ALTERNATOR
Misliti znanost.

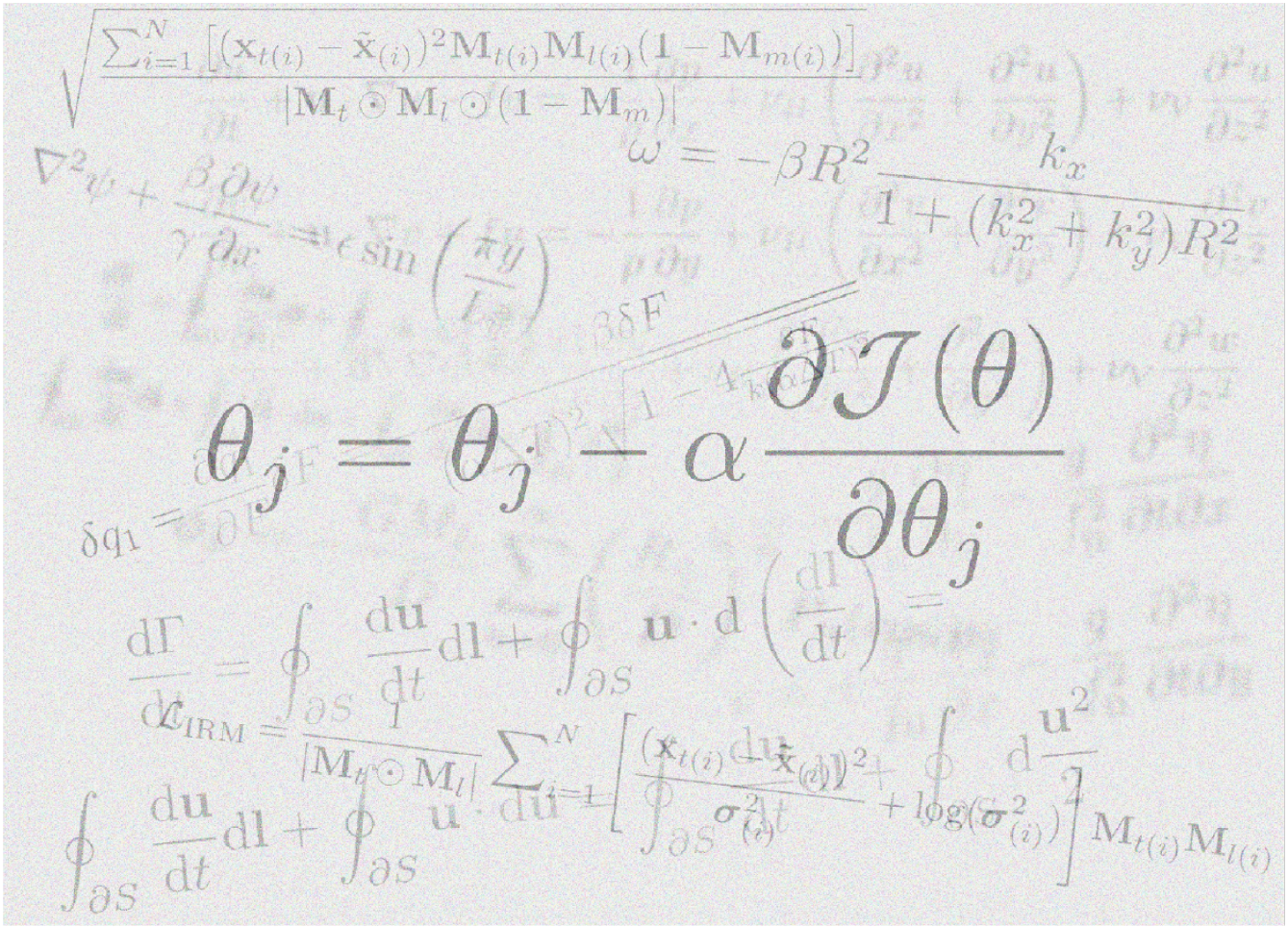
Nelagodje znanosti zaradi umetne inteligence – prvih dvesto let

29. 1. 2026

Številka: 02/2026

Avtor:

- Matjaž Ličer



Slika: Matjaž Ličer

Jorge Luis Borges v svoji zgodbi *O natančnosti v znanosti* (orig. *Del rigor en la ciencia*, naslov je prevod avtorja, op. a.) opisuje cesarstvo, v katerem je veščina kartografije dosegla tolikšno natančnost, da so bili cesarski zemljevidi provinc veliki kot celotna mesta, zemljevid cesarstva pa je bil velik kot vsa provinca. A kartografi so šli še dlje in izdelali najveličastnejši zemljevid cesarstva, katerega platno je v naravi povsem sovpadalo s cesarstvom samim. Kasnejše generacije so hitro ugotovile, da je tovrstni zemljevid neuporaben, ter velikansko platno prepustile soncu in zimam, njegove ostanke pa lahko danes najdemo le še v najodročnejših puščavah cesarstva.

