zvezo, sploh pa korejska vojna, napihnila zvezno financiranje zasebnih raziskav in razvoja, kot je James W. Cortada lepo razložil v svoji nedavni zgodovini podjetja z naslovom *IBM: The Rise and Fall and Reinvention of a Global Icon* (*IBM: Vzpon, padec in preobrazba globalne ikone*). Pomembno vlogo so odigrali tudi inženirji, zaposleni v IBM, najprej, ker so pritiskali na vodstvo, naj že sprevidi obete nove tehnologije, in nato s preobrazbo zapletenih računskih sistemov v komercialne izdelke. Kupci so začeli zahtevati nove naprave. No, zlahka bi se bilo obrnilo tudi drugače. Kljub prednosti pri računalnikih se je Remington Rand, veliki tekmec IBM v panogi tabelirnih strojev, raje osredotočil na električne brivske aparate, pisalne stroje in pisarniško pohištvo.

Biografija *The Greatest Capitalist Who Ever Lived* (*Največji kapitalist, kar jih je živelo*) tako kot vse biografije poudarja posameznike – še posebej Thomasa Watsona mlajšega – in tiste vodstvene delavce IBM, ki so v različnih obdobjih predstavljali njegove zaupnike, preroke in sovražnike. McElvenny kot Watsonov vnuk ponuja poznavalsko oceno družinske dinamike, povzete iz intervjujev

▽ **IBMov Sistem 360 opisujejo kot enega najboljših inovativnih izdelkov v ameriški zgodovini 20. stoletja, takoj za Fordovim modelom T.**

in zasebnih dokumentov. Še posebej izstopa, da so avtorji naredili korak dlje od večine poznavalcev, ko so opisali sinovo navdušenje nad računalniki kot zavračanje očeta. Zamere so bile obojestranske, so pojasnili: ko so Watsona mlajšega leta 1955 upodobili na naslovnici revije Time, kar je bil trženjski podvig za podjetje, stari ni rekel ničesar. Tekmovalnost med njima je še naprej gnala Watsona mlajšega, celo po očetovi smrti leto dni pozneje.

Lahko bi rekli, da je Watson mlajši ustanavljal novo družbo, ko je prevzel vodenje IBM, in ker se je moral dokazati, je podjetje vodil kot podjetnik, ne kot dedič. Ni se obdal s kimavci, temveč je raje imel bistre, robate, ostre, skoraj že neprijetne tipe, ki so imeli realen pogled in so mu ga tudi predstavili. Oblikoval je sistem 'bojevitega vodenja', ki je od direktorjev in njihovih podrejenih zahteval, da so se v primeru nesoglasja soočili pred odborom vodstvenih delavcev. Jamstvo doživljenjske zaposlitve naj bi spodbujalo odgovorno tveganje in pripomoglo k trenjem, produktivnim za družbo v hierarhiji. Richard Tedlow je pripomnil, da je Watson želel, da bi dinamika med njim in očetom pognala 'metastaze' in se razširila po vsem IBM.

Medtem ko je ta ojdipovski kompleks koristil poslovni bilanci IBM, je usodno načel odnos med obema Watsonoma. V slikovitem opisu McElvennyja in Wortmana sta postala Kennedyja poslovnega sveta, vključno z regatami na jahtah, zunajzakonskimi aferami, s smučarskimi konci tedna s pravimi Kennedyji in psihološkimi zlomi. Zgodba o Sistemu 360 je bila pogubna tudi za Dicka, Tomovega mlajšega brata. Dick Watson je bil precej manj uporniški duh od Toma, očetu je celo dovolil, da ga je spremljal na poročnem potovanju. Brata sta si bila v mladosti blizu in Dick je znal depresivnega Toma spraviti iz postelje, ko so vsi že obupali nad tem. Dick je uspešno vodil mednarodno mrežo IBM in Tom si je želel, da bi ga brat nasledil na direktorskem mestu.

A Tomova odločitev, da Dick prevzame vodenje proizvodnje in inženiring Sistema 360, njegov tekmec za direktorski položaj, T. Vincent Learson, pa njegovo prodajo, je prinesla nasprotni učinek od želenega. Zamude v proizvodnji so se kopičile in Dick je prenehal prihajati v službo; širile so se govorice, da je popival. Brata sta se komaj še pogovarjala, in ko je Tom tako rekoč odpustil Dicka, sta se povsem odtujila. Tom je leta 1970, pri svojih 56, doživel hud infarkt in se kmalu umaknil z direktorskega stolčka, uradno pa se je upokojil leta 1974. Istega leta je Dick umrl, star komaj 55 let, zaradi padca po stopnicah v svojem domu.

△ **IBM je danes še vedno »računalniško podjetje«, znano tudi po superračunalnikih.**

Ko je vodstvo IBM sklenilo poguben dogovor z Billom Gatesom, je Tom Watson mlajši v Sovjetski zvezi služil kot veleposlanik predsednika Jimmyja Carterja. Watson je napisal neobičajno iskrene spomine z naslovom *Father, Son & Co.* (*Oče, sin in d. o. o.*), ki je leta 1990 14 tednov vztrajala na seznamu največjih prodajnih uspešnic New York Timesa. Takrat je bil IBM – ki ga je še vedno imenoval »moje podjetje« – kot ranjen velikan. V 80. so preveč vlagali v velike računalnike in Watsonovi nasledniki niso znali unovčiti osebnih računalnikov ter programske opreme, s čimer so zapravili velik trg z izdelkom za široko porabo. Ko je IBM na začetku 90. tonil, se je Watson mlajši ponoči zbujal v solzah. Umrl je leta 1993 po kapi.

Lou Gerstner, direktor, ki je istega leta prevzel reševanje IBM, je spoštoval Watsonov način vodenja, a v svojih spominih ni olepševal škode, ki jo je povzročilo šest desetletij pod Watsoni. Sistem reševanja sporov je spodletel in pripomogel k slabemu občutku, ta pa k izogibanju sporom namesto k odkriti razpravi o drugih možnostih. Doživljenjsko zagotovljeno delovno mesto se je izjalovilo v občutek upravičenosti do njega in Gerstner je vztrajal, da je to načelo treba odpraviti.

Slabi dve leti po tistem, ko je Gerstner prevzel najvišji stolček v IBM, skoraj polovica zaposlenih, ki je bila leta 1987 na plačilni listi, ni več imela službe v tem podjetju. Stara watsonska kultura je životarila le še kot medel spomin. IBM so večkrat tožili zaradi odpuščanja delavcev, starejših od 40 let. (Podjetje je izjavilo, da nikoli ni sistematično diskriminiralo zaposlenih zaradi starosti.)

A ljudje tako in tako kmalu morda ne bodo več pomembni. Lansko pomlad je IBM predstavil svoj novi izdelek z umetno inteligenco, Watsonx, ki ga hvalijo kot najdragocenejšo inovacijo podjetja v zadnjih letih. Zmore olajšati kadrovanje, komentirati tenis v Wimbledonu in še marsikaj – je stvaritev, ki bi lahko pospešila avtomatizacijo, kot si ne znamo še niti predstavljati.

Od nekdaj je Watsona starejšega bolj kot sina skrbela možnost, da bi stroji prevzeli človekovo mesto. Hkrati IBM nikoli ni videl zgolj kot podjetje. Nekoč je oznanil: »IBM ni samo organizacija ljudi, je institucija, ki bo večno živela.« Ključno za to je po njegovem mnenju bilo, da ohrani dušo. Zagotovo bi tako očeta kot sina osupnilo, da bo ravno opuščanje humanosti v resnici morda skrivnost večnega življenja. ◀

IBM se je fanatično skrbno posvečal strankam in zaposlenim zagotavljal doživljenjsko delovno mesto (kar je tradicija iz obdobja velike depresije). Njegovo vodilo je bilo spoštovanje posameznika.

Googlov Gemini je res potreboval popravke

Če Googlovo generativno umetno inteligenco (UI) Gemini prosite, naj ustvari podobo ameriških vojakov med vojno za neodvisnost proti koncu 18. stoletja, vam bo mogoče ponudila temnopolto žensko, Azijca in Indijanko v modrem plašču Georgea Washingtona.

Chris Stokel-Walker, *Fast Company*

Takšna pestrost je pošteno pojezila nekaj ljudi, med njimi tudi Franka J. Fleminga, nekdanjega računalniškega inženirja in pisca za Babylon Bee. Fleming je na omrežju X objavil vrsto vse bolj nemogočih izmenjav z Geminijem, na primer, ko mu je naročil, naj izdela podobo svetlopoltih ljudi v položajih ali pri opravilih, kjer so zgodovinsko prevladovali (recimo srednjeveški vitezi). Njegovim pritožbam so se pridružili še drugi, ki trdijo, da je to raznolikost zaradi raznolikosti same, in navedli, kaj vse je narobe s sedanjim osveščenim svetom.

Kljub zgodovinskim dejstvom Fleming in njegovi razdraženi somišljeniki bijejo boj brez upanja na izboljšanje. »Ti sistemi tega ne omogočajo,« je pojasnila Olivia Guest, profesorica in asistentka za računalniške kognitivne znanosti na univerz Radboud. »Pri stohastičnih sistemih ne moreš zajamčiti odziva oziroma vedenjskih vzorcev.«

Sedanja generacija orodij generativne umetne inteligence sodi v stohastične sisteme. V neki odmevni akademski razpravi, objavljeni leta 2021, so jih opisali kot sisteme, ki ob enakih vhodnih podatkih naključno pridejo do različnih končnih rezultatov. Ravno zato je generativna UI vzbudila takšno pozornost javnosti: ne ponavlja vedno istega do onemoglosti.

Strokovnjaki dvomijo tudi, da rezultati klepetalnega robota, kot so jih na družbenih omrežjih predstavili jezni uporabniki, res razkrivajo celotno ozadje. »Težko je oceniti verodostojnost vsebin, ki jih lahko vidimo na platformah, kot je X,« je pojasnila Rumman Chowdhury, soustanoviteljica in direktorica neprofitne organizacije Human Intelligence. »So to skrbno izbrani primeri? Brez obsežne analize generiranja podob, ki jim je mogoče slediti in jih razvrstiti tudi glede na različne ukaze, ne moremo oceniti, ali je model pristranski.«

Google ve za nezadovoljstvo in je v izjavi sporočil, da ukrepa. »Zavedamo se, da Gemini pri zgodovinskih podobah ni vedno natančen, in se trudimo, da bi to čim prej popravili,« je na omrežju X pojasnil Jack Krawczyk, produktni vodja za Googlov Bard. Poudaril je tudi, da je prikaz zgodovinskih dogodkov prerez dveh interesov, ki se bojujeta za prevlado: interesa, da bi natančno predstavili zgodovinske dogodke, in interesa, da bi upoštevali globalni zbir uporabnikov.

Rešitev teh temeljnih težav morda ne bo tako preprosta. Popravljanje stohastičnih sistemov je zahtevnejše, kot se zdi na prvi pogled. Varnostni mehanizmi za zagotavljanje etičnih standardov so enaki za vse modele UI in jih je mogoče 'spodkopati', razen če bi se poslužili neusmiljenega blokiranja (Google je nekoč programski opremi za prepoznavo fotografij preprečil, da bi temnopolte ljudi prepoznala kot gorile, tako da preprosto ni več mogla prepoznati goril.) Potem to ni stohastičen sistem, kar pomeni, da izgubimo tisto, zaradi česar je generativna umetna inteligenca tako posebna.

Ob vsem tem se postavlja zanimivo vprašanje, je pripomnila Chowdhuryjeva. »Res je težko definirati, ali obstaja pravilen odgovor, kakšna podoba naj se generira. Opiranje na zgodovinsko natančnost bi lahko povzročilo vnovično uveljavitev izključevanja, s tem pa bi tvegali posredovanje napačnih dejstev.«

Yacine Jernite, vodja oddelka za strojno učenje in družbo v podjetju Hugging Face, meni, da težava ni omejena samo na Gemini. »To je strukturna težava, povezana s tem, da kar nekaj družb razvija komercialne izdelke, ne da bi zagotavljale dovolj transparentnosti, ko gre za predsodke in pristranskost,« je povedal. O teh vprašanjih je Hugging Face že pisal. »Pristranskost je posledica izbir na vseh ravneh razvojnega procesa in najzgodnejše izbire imajo lahko največji vpliv – na primer izbira temeljne tehnologije, vir podatkov in koliko jih bodo uporabili,« je pojasnil Jernite.

Boji se, da je to, čemur smo priče, posledica hitrih in razmeroma poceni rešitev, za katere se odločijo podjetja. Če podatki za učenje zajemajo preveč svetlopoltih ljudi, je mogoče v ozadju prilagoditi ukaze, da dodaš raznolikost. »Vendar to ne rešuje jedra težave,« je dodal.

Podjetja bi se namesto kozmetičnih popravkov morala odkrito posvetiti pristranskosti in predstavitvi. Razlaganje preostalemu svetu, kaj točno počneš za odpravljanje pristranskosti, je zahtevna naloga, saj se podjetje tako izpostavi in zunanji deležniki podvomijo o njegovih odločitvah ali pa prizadevanja označijo za nezadostna, morda celo neiskrena,« je razložil. »Vendar je vse to nujno, saj morajo vprašanja postaviti ljudje z bolj neposrednim interesom za nepristranskost, tisti, ki so bolj podkovani v tej temi – sploh pa družboslovci, ki jih v tehnološkem razvoju kronično primanjkuje. In, najbolj bistveno, ljudje, ki utemeljeno ne verjamejo, da bo tehnologija delovala. Le tako se lahko izognemo konfliktu interesov.« ◀

Vodnik po **zaposlitvi v UI**

Vlagatelji panogo umetne inteligence (UI) zasipajo z denarjem in ta sektor zdaj lahko ponudi nekaj najbolje plačanih delovnih mest v tehnološki industriji.

AJ Eckstein, *Fast Company*

Zagonska podjetja na področju umetne inteligence so lani zbrala skoraj 50 milijard dolarjev kapitala, je mogoče razbrati iz podatkov *Crunchbasa*, nekatera podjetja, kot sta OpenAI in Anthropic, pa so sama pridobila več milijard. Lanska raziskava javnega mnenja *Biz Report* ugotavlja, da delovna mesta, povezana z UI, iskalcem zaposlitve v povprečju nudijo 77 odstotkov višjo plačo kot drugi poklici in da nekatere začetne plače dosegajo tudi 450.000 dolarjev na leto.

A kako je v trenutnem razcvetu UI sploh mogoče priti do takšne službe? Kaj kadrovniki, ki zaposlujejo strokovnjake za UI, sploh iščejo, saj gre za razmeroma novo področje?

Pogovarjal sem se s tremi vodji kadrovske službe iz OpenAI, Uberja in Amazon Web Services, da bi odstrl to tančico skrivnosti in vam pomagal do nove zaposlitve. Povedali so mi tole.

Kakšne vrste delovnih mest sploh obstajajo?

Preden si naredite načrt, kako boste prišli do službe v panogi UI, morate vedeti, po kakšnih profilih se sploh povprašuje.

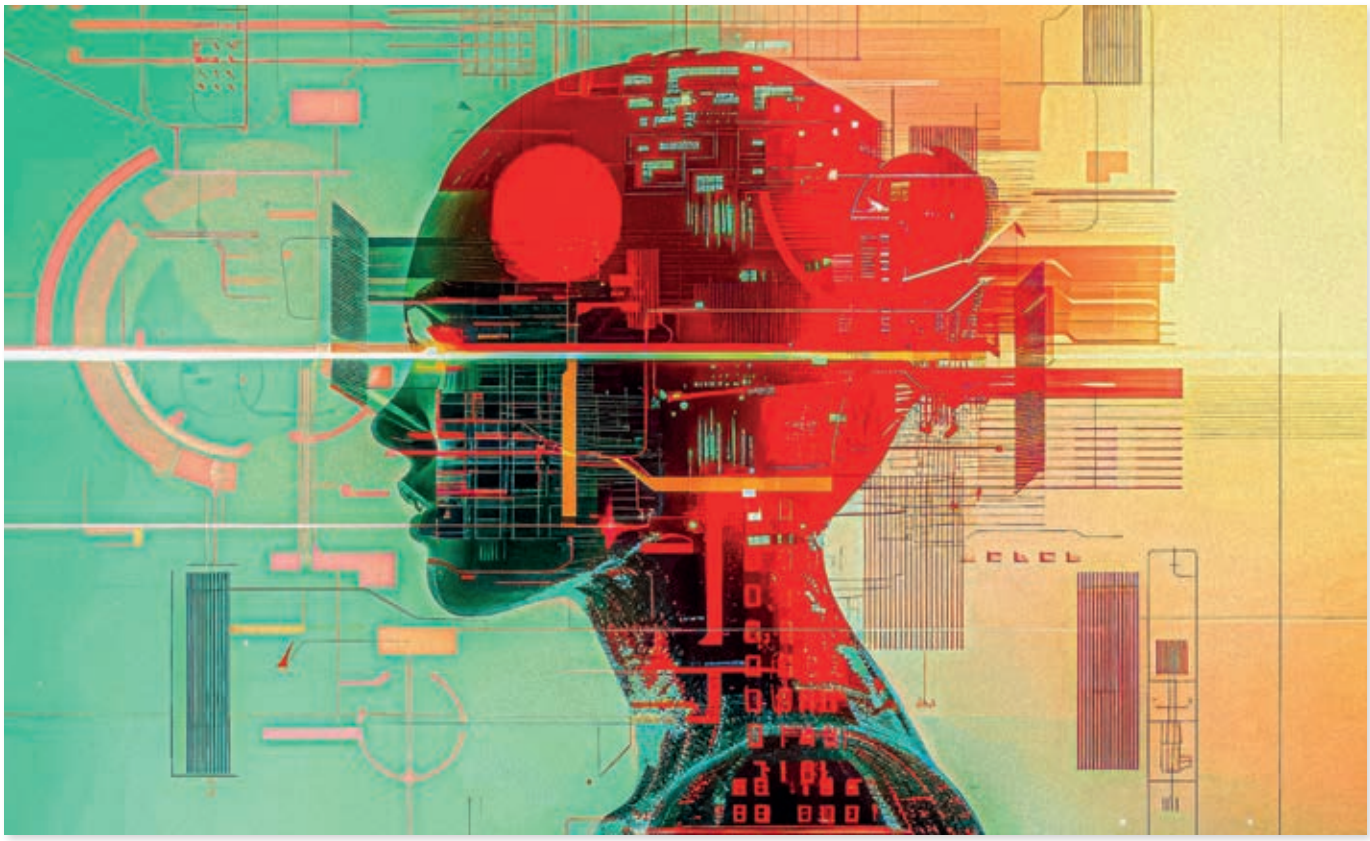
Robert Infantino je nadzoroval zaposlovanje raziskovalcev v OpenAI in pomagal pri iskanju raziskovalcev v Amazon Web Services za področje umetne inteligence. Pravi, da so za ta področja zanimivi predvsem znanstveniki in inženirji z izkušnjami z uporabnimi in s temeljnimi raziskavami, programski in podatkovni inženirji za infrastrukturo, vodje tehničnih izdelkov in programov ter inženirji za strojno opremo.

Nekatera podjetja, na primer Runway, ki je specializirano za uporabne raziskave, ki ob pomoči generativne UI vzdržuje platformo za ustvarjanje vsebin za umetnike in kreativce, sestavlja ekipe za strojno učenje in uporabne raziskave s poudarkom na računalniškem vidu in generativnih medijih, je pojasnila Lauren Saltus, kadrovica za tehnično osebje v tem podjetju z več kot sedmimi leti izkušenj.

Preden se zakopljete v internetne strani podjetij, kjer objavljajo prosta delovna mesta, bi po priporočilih Saltusove morali vedeti, da se nekatera podjetja posvečajo raziskavam in razvoju, druga pa so osredotočena na uporabo obstoječih raziskav pri konkretnih izdelkih, kar vpliva na to, kakšna delovna mesta podjetje poskuša zapolniti. »Najprej se morate vprašati, na katerem področju želite delati, saj boste tako zožili iskanje in lažje našli organizacijo, ki vam bo pisana na kožo.«

Ne potrebujete izkušenj z umetno inteligenco – so pa dobrodošle

Nikita Gupta, nekdanja vodja tehnične kadrovske službe v Uberju in ustanoviteljica Careerflow.ai, se pri iskanju in najemanju novih sodelavcev osredotoča na tri ključna področja: solidna izobrazba v matematiki in statistiki,



obvladanje programov, ki se običajno uporabljajo pri UI in strojnem učenju, ter dobro razumevanje algoritmov za strojno učenje.

Poleg konkretnih veščin in izkušenj so kadroviki pozorni tudi na tako imenovane mehke veščine, kot sta radovednost in optimizem, je pojasnila Saltusova. Te lastnosti pri prosilcih preceni z vprašanji, kot so: »Komaj čakate, da se boste učili in rasli? S kakšnimi projekti se ukvarjate v prostem času? Ste naravnani na iskanje rešitev, ko se soočate z izzivi?« Pojasnila je tudi, da njena ekipa premika meje inovacij in išče ljudi, ki so uspešni v takšnem okolju.

Infantino pa je povedal, da na splošno išče znamenja izjemnih sposobnosti, od zmage na matematični olimpijadi v gimnaziji do predsedovanja filharmonikom ali celo profesionalnega igranja pokra.

Kandidati izstopajo tudi, če lahko dokažejo hitro napredovanje na poklicni poti, delajo ob boku najboljšim na področju in so na najboljši konferencah oziroma v publikacijah objavili izjemno odmevne razprave, je še dodal. Infantino te lastnosti enači s »strmo krivuljo navzgor«, kar pomeni, da kandidat za delovno mesto kaže ambicije, gradi nekaj pomembnega, je pripravljen tvegati in tako spremeniti smer svoje panoge oziroma podjetja.

Povpraševanje po najboljših in najobetavnejših se hitro povečuje

Medtem ko v drugih panogah delavce odpuščajo, se delovnih mest, povezanih z UI, odpira več, kot je primernih strokovnjakov.

»Zadnjih nekaj let se je povpraševanje po strokovnjakih za UI občutno povečalo in postaja tudi vse bolj specifično,« je komentirala Guptova. »Ženejo pa ga napredek tehnologije, vse širša uporaba UI v različnih panogah in prepoznavanje vrednosti, ki jo UI lahko ima za posel.«

Nikita Gupta opaža tri glavne trende pri povpraševanju po strokovnjakih za UI: število objavljenih delovnih mest, povezanih z UI, narašča v skoraj vseh panogah, tekmovalnost med podjetji, da bi pritegnila in obdržala najboljše strokovnjake, je vse izrazitejša in pojavila so se že specializirana delovna mesta v zvezi z UI (na primer etika uporabe UI, njena varnost, infrastruktura in produktni menedžment).

Boj za najobetavnejše kadre se še bolj razvnema, ker se za dobro usposobljene poznavalce UI potegujejo od tehnoloških velikanov in zagonskih podjetij do raziskovalnih ustanov in drugih organizacij. Mamijo jih s konkurenčnimi plačami, z bonusi in ugodnostmi.

Izzivi, ki jih prinaša razcvet UI

Kadroviki so mi povedali, da se zaradi večjega povpraševanja po poznavalcih UI odpirajo novi izzivi, vključno s premajhno raznolikostjo in premalo primernega vodstvenega kadra.

»Glavni izzivi so premalo žensk na področju UI, pomanjkanje veščin ter znanja in huda konkurenca za kadre med velikimi tehnološkimi podjetji,« je pojasnila Guptova.

Saltusova pa je povedala, da si njena ekipa v Runwayu ne prizadeva le za čim večjo pestrost kadrovskega bazena, temveč razmišlja tudi o tem, kako podpirati raznolikost, enakost in vključevanje v okviru njihovega modela UI, da bi to pripomoglo tudi k čim bolj poglobljenemu prikazovanju UI v medijih.

Kadroviki uporabljajo nekonvencionalne metode

Ker je na trgu malo primerne delovne sile za delo z UI, morajo biti kadroviki zelo iznajdljivi.

Vemo, da so željni prosilcev z izkušnjami pri raziskovanju UI. Saltusova je povedala, da je najučinkovitejša strategija za privabljanje najboljših tovrstnih sodelavcev »uveljavitev imena njihovega podjetja v svetu raziskav in izdelki (ter raziskave), ki govorijo sami zase. To pomeni večji poudarek na objavljanju in sodelovanju v panogi.«

Ta tehnologija se je po podjetjih bliskovito razmahnila. Raziskava društva za upravljanje človeških virov je pokazala, da si pri iskanju kadra in zaposlovanju osem desetin organizacij pomaga z umetno inteligenco.

Povedano preprosto, po besedah Saltusove je prihodnost kadrovanja v naši zmožnosti prilagajanja, učenja in izkoriščenja novih orodij. Trdi, da je z njimi mogoče nadgraditi iskanje najprimernejših kadrov za zapolnitev potreb in da ne nadomeščajo človeške presoje.

