

SUPERINTELIGENCA

Poslednji izum človeštva?

Mitja Pavlič

Knjiga o grožnji, tveganju in meji, po kateri človek morda ne bo več najinteligentnejše bitje v svojem lastnem svetu.

»Razvoj umetne inteligence je tako temeljnega pomena, kot so bile iznajdbe mikroprocesorja, osebnega računalnika, interneta ali mobilnega telefona. Spremenil bo način, kako delamo, se učimo, potujemo, skrbimo za zdravje in komuniciramo med seboj. Cele panoge se bodo preoblikovale okoli njega.«

Bill Gates, soustanovitelj Microsofta, 2023

UVOD

Obožujem jo, a se je hkrati bojim

—
*»Mislim, da je povsem mogoče, da je človeštvo le prehodna faza v evoluciji
inteligence.«*

Geoffrey Hinton, boter umetne inteligence in dobitnik Nobelove nagrade, 2024

Obožujem jo, a se je hkrati bojim

• • •

Stojim pred občinstvom in predavam o umetni inteligenci. Na velikem zaslonu za menoj se izmenjujejo različni primeri uporabe sistemov, ki znajo pisati, analizirati ogromne količine podatkov, ustvarjati slike, glasbo in videe. Ljudje poslušajo, a nekaj je drugače kot pred nekaj leti. Pozornost je globlja, pogledi so resnejši. V zraku ni več le radovednost in navdušenje, temveč tudi nekaj, kar bi lahko opisal kot tiho napetost. Kot bi vsi slutili, da se pred nami dogaja nekaj večjega od običajnega tehnološkega napredka.

Nato se v zadnji vrsti dvigne roka.

Vprašanje, ki ga slišim vedno pogosteje:

»Mislite, da se moramo bati umetne inteligence?«

Vsi pogledi se obrnejo proti meni. Za trenutek obstanem. Ker je moj odgovor neprijetno zapleten. Resnica je hkrati pritrtilna in nikalna. Da. Umetne inteligence se moramo bati. Toda ne na način, kot si večina ljudi predstavlja.

Ne kot robotov iz filmov, ki se nekega dne odločijo uničiti človeštvo. Strah, o katerem govorim, je veliko bolj tih. Bolj podoben vodi, ki za jezom počasi narašča. Dolgo se zdi, da je vse pod nadzorom. Gladina se viša skoraj neopazno in večina ljudi tega sploh ne zazna. Potem pa nekega dne pogledaš proti jezu in ugotoviš, da voda ni več tam, kjer je bila še včeraj. In ko končno razumeš, kako visoko je narasla, je morda že prepozno, da bi jo ustavil.

To vprašanje ni nastalo čez noč. Raslo je počasi, neopazno, med vrsticami navdušenja in tehnološkega optimizma. Na predavanjih sem vedno znova opazal majhen, a vztrajen del ljudi, ki niso ravno navdušeno ploskali najnovejšim dosežkom, temveč so ob njih postajali vse tišji.

Ljudje, ki niso iskali odgovorov na vprašanje, kaj vse umetna inteligenca zmore, temveč so si zastavljali težje vprašanje – kaj vse to lahko pomeni.

Toda časa za ta vprašanja je bilo vedno premalo. Predavanja so se morala prilagajati večini, ki je v umetni inteligenci videla predvsem priložnost, napredek in učinkovitost. In prav zato je nastala moja prva knjiga.

Ne kot projekt. Ne kot ambicija. Temveč kot nuja.

Knjiga *Temna stran umetne inteligence* je bila moj prvi poskus, da zberem vse tisto, za kar na predavanjih ni bilo prostora. Da dam glas tistim dvomom, ki so ostajali neizrečeni. Da napišem knjigo za tiste, ki so želeli pogledati dlje, tudi tja, kjer pogled ni več udoben. Knjigo sem sprva nameraval deliti zgolj udeležencem svojih predavanj, kot dopolnilo, kot moje darilo, kot tiho opozorilo za tiste, ki so želeli slišati več in brez olepševanja.

Nisem imel namena, da bi knjiga prišla na police knjižnic in knjigarn. Še manj sem imel namen postati pisatelj.