Borgesova skica tako rekoč igraje zadane najgloblje bistvo znanosti, kot se ga je razumelo preteklih 200 let: ko postane znanstveni opis enako kompleksen kot pojavi, ki jih opisuje, postane nekoristen. Razsvetljenska znanost je svojo vrednost v veliki meri črpala iz dejstva, da je preprostejša od svojega objekta. Drugače rečeno, znanost se je do danes vedno pretežno razumelo kot karseda preprosto reprezentacijo sveta, ki operira z dovolj nizkim številom parametrov, da o njih ljudje še vedno lahko koherentno razmišljamo, a je obenem dovolj kompleksna, da opiše bistvo stvari. Znana je opazka slovitega fizika Johna von Neumanna – človeka, ki je formaliziral kvantno mehaniko in postavil temelje sodobnim računalniškim algoritmom –, da lahko s štirimi parametri opiše slona, s petim pa pomiga z njegovim rilcem. Von Neumannova poanta je preprosta: če v svoj opis sveta poljubno dodajamo proste parametre, bomo sčasoma sicer lahko

opisali sleherni proces, razumeli pa ne bomo nobenega. Pomislimo samo na ptolemajsko geocentrično sliko Osončja, ki planete opisuje z nizom ekscentričnih krožnic, ki krožijo vzdolž drugih krožnic, ki krožijo vzdolž tretjih krožnic in tako dalje. Najbolj izpopolnjeni ptolemajski modeli Osončja so uporabljali več kot šestdeset tovrstnih sestavljenih krožnih gibanj. Kopernikanski sistem se ni uveljavil zato, ker bi bil natančnejši – ptolemajska slika omogoča poljubno natančnost izračuna planetnih orbit. Heliocentrizem se je uveljavil, ker je bil pojmovno preprostejši in ker je bil skladen z Newtonovo mehaniko, ki je imela neverjetno razlagalno in napovedno moč. Geocentrizem lahko seveda uporabljamo še naprej, cena, ki jo zato plačamo, pa zajema ogromno število parametrov (radiji in obhodne dobe vseh krožnic, potrebnih za opis posamezne orbite planeta) in pomanjkanje osnovnega fizikalnega razumevanja Osončja.

Von Neumannova pripomba je torej vsekakor tehtna in utemeljena na znanstveni tradiciji. Ampak zagata je v tem, da je ost te pripombe uperjena natanko proti sodobni umetni inteligenci: sodobne globoke mreže so ogromne množice parametrov (https://doi.org/10.3986/alternator.2023.14) , ki jih v procesu učenja prilagajamo, dokler ne dobimo rezultata, ki ga želimo, ne omogočijo pa nam razumevanja, kako in zakaj so do tega cilja prišle. Izziv znanosti je bil v modernem obdobju ravno nasproten: razumeti karseda široke množice pojavov ob čim manjšem številu parametrov. V zadnjih nekaj stoletjih je šlo pri znanosti torej za iskanje kompromisa med univerzalnostjo uvida in preprostostjo opisa. Umetna inteligenca tovrstnih kompromisov, ki so napajali odpor fizikov do pretiranega števila parametrov, vsaj *a priori* ne potrebuje.

Omenjena redukcija kompleksnosti je bila znanosti do neke mere položena že v zibelko, in sicer z Galilejevo premiso, da je velika knjiga narave zapisana v jeziku matematike. Galilej je z njo predpostavil korespondenco med empiričnim svetom, v katerega smo vrženi, in matematičnim svetom naših misli. Navidezno pestrost vseh naravnih pojavov naj bi bilo možno opisati matematično in s karseda nizkim številom parametrov. Tovrstna matematizacija je abstrakcija od empiričnega sveta: forma, ki v pojavih, ki jih opisuje, zajema bistveno in opušča nebistveno. Fiziki in kemiji je to z uporabo matematičnih orodij zadnjih nekaj stoletij izjemno dobro uspevalo, to sta namreč naši najnižjedimenzionalni reprezentaciji narave (tj. opisujeta naravo s karseda nizkim številom prostih parametrov in predpostavk). Na teh temeljih so zrasli znanstveni preboji 17. stoletja, ki so prispevali h kasnejšim družbenim transformacijam, kot je industrijska revolucija. Tedanja intenzivna matematizacija sveta in hkratna industrializacija sta sovpadli – seveda ne po naključju – z vznikom evropske romantike, tehnološkega skepticizma in ambivalentnega odnosa človeštva do lastnih stvaritev. Razsvetljenstvo, znanstvena »odčaranost sveta« ter vera v tehnologijo in znanost niso pripeljali le do tehnoloških refleksij, temveč tudi do vprašanja o naravi inteligence kot take ter do kritičnih razmislekov o možnostih umetne inteligence in njenih posledicah.