Tudi Infantino poskuša avtomatizirati čim več svojega rutinskega dela in uporablja številne izdelke, ki temeljijo na velikih jezikovnih modelih za naloge, kot so iskanje in izpeljevanje podatkov ter urejanje koledarja. Zagotovil je, da osebne interakcije s kandidati ne izpodriva UI, saj na prvo mesto postavlja kakovost, ne količine.

Kadroviki, tudi Infantino, trdijo, da UI pri iskanju primernih novincev ponuja številne prednosti, na primer hitrejše zaposlovanje, manjši vpliv predsodkov, nižje stroške, kakovostnejši izbor, dostop do kadrov v tujini, kar vse izboljša tudi izkušnjo iskanja novega delovnega mesta in preverjenega delodajalca.

Prihodnost kadrovanja je v zmožnosti prilagajanja, učenja in izkoriščenja novih orodij.

Podjetja uvajajo posebne programe, da bi privabila in izobrazila primerne kadre. Runway tako ponuja pospeševalni program – sodelovanje v polno plačanem programu, da ljudem z različno izobrazbo in izkušnjami ob podpori mentorja olajšajo prehod na delovna mesta, povezana s strojnim učenjem.

Saltusova je poudarila, da se modeli UI razvijajo zelo hitro, in njena ekipa je prepričana, da so močne tehnične veščine, zmožnost hitrega učenja in pogum za soočenje z neznanimi izzivi in tehnologijami veliko pomembnejši od formalne izobrazbe s tega področja.

Kadroviki bi morali vložiti čas v učenje in vsaj temeljno razumevanje tehnologije, preden začnejo iskati primerne nove sodelavce, je izpostavil Infantino. »Priporočam, da so na tekočem s hitrimi spremembami pri najnovejših tehnologijah prek platforme X in tehničnih medijev, nato pa s te odskočne deske brskajo nazaj, kdo stoji za novostmi. Seznaniti bi se morali z analizo citiranja in metodami ocenjevanja profilov ter razprav raziskovalcev, pri čemer si lahko pomagajo s platformami, kot so Google Scholar, Aminer, Hugging Face in Github. Tako bodo kadre našli prek njihovega konkretnega dela.«

Kadroviki delovna mesta, povezana z UI, zapolnjujejo s podporo UI

Iskalci zaposlitve se morajo zavedati, da kadroviki uporabljajo UI za zapolnitev delovnih mest v zvezi z UI.

Nasveti kadrovikov za prodor v panogo UI

Zdaj ko veste, kakšna je vloga UI in kako kadroviki iščejo primerne sodelavce, lahko naštejemo še najučinkovitejše zvijače za uspešen prodor v ta ekskluzivni segment zaposlovanja.

Saltusova priporoča aktivno vlogo v skupnosti. Med najpomembnejšimi razlogi, da se je panoga UI razvila tako hitro, je njena odprtost pri raziskavah. Tisti, ki sodelujejo v razpravah, s katerimi se poglablja naše poznavanje tehnologije in njenih dosežkov, so tudi prvi, s katerimi bi se ona rada pogovorila.

Infantino kandidatom svetuje, naj se vključijo v tovrstne projekte in sodelovanje zunaj delovnega časa, če jim trenutno delovno mesto ne omogoča delati z najnovejšo tehnologijo. Mreženje in sklepanje navez na področju UI morata biti načrtna, tako da se redno udeležujete konferenc, dejavno prispevate k najbolj razširjenim repozitorijem računalniških kod in hodite na srečanja, je dodal.

Po Infantinovih besedah za vsa delovna mesta, tudi zunaj sektorja UI, velja, da morajo kandidati ravnati strateško in izkoristiti vsako priložnost za razkazovanje svojih veščin pred najuglednejšimi predstavniki področja. Njihov podpis na projektu vas lahko pri iskanju novega delovnega mesta daleč pripelje. ◀

Kako bo generator filmov Sora pretresel svet. **Ali pa tudi ne.**

Po letu dni hitre razvojne poti ChatGPT, ki ga je javnost spoznala predlansko jesen, se mnogim zdi, da umetna inteligenca ne more več šokirati. A sredi februarja je prišla demo različica Sore, modela, ki besedilo pretvarja v posnetek in v hipu izdela žive, realistične prizore na podlagi zapletenih ukazov.

Joe Berkowitz, *Fast Company*

Sora je presunila je gledalce po vsem svetu, še posebej v Hollywoodu. Filmski ustvarjalec Tyler Perry, na primer, je bil tako pretresen, da je sklenil počakati z 800 milijonov dolarjev vredno širitvijo svojega studia v Atlanti. Zakaj bi se trudil s sceno, če pa jo računalniški program nasnuje s pritiskom na tipkovnico.

A vsi v industriji zabave naskoka umetne inteligence (UI) na to panogo ne pozdravljajo z odprtimi rokami. Skupina prvovrstnih scenaristov – vključno z avtorico scenarija za M3GAN Akelo Cooper in avtorjem Iron Mana 3 Shanom Blackom – je 27. februarja predstavila tehnološko platformo Gauntlet, ki so jo razvili skupaj. Njen cilj je doseči, da ljudje ostanejo v osrčju razvoja, zdaj ko se vsi pomembni odločevalci pri ocenjevanju scenarijev opirajo na umetnointeligenčna orodja. To je le eden v vrsti izrazov mnenja številnih v Hollywoodu, da je človeški element ustvarjalnosti nepogrešljiv in neponovljiv, zato ga je treba za vsako ceno ohraniti.

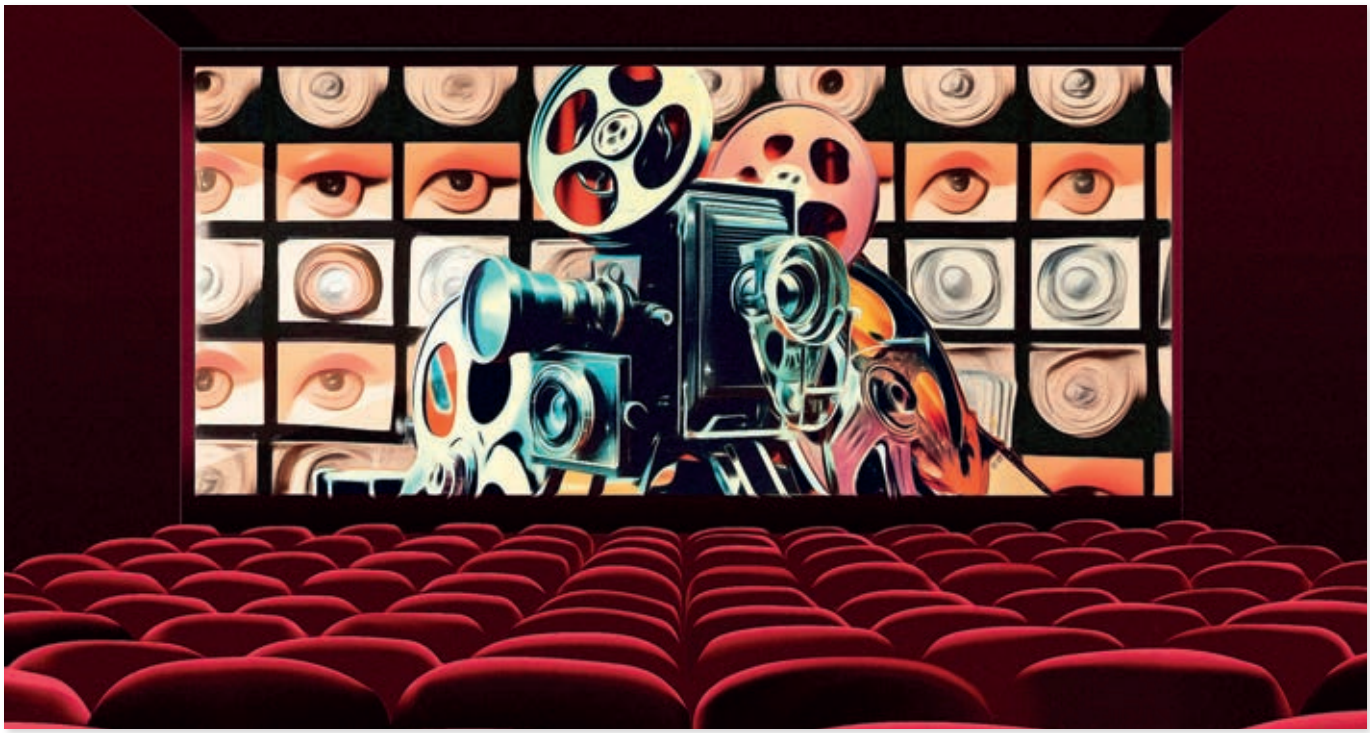
»Umetna inteligenca nas bo stisnila v matrico, v kateri nihče ne bo vedel, kaj je resnično,« je na podelitvi nagrad združenja *Screen Actors Guild* 24. februarja potožil Fran Drescher, predsednik Sag-Aftra. »Če izumitelju manjka sočutja in duhovnosti, morda ne potrebujemo njegovega izuma.«

Drescher je večino lanskega leta preživel na pogajanjih o novi pogodbi za združenje SAG, na katerih se je umetna inteligenca pokazala kot osrednja sporna točka, enako kot ob hkratni stavki scenaristov. Igralcem ni všeč, da bi v prihodnosti brez njihovega dovoljenja uporabili njihove umetno generirane dvojnike, scenaristi pa nočejo dopolnjevati in nadgrajevati zgodb, ki bi jih zasnovali algoritmi. Obe združenji imata še druge skrbi. Na koncu sta si obe zagotovili predpise, kako se bo umetna inteligenca uporabljala na teh dveh področjih, čeprav se že oglašajo kritiki, da so ti predpisi preblagi.

Agencija za iskanje novih igralcev WME se je pred kratkim povezala s podjetjem za UI Vermillio, ker želi dodatno zavarovati svoj zbir strank pred neavtorizirano uporabo njihovih obrazov. Mogoče bodo sledile še druge varovalke in bo Hollywood ohranil več svoje človeškosti. Vsaj tako so upali, preden jim je pred oči prišla Sora.

Najnovejša ponudba podjetja OpenAI je naprednejša od vseh podobnih izdelkov doslej, na primer Gen-2 podjetja Runway. Po trditvah OpenAI model Sora, poimenovan po japonski besedi za nebo, v katerokoli izmišljeni prizor lahko doda več izdelanih junakov, dinamične premike in srhljivo natančne ter zelo razločne podrobnosti. Njen veter piha kot pravi, njena voda teče in se peni in vse, česar se elementi dotaknejo, se ustrezno odzove. Na spletni strani Soro opisujejo: »Model ne razume le tega, česar ga prosi uporabnik v ukazu, temveč tudi, kaj se s tem dogaja v fizičnem svetu. Za nameček njena maksimalna dolžina 60 sekund daleč prekaša Runwayev čas od štirih do 18 sekund. Čeprav nekateri ambiciozni scenaristi že uspešno eksperimentirajo s platformami, kot je Midjourney, da bi ustvarili kratke odlomke, model OpenAI obeta, da bo odprl svet dodatnih možnosti.«

Sorine najpomembnejše posledice za filmsko in televizijsko industrijo presegajo



Filmski ustvarjalec Tyler Perry je bil tako pretresen, da je sklenil počakati z 800 milijonov dolarjev vredno širitvijo svojega studia v Atlanti. Zakaj bi se trudil s sceno, če pa jo računalniški program nasnuje s pritiskom na tipkovnico.

trenutno zmožnost, da hitro ustvari 'čedno žensko v Tokiu'. Očitno omogoča tudi vstavljanje enega posnetka v drugega, kar pomeni, da uporabniki lahko v simulirane svetove vključijo vse, kar so posneli v resničnem svetu. Model zaradi prožnosti pri vzorčenju omogoča tudi, da isti ukaz vidijo z drugega zornega kota, kot nekakšen digitalni panoptikum. Glede na takšne dosežke ima Tyler Perry prav, saj sluti možnosti za varčevanje pri proračunu, ki jih odpira Sora. Z njo je mogoče znižati stroške produkcije ne le zaradi posebnih učinkov in predhodne vizualizacije, temveč tudi pri sekundarnih nalogah, kot je snemanje ustanovitvenih in glavnih posnetkov.

Seveda bo vse to zniževanje stroškov povzročilo, da bo veliko ljudi izgubilo službo.

Poleg vprašanj, o katerih so se lani pogajali igralci in scenaristi, se poznavalci razmer v Hollywoodu bojijo, da bo revolucija zaradi UI izrinila tudi umetnike vizualnih učinkov, zvočne tehnike in številne druge. Po ocenah novejše raziskave med 300 direktorji in šefi v industriji zabave je v prihodnjih treh letih pričakovati izgubo skoraj 204.000 delovnih mest. Po predstavitvi Sore so zaposleni in umetniki v tej panogi še bolj prestrašeni, pa tehnologija sploh še ni na voljo širši javnosti.

Nihče ne ve, koliko se bo Sora še spremenila do uradne uvedbe na trgu. V OpenAI trdijo, da njihovi inženirji trenutno sodelujejo s politiki, z učitelji in umetniki z vsega sveta, da bi razumeli njihove pomisleke in prepoznali vse mogoče koristne možnosti uporabe za to novo tehnologijo. Morda bo nekaj njenih zmogljivosti sčasoma, ko bo na voljo vsem, omejenih ali pa bodo njene napake in nenatančnosti postale vpadljive šele, ko bo dostop dobilo na milijone uporabnikov. Četudi se bo izkazala tako enkratno, kot nakazuje demo različica, in bo oklestila toliko tehničnih delovnih mest, raje ne računajte, da bo Sora demokratizirala zabavo.

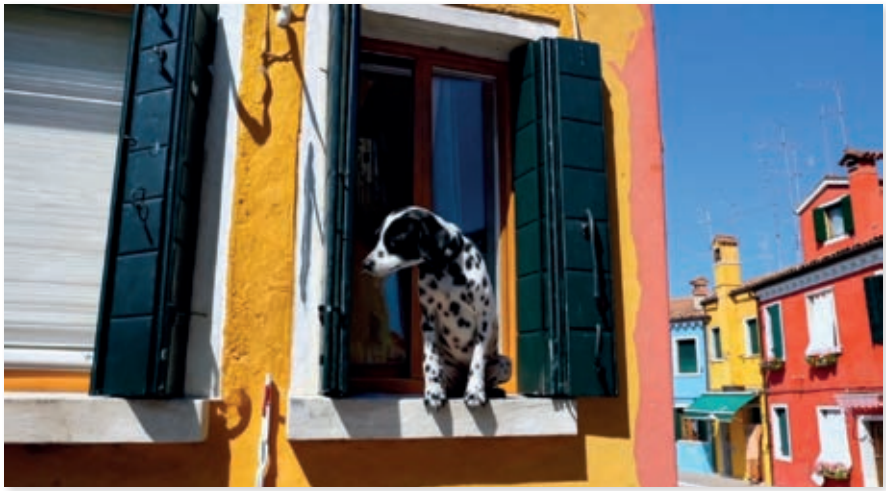
Ko se bo ta tehnologija razvijala, se utegne že kmalu zgoditi, da si bo Sora izmislila celovečerno grozljivko v slogu Jordana Peeleja, vendar to ne pomeni, da bodo tovrstni ukazi prinesli nekaj, kar bi si morali ogledati, razen seveda kot zanimivo novost. Ko bodo ljudje dobili orodje, s katerim bodo lahko naredili vse, kar bi radi videli, bodo ugotovili tudi, kako zahtevno je posneti film ali predstavo, ki bi bila zanimiva še komu – kaj šele nekaj, kar bi vsi radi videli.

Zmožnost izpolnjevanja ukazov ne pomeni nujno zmožnosti, da dosegamo pričakovano kakovost. Če bi uporabnik prosil Soro, naj posname čuden in smešen film o vzponu in padcu blackberryja – ob predpostavki, da bo nekoč znala posneti videe, daljše od 60 sekund –, kakšne so možnosti, da bi model izbral enake dialoge, tempo zgodbe in razvijanje glavnih igralcev kot v lanski čudaško kritični filmski uspešnici o prvem pametnem telefonu? Kako bi upodobila svet, tako poseben kot v lanski filmski komediji Poletni kamp, ne da bi ji pomagale življenjske izkušnje ekipe piscev in režiserja?

Kinematografska čarovnija, ki jo občuti večina gledalcev, je dosežek sodelovanja številčne ekipe z vizionarjem na čelu. Nastane po izločanju in nadgrajevanju celega kupa osnutkov scenarija, spontanosti in svežega navdiha v montaži. Če aplikacija zna generirati čedno žensko v Tokiu, to še ne pomeni, da zna narediti lep film o ženski iz Tokia. Morda zmore prikazati začudenje animirane pošasti na Pixarjevi kakovostni ravni, toda ali si zna izmisliti zgodbe, ki takšna občutenja vzbudijo tudi pri gledalcih?

Četudi so nekateri pomisleki o kakovosti filmov, generiranih z UI, upravičeni, je povpraševanje po njih za zdaj mlačno. Po nedavni raziskavi *Morning Consulta*, pri kateri je sodelovalo 2.201 odraslih Američanov, gledalce še vedno dvakrat bolj zanimajo filmi in televizijske oddaje, ki jih v celoti ustvarijo ljudje kot izdelki UI. In to še preden jih je občinstvo sploh imelo priložnost videti, če odštejemo lanske nerodne epizode South Parka, ki jih je pripravila umetna inteligenca.

Če bo zdaj, ko smo seznanjeni z demo različico Sore, katero od hollywoodskih velikih imen začelo razmišljati o projektih, generiranih z umetno inteligenco, potem bi ta mogoče prej morala nadomestiti katerega od šefov, kot se je pred kratkim pridušal igralec, producent in scenarist Joel Kim Booster. ◀



△ Dva primera iz video "posnetkov", ki jih je ustvarila Sora.

Kakšen je donos naložbe v generativno umetno inteligenco

V informacijsko-tehnološkem sektorju delam več kot 25 let in doživel sem razcvet internetnih podjetij, vzpon pametnega telefona in nenaden prehod na hibridno delo oziroma na delo od doma zaradi pandemije. A kar zdaj prinaša revolucija zaradi generativne umetne inteligence (UI), je nekaj čisto drugega od dosedanjih izkušenj.

David McCurdy, *Fast Company*

V Insightu, podjetju za povezovanje rešitev in sistemov, že poldrugo leto raziskujemo nove možnosti, odkar je OpenAI javnosti omogočil uporabo ChatGPT in pretresel tako upravne odbore kot celotne panoge, od proizvodnje in zdravstva do akademskih krogov. OpenAI je tudi poročal, da kar 92 odstotkov podjetij na seznamu 500 največjih, ki ga sestavlja revija *Forbes*, tako ali drugače uporablja njegovo platformo. A še vedno ostaja veliko vprašanj.

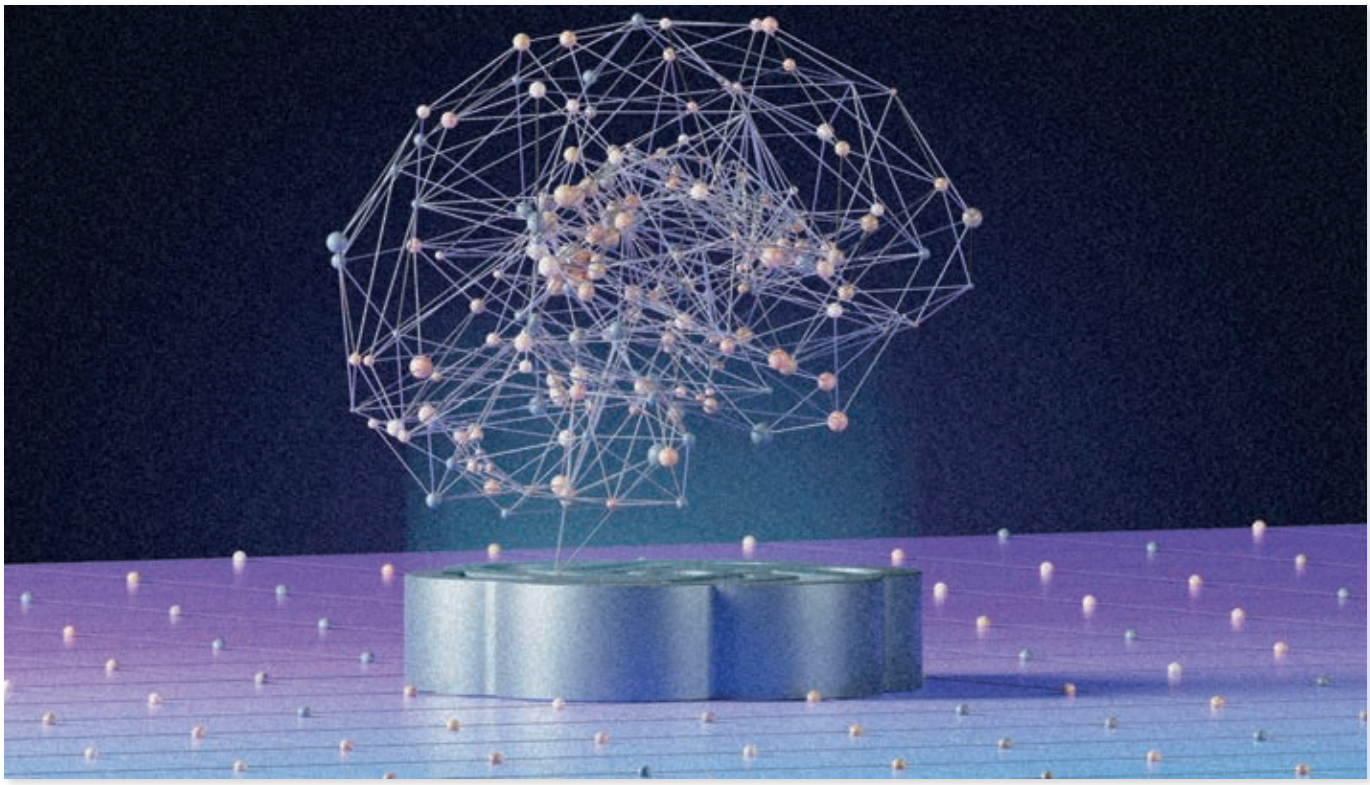
Po naših raziskavah je 90 odstotkov poslovnežev prepričanih, da bo generativna umetna inteligenca nadgradila širok razpon funkcij in nalog, kar pomeni, da bi ta tehnologija lahko koristila skoraj vsem ljudem in opravilom. To potrjujejo tudi raziskave ameriškega podjetja za tržne raziskave *Harris Poll*, ki kažejo, da 74 odstotkov poslovnih vodij uporablja generativno umetno inteligenco za analizo podatkov, 66 odstotkov kot pomočnika pri opravilih, 63 odstotkov pa si jih z njo pomaga pri pisanju poročil in predstavitev.

Tehnološka panoga je trenutno žejna podatkov. Večina podjetij, s katerimi sodelujemo pri uporabi novih digitalnih rešitev, to z generativno UI še vedno počne poskusno. Mogoče načine uporabe še vedno odkrivajo s poskusi in sprotnim učenjem, številni poslovneži postavljajo enaka vprašanja.

Vse je osredotočeno na donos naložbe in po naših raziskavah skoraj dve tretjini uporabnikov navajata, da so za to zadolžili svoj oddelek.

A tehnologija se spreminja tako hitro, da je ocenjevanje sedanje in prihodnje donosnosti izjemno zahtevno. V nadaljevanju je opisanih nekaj načinov, kako poslovni vodje preverijo, kaj deluje med uvajanjem njihove generativne UI in, kar je mogoče še pomembnejše, kaj ne.

Ugotovite, kako generativna UI lahko pomaga pri konkretnih poslovnih izzivih.



Trdno sem prepričan, da se je ne glede na panogo treba vprašati, zakaj uvesti in uporabljati generativno UI. Katere težave lahko premagamo z njo? Kaj je razlog za uporabo te tehnologije kot najboljše rešitve (oziroma zakaj je naložba v generativno UI pomembna za dosego naših ciljev)?

Kljub zmogljivostim – od analize podatkov in zasnove izdelka do besedil in personaliziranih klepetalnih robotov – generativna UI ni univerzalna rešitev za vse težave. In prav gotovo ni brez napak. Obstoječa digitalna orodja, odvisno od poslovnega izziva, resda učinkoviteje rešijo nekatere težave. Oziroma še verjetneje je, da bo nujna povezava velikih jezikovnih modelov z drugimi tehnologijami (bolj tradicionalnimi modeli UI in s strojnim učenjem), sicer njena uvedba ne bo pripomogla k smiselni uporabi v vsem podjetju.