A ti razmisleki se prvenstveno niso zgodili na področju znanosti, temveč v literaturi, kjer so pot najprej in najnatančneje utirale zgodbe E. T. A. Hoffmanna, ki je pred več kot dvesto leti odprl številna vprašanja, ki nas danes v razmerju do umetne inteligence prav posebno zaposlujejo. Nekateri nastavki teh vprašanj so prisotni v znanem Hoffmannovem *Peskarju* (*Der Sandmann*, 1816), v *Avtomatih* (*Die Automate*, 1819) pa so ta vprašanja postavljena še nekoliko natančneje. *Peskar* se, spomnimo, odvija skozi pisemsko korespondenco glavnega protagonista Nathaniela in njegove zaročenke Klare, ki razkriva, na kakšen način se je Nathaniel v različnih obdobjih življenja spopadal s svojo obsedenostjo z likom Peskarja, grozljive fantazijske figure, ki porednim otrokom v očesne dupline tlači pesek, dokler jim očesna zrkla ne izpadejo, da jih lahko Peskar odnese in z njimi nahrani svoje otroke. Eno od digresij v zgodbi predstavlja zaplet, ko se glavni junak Nathaniel zaljubi v Olimpijo, dekle, ki ga očara s svojim popolnim posluhom, molčečo zadržanostjo in avro nedoločljive skrivnostnosti. Naposled se izkaže, da je dekle izpopolnjen mehanski stroj, ki – ne moremo mimo te opazke – uteleša mokre sanje današnje moškosfere: to je dekle, ki posluša brez prekinjanja, govori le, ko je vprašana, ne ugovarja in vedno zrcali čustva glavnega junaka. Z drugimi besedami, Olimpija se obnaša kot današnji jezikovni modeli in je idealna partnerica za nekoga, ki ne išče zveze, temveč potrditev. Hoffmann interakcijo med Nathanielom in Olimpijo opisuje s slabo prikritim norčevanjem, ki ga naposled razširi na celotno sfero zmenkarij: da bi se prepričali, da niso zaljubljeni v antropomorfne robote, so mnogi ljubimci po razkritju, da je Olimpija stroj, od svojih izvoljenk zahtevali, naj v svojih glasbenih nastopih pojejo narobe, predvsem pa naj v pogovorih ne le poslušajo, temveč kdaj tudi kaj povejo oziroma izrazijo kako misel in čustvo. Lahko si torej mislimo, s kakšnimi občutki bi Hoffmann gledal na današnjo, bodimo iskreni, moško klientelo serijske proizvodnje umetnointeligenčnih spolnih robotk.

V delu *Avtomati*, ki je izšlo nekaj let po *Peskarju*, Hoffmann nadaljuje v isti smeri. Opisuje inteligentnega antropomorfnega robota, ki se v zgodbi imenuje Turek. Turek je sposoben inteligentne interakcije z ljudmi: ljudje si ga hodijo ogledovat in se z njim pogovarjajo, on pa jim daje večinoma koristne odgovore na vprašanja z najrazličnejših področij in odgovarja v številnih svetovnih jezikih. Hoffman tako leta 1819 v dobrem in slabem predvideva današnje velike jezikovne modele: ko ga ljudje kaj okorno ali nespretno vprašajo, Turek pogosto daje enako samozavestne, a povsem napačne odgovore, ki jim danes nekoliko neposrečeno pravimo halucinacije. Podobno kot veliki jezikovni modeli Turek nekoč za nekaj tednov izgine, nato pa se zopet pojavi v bolj prefinjeni obliki in dajajoč boljše odgovore. Temu bi danes rekli, da je bil stroj doučen, dotreniran. Kaj se dogaja s Turkom, z gotovostjo ne ve nihče, kajti njegov izumitelj profesor X, podobno kot številna UI tehnološka podjetja, z njim upravlja v popolni tajnosti, ga je pa, spet podobno kot številna podjetja, pripravljen komurkoli pokazati in ponuditi za pogovor. Glavna protagonista zgodbe se na neki točki odpravita k profesorju X, da bi bolje razumela, zakaj Turek daje takšne odgovore, kot jih daje. To je podobno, kot če bi od tehnoloških podjetij zahtevali razlago, zakaj njihovi modeli dajejo takšne odgovore, kot jih. In prav kakor je skrivnost delovanja ChatGPT skrivnost za OpenAI, tako tudi Hoffmannova protagonista razočarana spoznata, da profesor X nima odgovorov, ki jih pričakujeta od njega. Turek je tudi za svojega izumitelja netransparentna črna škatla.