Na nas se nenehno obračajo stranke, ki potrebujejo pomoč pri uvajanju napredne analitike v lastne podatkovne zbirke. Generativna UI lahko opravi napredno analitiko, vendar je omejena glede podatkov, ki jih zmore predelati hkrati. Za velike količine podatkov in repozitorije so veliko primernejše druge tehnologije, pomagajo pa tudi pri razdeljevanju pravih podatkov pravim uporabnikom, s čimer se izognemo halucinacijam UI.

Pred odločitvijo za najprimernejšo tehnologijo je treba torej oceniti količino in zahtevnost podatkov. Za manjše in preprostejše podatkovne zbirke bi bila idealna generativna UI. Če se zavedate njenih omejitev in izberete ustrezno tehnologijo, podjetju zagotovite učinkovito analizo in mu s tem pomagate sprejemati pretehtane poslovne odločitve.

Pri generativni UI se mi najbolj vznemirljiva zdi njena lastnost, da izboljša produktivnost, ker so postopki so z njo učinkovitejši, saj pogosto pripomore k preobrazbi in osvežitvi delovnega procesa.

Večina vodstvenih delavcev si postavlja vprašanja, kot so:

- Kako generativna UI mojim podrejenim prihrani čas pri banalnih nalogah?
- Kako generativna UI pripomore k boljši storitvi za stranke?
- Kako nam generativna UI pomaga pripraviti novo ponudbo, kot pred letom dni še ni bilo mogoče?

V nedavni raziskavi massachusetskega tehnološkega inštituta so ugotovili, da so zaposleni ob pomoči klepetalnih robotov kot asistentov kar za štiri desetine skrajšali čas za nekatera opravila, kot sta pisanje spremnih pisem in uvidevnih elektronskih sporočil ter sestavljanje analiz o stroških in koristih. Tudi v Insightu ugotavljamo podobne rezultate.

Primer je podatkovna zbirka, ki jo poganja generativna UI in predeluje na tisoče pogodb s strankami. Tako zaposleni občutno skrajšajo čas, ki so ga sicer porabili za branje dolgih pogodb. V tem primeru donosnost naložbe v orodje preračunamo glede na čas, energijo in vire, ki jih tehnologija prihrani zaposlenim. Ni bilo težko dokazati, da se naložba v orodje izplača, zato so se tudi naše stranke odločile, da uvedejo svojo različico.

74 odstotkov poslovnih vodij uporablja generativno umetno inteligenco za analizo podatkov, 66 odstotkov kot pomočnika pri opravilih, 63 odstotkov pa si jih z njo pomaga pri pisanju poročil in predstavitev.

Ne gre za nadomeščanje ljudi, saj so ti še vedno ključni ne glede na nalogo. Dodati moram, da bi koncept »odstranjevanja posrednika« lahko predstavljal prelomno spremembo tako za zaposlene kot stranke – in bi bil koristen dejavnik pri presojanju donosnosti.

Vse bolj je razširjena nekakšna naveličanost; odločevalce lahko premaga malodušje in ne morejo nadaljevati pobud za preobrazbo, saj so možnosti za uporabo neskončne, sama tehnologija pa se prehitro spreminja. Katero pot izbrati, če greste lahko v skoraj vse smeri, vaši viri za naložbe pa so omejeni?

Generativna tehnologija ni statična, nenehno se uči in razvija. Tega se je treba zavedati, sploh ker bi nas lahko zavedel pogovorni, intuitivni tip klepetalnega vmesnika, kot ga lahko spoznamo v orodjih, podobnih GPT. Meni se zdi bolj vznemirljivo tisto, kar dejansko poganja ta vmesnik.

Ker nista več nujna ročno vnašanje in sodelovanje, orodje GPT z naprednimi zmogljivostmi UI pomaga avtomatizirati ponavljajoče se naloge, s čimer prihrani čas in denar. Vgraditi ga je mogoče tudi v druge tehnologije za večjo učinkovitost, na primer v robotsko avtomatizacijo delovnih procesov in razumevanje naravnega jezika. Delovni procesi se tako prečistijo, splošna učinkovitost podjetja in drugih organizacij pa se izboljša, kar prinaša stokrat večjo vrednost od uporabe na vidni ravni.

Poleg tega se bo vmesnik, kot ga poznamo danes, v prihodnosti nedvomno spreminjal. Ljudem bo pomagal doseči, kar jim je bilo nekoč nedosegljivo. Ravno na to bi ljudje morali postati pozornejši in prav na tem področju se lahko nadejamo največje donosnosti naložbe.

Zdaj smo šele na začetku te poti in potenciali so v veliki meri še neizkoriščeni. Večina podjetij dejansko ne zmore niti okvirno predvideti čiste donosnosti naložbe, ki jo bo nekoč omogočila generativna UI. A tega ne vidim kot ovire, temveč poslovne voditelje spodbujam, naj ta prazni prostor dojemajo kot priložnost za ustvarjalno odkrivanje in iskanje novih načinov uporabe, kot si jih niso znali niti predstavljati. ◀

David McCurdy je vodja strokovnjakov za arhitekturo podjetij in tehnologijo v Insight Enterprises.

Kaj je inženir kozmične resničnosti?

Novo Amazonovo poročilo o delovnih mestih, povezanih z umetno inteligenco, je odškrnilo okno v prihodnost dela

A. J. Hess, *Fast Company*

Medtem ko po trgu z delovno silo pometa val odpuščanja, tiste, ki še imajo službo, skrbi, da jih bo nadomestila umetna inteligenca (UI). In to utemeljeno. Na svetovnem gospodarskem forumu so lani ocenili, da kar tri četrt podjetij dejavno razmišlja o uvedbi tehnologij, kot so velike podatkovne zbirke, računalništvo v oblaku in umetna inteligenca, in da bo zaradi avtomatizacije do leta 2027 ukinjenih 26 milijonov delovnih mest.

Poslovneži z velikimi naložbami v UI strahove zaradi nove tehnologije pogosto poskušajo pomiriti z argumentom, da bo ta pripomogla k odpiranju novih delovnih mest in ne bo le zapirala starih.

Najbrž ni presenečenje, da se s tem strinja tudi Victor Reinoso, globalni direktor filantropije za izobraževanje pri Amazonu: »Ko je nastal Amazon, profili, kot jih zaposlujemo danes, sploh niso obstajali,« je pojasnil. »Podjetje se zaveda, da so inovacije nenehne, in ko iščemo načine za ugoden vpliv na skupnost, ji želimo prenesti ravno to zavedanje, da danes šele slutimo, kakšne bodo poklicne poti prihodnosti, vendar bodo tudi zanje nujne temeljne veščine in pismenosti.«

Reinoso je med drugim zadolžen za nadzor nad pobudami za najmlajše, s katerimi naj bi otrokom in mladostnikom iz neprivilegiranih okolij olajšali dostop do računalniškega opismenjevanja. Novembra lani je Amazon napovedal novo pobudo, imenovano *AI Ready* (Pripravljeni na UI), ki obeta, da bo do leta 2025 dvema milijonoma ljudi ponudila brezplačno izobraževanje o umetni inteligenci in drugih veščinah. Pozneje je Reinosova ekipa objavila tudi, da po njeni novi raziskavi več kot šest desetin učiteljev meni, da bodo za njihove učence veščine, povezane z UI, nujne za dobro plačana delovna mesta.

»Domnevamo, da bo znanje o UI olajšalo poklicno pot,« je izjavil Reinoso. »Računalniško izobraževanje se je dolgo časa osredotočalo na razvijanje.« Danes poudarek ni več le na računalniškem programiranju. »Trenutka točka preloma z UI in s tehnologijo nasploh izpostavlja, da ni pomembno, za kakšen poklic se boste odločili, temveč bo pomembna osnovna uporaba UI in tehnologije. Zato se trudimo poiskati načine, kako otrokom predstaviti, da programiranje ni edini način za dobro znanje računalništva oziroma edini način za praktično uporabo tega znanja.« Podatki kažejo, da se spremembe že kažejo na delovnih mestih. Medtem ko je programiranje nekoč veljalo za najzanesljivejšo in najbolje plačano delovno mesto, so ravno razvijalci programe opreme predstavljali glavnino lanskega vala odpuščanja.

Poleg tega lansko poročilo PwC in Svetovnega gospodarskega foruma opozarja, da bi

zaradi pomanjkanja kandidatov s primernimi veščinami za poklice prihodnosti do leta 2030 lahko ostalo praznih kar okoli 85 milijonov delovnih mest – zaradi tega bi gospodarstvo lahko bilo ob 8,5 bilijona letnega prihodka. Če ti podatki držijo, bi bilo pametno, da podjetja, kot je Amazon, vlagajo v dodatno izobraževanje in usposabljanje mladine, ki bo kmalu prišla na trg delovne sile.

Amazon je objavil tudi poročilo futuristke Tracey Follows, ki napoveduje, katera delovna mesta, povezana z UI, se napovedujejo. Followsova in Amazon sta ugotovila, da bodo med profili prihodnosti analitik preciznega kmetijstva, producent virtualnega turizma, specialist za restavriranje, inženir kozmične resničnosti in umetnointeligenčni zdravstveni tehnik.

Tu je opis petih profilov, ki temeljijo na UI in za katere Amazon pričakuje, da bodo zaželeni v prihodnosti.

Med profili prihodnosti bodo analitik preciznega kmetijstva, producent virtualnega turizma, specialist za restavriranje, inženir kozmične resničnosti in umetnointeligenčni zdravstveni tehnik.

▶ **Analitik preciznega kmetijstva.** Kmetijstvo postaja vse bolj tehnična znanost in analitiki, ki so usposobljeni tudi za delo z UI, bodo to panogo še občutneje spremenili, povečali donose, a hkrati učinkovito porabljali vire za proizvodnjo hrane. Modeli UI bodo analitikom pomagali napovedovati in omejevati vpliv podnebnih sprememb, saj bodo izbrali najprimernejše pridelke in dodeljevali vire, kot so voda in umetna gnojila. Umetna inteligenca bo integrirana z roboti, ki znajo saditi in žeti, hkrati pa v realnem času spremljati rast pridelka.

▶ **Producent virtualnega turizma.** Predstavljajte si, da načrtujete poletne počitnice in si lahko privoščite predogled z virtualno resničnostjo. Producenti virtualnega turizma bodo poskrbeli za izkušnje potopljene virtualne resničnosti ob pomoči UI in stranki razkazali potovalne cilje ter aktivnosti. Urejali in dopolnjevali bodo virtualno resnično vsebino, izpostavili najnovejše novice in kulturno dogajanje v različnih predelih. Sklepali bodo partnerstva s turističnimi organizacijami za trženje potovalnih izkušenj.

▶ **Specialist za rokodelsko restavriranje.** Mojstri, specializirani za restavriranje, bodo izkoristili napredne sisteme UI za popravila in obnovo luksuznih izdelkov, od visoke mode do starinskega pohištva. Ti specialisti bodo z UI prepoznali prvotne materiale in jih poiskali, nato pa ocenili najučinkovitejšo tehniko restavriranja za ohranitev izvirne estetike, zgodovinskega pomena in vrednosti konkretnega predmeta.

▶ **Inženir kozmične resničnosti.** Strokovnjak na področju medzvezdnih študij bo ob pomoči človeške domišljije in naprednih veščin UI vizualiziral oddaljene in včasih nevidne dele vesolja, od zvezd in planetov do celotnih galaksij. Z globokim razumevanjem astrofizike, kozmologije in astronomije bo z UI generirane podatke postavil v kontekst in povzetek prenesel v vizualno osupljive simulacije, s katerimi se bodo drugi lahko učili o vesolju.

▶ **Umetnointeligenčni zdravstveni tehnik.** Umetna inteligenca bo igrala pomembno vlogo kot pomočnica zdravstvenim tehnikom in drugim zdravstvenim delavcem pri diagnozi, spremljanju življenjskih znakov, sledenju zdravilom in načrtih za okrevanje bolnikov. Od dobrega zdravstvenega tehnika se že od nekdaj pričakujejo dobre komunikacijske veščine in zanesljivo poznavanje medicine. Zdravstveni tehniki prihodnosti pa bodo poznali orodja UI, da si bodo znali razložiti vpogled, kot ga omogoča UI, in zapletene analize v zvezi z diagnozo in zdravljenjem zmogli prenesti v jezik, ki ga bodo razumeli tudi bolniki.

Iz teh novih vlog lahko izluščimo, da temeljijo na manj tehničnih poklicih, kot obstajajo trenutno, in so prilagojene za uporabo UI, da bi izboljšale učinkovitost, vsebine in materiale ter tudi analizirale podatke. Strokovnjaki najpogosteje napovedujejo, da bo revolucija zaradi UI oziroma razvoj uporabnikom prijaznih velikih jezikovnih modelov, kot je ChatGPT, pripomogla, da bo širok krog delavcev – tudi tisti, ki ne delajo v tehnološki panogi – uporabljal avtomatizirana orodja. V nedavnem poročilu *McKinseyja* so se izrazili tako: »Generativna umetna inteligenca pospešuje avtomatizacijo in vzpostavlja povsem nov izbor poklicev. Na delovna mesta, ki jih opisujeta Amazon in Follows, bomo morali še počakati, vendar podjetja, kot je Amazon, že izkoriščajo UI in ukinjajo nekatera opravila, ki se zdijo idealni kandidati za to, le da se vse skupaj dogaja manj spektakularno. Lani je v skladiščih uvedel nove robotske roke, sortirne naprave poganja UI. Takrat je podjetje poudarjalo varnostni vidik za delavce, vendar je glavni namen avtomatiziranega robotskega sistema povečati produktivnost in skrajšati dobavni rok, s čimer bi čas izpolnitve naročila skrajšali za četrtino, zaloge pa obrnili 75 odstotkov hitreje kot doslej.«

Reinoso je na vprašanje o strahovih delavcev, ko spremljajo, kako UI opravlja delo, ki so ga nekoč oni, odgovoril sočutno, skoraj filozofsko: »Naše izkušnje so takšne, da neznano prinaša negotovost, negotovost pa včasih človeka spravlja v stisko.« ◀

Strojno pozabljanje

Na spletu objavljam že več kot dve desetletji. Kot najstnik sem za seboj pustil kup blogov in objav na družbenih omrežjih, od banalnih do takih, ki me danes spravljajo v zadrego. Zdaj pa kot novinar med drugim pišem zgodbe o družbenih omrežjih, zasebnosti in umetni inteligenci (UI). Ko mi je torej ChatGPT povedal, da moje pisanje nemara vpliva na njegove odgovore drugim uporabnikom, sem se jadrno namenil zbrisati svoje podatke iz njegovega spomina. A kot sem kmalu ugotovil, gumb za brisanje ne obstaja.

Shubham Agarwal, *New Scientist*

Klepetalni roboti, ki delujejo na podlagi umetne inteligence in so jih učili s podatkovnimi zbirkami, vključno z ogromnim številom spletnih strani in člankov, nikoli ne pozabijo, kar so se naučili.

To pomeni, da stvaritve, sorodne ChatGPT, razkrivajo občutljive osebne podatke, če so se ti pojavili kje na spletu, in da se bodo podjetja, ki stojijo za temi modeli UI, morala pošteno potruditi za uresničevanje zakonov v zvezi s 'pravico do pozabljenja', po katerih morajo podjetja na zahtevo uporabnika odstraniti njegove osebne podatke. A hkrati to pomeni, da smo brez moči, ko bi bilo treba ustaviti hekerje, ki na proizvode UI vplivajo tako, da ji med učenjem podtikajo lažne podatke in dajejo zlonamerne ukaze.

To je hkrati odgovor, zakaj si številni računalničarji prizadevajo, da bi UI naučili tudi pozabljati, kar je izjemno težko. Kljub temu se že pojavljajo rešitve za 'strojno odučenje' (dobesedni prevod *machine unlearning*). Ta prizadevanja bodo morda pomembna še iz drugih razlogov, ne le zaradi pomislekov o varovanju zasebnosti in dezinformacij. Če si resno prizadevamo za sisteme UI, ki se bodo učili in razmišljali kot ljudje, jih bomo morda morali sestaviti tako, da bodo zmožni tudi pozabljati.

Novo generacijo klepetalnih robotov, ki jih poganja UI, kot sta ChatGPT in Googlov Bard, ki se na naše ukaze odzoveta z besedilom, podpirajo veliki jezikovni modeli. Učijo jih s kupi podatkov, ki jih večinoma pridobijo z interneta, od objav na družbenih omrežjih do okoli četrt milijona knjig in skoraj vseh javno dostopnih podatkov, vključno z novičarskimi stranmi in Wikipedijo.

Iz tega materiala se naučijo prepoznavati statistične vzorce, kar pomeni, da lahko napovedo, katera beseda se bo najverjetneje kot naslednja pojavila v stavku. Delujejo osupljivo dobro in pripravijo smiseln odgovor na vsako vprašanje.

Žal se klepetalni roboti z UI lahko le učijo, pozabljajo pa ne. Veliki jezikovni modeli odgovore sestavijo na podlagi zbranih podatkov, zato sploh ne bi bilo preprosto doseči, da bi nekatere podatke zbrisali s svojega seznama znanja – pri iskalnikih, recimo Googlu, je to preprosto. Prav tako posameznik ne more preveriti, kaj točno aplikacija z UI ve o njem, je pojasnil David Zhang, raziskovalec UI in inženir v avstralski državni znanstveni agenciji Csiro.

Hekerji lahko klepetalne robote prelisičijo, da delujejo drugače, kot so si zamislili njihovi snovalci. S tem nastanejo velike težave z varovanjem zasebnosti, kot so Zhang in njegovi kolegi potrdili v nedavni raziskavi. Z njo so osvetlili, kako težko bodo podjetja na področju UI spoštovala 'pravico do pozabe', kar je Evropska unija leta 2014 razglasila za eno od človekovih pravic.

Po splošni uredbi EU o varstvu podatkov (GPDR) imajo ljudje pravico zaprositi, da njihove podatke odstranijo iz arhiva neke organizacije ali podjetja. Na internetu to pravico običajno uveljavljajo tako, da izbrišejo vsebino, ki so jo objavili na družbenih omrežjih ali na straneh javnih občil, poleg tega pa podjetja, kot je Meta, lahko prosijo za povzetek podatkov, ki so jih zbrala o njih. A takšne rešitve v primeru klepetalnih robotov ne delujejo, je pojasnil Zhang. »Posameznik izgubi pravico do izbrisa, saj ne obstaja postopek za izbris podatkov iz spomina modelov UI.«

Podjetja, ki razvijajo klepetalnike z UI, bodo morala poiskati rešitev, sploh ker velike jezikovne modele začenjajo učiti z občutljivejšimi podatki, recimo zdravstvenimi, podatki iz elektronskih poštnih nabiralnikov in tako dalje, je izpostavil Florian Tramer, informacijski strokovnjak, zaposlen v ETH v Zürichu.

Še bolj pereče pa je, da so klepetalni roboti ranljivi za napade, v katerih so informacije skrite v podatkih za učenje, in tako model pripravijo do nepredvidenega obnašanja. Strokovnjaki za varnost so dokazali, da je s tehniko, imenovano posredno vstavljanje ukazov, mogoče doseči, da klepetalni robot na daljavo zažene kode na uporabnikovem računalniku ali ga prosi za podrobnosti o bančnem računu. Britanska obveščevalna agencija GCHQ, je opozorila na to težavo, saj se bo v prihodnje še stopnjevala.

Glede na vse tveganje pa je dobra novica, da so že začeli tuhtati, kako selektivno brisati podatke iz podatkovne zbirke UI. Žal je to zelo zapleteno.

Umetno povzročena amnezija

Podjetja tako trenutno lahko le gasijo požar, kar pomeni, da svojo storitev sprogramirajo za preprečevanje dostopa do določenih podatkov in zavrnitev odgovora. ChatGPT se je opravičil, da mi ne more pomagati, ko sem ga prosil, naj sestavi osebni dosje o meni. Luciano Floridi, direktor centra za digitalno etiko, ki deluje na univerzi Yale, meni, da je takšen pristop lahko do neke mere učinkovit, vendar ciljni podatki ostajajo, je poudaril, kar pomeni nenehno tveganje, da se bodo pojavili v odgovorih kot posledica napak ali zlonamernega posredovanja.

Potuhtati moramo, kako izbrisati podatke, ne da bi bilo treba UI učiti od začetka.

Žal je najprimernejši način za povzročitev pozabe pri velikih jezikovnih modelih, to je, da se jih 'oduči' in odstrani izbrane podatke, zelo nepraktičen, saj zahteva tedne računalniškega programiranja. Ugotoviti torej moramo, kako odstraniti ali vsaj skriti določene podatke, ne da bi morali model učiti od

začetka, je povedal Yacine Jernite iz podjetja za UI Hugging Face. »To je res poseben raziskovalni problem.«

Iskanje rešitve se je začelo leta 2014, ko se je Yinzhi Cao, ki je takrat delal na univerzi zveznega okrožja Kolumbija v New Yorku (danes je zaposlen na univerzi Johnsa Hopkinsa v Baltimoru), domislil preproste zvijače: namesto da bi algoritem učili z vsemi razpoložljivimi podatki, bi to, česar se je naučil, lahko razdelili na vrsto manjših segmentov; ko bi uporabnik zaprosil za odstranitev podatkov, bi morali skrbniki spremeniti le segment s temi podatki, kar bi občutno znižalo stroške dela.

Načelo je dobro, vendar je metoda delovala le pri bistveno preprostejših modelih od velikih jezikovnih modelov, na katerih temelijo današnji umetnointeligenčni klepetalni roboti. V novih jezikovnih modelih so deli podatkov tako tesno prepleteni, da jih ni mogoče tako ločevati, priznava Cao.

Leta 2019 je Nicolas Papernot s kanadske univerze v Torontu skupaj s kolegi predlagal drugačno metodo s kratico sisa (*sharded isolated sliced aggregated* – razdrobljeno, ločeno, razrezano, povezano), ki deluje tudi pri zapletenejših umetnih nevronskih mrežah, skritih za številnimi velikimi jezikovnimi modeli. S to metodo je lažje poiskati in izbrisati posamezne podatkovne točke. Podatkovne sklope se razdeli na več manjših kosov in model učijo z vsakim posebej, preden združijo rezultate. Napredek pri učenju shrani podobno kot kontrolne točke v videoigri in nato napreduje k naslednjemu kosu. Ko dobi prošnjo za izbris, se lahko vrne na kontrolno točko, odreže tisti kos s podatki, ki jih mora izbrisati, in vnovič začne učenje s te točke.

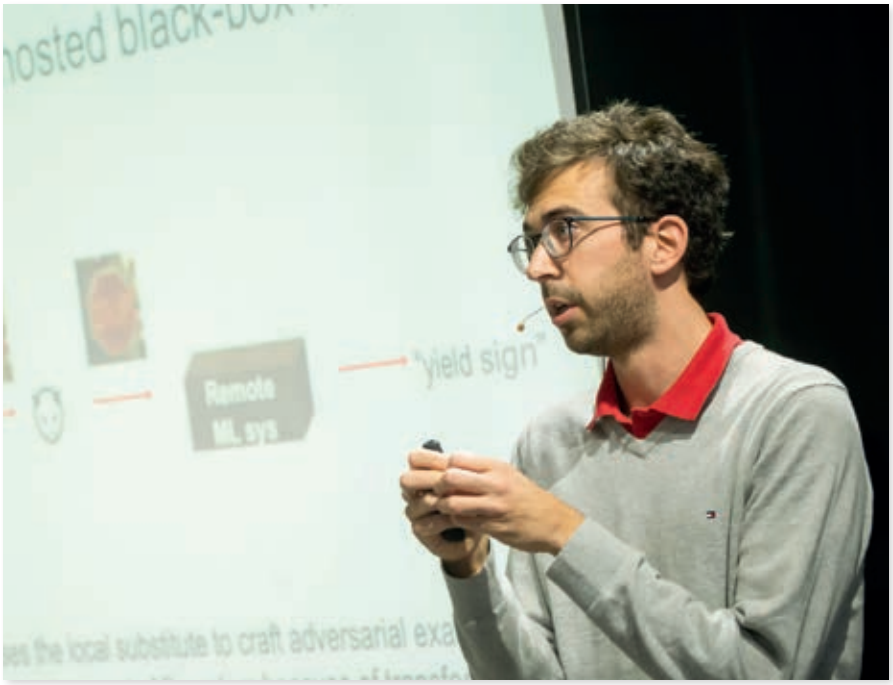
Ko je Papernotova ekipa preverila učenje po sistemu sisa z dvema velikima podatkovnima sklopoma – v prvem so bile podrobnosti o približno 600.000 domačih naslovih, v drugem pa zgodovina 300.000 nakupov –, je ugotovila, da je tako občutno pospešila učenje v primerjavi z učenjem čisto od začetka.

Tudi metoda sisa ima svoje težave, na primer to, da bi lahko izrazito neugodno vplivala na učinkovito delovanje UI. A goli princip je navdihnil številne posnemovalce. Leta 2021 je Min Chen iz nemškega Helmholtzevega centra za informacijsko varnost CISPA metodično razdelil in spet združil podatke, torej ne naključno tako kot sisa, kar je olajšalo njihovo brisanje, ne da bi preveč ogrozilo samo delovanje UI.

Druge skupine strokovnjakov ubirajo nekoliko drugačne pristope. Ker brisanje podatkov včasih močno zavira učinkovito delovanje modela strojnega učenja, so se nekateri raje odločili za skrivanje ali zatemnjevanje spornih podatkov, da jih ni mogoče pridobiti. Raziskovalci v Microsoftu in na državni univerzi v Ohiu so v podatke, s katerimi učijo model, vnesli šum oziroma naključno nihanje podatkov, tako da težje prepozna ciljne vzorce in so rezultati manj specifični ter zanesljivi. »To je teoretično jamstvo, da model ne bo razkril podrobnih podatkov o posamezniku, ki so vključeni v podatke za učenje,« je pojasnil član te ekipe Xiang Yue.

Z generalizacijo podatkov se delno okrni statistična zmožnost učenja, zaradi katere so klepetalni roboti tako zmogljivi. Tej težavi se je Minjoon Seo skupaj s sodelavci z južnokorejskega inštituta za znanost in tehnologijo poskušal izogniti s *post hoc* pristopom. Po tej metodi naj bi zmanjšali vpliv nekega podatka na algoritem, namesto da ga izbrišejo, tako, da ga klepetalni robot nikoli ne omeni.

Žal je ‚odučenje' in odstranitev izbranih podatkov iz velikih jezikovnih modelov zelo nepraktično, saj zahteva tedne računalniškega programiranja.



△ **Leta 2019 je Nicolas Papernot s kanadske univerze v Torontu predlagal metodo s kratico sisa (sharded isolated sliced aggregated – razdrobljeno, ločeno, razrezano, povezano). S to metodo je lažje poiskati in izbrisati posamezne podatkovne točke.**

Velja za eno najobetavnejših na področju, saj je zanjo potrebne precej manj računske podlage in manj časa, deluje pa na rahlo starejši ustreznici zasnove od te, na kateri temelji ChatGPT.

V resnici v tekmi za razvoj metode, po kateri bi jezikovni modeli znali kaj tudi zamolčati, še ni favorita. Google je celo organiziral natečaj in ponudil nagrado tistim, ki se bodo domislili učinkovite rešitve, kar ne dokazuje le pomembnosti tega izziva, temveč tudi, da bo že kmalu bolj jasno, katere metode bodo pripomogle, da bomo dobili novo generacijo velikih jezikovnih modelov, zmožnih pozabljanja tistega, česar so se naučili.

Selektivni spomin

V ta cilj je vredno vlagati trud, saj bi lahko vplival še na marsikaj drugega razen na varovanje podatkov in zlorabo klepetalnih robotov, razmišlja Ali Boyle, filozofinja na londonski ekonomski fakulteti, ki se ukvarja z UI. Čeprav ima človekovo nagnjenost k pozabljanju marsikdo za kognitivno napako, je včasih dobrodejna, saj si nam res ni treba zapomniti vsakega podatka, na katerega naletimo. S pozabljanjem pripomoremo k učinkovitejšemu priklicu koristnih spominov. Enako morda velja za sisteme UI, pravi Boylova.

Trdi, da so to prikazali leta 2017, ko so raziskovalci v Googlovem DeepMindu razvili umetno inteligenco, ki je lahko igrala več Atarijevih videoiger. Učinkovito je generalizirala svoje znanje, ker se ni učila v realnem času iz izkušenj, temveč je shranjevala spomine na odigrane igre in jih pozneje lahko priklicala ter se učila iz njih. Tako seveda deluje spomin. A raziskovalci so nato model nadgradili, da je prednost dajal shranjevanju in priklicu nepričakovanih dogodkov, ki se niso ujemali z njegovimi napovedmi – preostale podatke pa je pozabil –, in sistem je deloval učinkoviteje.

Pozabljanje ni napaka v zasnovi, je lastnost učinkovitega in gladko delujočega spominskega sistema.

Iz vsega napisanega sledi, da selektivno pozabljanje umetne inteligence lahko izboljša njeno delovanje. Ključno je najti ravnovesje med prevelikim in premajhnim spominom. A če je končni cilj strokovnjakov razviti sisteme, ki se bodo učili in razmišljali kot ljudje, kar je gotovo eden prvotnih namenov tega področja, potem jih morajo zasnovati tako, da bodo zmožni selektivnega pozabljanja. »Pozabljanje ni napaka v zasnovi,« je poudarila Boylova, »je lastnost učinkovitega in gladko delujočega spominskega sistema.« ◀

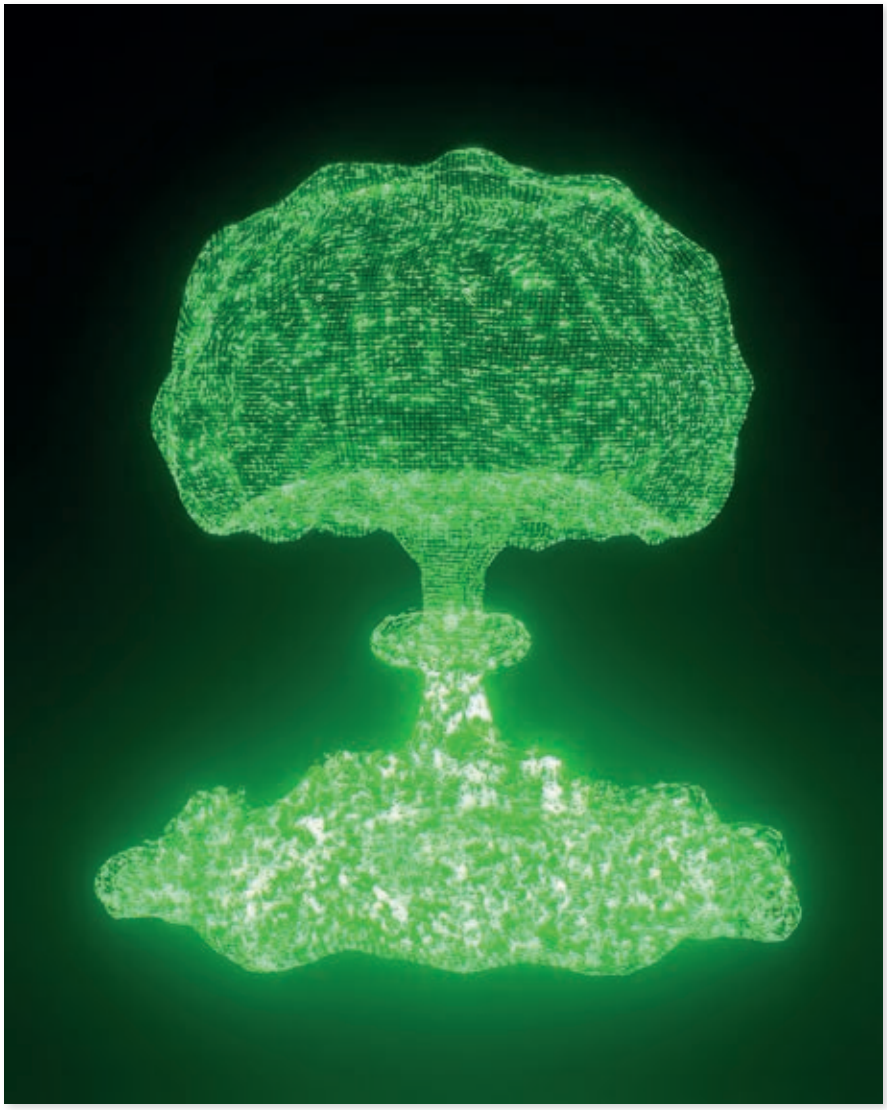
Jedrske šifre niso primerne za umetno inteligenco

Vse od jedrske bombe ni bilo tehnologije, ki bi tako močno spodbudila apokaliptično domišljijo kot umetna inteligenca (UI). Predstavljajmo si, kako hudo bi se lahko zapletlo, če bi inteligenca, ki bi bila šele v povojih, svet in lastne cilje razumela malenkost drugače od svojih človeških stvarnikov. Prvi koraki so potrjeno že narejeni – postopno vgrajevanje UI v najbolj uničevalne tehnologije.

Ross Andersen, *The Atlantic*

Največje svetovne vojaške sile so začele tekmovati tudi v vgrajevanju UI v oboroževanje in bojevanje. Trenutno to večinoma pomeni, da algoritmom predajajo nadzor nad posameznim orožjem in rojem dronov. Nihče ne vabi UI, naj zasnuje veličastno strategijo ali se pridruži sestanku generalštaba. A ista zapeljiva logika, ki je pospešila jedrsko oborožitveno tekmo, bi v nekaj letih lahko UI potisnila navzgor po hierarhiji poveljevanja. Kako hitro bi se to lahko zgodilo, je delno odvisno od hitrosti tehnološkega napredka; za zdaj lahko rečemo, da je hiter. Kako visoko se bo prebila, pa je odvisno od naše zmožnosti predvidevanja in kolektivne zadržanosti pri ukrepanju.

Jacquelyn Schneider, direktorica pobude za vojne igre in simulacijo kriz v *Hoover Institution*, ki deluje na Stanfordu, mi je pred kratkim pripovedovala o igri, ki jo je razvila leta 2018. V njej se pospešeno odvija jedrski konflikt in ljudje, katerih odziv zasluži največjo pozornost – nekdanji premierji, zunanji ministri in visoki Natovi uradniki – so jo odigrali 115-krat. Ker je jedrska hoja po robu zgodovinsko gledano na srečo redka, igra Schneiderjeve omogoča enega nazornejših vpogledov v to, kakšne odločitve bi sprejemali odgovorni v najbolj tveganih situacijah.



Igra se odvija nekako tako: ameriškega predsednika in člane njegovega kabineta v naglici pokličejo v klet v Zahodnem krilu, kjer jim poročajo o srhljivih dogodkih. Ozemeljski spor se je zaostril in sovražnik premleva, ali naj sproži napad z jedrskim orožjem na Združene države Amerike. Vzdušje v situacijski sobi je napeto. Zagovorniki uporabe sile svetujejo takojšnje priprave na povračilni udarec, a kabinet kmalu izve za neprijeten zaplet. Sovražnik je namreč razvil novo kibernetično orožje in najnovejši podatki obveščevalnih služb kažejo, da lahko vdre v komunikacijski sistem med predsednikom in njegovim jedrskim orožjem. Povelja, ki bi jih poslal, ne bi nujno dosegla poveljnikov, odgovornih za njihovo izvršitev.

V takšnih razmerah ni ugodnih možnosti. Nekateri igralci pristojnost za sprožitev orožja preložijo na vojaške poveljnike na lokaciji orožja in ti morajo potem sami presoditi, ali je povračilni jedrski napad upravičen – to je strašna možnost. Schneiderjeva mi je povedala, da jo najbolj skrbi druga strategija, po kateri so igralci posegli presenetljivo pogosto. Številni igralci, ki se bojijo totalnega razpada hierarhije poveljevanja in izgube nadzora, hočejo popolnoma avtomatizirati možnosti za sprožitev jedrskega orožja. Zagovarjajo opolnomočenje algoritmov, da bi ti določili, kdaj je jedrski povračilni napad upravičen. Umetna inteligenca bi sama odločala, ali bo sodelovala v jedrskem spopadu.

Igra Schneiderjeve je načrtno kratka in napeta. Igralci navodil o avtomatizaciji običajno ne oblikujejo z inženirsko preciznostjo – kako natančno naj bi to potekalo? Bi sploh lahko vzpostavili avtomatiziran sistem, preden bi se kriza zaostrila? Vzgib pa je kljub temu zelo poveden. »Veliko pobožnih želja je

v zvezi s to tehnologijo,« je povedala Schneiderjeva. »Skrbi me, da bodo voditelji, ki se ne zavedajo nezanesljivosti samih algoritmov, ravno z UI želeli zmanjšati negotovost.«

Umetna inteligenca namreč vzbuja privid trezne preciznosti, sploh v primerjavi s človekom, ki je nagnjen k napakam in je za nameček lahko tudi labilen. A najsodobnejši sistemi UI so danes črne skrinjice in ne razumemo povsem, kako delujejo. V zapletenih in pomembnih napetih razmerah morda ne bo mogoče razbrati, kaj si UI predstavlja pod zmago, ali pa se to ne bo ujemalo z našimi predstavami. Umetna inteligenca na najglobljih in najpomembnejših ravni mogoče ne bo razumela, kaj sta Ronald Reagan in Mihail Gorbačov mislila z besedami, da v jedrski vojni ni zmagovalca.

Seveda že obstaja primer avtomatizacije Armagedona. Ko so Združene države Amerike in Sovjetska zveza dobile drugo svetovno vojno, je kmalu kazalo, da že rožljajo z orožjem pred tretjo, in tej usodi sta se obe strani izognili le z vzpostavljanjem infrastrukture za zagotovljeno medsebojno uničenje. Ta sistem temelji na elegantni, vendar srhljivi simetričnosti, a se zamaje vedno, ko ena od strani doseže nov tehnološki napredek. V zadnjih desetletjih hladne vojne je sovjetske voditelje skrbelo, da bi lahko bila ogrožena njihova zmožnost odgovoriti na ameriški jedrski napad, zato so razvili program *Dead Hand*, znan tudi kot Perimeter.

Bil je tako preprost, da bi mu komajda lahko rekli algoritem. Če bi ga v jedrski krizi aktivirali in če bi se prekinila komunikacija med poveljniškim ter nadzornim središčem zunaj Moskve, bi posebna naprava preverila atmosferske razmere nad prestolnico. Če bi odkrila sledi zaslepljujočih bliskov in izrazit dvig radioaktivnega sevanja, bi vse preostale sovjetske izstrelke usmerili v ZDA. Rusija je skrivnostna, ko gre za podatke o sistemu, vendar je leta 2011 poveljnik ruskih enot za strateške izstrelke povedal, da še obstaja in deluje. Leta 2018 pa je nekdanji poveljnik teh enot razkril, da so ga celo izboljšali.

Curtis McGiffin, ugleden predavatelj na letalskem tehnološkem inštitutu, in Adam Lowther, takratni direktor raziskav in izobraževanja na louisianskem tehnološko-raziskovalnem inštitutu, sta leta 2019 objavila članek, v katerem sta predlagala, da bi ZDA morale razviti lastni jedrski program *Dead Hand*. Zaradi novih tehnologij se je skrajšal čas med trenutkom, ko se zazna začetek napada, in zadnjim trenutkom, ko predsednik še lahko ukaže povračilno salvo. Če se bo okno za odločitev še bolj skrčilo, bi bila ogrožena ameriška zmožnost za protiudarec. Njihova rešitev je bila podpora ameriški jedrski obrambi z UI, ki z računalniško hitrostjo sprejema odločitve za izstrelitev raket.

McGiffin in Lowther imata prav glede časovnega okna za odločanje. Na začetku hladne vojne so bili bombniki, kot so jih uporabili nad Hirošimo, prva izbira za začetni napad. Ta letala so potrebovala dolgo časa za prelet od Sovjetske zveze do ZDA, in ker so jih pilotirali ljudje, jih je bilo mogoče odpoklicati. Američani so prek kanadskega dela arktičnega kroga, Grenlandije in Islandije vzpostavili obod radarskih postaj, da bi predsednik prejel opozorilo vsaj eno uro, preden bi se prvi gobasti oblak razcvetel nad katerim od ameriških mest. To je dovolj časa za komunikacijo s Kremljem in za razstrelitev bombnikov, če pa to ne bi delovalo, bi še vedno ostalo dovolj časa za odziv z vsemi sredstvi.

Nekateri zagovarjajo opolnomočenje algoritmov, da bi ti določili, kdaj je jedrski povračilni napad upravičen. Umetna inteligenca bi sama odločala, ali bo sodelovala v jedrskem spopadu.

Medcelinski balistični izstrelki, ki jih je prva izstrelila Sovjetska zveza leta 1958, so skrajšali to časovno okno, v naslednjem desetletju pa so jih na stotine namestili po tleh Severne Amerike in Evrazije. Takšen izstrelek severno poloblo preleti v manj kot pol ure. Za čim boljši izkoristek teh minut sta obe velesili izstrelili flote satelitov, ki bi lahko zaznali poseben infrardeči podpis izstrelitve za spremljanje natančne krivulje in cilja.

Ko so v 70. letih nadgradili jedrsko oborožene podmornice, so v svetovne oceane še bliže svojim ciljem namestili na stotine dodatnih izstrelkov, opremljenih z bojnimi glavami, tako da se je časovno okno za odločanje skrčilo še za pol, na 15 minut ali še manj, kar je grozovito malo časa za pretehtan človekov odziv. Razvijajo novo tehnologijo za izstrelke, vključno s hipersoničnimi izstrelki, ki jih Rusija že uporablja v Ukrajini za hiter napad in izogibanje obrambnim izstrelkom. Tako Rusija kot Kitajska naj bi hipersonične izstrelke želeli opremiti tudi z jedrskimi glavami in ti tehnologiji bi lahko časovno okno vnovič skrajšali za pol.

Preostale minute bodo minile hitro, sploh če Pentagon ne bo mogel takoj ugotoviti, ali izstrelek leti proti Beli hiši. Predsednika bo mogoče treba zbuditi, morda bodo težave z iskanjem pravih šifer za izstrelitev raket. Usodni napad bi lahko bil zaključen brez ukaza za povračilno salvo. Zunaj Washingtona bi se poveljniško in nadzorno središče trudilo iskati naslednjega civilista v hierarhiji poveljevanja, medtem pa bi po ameriških vojaških oporiščih, lokacijah z vojaško obrambo in ključnih infrastrukturnih vozliščih padala toča izstrelkov.

Takšen napad bi kljub vsemu še vedno bil norost, saj bi vsaj del ameriških jedrskih sil najverjetneje preživel prvi val, sploh podmornice. A kot smo v zadnjih letih večkrat izkusili, se na čelu jedrskih sil včasih znajdejo nepremišljeni ljudje. Četudi so zaradi manjšega časovnega okna uničevalni napadi le rahlo mamljivejši, bi si države najbrž želele vzpostaviti sistem za njihovo odvračanje.

ZDA še ni med temi državami. Po objavi članka McGiffina in Lowtherja so generalporočnika Johna Shanahana, direktorja Pentagonovega skupnega središča za umetno

Najsodobnejši sistemi UI so danes črne skrinjice in ne razumemo povsem, kako delujejo. V zapletenih in pomembnih napetih razmerah morda ne bo mogoče razbrati, kaj si UI predstavlja pod zmago, ali pa se to ne bo ujemalo z našimi predstavami.

inteligenco, povprašali o avtomatizaciji in jedrskem orožju. Odgovoril je, čeprav ne pozna bolj vnetega zagovornika UI za vojaško uporabo, kot je on sam, da sta ravno jedrsko poveljevanje in nadzor področji, kjer potegne mejo.

Pentagon si je na splošno z vsemi silami prizadeval avtomatizirati ameriški vojaški stroj. Kot piše v poročilu za leto 2021, je takrat imel odprtih vsaj 685 projektov v zvezi z UI in odtlej tudi nenehno povečuje sredstva zanjo. Vsi projekti niso znani, vendar se počasi izoblikuje vsaj delna podoba ameriških avtomatiziranih sil. Tanki, ki bodo ameriške oborožene sile popeljali v prihodnost, bodo samostojno spremljali razmere in odkrivali grožnje, tako da bodo upravljalci lahko samo pritisnili na osvetljeno točko na zaslonu in izbrisali morebitne napadalce. Nad njimi bodo krožila letala F-16, pilotom pa se bodo v kabini pridružili algoritmi, ki se bodo dobro znašli tudi v zahtevnih bojnih manevrih. Piloti se bodo zato lahko osredotočili na sprožitev orožja in usklajevanje z jatami avtonomnih dronov.

Pentagon je januarja dopolnil svojo doslej nejasno politiko in pojasnil, da bo dovolil razvoj umetnointeligentnega orožja, zmožnega samostojnega ubijanja. Že sposobnost sama po sebi vzbuja tehtna moralna vprašanja, a celo ti sistemi UI bodo v bistvu delovali kot vojaki. Vloga UI pri poveljevanju na bojišču in strateškem delovanju ameriške vojske je bolj ali manj omejena na obveščevalne agoritme, ki hkrati prečiščujejo podatkovne tokove iz več sto senzorjev – podvodnih mikrofonov, kopenskih radarskih postaj in vohunskih satelitov. UI ne bodo že v zelo bližnji prihodnosti prosili, naj spremlja premike vojakov ali sproži usklajene napade, a vojna bo morda postala še hitrejša in bolj zapletena, delno tudi zaradi orožja z UI. Če se bo ameriškim generalom zdelo, da so jih nadkrilili kitajski sistemi UI, ki lahko tedne in tedne brez prestanka in vsaj kratkega dremeža spremljajo dinamične strateške razmere z milijon spremenljivkami, bo umetna inteligenca lahko prevzela tudi sprejemanje pomembnih odločitev.

Natančen ustroj ameriškega poveljevanja in nadzorovanja je državna skrivnost, vendar enkratne procesne zmogljivosti UI že koristno izrabljajo v ameriških sistemih za zgodnje obveščanje. No, tudi v tem primeru avtomatizacija predstavlja precejšnje pasti. Leta 1983 je sovjetski sistem za zgodnje obveščanje svetlikajoče se oblake nad ameriškim srednjim zahodom zamenjal za izstrelke. Katastrofo so preprečili le zato, ker je podpolkovniku Stanislavu Petrovu – treba bi mu bili postaviti par spomenikov – šesti čut govoril, da gre za lažni alarm. Današnji algoritmi računalniškega vida so bolj izpopolnjeni, vendar njihovega delovanja pogosto ni mogoče razumeti. Leta 2018 so raziskovalci UI prikazali, da drobne napake na fotografijah živali lahko preslepijo nevronske mreže, da pando napačno prepoznajo kot gibona. Če umetna inteligenca naleti na nov pojav v ozračju, ki ga njeni podatki za učenje niso zajemali, bi lahko halucinirala in napovedala skorajšnji napad.

A pustimo halucinacije ob strani. Veliki jezikovni modeli se nenehno izboljšujejo in nekoč jih utegnejo prositi, naj sestavijo verodostojne zgodbe o naglo stopnjujočih se krizah v realnem času, vključno z jedrsko krizo. Ko bodo te zgodbe presegle preproste izjave o številu in kraju bližajočih se izstrelkov, bodo postale bolj podobne izjavam svetovalcev, saj bodo tudi interpretirale in prepričevale. Sistemi UI se bodo morda izkazali za odlične svetovalce – hladnokrvne, dobro obveščene

▽ **Preprost sistem, ki so ga razvili v OpenAI, je v prilagojeni različici igre Dota 2, simulaciji vojne, premagal človeške igralce s strategijami, ki njim niso niti padle na pamet. (Dodati je treba, da ni imel zadržkov pri žrtvovanju lastnih vojakov.)**

in vedno zanesljive. Upajmo, da bo tako, saj četudi jih nikoli ne bi prosili za priporočila, kako se odzvati, bi njihove slogovne nianse nedvomno vplivale na predsednika.

Tudi če umetna inteligenca ne bi imela pristojnosti za jedrsko orožje, bi glede na dovolj širok manevrski prostor v primerjavi s konvencionalnim načinom bojevanja sledila taktiki, ki bi nepovratno in bliskovito zaostrila spor do te mere, da bi v paniki sledil jedrski napad. Lahko bi tudi namerno poskrbela za razmere na bojišču, ki bi sprožile izstrelitev, če bi menila, da bi z jedrskim orožjem dosegla želeni cilj. Umetnointeligenčni poveljnik bo ustvarjalen in nepredvidljiv: preprost sistem, ki so ga razvili v OpenAI, je v prilagojeni različici igre Dota 2, simulaciji vojne, premagal človeške igralce s strategijami, ki njim niso niti padle na pamet. (Dodati je treba, da ni imel zadržkov pri žrtvovanju lastnih vojakov.)

Ti manj verjetni scenariji niso neizogibni. Do umetne inteligence smo danes sumničavi, in če bo njena razširjena uporaba povzročila zlome delniških trgov ali druge krize, bodo možnosti uporabe vsaj začasno omejene. A recimo, da se bo umetna inteligenca po začetnem opotekanju v desetletju ali več lepo izkazala in s takšnimi referencami bi ji morda lahko dovolili tudi upravljanje jedrskega orožja in nadzor v kriznih trenutkih, kot so si to predstavljali sodelujoči v vojni igri Schneiderjeve. Predsednik bi nekoč lahko že prvi dan službe naložil algoritme za ukaze in nadzor ter morda umetni inteligenci celo odobril improviziranje na temelju njenega vtisa o poteku napada.