Hoffmannove zgodbe stavijo na sodobno mehanično sliko biološkega življenja, ki je pomenila temeljni premik v novoveški in kasnejši znanosti, ampak izkoristijo jo na način, ki v pripoved vnaša določeno nedomačnost (seveda ni naključje, da je

Peskarja v svojem besedilu o nedomačnosti (*Das Unheimliche)* obravnaval tudi Freud). Temeljne predpostavke Hoffmannovega mehanicizma so še danes vseprisotne v znanosti povsod tam, kjer se opisovanje narave reducira na razumevanje delovanja narave kot stroja, na razumevanje zakonov mehanike njenih gradnikov. V določenih obdobjih, v podobnem kontekstu opozarja Paolo Rossi (https://www.zalozbacf.si/index.php/rojstvo-moderne-znanosti-v-evropi.html) , je prevladovalo stališče, da je vesolje ne le narejeno za človeka, temveč mu je podobno tudi po strukturi. Mehanicizem pa je, nasprotno, človeka in vesolje interpretiral kot bolj ali manj zapletena stroja ter s tem iz razmišljanja o naravi odstranil vsakršni temeljni antropomorfizem. Človeško in živalsko fiziologijo je v okviru mehanicizma mogoče razumeti z metaforo stroja. Pri Descartesu so živalski živci cevi, mišice in kite pa vzmeti. La Mettrie je to metaforo razširil na človeka – stavil je na to, da je mogoče sleherni fiziološki proces razumeti skozi mehanske interakcije, ki se odvijajo v telesu. Zavrnil pa je Descartesov dualizem duše in telesa ter duševne procese ravno tako pripisal fizikalnim procesom v možganih. Ta predpostavka, ki sodobni medicini vsekakor ni tuja – v veliki meri je za sodobno znanost celo nujna –, je v različnih obsegih prisotna v likih Turka in Olimpije. Hoffmanovi avtomati se zdijo nekakšna interpolacija med okornimi stroji in živimi organizmi – kot bi Hoffmann v mehanicizem vpeljal nekakšno gradacijo, ki sicer izhaja iz povsem klasičnih kolesij in vzmeti mehanskih strojev, katerih kompleksnost nato raste preko vseh meja, dokler se na koncu ne iztečejo v inteligentne, antropomorfne organizme, s katerimi lahko govorimo in v katere se lahko zaljubimo.

Hoffmann je potemtakem izumil hibridne kibernetične organizme, ki zvezno prehajajo med poloma mehaničnosti in inteligentnosti. Njegovi mehanizmi igrajo vlogo nekakšnih strojnih orakljev, ki jih ljudje uporabljajo tako, kot mi uporabljamo velike jezikovne modele umetne inteligence – za raziskovanje samih sebe in narave. S tem smo prek Hoffmanna in mehanicistov pristali na točki, kjer je naš znanstveni opis narave prav tako zapleten kot narava sama – vrnili smo se torej k Borgesovemu zemljevidu vseh zemljevidov. Napetost med klasično znanstveno težnjo po poenostavitvah in rastočo kompleksnostjo globokih modelov umetne inteligence torej ni nekaj sodobnega, temveč ima svoje korenine v evropski romantiki. Ta napetost je neposredni produkt novoveške znanosti in razsvetljenstva – Hoffmann vzame v svojih besedilih razsvetljensko vodilo o moči razuma povsem zares in ga prižene do skrajnih meja, do umetne inteligence, kot jo poznamo danes. Globoke mreže so s svojimi milijardami prostih parametrov pravzaprav nekakšen Borgesov zemljevid, kompleksen prav toliko kot cesarstvo vse človeške vednosti. Vprašanje, ki se s tem odpira, pa je, do kolikšne mere je bogastvo parametrov in uboštvo razumevanja za znanost dejansko relevanten epistemološki problem.