Veliko bi bilo odvisno od tega, kako bo umetna inteligenca razumela svoje cilje v okviru jedrskega spopada. Raziskovalci, ki so UI učili igrati različne igre, so se večkrat srečali z različico te težave: občutek UI, kaj pomeni zmaga, ni premočrten. V nekaterih primerih so se sistemi UI obnašali predvidljivo, dokler majhna sprememba v njihovem okolju ni povzročila nenadne preusmeritve strategije. UI so naučili igrati igro, v kateri igralci iščejo ključ, da bi odklenili skrinjo z zakladi, in tako upravičeno pričakovali nagrado. Umetna inteligenca je temu sledila, dokler inženirji niso priredili okoliščin, tako da je bilo več ključev kot skrinj. Takrat je začela kopičiti ključe, čeprav so bili številni neuporabni, in le tu pa tam poskušala odkleniti skrinje. Kakršnakoli inovacija jedrskega orožja – oziroma obrambe – bi UI lahko napeljala na podobno dramatične spremembe.

Država, ki bo UI vključila v poveljevanje in nadzor, bo druge spodbudila k posnemanju že zato, da bodo prepričljivo odvračale sovražnika. Michael Klare, predavatelj študij o miru in varnosti na svetu na kolidžu Hampshire, opozarja, da bi lahko prišlo do »hitre vojne« po zgledu »hitrega zloma« na finančnem trgu, če bi več držav avtomatiziralo odločitve za izstrelitev. Predstavljajte si, da bi si

Umetna inteligenca na najglobljih in najpomembnejših ravni mogoče ne bo razumela, kaj sta Ronald Reagan in Mihail Gorbačov mislila z besedami, da v jedrski vojni ni zmagovalca.

ameriška umetna inteligenca zvočno nadzorovanje podmornic v Južnokitajskem morju napačno razlagala kot premike, ki napovedujejo jedrski napad. Priprave na protinapad bi opazila kitajska umetna inteligenca, ki bi začela pripravljati izstrelitvene ploščadi, in tako bi se sprožilo zaostrovanje razmer, ki bi lahko pripeljalo do obstreljevanja z jedrskim orožjem.

Na začetku 90. let, ko je nastopil trenutek relativnega miru, sta se George H. W. Bush in Mihail Gorbačov zavedela, da bi tekma v razvijanju orožja lahko vodila v neskončno dodatno oboroževanje z jedrskimi konicami. Bila sta dovolj preudarna, da se nista prepustila orožarski dirki, temveč sta podpisala sporazum o zmanjšanju strateškega orožja, prvega v vrsti nenavadnih dogovorov, s katerimi se je arzenal v obeh državah zmanjšal na manj kot četrtino prejšnjega obsega.

Zgodovina se ponavlja. Nekatere pogodbe so prenehale veljati, druge so zvodenele, ko so se odnosi med ZDA in Rusijo ohladili. Državi sta zdaj bliže neposredni vojni, kot sta bili več generacij prej.

Ruski predsednik Vladimir Putin je 21. februarja lani, manj kot 24 ur po tistem, ko se je ameriški predsednik Joe Biden sprehodil po kijevskih ulicah, izjavil, da bo njegova država zamrznila sodelovanje pri zadnji, še veljavni pogodbi o zmanjševanju vojaškega arzenala Novi START. Kitajska ima danes verjetno dovolj izstrelkov za uničenje vseh velikih ameriških mest, tamkajšnji generali pa so vse bolj naklonjeni polnemu arzenalu, saj vidijo, kakšen vzvod moči ima Rusija v Ukrajini zaradi jedrskega orožja. Zagotovljeno vzajemno uničenje je tako postalo tristranska težava in vse udeležene strani razvijajo tehnologije, ki bi lahko zamajale njeno logiko.

Naslednje obdobje relativnega miru je morda daleč, a če se vrne, bi nam morala biti Bush in Gorbačov v navdih. Njuni razorožitveni sporazumi so bili genialni, ker so predstavljali okrevanje človeškega odločanja, kar bi dosegli tudi z globalnim sporazumom, da umetna inteligenca za vedno ostane izločena iz jedrskega poveljevanja in nadzora. Nekateri od scenarijev, opisani v tem članku, morda zvenijo nemogoči, vendar je to še dodaten razlog, da razmislimo, kako bi se jim izognili, preden se umetna inteligenca izkaže z nizanjem uspehov na bojišču in bo njena uporaba postala preveč mamljiva.

Sporazum je vedno mogoče kršiti in njegovo spoštovanje bo v primeru umetne inteligence še veliko težje nadzirati, saj za njen razvoj niso potrebni sumljiva raketna izstrelišča in obrati za bogatenje urana. A sporazum lahko pripomore k utrditvi močnega tabuja in v kraljestvu UI je trdovraten tabu morda največ, česar se lahko nadejamo. Zemljinega površja ne moremo obdati z avtomatiziranim jedrskim arzenalom, s katerim bi bili le korak oddaljeni od apokalipse. Če bi se v jedrski vojni znašli zaradi napake, naj bo ta vsaj človeška. Prepustitev najresnejše vseh odločitev dinamiki tehnologije bi pomenila dokončno odpoved človekovi izbiri. ◀

Številni svetovi umetne inteligence

Če klepetalnega robota ChatGPT vprašate, kaj njen vzpon pomeni za prihodnost človeštva, ponudi umirjen odgovor, da prihodnost človeštva z UI ni vnaprej določena in da bo njen vpliv odvisen od tega, kako se bo razvijala, kako bo zakonsko urejena in vključena v različne vidike družbe.

Michael Brooks, *New Scientist*

Prametne besede, a bodimo odkriti – tudi izmikajoče. Če se res želite spopasti z vprašanjem, kako bo obdobje UI vplivalo na nas, se raje obrnite na računalniška strokovnjaka Scotta Aaronsona s teksaške univerze v Austinu in Boaza Baraka s Harvarda, ki sta dolge razprave že strnila v peščico mogočih izidov. Vsakemu sta tudi namenila nenavadno ime. »Najin cilj je bil orisati razplete, o katerih ljudje trenutno govorijo,« je pojasnil Aaronson.

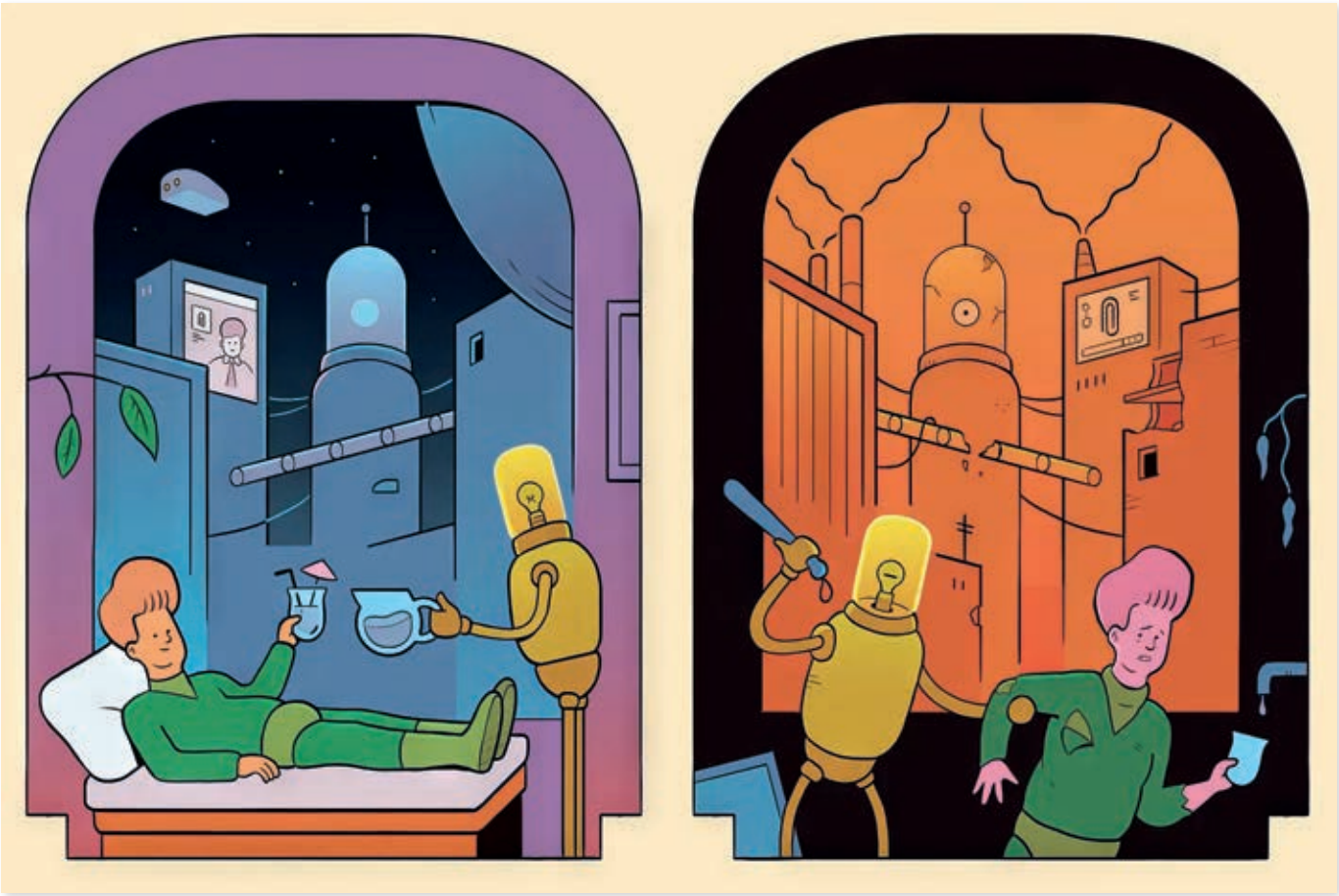
Po izčrpni razpravi sta bila prepričana, da le pet scenarijev pokriva vse mogoče končne dosežke našega razvijanja UI. Znanstvenika sta nato na Aaronsonovem blogu objavila prispevek in upala, da bo Pet svetov umetne inteligence strokovnjakom na tem področju pomagalo začeti tehtne razgovore o končnih ciljih in predpisih. »Napisala sva ga v šaljivem tonu, čeprav z razpravo misliva resno,« je poudaril Barak.

Tu je torej njun skoraj otroški opis teh petih svetov. Rahlo je zabaven, le da bi odvisno od tega, kako bomo ravnali, lahko ali končali v utopiji ob podpori UI – ali pa izbrisali človeško raso.

Potovanje v teh pet svetov se začne s preprostim vprašanjem, ali se bo potencial UI razblinil kot milni mehurček. Trenutno nič ne kaže na to, vendar bi njen pohod lahko zaustavila njena volčja lahkota po virih. Mogoče je, da bi zaradi elektrike in vode, nujno potrebnih za poganjanje in hlajenje strežnikov z UI, sčasoma morali prepovedati njen nadaljnji razvoj. Lahko pa bi se tudi zgodilo, da bi nam zmanjkalo podatkov za učenje prihodnjih modelov UI, in tehnologija tako nikoli ne bi dosegla prelomne zmogljivosti, kot si jo obetamo, in ne ustvarila sveta, ki ga Aaronson in Barak imenujeta umetnointeligenčno mehurčkanje (angleško *fizzle*). »Menim sicer, da se to ne bo zgodilo,« je komentiral Barak. »Prepričan sem, da bo umetna inteligenca temeljito spremenila svet.«

Če se ne bomo razblinili kot mehurčki, se bomo morali soočiti s še enim vprašanjem: se civilizacija še vedno razvija na prepoznaven način? Če je odgovor pritrdilen, Aaronson in Barak menita, da nas čaka eden od dveh nadaljnjih scenarijev. Prvi je ugoden in poimenovala sta ga po animirani televizijski seriji Futurama, v kateri predstavljajo večinoma prijazne prihodnje civilizacije. Drugi ni tako prijeten in poimenovala sta ga Umetnointeligenčna distopija.

Distopijo poznamo iz številnih znanstvenofantastičnih zgodb, od katerih je verjetno



najslavnejši roman Georgea Orwella 1984. Država z globokim nadzorom doseže stoodstotno ubogljivost državljanov brez pomislekov. Neenakost in predsodki se zakoreninijo, delovno okolje je turobno, zaposleni razen neznatne elite so slabo plačani. »Ljudje pravijo, da usoda, kot je opisana v 1984, nikoli ni bila zares mogoča, saj ne bi mogli dobiti dovolj ljudi za nadzorovanje vseh zaslonov. A z UI to postane izvedljivo,« je poudaril Aaronson.

Ljudje pravijo, da usoda, kot je opisana v 1984, nikoli ni bila zares mogoča, saj ne bi mogli dobiti dovolj ljudi za nadzorovanje vseh zaslonov. A z UI to postane izvedljivo.

Če bomo res prišli tako daleč, tega ne bomo mogli očitati UI, saj je zmogljiva tehnologija zgolj in samo ojačevalka obstoječih težav, je pojasnil. »Nekateri ljudje bodo ugovarjali, da niti ne bi smeli ustvariti tehnologije, če je jasno, da bo uporabljena za slabe namene. A če ne bi izumljali tega, kar bi lahko zlorabili za nekaj slabega, preprosto ne bi več mogli ničesar izumiti.«

Barak upa, da bo razvoj UI pripomogel k boljši resničnosti v slogu Futurame. To je po njegovem mnenju najboljši od vseh mogočih svetov. »To je pravzaprav stanje, kot ga poznamo danes, le izboljšano,« je dodal. Z UI bi zmanjšali revščino in zagotovili, da bi več ljudi dobilo dostop do hrane, zdravstvene oskrbe, izobrazbe in poslovnih priložnosti. Še vedno bi koga doletela tudi nesreča zaradi človeške zlobe ali malomarnosti, a na splošno bi ljudje v tem svetu živeli zadovoljno.

In to mogoče še ne bo vse. Aaronson upa, da bo nekega dne živel v utopiji, ki jo bo omogočila umetna inteligenca. Z Barakom sta jo poimenovala Singularija. Do nje bo prišlo, če bo vpliv UI tako velik, da bo naš prihodnji svet neprepoznaven, in to v ugodnem smislu. Živeli bi z modeli UI, ki bi bili nova superinteligenčna vrsta, ki bi prevzela usmerjanje civilizacije. Na srečo bi bili to dobronamerni bogovi, strpni do ubogih, šibkih ljudi na planetu. Reševali bi naše materialne težave, nam zagotavljali neomejene dobrine, prevzeli marsikatero od enoličnih zadolžitev in skrbeli za kratkočasenje našega po novem premalo stimuliranega uma. »To bodo tako rekoč nebesa, ki jih bo ustvarila umetna inteligenca,« je komentiral Aaronson.

Fino. Kaj pa, če vsemogočni modeli UI ne bodo tako velikodušni? Dobrodošli v Sponkalipsi. Razlog za nenavadno ime tega sveta je tehten. Leta 2003 je matematik in etik Eliezer Yudkowsky, soustanovitelj inštituta MIRI za raziskave strojne inteligence, opisal miselni eksperiment, med katerim so ljudje UI naložili, naj izboljša proizvodnjo nekega vsakdanjega izdelka, na primer sponke za papir. Če specifikacije niso dobre premišljene, bi umetna inteligenca lahko nazadnje izbrisala človeštvo, ker je naša vrsta v napoto pri izrabljanju zemeljskih virov in nemara tudi virov drugih svetov, iz katerih bi lahko izdelali več boljših sponk za papir.

Yudkowskyjev argument je bil bolj poglobljen od tega primera, a izhodišče je preprosto: ni si težko zamisliti scenarijev, po katerih bi umetna inteligenca lahko postala zadnji človekov izum. Temeljito je zato treba razmisliti, s čim se sme ubadati umetna inteligenca, o njeni povezavi s fizičnimi viri, kot so elektrarne in jedrsko orožje, ter o transparentnosti in odgovornosti njenih razvijalcev. In morali bi pohiteti. »Za spremembo politike in predpisov je nujen čas, državno kolesje pa melje počasi,« opozarja Gretta Duleba, vodja stikov z javnostjo v MIRI. »Skrbi nas, da nam bo zmanjkalo časa, preden bodo začeli veljati pomembni zakoni.«

Aaronsona in Baraka pa možnost Sponkalipse ne skrbi pretirano. Medtem ko se Yudkowskyjemu katastrofa zdi najverjetnejši in tako rekoč zagotovljeni rezultat slabo zasnovanih raziskav UI, Aaronson meni, da to ni verjetno. Barak je nekoliko previdnejši. »Ne verjamem v čisto skrajne stvari, kot sta Singularija in Sponkalipsa, vseeno pa menim, da je treba biti odprt za vse.«

Najbrž je to dobra zamisel, in sicer predvsem zaradi zagotavljanja, da bomo previdni »pri razvoju umetne inteligence, zakonodaje o njej in njenem vključevanju«, kot se je izrazil ChatGPT. Če ne bomo, se lahko zgodi, da ne bomo mogli izbirati, kako se bo razpleta naša zgodba z UI.

V resnici se lahko tudi z dobronamerno zakonodajo primeri, da bomo pristali v enem od svetov, ki nam jih bo krojila umetna inteligenca, čeprav bi se mu najraje na široko izognili. Navsezadnje bomo morda nekega dne morali iskati izhod iz umetnointeligenčne distopije in se zateči v Futuramo oziroma – še bolje – v Singularijo. »Želim živeti v svetu, kjer bi nebrzdano uspevala čuteča bitja v simuliranem paradižu po naših željah?« se je vprašal Aaronson. »Ja, v bistvu to podpiram.« ◀

Ko prisluškuje **država**

Izraelski Pegasus je najbolj znano programsko orodje, ki ga za prisluškovanje uporabljajo državne inštitucije širom sveta. Med strankami so tako avtoritarne kot načeloma demokratične države.

Alex Pasternack, *Fast Company*

V mobilnik se Pegasus naseli z nezaznavnim SMSom in snema vse – sporočila, fotografije, šifrirane klepete, slikovne in zvočne posnetke. Pogosto ni znano, kam potujejo podatki, ki jih je zajel, saj se sledi izgubijo v labirintu strežnikov. Kljub temu je usodni vpliv Pegasusa in drugega kibernetičnega orožja – s katerim rožljajo vlade od Španije do Savdske Arabije, in sicer z njim grozijo predvsem zagovornikom človekovih pravic, novinarjem, odvetnikom in drugim – danes dobro dokumentiran. Z valom podrobnega opazovanja in sankcij so razkrili tajno, le napol legalno industrijo za temi orodji in finančno pritisnili na podjetja, kot je izraelska NSO Group, ki je Pegasus razvila.

Kljub temu posel cveti. Nova raziskava, ki sta jo pred nekaj meseci objavila Google in Meta, nakazuje, da se kibernetični napadi kljub novim omejitvam povečujejo, postajajo nevarnejši, podpirajo vlado pri njenem nasilju in represiji ter načenjajo demokracijo po vsem svetu.

»Ta panoga je izjemno uspešna,« je povedala Maddie Stone, raziskovalka v Googlovi skupini za analize groženj, ki išče ranljivost in varnostne luknje, torej programske hrošče, da jih odpravi. Za prodajalce vohunske programske opreme so ti hrošči včasih vredni tudi več sto milijonov dolarjev. »Pojavlja se vse več podjetij in njihovi vladni kupci so odločeni skleniti posel. Zahtevajo takšne možnosti in jih tudi uporabijo.«

Prvič v zgodovini polovico teh varnostnih lukenj v Googlovih in Android izdelkih odkrijejo oziroma jih plačajo zasebna podjetja, ugotavljajo v poročilu, ki ga je objavila ekipa Stonove. Googlovi raziskovalci poleg znanih podjetij, kot sta NSO in Candiru, sledijo še okoli 40, ki so vpletena v razvoj orodij za vdor, uporabljenih pri 'uporabnikih z visoko stopnjo tveganja'.

Med 72 luknjami, ki jih je Google odkril med letoma 2014 in 2023, jih je 35 pripisal njim in drugim udeležencem v panogi, ne pa akterjem z državno podporo.

»Če so vlade res kdaj imele monopol nad najsodobnejšimi zmogljivostmi, je tega obdobja nepreklicno konec,« tudi piše v poročilu.

Googlove ugotovitve in Metino poročilo o grožnjah zaradi vohunske programske opreme, objavljeno teden dni pozneje, kažejo vse večje težave velikih tehnoloških družb pri odzivu na panogo, ki kuje dobiček z vdiranjem v njihove sisteme. Poročili pomenita tudi dodaten pritisk na Združene države Amerike in druge, naj ukrepajo proti trgovanju in poslovanju tako rekoč brez zakonov in predpisov.

Shane Huntley, eden iz vodstva skupine za analizo groženj, je na blogu omenil vrsto vladnih dogovorov za omejitev škode zaradi vohunske programske opreme, hkrati pa izrazil tudi upanje, da bodo tem uvodnim korakom sledili konkretnejše ukrepanje širše skupnosti držav, reforma panoge in glasnejše opozarjanje na zlorabe.

Kibernetični pohlepneži in višji predstavniki obveščevalnih služb opozarjajo na uporabo teh orodij pri kazenskem pregonu ter v boju proti terorizmu in na desetine držav jih skrivaj uporablja, da z njimi sledijo zločincem ter osumljenim terorizma, vključno z mamilarskim kraljem, znanim pod vzdevkom El Chapo. A osupljivo veliko raziskav in poročanja je dokazalo, da ta orodja pogosto uperijo proti novinarjem, zagovornikom človekovih pravic, pripadnikom opozicije, včasih v okviru neusmiljenega pogroma proti njim.

S temi orodji države sledijo zločincem ter osumljenim terorizma, a osupljivo veliko raziskav in poročanja je dokazalo, da jih pogosto uperijo tudi proti novinarjem, zagovornikom človekovih pravic in pripadnikom opozicije.

Raziskovalci ocenjujejo, da državne agencije v kar 46 državah – med njimi v Španiji, Indiji in Savdski Arabiji – tako ali drugače uporabljajo Pegasus. Novinarji so leta 2020 pridobili seznam 50 tisoč telefonskih številk, ki so jih stranke NSO izbrale za svoje tarče, in na njem so bili vsaj trije predsedniki, deset premierjev, kralj in več članov arabske kraljeve družine, najmanj 65 direktorjev, 85 borcev za človekove pravice, 189 novinarjev in več kot 600 politikov ter vladnih uradnikov, med njimi ministri, diplomati ter vojaški in varnostni predstavniki. (Pegasus Project, novinarski konzorcij pod vodstvom francoske medijske neprofitne organizacije Forbidden Stories, je potrdil virus na več deset telefonih s seznama.) Pegasus so našli na telefonih novinarjev iz Indije, Jordanije, Armenije, Toga in Dominikanske republike, vsaj šestih palestinskih zagovornikov človekovih pravic, od katerih je eden ameriški državljan, in na telefonu dveh žensk, ki sta bili zelo blizu umorjenemu savdskemu odpadniku Džamalu Hašokdžiju. Po oceni analize raziskovalne skupine *Forensic Architecture*, ki se na londonski univerzi ukvarja z raziskavami človekovih pravic, je bila vohunska programska oprema vpletena v najmanj 300 primerov telesnega nasilja.

»Škoda ni samo hipotetična,« poudarja Stonova.

Bidnova administracija je po letu 2021 na črni seznam uvrstila več podjetij za vohunsko programsko opremo in prepovedala prenos ameriške tehnologije nanje, letos pa je še zaostrila stališče do udeležencev v tem poslu. Dan pred objavo Googlovega poročila je ameriško zunanje ministrstvo napovedalo nove omejitve pri vizumu za ljudi, ki sodelujejo v zlorabi komercialne vohunske programske opreme, in njihove družinske člane ter dodatne omejitve pri dostopu razvijalcev te opreme do ključnega tehnološkega sektorja. Zunanji minister Antony Blinken je v izjavi za javnost poudaril, da vohunska

Leta 2019 je Meta tožila skupino NSO, potem ko je odkrila, da je bilo 1.400 uporabnikov Whatsappa žrtev vohunskega orodja NSO. Apple je tožbo vložil novembra 2021.

programska oprema »ogroža zasebnost in svobodo izražanja, mirnega zbiranja in druženja, povezana pa je tudi s samovoljnimi priprtji, z izginotji in celo zunajsodnimi poboji«.

Raziskovalci varnostnih lukenj trenutno boj proti vohunski programski opremi še vedno opisujejo kot globalni 'kdo bo koga' – ko ujamejo enega krivca, se že pojavita dva nova. In odločni hekerji bodo seveda naredili vse, kar je mogoče, da se izognejo zaščitnim ukrepom. Podjetja se zato iz Izraela selijo v druge države, kjer ne velja tako strog nadzor nad izvozom.

Kljub neugodnim posledicam ameriških kazni se nekateri podjetniki in podjetja ne morejo upreti donosnemu poslu. Po neki oceni se je leta 2022 na globalnem trgu z vohunsko programsko opremo obrnilo 12 milijard dolarjev.

»Če bi skupina NSO jutri bankrotirala, bi praznino poskušale zapolniti druge družbe, morda tudi takšne, ki bi bile podprte z ameriškim tveganim kapitalom,« je na lanskem zaslišanju v kongresu povedal John-Scott Railton iz Citizen Laba. »Dokler ameriški vlagatelji na panogo z vohunsko programsko opremo gledajo kot na rastoči trg, bo ameriški finančni sektor samo podžigal to problematiko in pri tem poteptal našo kibernetično varnost in zasebnost.«

Google v svojem poročilu opisuje povečanje vohunskih rešitev po meri, ki jih ponuja na desetine sumljivih podjetij. Najbolj so pod drobnogledom izraelska podjetja, ki izvajajo kibernetične napade, kot sta NSO in Candiru, vendar Googlova analiza opozarja tudi na pohod manjših tovrstnih podjetij s sedežem v Evropi in Aziji. Med njimi so tudi butična podjetja, 'popoldanski' hekerji, prodajalci podatkov o varnostnih luknjah in podobni, ki so jih pri možganskem trustu *Carnegie Endowment for International Peace* poimenovali sekundarna stopnja te panoge. V Googlovem poročilu je omenjenih 11 podjetij in devet hčerinskih družb; vsi so se pojavljali že v starejših poročilih. Stonova in tiskovni predstavnik podjetja nista pojasnila, zakaj niso našteli še drugih imen.

Podjetja za vohunsko programsko opremi predstavljajo le majhen del globalne panoge hekerjev, ki jih je mogoče najeti. (Iz I-Soona, prodajalca programov za kibernetično varnost, ki je povezan z vlado, je pricurljalo, da so nepridipravi leta 2021 in2022 izvedli napade na vrsto zelo pomembnih vladnih ciljev in odpadnikov po vsem svetu, delno z vdorom v brezžična omrežja in naprave.)

Nekatera podjetja nimajo spletnih strani in se skrivajo v zapletenih strukturah koncernov, da prikrijejo lastnike, vlagatelje in stranke. Candiru s sedežem v Tel Avivu je spremenil ime v Grindavik, potem v Taveta, nato v Saito Tech. Drugo podjetje, ki ga je prav tako izpostavil Google, DSRIF s sedežem na Dunaju, je avgusta lani nehalo poslovati, vendar njegovo delo vsaj delno menda nadaljuje podružnica Machine Learning Solutions.

Podjetja se niso odzvala na novinarska vprašanja ali pa niso bila dosegljiva. Podjetja na Googlovem seznamu:

▶ **Candiru**, ustanovljen leta 2014, naj bi bil drugi največji izraelski razvijalec vohunske programske opreme za NSO. Financirajo ga vlagatelji v NSO in katarska vlada, njegove sisteme pa uporablja vrsta držav, med drugim Savdska Arabija, Izrael, Združeni arabski emirati, Madžarska, Indonezija in Uzbekistan. Novembra 2021 je ameriško ministrstvo za trgovino Candiru in NSO dodalo na svoj črni seznam.

▶ **Cy4Gate**, ustanovljen 2014, je specializiran za tehnologijo za ‚zakonito prestrezanje', kot je programska oprema Epeius za vdiranje v sistema Android in iOS. Leta 2022 je kupilo sorodno italijansko podjetje RCS Lab, znano po vohunskih orodjih Hermit. Google in Meta sta delovanje RCS Lab zaznala v Italiji, Kazahstanu, Azerbajdžanu in Mongoliji.

▶ **The Intellexa Alliance** deluje kot stičišče različnih podjetij za nadzor, med njimi so **Cytrox**, razvijalec Pegasusu podobnega orodja Predator, podjetje za prestrezanje

brezžičnih povezav in mobilnikov **WiSpear** in **Senpai**, ki je specializiran za zbiranje odprtokodnih podatkov. Junija 2021 so štiri vodstvene delavce **Nexa Technologies**, članice Intellexe, na pariškem sodišču obsodili zaradi sodelovanja pri mučenju. To se je zgodilo desetletje po razkritju Wall Street Journala, da je podjetje prodajalo nadzorno programsko opremo libijski vladi. Leta 2021 je Meta Cytroxu prepovedala uporabljati njene platforme, ameriško ministrstvo za trgovino ga je na črni seznam uvrstilo leta 2023.

▶ **NSO Group** s sedežem v izraelski Herzliji, so razkrinkali leta 2016, ko je Citizen Lab na telefonu Ahmeda Mansurja, zagovornika človekovih pravic iz Združenih arabskih emiratov, odkril Pegasus. Po ocenah Pegasus Projecta naj bi ta vohunska programska oprema vohunila za več sto člani civilne družbe.

▶ **Negg Group**, italijansko podjetje za kibernetično varnost, je Kaspersky leta 2017 razkrinkal kot razvijalca zlonamerne programske opreme VBiss in Skygofree za sistem Android. Že samo z enim klikom ali prek brskalnika okužita mobilne naprave, tarče pa so odkrili v Italiji, Maleziji in Kazahstanu.

▶ **Paris Defense** je podjetje za kibernetično varnost s sedežem v Istanbulu, ki pomaga strankam reševati forenzične izzive v mobilnem svetu, kot dejansko piše na njegovi spletni strani. Povezano je z izrabljanjem dveh nedavnih ranljivosti v iOS.

▶ **QuaDream** je leta 2014 ustanovila skupina, v kateri sta bila tudi nekdanja zaposlena NSO. Razvil je vohunsko programsko opremo s kratico REIGN, ki med drugim omogoča snemanje klicev v realnem času ter aktivacijo kamere in mikrofona, piše v brošuri. Aprila 2023 se je podjetje nenadoma zaprlo po tem, ko mu je izraelska vlada preprečila, da bi svoje orodje izvažalo v tuje države, vključno z Marokom, in po tem, ko so raziskovalci v Microsoftu in Citizen Labu poročali, da so z Reignom zalezovali novinarje, pripadnike opozicije in organizacije za človekove pravice po vsem svetu.

▶ **Variston Information Technology** je nastal leta 2018 v Barceloni in je le malo pozneje kupil **TrueL IT**, italijansko podjetje, specializirano za ranljivosti. Sodeluje z ironično poimenovanim **Protect Electronic Systems** iz Abu Dabija in v paketu prodaja svojo vohunsko programsko opremo ter infrastrukturo. Aprila lani je strokovna publikacija *Intelligence Online* poročala, da je Variston stkal trdne vezi s kibernetično podružnico obrambne družbe Edge Group v lasti lastnikov iz Združenih arabskih emiratov.

▶ **Wintego Systems**, ki so jih ustanovili nekdanji zaposleni v Verintu, še ena izraelska družba, razvija napredne rešitve za komunikacije, obveščanje in dekodiranje podatkov za vlado in domovinsko varnost. V brošuri družbe piše, da njena vohunska programska oprema uporablja brezžično omrežje za pridobivanje zaščitenih podatkov s spletnih računov (storitev v oblaku) in aplikacij, vključno s popolno vsebino elektronskih sporočil, fotografijami, z dokumenti, s klepeti, z dejavnostjo na družbenih omrežjih, s seznamom stikov in koledarjem.

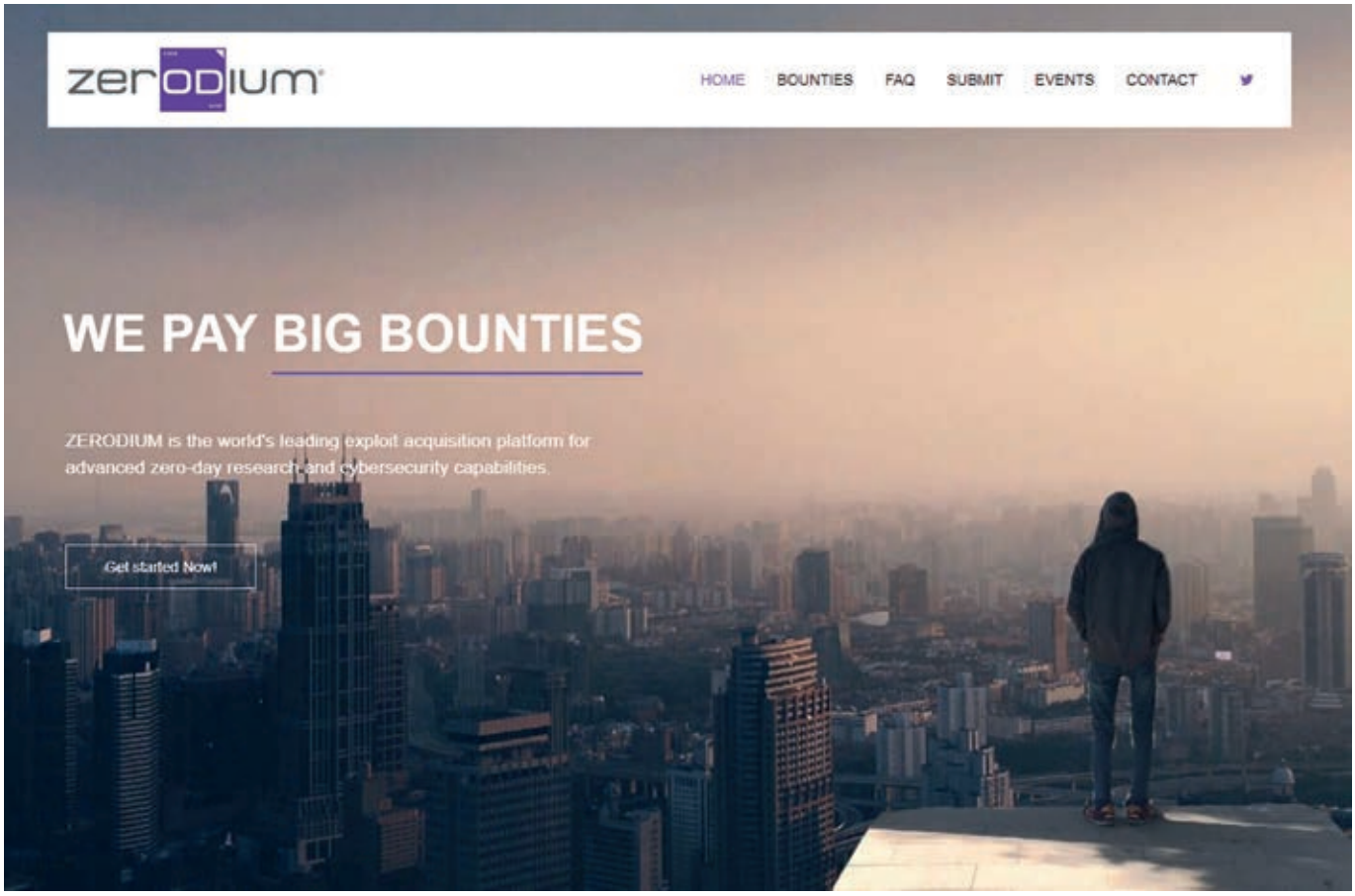
Metina ekipa za odkrivanje groženj je nedavno ukrepala proti nekaj deset računom, ki jih je odprlo osem podjetij za vohunsko programsko opremo iz Španije, Italije in Združenih arabskih emiratov. V četrtletnem poročilu, ki so ga objavili 14. februarja, so našteti italijanska podjetja Cy4Gate, RCS Labs, Negg Group, IPS Intelligence, španski Variston in njegovo podružnico TrueL IT pa Mollitiam Industries in Protect Electronic Systems iz Združenih arabskih emiratov. V Meti so povedali, da so podjetja z lažnimi računi pobirala podatke, izvajala družbeni inženiring in preizkušala svoje vohunske sposobnosti.

Med najbolj izstopajočimi vidiki Googlovega poročila so profili šestih žrtev zlonamerne opreme, ki so jih zbrali raziskovalci v njegovem možganskem trustu Jigsaw. Med njimi je Galina Timčenko, urednica izgnanega ruskega neodvisnega novičarskega medija Meduza, ki je junija 2022, ko je bila doma v Rigi, dobila nenavadno sporočilo Appla. V njem je pisalo, da je tarča napadalcev po naročilu države, ki so si jo za tarčo verjetno izbrali zaradi njenega poklica in dela. Ker je bila vajena poskusov ribarjenja in onemogočenega dostopa do storitev, je o tem preprosto obvestila svojo tehnično ekipo in pozabila na vse skupaj.

A v preiskavi, ki so jo začeli neprofitna organizacija Access Now in kibernetični detektivi iz Citizen Laba, so kmalu ugotovili, da je njen telefon okužil Pegasus, in sicer že mesece prej, le dan preden se je Timčenkova udeležila tajnega sestanka ruskih izgnanih medijev v Berlinu, da bi razpravljali o ruskem dopolnjenem zakonu o tujih agentih. Napadalci, katerih tarča sta bila Applov HomeKIt in iMessage, so dostop do njene naprave menda dobili med sestankom in ga verjetno imeli še nekaj tednov. Kompromitirani bi lahko bili občutljivo dopisovanje, podatki in viri, medtem ko se Timčenkovi ni niti sanjalo, da je kaj narobe.

Celo en sam zahrbten vdor – oziroma že samo grožnja – predstavlja splošno

Mesece po tistem, ko je ameriško ministrstvo za trgovino NSO in Candiru uvrstilo na črni seznam, so v *New York Timesu* poročali, da je FBI Pegasus nabavil za ‚testiranje‘.



△ **Na trgu z varnostnimi luknjami Zerodium je za zelo uporabno luknjo v Androidu mogoče iztržiti poltretji milijon dolarjev, kar je pol milijona več kot za podobno luknjo v sistemu iOS.**

nevarnost. »Ko ciljajo na naše politične nasprotnike, to škodi tudi nam, saj meče dvom na svobodne in poštene volitve,« je poudarila Stonova. »In na vse nas vpliva, če so žrtve novinarji in jih je zato strah razkrivati resnico.«

Kdo je vdrl v telefon Timčenkove? Osumljeni sta dve Pegasusovi stranki, Kazahstan ali Azerbajdžan, ki bi lahko delovali na prošnjo Moskve, so povedali preiskovalci organizacije Access Now. A kolikor je znano raziskovalcem, nobena od teh držav Pegasusa ni uporabila v Evropi, Timčenkova pa je bila v času vdora v Berlinu. Krivec bi utegnila biti evropska država – Nemčija, Latvija in Estonija so prav tako med potrjenimi Pegasusovimi kupci.

Tudi ameriške agencije so kupovale Pegasus in podobno vohunsko opremo. Mesece po tistem, ko je ministrstvo za trgovino NSO in Candiru uvrstilo na črni seznam, so v *New York Timesu* poročali, da je FBI Pegasus nabavil za 'testiranje' in je razmišljal o izdelku za ameriške telefonske številke, imenovanem Phantom, ki ga je NSO pred tem ponujal policiji. Cia je Pegasus kupila za vlado Džibutija, so napisali v *Timesu*, in to kljub vztrajnim pomislekom zaradi kršenja človekovih pravic v tej državi.

Ameriški urad za boj proti drogam je z Graphitom, izdelkom, podobnim Pegasusu, preganjal mamilarske kartele. Njegov razvijalec Paragon, ki ga podpirata nekdanji izraelski premier Ehud Barak in vsaj eno podjetje za tvegani kapital s sedežem v ZDA, Battery Ventures, je želel zakuhati škandal in je leta 2021 najel dobro povezane WestExec Advisors, da bi hitreje prodrl na ameriški trg. Podjetje je celo spraševalo za nasvet, katere druge države bi bile sprejemljive za ameriške uradne predstavnike kot stranke, je poročal Financial Times.

NSO je svoj ugled v Washingtonu poskušal zloščiti s sodelovanjem vplivnih lobistov, tudi z nekdanjim svetovalcem Nacionalne varnostne agencije. NSO je maja lani v pismu ameriškemu pravniškemu združenju opozoril na nevarnost zveznega moratorija na komercialno vohunsko programsko opremo. »Popolnoma podpiramo prizadevanja za oblikovanje zakonskih okvirov za prodajo in uporabo komercialne obveščevalne tehnologije, vendar se bojimo, da bi z moratorijem prevlado v panogi prevzela podjetja, ki poslujejo v okoljih z ohlapnejšo zakonodajo, s slabšim nadzorom in z manjšo motivacijo za spoštovanje človekovih pravic, vključno s podjetji s sedežem v Rusiji in na Kitajskem,« je zapisal svetovalec NSO Šmuel Sunraj.

Sporočila NSO so omenjala tudi vojno med Izraelom in Hamasom: po napadu 7. oktobra lani je izraelski časopis *Haaretz* poročal, da so izraelske varnostne službe začele vključevati podjetja, kot so NSO, Candiru in Paragon, da bi izsledile talce v Gazi.

»Tehnologija NSO podpira trenutni globalni boj proti terorizmu v vseh oblikah,« je neki lobist podjetja novembra lani zapisal v pismu zunanjemu ministru Antonyju Blinknu in prosil za izredni sestanek. »Ta prizadevanja so v skladu z večkrat ponovljenimi sporočili in ukrepi Bidna in Kamale Harris v podporo izraelski vladi.« Strokovnjaki za kibernetično varnost niso bili prepričani o primernosti zamisli o sledenju talcem in neki

Ko je skupina NSO po Hašokdžijevem umoru Savdski Arabiji onemogočila dostop do Pegasusa, je Savdski prestolonaslednik nemudoma poklical premierja Benjamina Netanjahuja in dostop je bil spet omogočen.

ameriški uradni predstavnik je za *Fast Company* izjavil, da ameriška vlada ne načrtuje spremembe statusa skupine NSO na seznamu tovrstnih poslovnih subjektov.

Kljub vladnim omejitvam razmah kibernetičnih trgovinskih poslov kaže simbiozo. Tako kot pri načrtnem širjenju dezinformacij je tudi panoga vohunske programske opreme pogosto odvisna od znanja in domiselnosti nekdanjih vladnih hekerjev, prodajalci pa vladam nudijo priročno oporo pri zanikanju. Pri Googlu pravijo, da finančno breme in tveganje za omadeževano ime ob morebitnem razkritju namesto na kupko, to je vlado, padeta na prodajalca, zaradi česar se morda poveča verjetnost uporabe teh orodij.

Izraelska panoga kibernetičnih napadov, ki velja za najboljšo na svetu, ni le v ponos uradnim predstavnikom, temveč tudi orodje realpolitike. Tako kot prodajo konvencionalnega orožja tudi izvoz orodij za hekerje nadzoruje država. Izrael je prodajo Pegasusa rutinsko uporabljal za zbiranje podpore med drugimi državami v OZN in v svoji kampanji proti Iranu. Ko je skupina NSO po Hašokdžijevem umoru Savdski Arabiji onemogočila dostop do Pegasusa, je Savdski prestolonaslednik nemudoma poklical premierja Benjamina Netanjahuja in savdski dostop je bil spet omogočen. (Malo pozneje so raziskovalci Citizen Laba ugotovili, da je Savdska Arabija Pegasus uporabljala za vdor v telefone 36 novinarjev Al Jazeere, katarske konkurenčne novičarske mreže.)

Ko je Rusija napadla Ukrajino, je Izrael blokiral prodajo Pegasusa Ukrajini in izklopil možnost uporabe Estoniji, čeprav je ta za program, s katerim se je nameravala vtihotapiti v ruske telefone, plačala 30 milijonov dolarjev.

Izrael je uporabo Pegasusa omejil tudi iz strahu, da bo razjezil druge zaveznike, vključno s Kremljem. Ko je Rusija napadla Ukrajino, je Izrael blokiral prodajo Pegasusa Ukrajini in izklopil možnost uporabe Estoniji, čeprav je ta za program, s katerim se je nameravala vtihotapiti v ruske telefone, plačala 30 milijonov dolarjev.

Na začetku dobavne verige kibernetičnega orožja je pogosto raziskovalec, ki je odkril dotlej neznana stranska vrata v operacijski sistem. Lahko bi pomagal izboljšati sistem, na primer tako, da bi skrbnikom sporočil ta podatek ali ga posredoval programu, ki za takšno odkritje ponuja nagrado. Tretja možnost je, da najdbo proda posrednikom ali prodajalcem, kot je NSO. Na trgu z varnostnimi luknjami Zerodium je za zelo uporabno luknjo v Androidu mogoče iztržiti poltretji milijon dolarjev, kar je pol milijona več kot za podobno luknjo v sistemu iOS. Orožje, ki ga razvijejo na podlagi takšne luknje, stranke stane še par milijonov več, dodane pa so mu še druge funkcije po meri uporabnika.

Zaradi številnih prijav zlorabe so Google in druga podjetja morali začeti bolj proaktivno ukrepati proti varnostnim luknjam. Pridružili so se raziskovalcem ter novinarjem pri razkrivanju prodajalcev in pojasnjevanju, kako deluje njihovo orožje. Apple je leta 2021 začel pošiljati opozorila morebitnim tarčam, kot je bila Timčenkova, lani pa

▽ **Pegasus, poimenovan po krilatem konju iz grške mitologije, v mobilniku pristane s kratkim sporočilom, se lahko brez vašega vedenja zakoplje v vašo napravo, se v njej skriva dneve ali celo tedne, vmes pa v realnem času skrbno snema vse – sporočila, fotografije, šifrirane klepete, slikovne in zvočne posnetke.**

Po neki oceni se je leta 2022 na globalnem trgu z vohunsko programsko opremo obrnilo 12 milijard dolarjev.

uvedel zaklenjeni način, dodatno opcijsko funkcijo za iOS in MacOS, s katero je mogoče zmanjšati zlorabe zaradi varnostnih luknj. Leta 2019 je Meta tožila skupino NSO, potem ko je odkrila, da je bilo 1.400 uporabnikov Whatsappa žrtev vohunskega orodja NSO. Apple je tožbo vložil novembra 2021 in NSO opisal kot nemoralnega lakomneža 21. stoletja, ki je razvil izjemno napredno mašinerijo za kibernetično rutinsko in nebrzdano nadzorovanje ter s tem zlorabljanje uporabnikov. (Kalifornijski sodnik je zavrnil vlogo NSO za opustitev primera.)

Stonova velike tehnološke družbe lahko pohvali, ker so odkritejše o pomanjkljivostih v svojih sistemih. S tem ščitijo uporabnike in zvišujejo stroške podjetjem, ki prodajajo vohunsko programsko opremo, kar jih prisili, da morajo svojim varnostnim luknjam posvetiti več časa.

»Še pred nekaj leti smo kot panoga in branilci imeli veliko jasnejšo podobo o tem, kaj se dejansko dogaja in kaj počnejo hekerji,« je pojasnila Stonova.

ZDA so tudi naredile vrsto korakov proti posameznim družbam in načinom uporabo vohunske programske opreme. Ministrstvo za trgovino je lani na svoj seznam dodalo še Cytrox in Intellexo, predsednik Biden pa je izdal uredbo, s katero je ameriški vladi prepovedal uporabo komercialne vohunske programske opreme, sploh tiste, pri kateri je veliko možnosti, da jo bodo neprimerno uporabljale tuje vlade ali tujci. (Bidnova administracija je povedala, da je v tujini žrtev zlonamerne opreme preverjeno postalo več kot 50 ameriških državnih uslužbencev.)

Ukrepale so tudi druge države. Dan po tistem, ko so ZDA predstavile nove omejitve pri vizumih za proizvajalce vohunske programske opreme, je skupina 35 držav pod vodstvom Združenega kraljestva in Francije podpisala sporazum o preprečevanju širjenja in neodgovorne uporabe komercialnih kibernetičnih orodij in storitev za vdiranje v tuje naprave. Predstavniki iz Madžarske, Mehike, Španije in Tajske – držav, ki so povezane z zlorabami vohunske programske opreme – sporazuma niso podpisali, Izrael se srečanja ni udeležil.

Izraelska vlada je zaradi sankcij pripravila lastne ukrepe za omejevanje izvoza te opreme in je število dovoljenih držav odjemalk omejila na 37 s prejšnjih 110, zaradi česar se je nekaj tamkajšnjih podjetij znašlo v finančnih težavah, najmanj tri pa so že bankrotirala. NSO se je prav tako opotekal na robu bankrota in leta 2022 je soustanovitelj Šalev Hulio v valu odhodov iz podjetja odstopil s svojega položaja.

No, NSO nadaljuje razvijanje svoje vohunske programske opreme, ima nove lastnike in vnovič upa na dobiček. *Wall Street Journal* je maja lani poročal, da so upniki NSO – med katerimi sta tudi Credit Suisse in Senator Investment Group – posodili dodatne milijone in sodelujejo s soustanoviteljem NSO Omrijem Laviejem. Po podatkih v registru je edini delničar v matični družbi NSO luksemburški holding, ki je v lasti Lavieja. Menda je prevzel tudi vodenje podjetja in odpustil vrsto vodstvenih delavcev.