S tem merimo na dejstvo, da si je znanost doslej morda lahko privoščila opisovanje narave s preprostimi modeli z nizkim številom prostih parametrov, ker se ni pogosto srečevala s problemi, kjer tovrstne poenostavitve niso mogoče. A v zadnjih desetletjih so v znanosti vzniknili problemi doslej neznane kompleksnosti, za katere je postajalo vedno bolj jasno, da bodo z ustaljenimi metodami novoveške znanosti zelo težko ali pa nikoli rešljivi. Zdelo se je, da bo tovrstne sisteme nemara lažje narediti kot simulirati – od tod fizični vetrovniki v avtomobilskih tovarnah in kvantni računalniki, ki fizično udejanjijo kompleksne kvantne sisteme, ki so klasičnim algoritmom na klasičnih procesorjih povsem nedostopni. Tovrstnih sistemov se morda načeloma niti ne da omejiti na opis z majhnim številom prostih parametrov. Z drugimi besedami – pri nerazumevanju teh procesov morda ne gre za naše metodološke pomanjkljivosti, temveč so ti problemi nemara v svojem bistvu prekompleksni, da bi jih lahko ljudje, navajeni razmišljanja v nizkodimenzionalnih pojmovnih prostorih, kadarkoli zares dojeli na način, kot smo ga bili vajeni v zadnjih stoletjih.

V nekaterih tovrstnih primerih se je umetna inteligenca že izkazala kot odlično komplementarno orodje klasični znanosti. Tak primer je globoka mreža AlphaFold2 (https://doi.org/10.1038/s41586-021-03819-2) , za katero je bila leta 2024 v zvezi z določanjem strukture proteinov podeljena Nobelova nagrada za kemijo. Šlo je za problem, ki so ga biokemiki desetletja eksperimentalno in teoretično reševali, z umetno inteligenco pa je problem danes, zahvaljujoč preteklim meritvam, rešen. Skratka, Borgesov cesarski zemljevid je morda pretrd oreh za človeške kartografe, ustrezni globoki mreži pa orientacija v visokodimenzionalni pokrajini proteinskih oblik ne predstavlja nobenih težav. Epistemološki odpor do mnoštva parametrov, ki je bil v preteklosti ukoreninjen v fiziki in marsikje drugje v naravoslovju, danes ni več tako prisoten. Marsikdo v znanstveni skupnosti se je z nerazložljivostjo umetne inteligence nemara sprijaznil iz povsem pragmatičnih razlogov. Ampak poanta pričujočega besedila je ravno nasprotna: znanost danes obravnava bistveno kompleksnejše procese kot pred sto leti in zdi se vsaj verjetno, da nekaterih opisov teh procesov ne bo možno ubesediti na nizkodimenzionalni način – nerazložljivost v teh primerih ni metodološki problem, temveč je v svojem bistvu povezana s kompleksnostjo dotičnih procesov. In umetna inteligenca je komplementarno hevristično orodje, ki lahko klasični znanosti pomaga rešiti, kar se rešiti da. Poleg tega, da nerazložljivost ne bi smela biti ovira pri uporabi algoritmov, ki rešujejo probleme, ki jih drugače ne znamo rešiti, je nujno omeniti, da se na področju umetne inteligence mnogo napora vlaga v zasnovo razložljive oz. interpretabilne umetne inteligence, ki bo nudila določen vpogled v način, kako je bila rešitev oblikovana, in kot taka ne bo več povsem črna škatla. Temu razvoju skuša slediti tudi evropska zakonodaja.

Morda bi bilo zato bolj smiselno, da bi umetno inteligenco kritično obravnavali z neke druge perspektive. Resno je treba vzeti precej utemeljeno skrb, da bodo tovrstne tehnologije močno povečale človekov ogljični odtis in družbene neenakosti, vodile bodo v neslutene možnosti družbenega nadzora in manipulacije ter s tem spodkopale demokratične ureditve, ki so se doslej bolj ali manj uspešno trudile s političnim zastopanjem svojih prebivalcev. Pretirano osredotočanje na umetno inteligenco kot epistemološko črno škatlo ter hkratno zanemarjanje političnih in ekonomskih tveganj, ki jih ta tehnologija prinaša s seboj, se zato zdita kot osredotočanje na napačen problem oziroma, če se še zadnjič naslonimo na *Peskarja*, metanje peska v oči.