Googlov predstavnik za javnost je pohvalil nove vizumske omejitve v ZDA, ki so še posebej uperjene proti posameznikom, kot je Lavie. Vseeno so potrebni še odločnejši ukrepi, pravijo pri Googlu, in zato vlade, vključno z ameriško, poziva, naj razkrijejo svojo zgodovino uporabe teh orodij.

»Dokler je povpraševanje po nadzorovanju, bodo podjetja motivirana za nadaljnji razvoj in prodajo orodij ter sodelovanje v panogi, ki ogroža uporabnike s tako imenovano visoko stopnjo tveganja in družbo na sploh,« še dodajajo pri Googlu. ◀

Umetna inteligenca mora dobiti stik **z resničnim svetom**

Kdo ima glavno besedo, vaši možgani ali telo? Odgovor se morda zdi na dlani, vendar veliko dokazov namiguje, da je naša fiziologija močno vplivala na naš način razmišljanja. Predstava o utelešeni kogniciji morda skriva pomembna spoznanja za vse, ki si prizadevajo sestaviti dejansko inteligentne naprave – oziroma umetno inteligenco (UI), ki se uči, razmišlja in svoje znanje tako kot človek lahko posploši na vse mogoče naloge.

Edd Gent, *New Scientist*

Dosh Bongard, robotik na univerzi v Vermontu, sodi med tiste, ki so prepričani, da bo UI svoje obljube izpolnila le, če bo neposredno zaznavala fizični svet in vzajemno delovala z njim. Sistemi UI, kot je ChatGPT, ki s svetom sodeluje le prek abstraktnega medija jezika, so daleč od tega. Kakorkoli že, področje utelešene umetne inteligence si prizadeva za združitev UI in robotike, na čelu teh prizadevanj pa je Bongard.

Po njegovem mnenju bi morali vnovič razmisliti o pristopu do obeh disciplin. Če bi klepetalnega robota na osnovi UI preprosto zvezali z robotsko roko, kot je to naredil Google s svojim sistemom Palm-e, najbrž to ne bi bilo dovolj. Bongard se zato raje posveča evolucijski robotiki, ki načela naravne selekcije posnema v zasnovi robotov, pogosto izdelanih iz mehkih materialov. Poleg tega je tudi član ekipe, ki iz živih celic sestavlja preproste biološke robote, znane kot ksenobote, ki ne opravljajo le temeljnih nalog, temveč lahko tudi vzajemno delujejo s svojim okoljem in se odzivajo nanj.

Bongard je za *New Scientist* pojasnil, da ta prizadevanja osvetljujejo čisto drugačen način razmišljanja o utelešeni kogniciji in razlago, kaj šteje kot robot. To bi lahko vplivalo tudi na naš pristop do sestavljanja inteligentnih naprav.

▶ Kako telo vpliva na način razmišljanja in učenja?

Josh Bongard: Kako ljudje dojemamo svet, je večinoma posredno ali neposredno pod vplivom naše fiziologije. Svoja telesa uporabljamo, da se odzivamo na svet, s čutili pa opazujemo, kako se svet odziva na nas. Ta zanka povratnih informacij razkriva, kako se učimo o svetu, pomeni pa tudi, da različna telesa pridejo do različnih posledic akcije in reakcije in se zato o svetu učijo drugače. Velika kopenska žival, na primer, visoke pečine in vodo dojema kot nevarnost, leteče živali pa imajo pečino za odlično zaletišče, žuželke vodo vidijo kot nekaj, po čemer lahko hodijo.

Nekaj najpreprostejših izhodišč o povezavi med telesom in načinom razmišljanja se skriva v jeziku. Ljudje imamo oči uperjene naprej, ko hodimo in opazujemo predmete ter položaje, ki bodo čez hip naša prihodnost. To pomeni, da ljudje gledamo tedne ali leta vnaprej oziroma se z obžalovanjem ali hrepeneče spominjamo preteklih dogodkov. Celo pri abstraktnem razmišljanju, kot je matematika, govorimo o manipulaciji enačbe. Koren te besede se nanaša na roko (*manus* v latinščini pomeni roka, op. prev.) in veliko matematikov poroča, da si predstavljajo, kako z rokami razstavljajo in vnovič kombinirajo matematične koncepte.

Povezava med telesom in načinom razmišljanja ni vedno očitna in je nekaj, česar nismo zares vgradili v naprave. Po mojem mnenju je to manjkajoča sestavina v vseh teh izjemno zmogljivih, pa kljub temu presenetljivo pomanjkljivih inteligentnih tehnologijah: pomanjkanje telesa in vpogleda v povezavo med resničnim svetom in svetom inteligence, abstrakcije, sklepanja in kognicije.

Zaznati je mogoče, da imajo nekateri umetnointeligenčni sistemi temeljne pomanjkljivosti. V njihovem znanju in razumevanju so luknje v zvezi z resničnim svetom. Tisti, ki delamo z utelešeno umetno inteligenco, lahko izpostavimo, da ChatGPT ni zmožen poseči v fizični ali družbeni svet ter izkusiti njunega odziva na to.

Trenutna UI se obnese pri sestavljanju podob, posnetkov in odgovorov na vprašanja, ne zmore pa storiti ničesar oprijemljivega. Nima dostopa do resničnega sveta, ne ve, kaj deluje in kaj ne, kar je pomemben vidik inteligence. Snovanje novih zamisli je razmeroma preprosto, a prava vrednost človeške inteligence je ravno v tem, da smo zmožni prikazati, kako bi zamisli dejansko delovale v praksi.

Če je načrt za preprostega robota mogoče v pol minute izdelati na prenosniku, lahko začnemo razmišljati, kako bi z enakim postopkom zasnovali zapletenejše in zmogljivejše robote.

▶ Bi zadostovalo, če bi velik jezikovni model, podoben kot sistem UI, na podlagi katerega delujejo ChatGPT in podobni klepetalni roboti, povezali z robotskim telesom, kot je Google storil s Palmom?

Mislim, da odgovora trenutno ne pozna nihče. Poglablja se zavedanje, da moramo utelesiti velike jezikovne modele ali pa mora UI dobiti stik z resničnim svetom. Palm je ena od metod, po kateri bi, povedano preprosto, naredili velike možgane v ohišju in jih povezali z zapleteno napravo (v konkretnem primeru z vzvodom s kavljem, ki lahko pobira predmete). Samo ugibamo lahko, ali bo to delovalo dolgoročno, zagotovo pa mati narava to počne drugače.

Prva leta letalstva se je večina zamisli sukala okrog tega, ali bi izdelali velik trup, ki bo mahal s krili, ali majhnega. Odgovor bratov Wright je seveda bil, da ne eno ne drugo.

Začne preprosto. Večina živali življenje začne kot posamezne celice, ki se nato delijo in se specializirajo, tako da žival počasi dobiva tako bolj zapletene možgane kot telo. Narava ima dobre razloge za to. Ko otroka učimo kolesariti, ga ne posadimo na veliko kolo. Kolesu dodamo pomožna kolesca, začetniku olajšamo nalogo. Vaš razvoj so vaša pomožna kolesca. Po mojem mnenju naveza zapletenega telesa in možganov, ki jih bodo nato povezali v celoto ter držali pesti, najbrž ni primeren obrazec za dolgoročno uspešnost.

▶ Kakšno telo naj dobijo roboti?

To sodi k zabavnemu delu naslednje stopnje ustvarjanja inteligentnih naprav. Če je telo pomembno pri inteligentnem in varnem vedenju v resničnem svetu, kakšno bi bilo najprimernejše za posamezne naloge? Prva leta letalstva se je večina zamisli sukala okrog tega, ali bi izdelali velik trup, ki bo mahal s krili, ali majhnega. Odgovor bratov Wright je seveda bil, da ne eno ne drugo. Narava je z vsemi svojimi fizičnimi oblikami, ki jih razkriva, izredno ustvarjalna, vendar nam je ni treba nujno posnemati.

Delam na področju evolucijske robotike in UI prosimo, naj zasnuje ne le robotovih možganov, temveč tudi njegov telesni ustroj – kombinacijo telesa pod nadzorom desne polovice možganov za neko nalogo. Večina robotikov sestavlja evolucijske izdelke, v evolucijski robotiki pa poskušamo poustvariti evolucijo. In med tem postopkom nastajajo nova robotska telesa in možgani, kot jih v naravni še nismo videli.

▶ Evolucija traja dolgo, zato jo pospešimo s simulacijami. Postopek je uspešen za preproste robote in okolja, pa bi si z njim lahko pomagali tudi pri reševanju bolj zapletenih nalog in zagat?

Ni nujno, da posnemamo evolucijo, kot se dogaja v resničnem svetu. Biološka evolucija je slepa in napreduje z naključnimi mutacijami. Oktobra lani smo objavili razpravo, v kateri smo prikazali, da je napake pri metodi s poskusi mogoče omejiti. Mogoče je razviti UI, ki svoje robotske zasnove spremlja v simulacijah, in če z njimi ne doseže cilja, UI lahko postopku sledi v obratnem vrstnem redu in v robotovo telo ter odkrije točno tisto točko, na kateri se je zalomilo.

Dokazali smo, da je s tem mogoče pospešiti razvoj robotov v simulaciji, in sicer z dveh tednov na superračunalniku na pol minute na prenosniku. Algoritem, ki tako nastane, resda snuje razmeroma preproste robote, ki se pač malo premikajo naokrog. A če je načrt za preprostega robota mogoče v pol minute izdelati na prenosniku, lahko začnemo razmišljati, kako bi z enakim postopkom zasnovali zapletenejše in zmogljivejše robote.

▶ Kakšni bi bili videti takšni roboti?

Veliko mojih kolegov dela na razmeroma novem področju, znanem kot mehka robotika. Roboti tako niso iz kovine in keramike, temveč gume, silikona in drugih mehkih materialov. So votli, da jih je mogoče napihniti in tako povečati.

Mehka robotika je zelo obetavna tehnologija resničnega sveta, ki nam omogoča izdelavo varnejših robotov že samo zato, ker so mehki. To hkrati pomeni, da je mogoče spreminjati njihovo obliko in velikost. Mogoče jim je izvleči roke, noge, prste in orodje, nato pa jih vnovič skrčiti, ko jih ne potrebujemo več. Vidim torej veliko napredka v znanosti o materialih, zaradi katerega lahko robote

izdelujemo iz bolj eksotičnih materialov in mešanice mehkih ter togih sestavin.

▶ V zadnjih letih raziskujete tudi možnosti sestavljanja robotov iz živih materialov. Kako je to delo povezano z utelešeno umetno inteligenco in evolucijsko robotiko?

Sodelujem z Michaelom Levinom s Tuftsove univerze v Massachusettsu, ki je prikazal, da je s telesno zasnovo živali laže manipulirati, kot smo mislili. V mislih imam kirurško presajanje tkiva v živalskem zarodku, ki bo v številnih primerih zraslo v odrasel organizem, po videzu in vedenju precej drugačen od običajnih primerkov vrste. Mike je prikazal, da je mogoče kirurško prestaviti oči v zarodku žabe, tako da ima odrasla žival oči na hrbtu.

Pred okoli šestimi leti sem začel sodelovati z Levinovim laboratorijem in se vprašal. »Če si človeški biolog lahko izmisli način, kako prerazporediti živo tkivo in tako ustvariti nov organizem, bi lahko tudi UI pripravili do tega, da bi si ga izmislila? Tako smo UI prosili, naj celice žabje kože in srca poveže v takšnem vzorcu, da bo nastalo nekaj, kar hodi po dnu petrijevke. Srčno tkivo se krči in razteza, zato lahko deluje kot majhni bati ali motorček. Umetna inteligenca je približno dva tedna na superračunalniku preizkusila na milijone različnih razporeditev. Na koncu nam je ponudila kratek posnetek virtualnega žabjega robota. Posredovali smo ga mikrokirurgu v Mikovem laboratoriju, ki je nekaj ur pedantno presajal celice tja, kamor mu je naročila umetna inteligenca. Nato je svoje bitjece izpustil na prostost in ksenobot se je sprehajal po dnu petrijevke.

▶ Dosežek je nedvomno dih jemajoč. Pa roboti iz biološkega materiala sploh lahko postanejo kaj več od raziskovalne zanimivosti?

Še vedno gre za razmeroma preprosto znanost. Kljub temu bi me presenetilo, če organizmi, kot jih sestavi UI, ni bi bili nekoč uporabni. So zelo majhni in takšne robotke bi lahko poslali marsikam, da bi se razgledali, si vse zabeležili in se vrnili z uporabnimi podatki. Vendar je iz običajnih sestavnih delov težko izdelati zelo majhne naprave, ki pa so tudi biološko kompatibilne in razgradljive, zato lahko govorimo o ekološko prijazni tehnologiji.

▶ Kaj nam lahko povedo o vlogi, ki jo pri inteligenci igra morfologija?

Že zdaj nam lahko veliko povedo o naravi utelešenja in inteligenci. Kažejo zametke inteligence, zanimajo jih določene stvari v neposredni okolici, medtem ko jih druge odbijajo. A so samo koža žabe in srčne celice, nimajo možganov. Posamezne celice se še nikoli prej niso znašle v takšni razporeditvi, pa vseeno nekako doženejo, kaj naj storijo. Sporazumevajo se med seboj, in sicer električno, kemično in mehanično, se medsebojno odrivajo in privlačijo. Kot kaže, tudi sodelujejo, da naredijo, kar od njih hoče UI.

Inteligenca se skriva v celicah. To je naprava, sestavljena iz inteligentnih naprav. In tako predlaga popolnoma nove metode za izdelavo inteligentne tehnologije. Ljudje nas pogosto sprašujejo, kako programiramo te robote, ob čemer se samo nasmehnemo in odgovorimo: »Ravno to je bistvo.« Ni programiranja, v teh robotih ni drobnega biološkega računalnika. Ni razlike med krmilnikom naprave in samo napravo.

Torej imamo morda že od začetka napačen pristop. Čemur običajno rečemo možgani, je morda le podporna struktura, ne pa središče za ukaze in krmiljenje. Možgani sesalcev so med najnovejšimi izumu matere narave. Telo pa obstaja že zelo dolgo, zato je mogoče, da večina inteligence tiči v telesu, možgani pa zgolj omogočajo ali okrepijo to latentno inteligenco. ◀

»Presenetilo bi me, če organizmi, kot jih sestavi UI, ni bi bili nekoč uporabni. So zelo majhni in takšne robotke bi lahko poslali marsikam.«

Holografski usmerjevalnik Wi-Fi

V prazni okrogli stekleni posodi, ne veliko večji od bučke za ribe, se pojavi tridimenzionalni obraz in se nasmehne, kot bi ga pričarali iz lebdečega prahu slikovnih točk. Spominja na klasični trenutek s hologramom iz izvirnega filma Vojna zvezd, le da namesto posnetega videosporočila hologram spremljamo v živo, kot 'holoklic'.

Nate Berg, *Fast Company*

Naprava, ki omogoča takšne futuristične klice, je več kot znanstvena fantastika, je delujoč prototip, ki so ga razvili v londonskem oblikovalskem podjetju Layer ob sodelovanju z *Deutsche Telekom Design & Customer Experience*. Pred kratkim so jo kot del koncepta T, linije konceptnih izdelkov, namenjenih seznanjanju javnosti z zmogljivostmi telekomunikacijskih naprav, predstavili na svetovnem kongresu mobilne telefonije. Linija poleg holografske naprave obsega tudi prilagodljiv in modularni sistem za nadzorovanje internetnih omrežij znotraj doma in robotskega tovariša, ki ga poganja umetna inteligenca.

Deutsche Telekom, eden največjih ponudnikov telekomunikacijskih storitev v Evropi in večinski lastnik ameriškega operaterja T-Mobile, je jeseni 2022 podjetje Layer prosil, naj 'preobrazi' enega bistvenih, vendar nekoliko premalo cenjenih elementov telekomunikacij – usmerjevalnik.

»Vsako gospodinjstvo potrebuje usmerjevalnik, zato je postal izvrstno izhodišče za projekt konceptne naprave, da preverimo, kaj je mogoče storiti z njim, da bi pridobil boljše mesto v ospredju, namesto da se praši nekje v ozadju,« je pojasnil ustanovitelj in kreativni direktor Layerja Benjamin Hubert. »Običajno je nekje v kotu, za televizorjem, v omari. Upravičeno, saj tiho opravlja svoje delo.«

Deutsche Telekom je v svojih navodilih razvijalcem izpostavil, naj si zamislijo različice usmerjevalnika, ki zmore nekaj več, naj mu dodajo funkcije in izboljšajo uporabniško izkušnjo. »Trenutno uporabniki z usmerjevalnikom ne počnejo nič, razen ko morajo pritisniti na gumb, da ga ponastavijo,« je povedal Hubert. »Če bi bil usmerjevalnik naprava, s katero bi lahko počeli še kaj, kako bi to vplivalo na njegovo zasnovo in s tem na obliko ter tudi na končni videz?«

Iz takšnega razmišljanja so se porodili trije futuristični usmerjevalniki koncepta T. Prvi je domači pomočnik Buddy, ki združuje usmerjevalnik na podlagi umetne inteligence in robota na glasovno upravljanje.

Lahko se premika po hiši in namesto stanovalcev opravi avtomatizirane malenkosti, kot je uravnavanje temperature in svetlobe. Robot bo opremljen z napredno senzorsko tehnologijo, zato bo vedel, kdaj je treba zaliti rožo in prepoznal primer za nujno medicinsko pomoč.

Drugi bolj eksperimentalni usmerjevalnik je Level, sistem, ki ga je mogoče prenastavljati, sestavljajo pa ga modularni elementi, s čimer imajo uporabniki večji nadzor nad svojo internetno povezavo. Preprosti osnovi je mogoče dodati več modulov, kot so senzorji gibanja, dihanja in drugih biometričnih meritev na osnovi Wi-Fija in mrežnega ponavljalnika. Vsak skulpturni modul bo mogoče nadgraditi in nadomestiti, ko bo tehnologija napredovala, poleg tega pa je tudi ličen okras na polici.

Najbolj futuristični usmerjevalnik je View, hologramska naprava, ki omogoča holoklice in v svoji zaobljeni obliki prikaže tridimenzionalne podobe. Uporabljati jo je mogoče za projekcijo podatkov o usmerjevalniku, recimo o uporabi omrežja, a tudi za naprednejše prikaze, na primer prijateljeve podobe ali točno določenega mesta na živahni železniški postaji, kjer se je uporabnik dogovoril za srečanje.

Nenavadno, a zadnji Layerjev koncept ima največ možnosti, da se uresniči. Oblikovalska ekipa Layerja in Deutsche Telekoma je sodelovala z veliko skupino tehnologov in partnerskih podjetij ter zasnovala delujoč prototip naprave, za katero Hubert pravi, da lahko vse funkcije opravlja z že razpoložljivo tehnologijo. Zaobjema svetove oblikovanja, uporabniškega vmesnika, uporabniške izkušnje in znanstvene fantastike.

»Če razmišljamo o načinih uporabe, je mogoče res še v oblakih, vendar je že

Če bi bil usmerjevalnik naprava, s katero bi lahko počeli še kaj, kako bi to vplivalo na njegovo zasnovo in s tem na obliko ter tudi na končni videz?

oprijemljiv,« je pojasnjeval. »Prav gotovo je več kot samo koncept, saj deluje, kot je predvideno.«

Naprava holoklice omogoča zaradi vse bolj razširjenih kamer, ki zaznavajo tudi globino in so že na voljo na nekaterih pametnih telefonih. Več projektorjev v krogli projicira video na zelo tanko in prosojno membrano, ki niha vertikalno znotraj okvirja. Projicirane podobe razbije na sloje, ki se jih projicira na membrano, medtem ko se premika, da poustvari tridimenzionalnost posnetka v približno takšni ločljivosti, kot so videoigre iz 80. let.

»Mogoče bi to lahko opisali kot tridimenzionalno tiskanje s svetlobo. Tridimenzionalno tiskanje poteka v slojih materiala, projiciranje pa v slojih svetlobe,« je svojo razlago dopolnil Hubert. »Tako dobimo duhovom podobno tridimenzionalno prosojno bitje, ki se pogovarja z vami in vam pomaha. Ob tem ne morete, da se ne bi nasmehnili.«

Naprava lahko prikazuje tudi tridimenzionalne digitalne modele. »Prikaže vam lahko zgradbo, ki bi jo radi najeli, posest, ki jo kupujte, igrišče, ki ga nameravate obiskati, ali tridimenzionalni zemljevid področja, ki bi ga radi raziskali,« je povedal Hubert. »A smo šele na začetku, zato še preizkušamo, za kaj vse je uporabna.«

Neoster hologram v stekleni posodi se morda zdi korak nazaj ali vsaj vstran od

vizualnih možnosti, ki jih ponuja oprema za razširjeno resničnost, kot je Applov naglavni komplet Vision Pro, vendar Hubert zatrjuje, da z zamislimi, ki jih preverjajo v okviru tega koncepta, iščejo nove načine sporazumevanja in deljenja. Tridimenzionalni prikaz z naglavnim kompletom je primeren za nekatere okoliščine, vendar je včasih morda primernejša vizualizacija, ko nismo ujeti v napravo ali povezani z njo. S tehnologijo, ki je mogoča danes, se bodo pojavljali novi načine uporabe, je poudaril.

»Pri razvijanju tehnoloških platform začetek nikoli ne napoveduje konca,« je še dodal Hubert. »Ne trdimo, da je ta tehnologija boljša od virtualne resničnosti, je le drugačen pristop.« ◀

Avtomobilske baterije **za mraz**

Prihodnost prometa bo električna, in to ne le za avtomobile, temveč za vse od dostavnikov do letal. A preden se bo to uresničilo, se mora tehnologija znebiti nekaterih bolečih točk, kot so polnjenje in baterije.

Kristin Toussaint, *Fast Company*

Slabost baterij za električna vozila je, da se ne izkažejo vedno v skrajnih vremenskih pogojih. To se je pokazalo to zimo, ko so se vozniki električnih vozil v Chicagu spopadali s težavami s polnjenjem, z manjšim dosegom in s krajšim trajanjem baterije v obdobju s temperaturami pod zmrziščem. Vozniki se takšnim razmeram lahko prilagodijo s spremembo navad, a še vedno ostaja dejstvo, da imajo litij-ionske baterije v skrajnem mrazu in vročini dodatne težave – poleg omejitev v zvezi z dometom in s težo.

Proizvajalec baterij Amprius morda pozna rešitev, do katere je prišel z uporabo silicija v celicah baterije. Čeprav se ta proizvajalec primarno osredotoča na baterije za polete na električno, načrtuje, da se bo razširil tudi na električna vozila. Napredek pri njegovih baterijah bi lahko pomenil, da bodo na boljšem vsi načini električnega prevoza.

Zakaj jim mraz ni všeč?

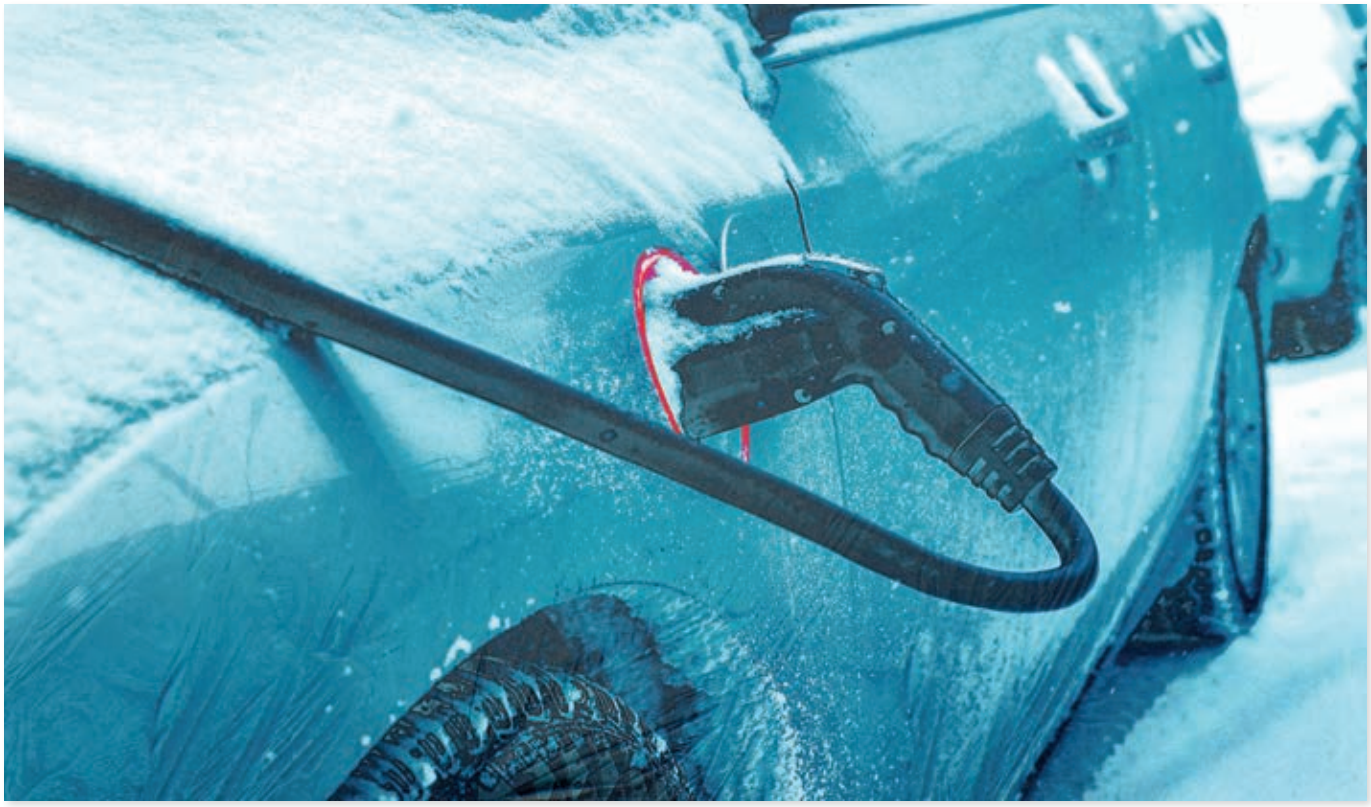
Večina električnih in hibridnih vozil ima litij-ionske baterije vgrajene z dobrim razlogom: imajo veliko energijsko gostoto, so razmeroma lahke in njihova življenjska doba je od 10 do 20 let. Poleg tega odlično delujejo tudi pri nizkih in visokih temperaturah – a v skrajnostih, kot so temperature pod ničlo in v hudi vročini, to ne velja več.

Baterije so kot ljudje: rade imajo zmerne temperature in ne marajo, če jih kdo sili, naj gredo prehitro ali prepočasi.

Osnovne sestavine baterije so pozitivna katoda, negativna anoda, separator in elektrolit – tekoča raztopina, ki prenaša ione sem ter tja; v primeru litij-ionskih baterij je to litij. Ko je baterija polna, je ves litij v anodi, če je prazna, pa se ves litij premakne v katodo.

Ionel Stefan, tehnološki direktor v Ampriusu, pravi, da so baterije kot ljudje: rade imajo zmerne temperature in ne marajo, če jih kdo sili, naj gredo prehitro ali prepočasi. V bateriji se nenehno odvijajo kemične reakcije; pri izjemno visokih temperaturah se hitreje, zaradi česar se baterija lahko hitreje stara. Njeno polnjenje je v vročini zahtevnejše, ker se te reakcije še bolj pohitrijo in vse to povzroča, da so baterija in materiali, iz katerih je, manj stabilni.

V hudem mrazu se zgodi ravno nasprotno: reakcije v bateriji se upočasnijo in zmanjša se naboj, ki je shranjen v njej, s tem pa tudi domet. Temu se je mogoče zoperstaviti z drugačnim vedenjem. Na Norveškem, v hladni



deželi, ki pa hkrati vodi pri uvajanju električnih vozil, se vozniki lahko naučijo, kako se prilagoditi, na primer, da baterijo najprej ogrejejo ali polnijo čez noč. To je tam preprosto, med drugim zaradi velikega števila domačih polnilnic. A celo ta prilagoditev ne pomaga, da bi električni avti v mrazu ostali priročni.

Stefan je opisal značilno reakcijo v hladni bateriji. Elektrolitski material se zgosti ali postane bolj viskozen, zato ionov ne prevaja enako hitro. »Pri nizkih temperaturah je škodljivo tudi, če jo poskušate napolniti, saj vam to ne bo ravno uspelo, ob tem pa lahko celo tvegate, da se bo litij premaknil v anodo in se nakopičil na njenem vrhu,« je pojasnil.

Tako lahko nastane dendrit, skup kovin, ki lahko seže do katode, če se dovolj dolgo povečuje, in tako pride do električne povezave med njima ter kratkega stika v bateriji. Kratki stiki pa lahko povzročijo okvaro baterije ali celo požar.

Kako silicij pomaga baterijam?

Amprius, ki je pravzaprav nastal kot zagonsko podjetje univerze Stanford, izdeluje ultra goste baterije, in sicer namesto grafita v anodi uporablja silicij. Grafit je prva izbira za baterijske anode že od 90. let, saj ga je dovolj in je energijsko gost, vendar ima silicij desetkrat večje shranjevalne zmožnosti. Težava je le, da se mu poveča prostornina, ker absorbira litij – in to tudi do 300 odstotkov. »Raztezanje in krčenje med praznjenjem in polnjenjem sodita med lastnosti, ki si jih pri bateriji ne želimo,« je povedal Stefan.

Veliko proizvajalcev baterij tuhta, kako vgraditi silicij brez neugodnih učinkov (dva primera sta Group14 in Sila Nanotechnologies). Nekateri so siliciju primešali grafit ali ogljik, da bi dosegli večjo energijsko gostoto brez takšnega napihovanja. Amprius pa je ustvaril trirazsežnostno strukturo, ki omogoča napihovanje. »Lahko se trudiš doseči, da se to ne bi dogajalo, ali pa to dopustiš, vendar ne tako, da se vidi,« je razlagal Stefan.

Silicij izboljša delovanje baterije v mrazu: ker ima rahlo višjo voltažo kot grafit, je polnjenje pri nizkih temperaturah varnejše in ne pride do dendritov, ki povzročajo kratke stike v bateriji. Elektroliti so v mrazu resda še vedno lenobni, a se ni treba bati ogrožanja varnosti, in medtem ko se baterija ogreva, se polnjenje pospeši. Ker so baterije s silicijevo anodo energijsko gostejše, se v hladnem vremenu izgublja manj energije. Medtem ko večina baterij zaradi nizkih temperatur v povprečju izgubi štiri desetine energije, jim je v Ampriusu ta delež uspelo oklestiti na petino.

Dodajanje silicija v baterijo bi lahko pripomoglo tudi k večjemu dometu in hitrejšemu polnjenju. Lani so v neodvisnem laboratoriju potrdili, da ima Ampriusova baterija energijsko gostoto 500 WH/kg; Tesline

Lani so v neodvisnem laboratoriju potrdili, da ima Ampriusova baterija energijsko gostoto 500 WH/kg; Tesline baterije dosegajo okoli 269 WH/kg.

baterije dosegajo okoli 269 WH/kg. V Ampriusu pravijo, da se njihove baterije z nič na 80 odstotkov napolnijo v osmih minutah, zdržijo pa dvakrat toliko kot grafitne baterijske celice.

Prvi Ampriusovi kupci so iz panoge vesoljskih in električnih poletov, saj sta to področji, kjer je energijska gostota še kako pomembna, je povedal Stefan. Baterije za letalski promet morajo biti čim lažje, a še vedno dovolj zmogljive. Novembra lani si je Amprius zagotovil naročila iz treh podjetij za električna letala. A napredek pri lažjih, energijsko gostejših baterijah se bo prej ali slej pokazal tudi v kopenskem prometu, torej v avtomobilih in tovornjakih.

Električni avti so težji od vozil, ki imajo motorje z notranjim izgorevanjem, kar prav tako vpliva na domet. Pri večini majhnih avtov to ni velika težava – proizvajalci električnih vozil so se najprej osredotočili na čim ugodnejšo ceno, lažje baterije ter s tem lažje vozilo pa niso bili na vrhu seznama s prednostnimi nalogami. Za tovornjake, sploh velike s težkim tovorom, ki ga morajo prepeljati na dolgih razdaljah, pa je to ovira. Če dodajo več baterij, dodajo tudi težo. »Ne moreš dodati za pet ton baterij, če pa je največja obremenitev tovornjaka med 20 in 25 tonami,« je pripomnil Stefan. »Vsak kilogram šteje … Domet tovornjakov je res majhen.« Te težave je imel tudi Fordov F150 Lightning: vozniki so poročali, da se je domet občutno zmanjšal, ko so na poltovornjak pripeli še prikolico. Energijsko gostejše baterije so za nameček tudi lažje.

Stefan se zaveda, da obstajajo še druge tehnologije za baterije, a zanj je odgovor jasen: silicij. »Mislim, da bo naslednja generacija električnih vozil v anodi imela vsaj malo silicija,« je komentiral. A težav s tem še ne bo konec, treba bo povečati tudi proizvodnjo. Ampriusove baterije za letala so že na voljo, medtem ko podjetje pričakuje, da jih bodo za avtomobile začeli prodajati leta 2027 ali pozneje. ◀

V Ampriusu pravijo, da se njihove baterije z nič na 80 odstotkov napolnijo v osmih minutah, zdržijo pa dvakrat toliko kot grafitne baterijske celice.

Natisnjeno raketno gorivo

Predstavljajte si, da se v dežju peljete v službo, in ko niste pozorni, se od nekod pojavi divjak in trešči v vaš avtomobil. Kot bi trenil, se sunkovito razpre varnostni meh in vam reši življenje. Varnostni meh se je tako hitro sprožil zaradi energijskega materiala, imenovanega natrijev azid, ki med kemično reakcijo, s katero se napolni meh, tvori dušik. A kaj so energijski materiali?

Monique McClain, *The Conversation, Fast Company*

Mednje sodijo pogonska goriva, pirotehnika, gorivo in eksplozivna sredstva, uporabljajo pa jih za različne namene, na primer za rakete, vžigalice, raketne pogone, potisni plin za orožje, termitno varjenje, ognjemet in eksplozivna sredstva za posebne učinke, kot jih vidite v najljubših akcijskih filmih.

Energijski materiali so najrazličnejših oblik in velikosti, najpogosteje pa v trdi obliki. Z gorenjem oziroma eksplozijo sprostijo veliko energije, odvisno od njihove oblike in razmer, v katerih jih uporabimo.

Sem predavateljica strojništva, ki preučuje energijske materiale. Njihova izdelava ni preprosta, vendar bi napredek pri tridimenzionalnem tiskanju lahko olajšal prilagajanje in povečal možnosti znanstvene uporabe.

Vloga geometrije

Raven energije v materialih vpliva na njihovo obliko in sproščanje energije. Trdo raketno gorivo je izdelano podobno kot torta – robot meša »testo«, ki je večinoma iz amonijevega perklorata, aluminija in gumastega veziva, nato pa ga prelije v ponev. Zmes se strdi v ponvi, medtem ko se peče v pečici.

Raketna goriva so največkrat valjaste oblike, le da imajo na sredini palico posebnega prereza v obliki kroga ali zvezde. Ko se gorivo strdi, palico odstranijo in na sredini ostane odprtina.

To jedro vpliva na izgorevanje goriva, kar lahko spreminja potisk v motorju. Že samo z drugačno obliko goriva je mogoče doseči, da motor pospešuje, upočasnjuje ali ohranja svojo hitrost.

Vendar je pri tradicionalni »peki« omejena oblika. Ko se gorivo strdi, mora ostati možnost odstranitve palice; če ima ta preveč zapleteno obliko, lahko zlomite gorivno pogačo in potem izgoreva neenakomerno.

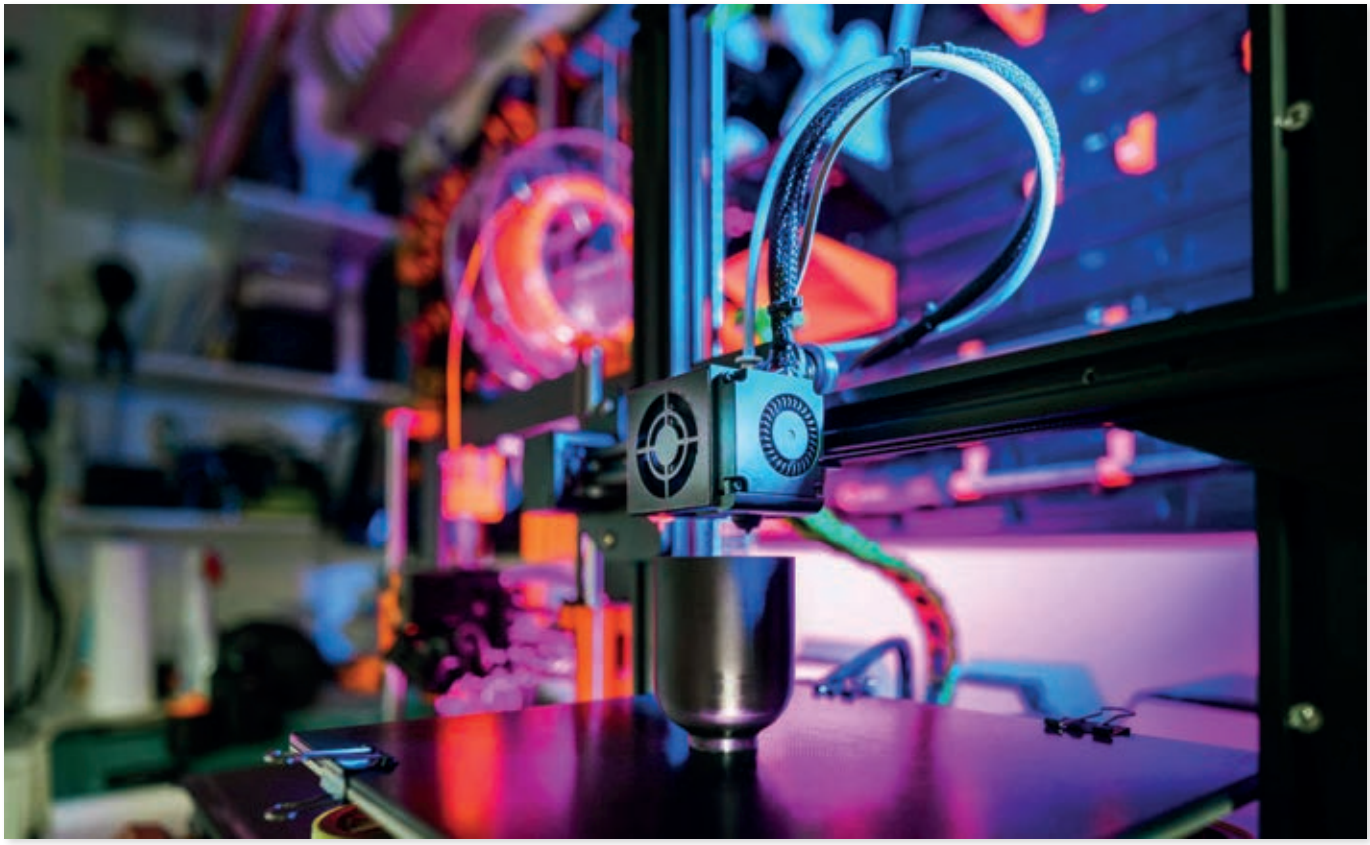
Oblikovanje goriva, da rakete letijo hitreje ali dlje, je živahno raziskovalno področje, vendar bi inženirji potrebovali nove proizvodne metode, da bi lahko zasnovali tudi bolj zapletene oblike.

Rešitev je tridimenzionalno tiskanje

Tridimenzionalno tiskanje je v več pogledih povzročilo revolucijo v proizvodnji. Raziskovalci poskušamo ugotoviti, kako lahko izboljša učinkovitost energijskih materialov. Pri tej vrsti tiskanja se material nalaga v slojih in tako nastane predmet. Omogoča oblike po meri, uporabi se lahko različne materiale tudi v istem predmetu, prihranita se denar in material.

Vendar energijski materiali tudi za tridimenzionalno tiskanje predstavljajo precejšen izziv. Nekateri so zelo viskozni in takšno mešanico je iz tube z majhno odprtino težko iztisniti, saj je, kot bi skozi injekcijo poskušali spraviti glino – pregosta je za ozko konico.

Poleg tega so energijski materiali lahko nevarni, če z njimi ne ravnamo pravilno.



Lahko se vnamejo, če se med proizvodnjo ali skladiščenjem temperatura preveč zviša, enako velja, če so izpostavljeni statičnemu električnemu šoku.

Najnovejši dosežki

Raziskovalcem je v preteklem desetletju kljub vsemu uspelo premagati nekaj ovir. Znanstveniki so tako začeli tiskati reaktivno črnilo na elektroniko, s čimer so omogočili samouničenje, če bi material prišel v napačne roke.

Teoretično bi to črnilo s tridimenzionalnim tiskalnikom lahko nanesli tudi na stare satelite ali ostarelo mednarodno vesoljsko postajo, in te naprave, ki krožijo v orbiti, bi razpadle v dovolj majhne smeti, da bi zgorele v atmosferi, preden bi lahko treščile na Zemljino površje.

Mnogi raziskovalci se posvečajo eksplozivu za orožje, ki bi ga tiskali s tridimenzionalnim tiskalnikom. Če bi spremenili obliko tega goriva, bi lahko izdelali krogle, ki bi letele dlje.

Drugi želijo s tridimenzionalnim tiskanjem omejiti vpliv streliva za orožje in vžigalnika, katerega proizvodnja ni mogoča brez ostrih topil, na naravo. Ta topila so nevarna, in ko odslužijo, jih je težko varno odvreči, saj lahko škodijo okolju ter človekovemu zdravju.

Dokazala sem, da je mogoče natisniti trdo raketno gorivo s podobnimi lastnostmi, kot jih imajo tradicionalno proizvedena. S temi raziskavami se zdaj odpira priložnost, da preverimo, kako izgorevajo goriva, narejena iz več materialov. To pa je novo področje.

Na primer, namesto palice za prerez v gorivu bi lahko s tridimenzionalnim tiskalnikom natisnili visoko reaktiven material, ki bi ga dodali v sredico. In namesto da bi nam bilo treba nato ta material na sredini odstraniti, bi novi zgorel tako hitro, da bi za njim ostal prazen prostor. Ta reaktivni material bi tudi dodal energijo gorivu, s tem pa ne bi več bilo potrebe po nameščanju in odstranjevanju palice za sredico.

Znanstveniki so začeli tiskati reaktivno črnilo na elektroniko, s čimer so omogočili samouničenje, če bi material prišel v napačne roke.

Večina tovrstnih raziskav je še v povojih, vendar podjetja, kot je X-Bow, že tiskajo goriva in opravljajo uspešne teste s temi motorji.

In končno, več raziskovalcev preučuje, kako detonirati natisnjena eksplozivna sredstva. Ko jih natisnejo v mrežo, se odzivajo različno, če so pore napolnjene z vodo ali zrakom. S takšnim postopkom dobijo varnejše eksplozivno sredstvo, ki reagira izključno pod točno določenimi pogoji.

Tridimenzionalno tiskanje energijskih materialov je novo področje in znanstvenike čaka še dolga pot, preden bodo povsem razumeli, kako tiskanje vpliva na varnost in učinkovitost. Vendar znanstveniki, kot sem jaz, vsak dan odkrivamo nove načine za uporabo natisnjenih energijskih virov za ključne in včasih življenjsko pomembne namene. ◀

Avtorica je profesorica asistentka strojnega inženirstva na univerzi Purdue.

Umetna inteligenca Deepminda je mojstrica iger

En sam sistem umetne inteligence (UI) lahko človeškega igralca premaga v šahu, goju, pokru in drugih igrah, v katerih ni mogoče zmagati brez vrste strategij. V Googlovi hčerinski družbi Deepmind so razvili UI, imenovano Student of Games. Ta naj bi bila po navedbah avtorjev korak proti umetni splošni inteligenci, ki bo z nadčloveško sposobnostjo opravila katerokoli nalogo.

Matthew Sparkes, *New Scientist*

Martin Schmid, ki je v Deepmindu razvijal UI, zdaj pa dela v zagonskem podjetju Euqilibre Technologies, pravi, da model Student of Games izvira iz dveh projektov. Prvi je bil Deepstack, umetna inteligenca, ki jo je razvila ekipa (njen član je bil tudi Schmid) z univerze v kanadski Alberti. To je tudi prvi model UI, ki je v pokru premagal profesionalnega človeškega kvartopirca. Drugi projekt je Deepmindov Alphazero, ki je premagal najboljše človeške igralce v igrah, kot sta šah in go.

Razlika med modeloma je, da se je prvi osredotočil na igre brez popolnih informacij, pri katerih igralci ne vedo, kakšno je stanje pri soigralcih, na primer igre s kartami, drugi pa na igre s popolnimi informacijami, kot je šah, kjer oba igralca vidita položaj vseh figur. Pristop do teh dveh tipov iger je namreč čisto drugačen.

Deepmind je najel ekipo Deepstacka, da bi razvila model, ki bi bil zmožen posploševanja pri obeh tipih iger, in tako je nastal Student of Games.

Kot pravi Schmid, se ta model začne kot 'navodila', kako se naučiti iger, in nato z vadbo in s prakso postaja vse boljši v njih. Začetniški model se lahko loti različnih iger in se nauči igrati proti drugi različici samega sebe, odkriva nove strategije in postopno postaja spretnejši. Medtem ko se je starejši sistem Deepminda, Alphazero, lahko prilagodil le drugim igram s popolnimi informacijami, se Student of Games prilagaja obema vrstama, zato je zmožen daleč večjega posploševanja.

Raziskovalci so Student of Games preverili pri šahu, goju, pokru Texas hold'em in namizni igri z imenom Scotland Yard, pokru Leduc hold'em in različici Scotland Yarda po meri z drugačno desko. Ugotovili so, da lahko premaga več obstoječih modelov UI in človeške igralce (objavljeno v *Science Advances*).

Kljub dih jemajočim sposobnostim čaka Student of Games še dolga pot, preden bi UI lahko pojmovali kot splošno inteligentno.

Lahko bi se naučil tudi drugih iger, meni Schmid. »Tudi v igrah, ki jih ni še nikoli igral, je res dober.«

Te obsežne zmogljivosti so mogoče zaradi učinkovitosti v primerjavi z bolj specializiranimi algoritmi Deepminda, vendar Studen tof Games v igrah, ki se jih nauči, z lahkoto premaga celo najboljše človeške igralce.

Michael Rovatsos z univerze v Edinburgu pravi, da Student of Games kljub dih jemajočim sposobnostim čaka še dolga pot, preden bi UI lahko pojmovali kot splošno inteligentno, saj so igre okolje z nedvoumnimi pravili in vnaprej znanimi vedenjskimi vzorci.

»To je nadzorovano, omejeno in umetno okolje, v katerem je popolnoma jasno, kaj nekaj pomeni in kakšne so posledice nekega dejana,« je poudaril Rovatsos. »Ne rečem, da igre niso zapletene, vendar to ni resnični svet.«

Monitor

Besedila so prevzeta iz revij New Scientist, Fast Company in The Atlantic, vse pravice pridržane. Distribucija Tribune Content Agency.

ODGOVORNI UREDNIK
Matjaž Klančar

UREDNIK
Uroš Mesojedec

LEKTURA
Simona Mikeln

PREVAJANJE
Petra Piber

LIKOVNA ZASNOVA
Peter Gedei

OBLIKOVANJE NASLOVNICE
Peter Gedei

RAČ. GRAFIKA IN STAVEK
Peter Gedei

FOTOGRAFIJE
Fotoarhiv Monitorja, Profimedia, Midjourney

NASLOV UREDNIŠTVA
Monitor, Dunajska 51, 1000 Ljubljana, tel.: (01) 230 65 00 faks: (01) 230 65 10 e-pošta: urednistvo@monitor.si

MONITOR V SPLETU
www.monitor.si

Nenaročenih rokopisov in fotografij ne vračamo. Vse gradivo v reviji Monitor je last družbe Mladina d.d. Kopiranje ali razmnoževanje je mogoče le s pisnim dovoljenjem izdajatelja.

IZDAJATELJ
Mladina časopisno podjetje d.d., Dunajska cesta 51, 1000 Ljubljana, dav. št. 83610405

IZVRŠNA DIREKTORICA
Denis Tavčar

PRODAJA OGLASNEGA PROSTORA
tel.: (01) 230 65 36, e-pošta: marketing@monitor.si

VODJA MARKETINGA IN OGLASNEGA TRŽENJA
Ines Markovčič, tel.: (01) 230 65 33

NAROČNINE IN PRODAJA
tel. (01) 230 65 00, e-pošta: narocnine@monitor.si

RAČUNOVODSTVO
e-pošta: racunovodstvo@monitor.si

TISK
Shwartz Print, Ljubljana

DISTRIBUCIJA
Izberi d.o.o., Ljubljana

Poštnina za naročnike plačana pri pošti 1102, Ljubljana. V ceno izvodov v maloprodaji s priloženim DVDjem je vključen DDV v višini 22%, v ceno ostalih izvodov pa DDV v višini 9,5%.
ISSN 1318-1017

BERITE MONITOR 20% CENEJE

Revijo Monitor lahko naročite tako, da plačate letno naročnino in jo od naslednje številke naprej prejemate na želeni naslov.

• Vse informacije lahko dobite po telefonu (01) 230 65 30 ali po elektronski pošti narocnine@monitor.si.

Izid je finančno podprla Javna agencija za raziskovalno dejavnost Republike Slovenije.