A nekaj dni po tem, ko sem iz tiskarne prejel prve izvode, so se začeli oglašati distributerji. Prepričevali so me, da je tema preveč aktualna, da bi ostala zaprta samo v predavalnicah, med manjšo skupino ljudi – bralcev. In tako je knjiga, mimo mojega prvotnega načrta, našla pot do širše javnosti.

Verjetno se zdaj, ko držite v rokah to mojo drugo knjigo sprašujete, zakaj druga knjiga – in to tako kmalu.

Odgovor ni strateški. Ni načrtovan. In ni marketinški.

Je preprosto iskren.

Občutek, da v svoji prvi knjigi nisem povedal vsega.

Ko sem zaprl prvo knjigo, sem vedel, da sem odprl vrata, nisem pa stopil skozi njih. Govoril sem o tveganjih, o pasteh, o temni strani. Nisem pa šel dovolj daleč in nisem dovolj jasno spregovoril o hitrosti razvoja.

O smeri, v katero gremo. Predvsem pa ne o tem, kam ta pot lahko pripelje – do nečesa, kar presega vse, kar smo kot ljudje kadarkoli ustvarili – do SUPERINTELIGENCE.

Ta knjiga je zato nadaljevanje, a hkrati upam tudi presežek prve. Je poskus, da naredimo korak naprej. Da ne govorimo več le o tem, kaj lahko umetna inteligenca naredi danes, temveč o tem, kaj se zgodi, ko preseže meje, ki jih še razumemo – ko ne bo več le orodje, temveč sistem, ki nas v določenih pogledih že prekaša.

Zato ta knjiga, enako kot prva, ni napisana iz strahu, temveč iz občutka odgovornosti. Verjamem, da je znanje najboljšo orožje proti slepemu optimizmu in nepotrebnemu strahu. Ko tehnologijo razumemo, lahko o njej razmišljamo trezno – in ko razmišljamo trezno, sprejemamo boljše odločitve.

Ta knjiga ni le o tehnologiji. Je o nas samih.

O vprašanju, ki si ga kot civilizacija še nismo dovolj jasno zastavili: kakšno prihodnost pravzaprav želimo ustvariti. Svet, v katerem tehnologija služi človeku – ali svet, v katerem človek začne, skoraj neopazno, služiti tehnologiji.

Dolgo smo verjeli, da so največji izumi tisti, ki širijo naše sposobnosti. Orodja, ki nas naredijo hitrejša, močnejša, učinkovitejša. Toda prvič v zgodovini stojimo pred možnostjo, da ustvarimo nekaj, kar ne bo le podaljšek človeka, temveč tudi njegov naslednik v sposobnosti razumevanja, odločanja in ustvarjanja.

In prav zato to ni več zgolj tehnološko vprašanje.

Je civilizacijsko vprašanje.

Morda ustvarjamo najmočnejša orodja v zgodovini. Morda pa ustvarjamo nekaj, kar bo zadnje, kar bomo kot človeška vrsta sploh ustvarili sami.

Ne zato, ker bi nas uničilo. Temveč zato, ker nas bo preseglo.

In ko enkrat obstaja inteligenca, ki lahko izboljšuje samo sebe hitreje, kot jo lahko razumemo mi, se vloga človeka nepovratno spremeni. Ne nujno izgine. A zagotovo ni več enaka.

Zato naslov te knjige ni provokacija.

Je vprašanje, na katerega še nimamo odgovora:

Je superinteligence poslednji izum človeštva?

In morda še pomembneje – bomo ta trenutek prepoznali, ko bo enkrat zares prišel.

1

—

Superinteligenca

Ko stroj začne razmišljati sam

—
»Ko bo ustvarjena superinteligenca, bo to zadnji izum, ki ga bo človeštvo kdaj potrebovalo narediti.«

Nick Bostrom, avtor knjige Superintelligence, 2014

Ko stroj začne razmišljati sam

• • •

Zgodovina človeštva je tudi zgodovina orodij. Vsaka velika civilizacijska sprememba se je začela z novim orodjem, ki je razširilo človeške sposobnosti. Kamen, iz katerega smo naredili prvo rezilo. Kolo, ki je premaknilo težo sveta. Tiskarski stroj, ki je znanje razširil hitreje, kot bi ga lahko prenašal katerikoli učitelj. Električna, ki je noč spremenila v dan. Računalnik, ki je pospešil računanje do hitrosti, ki si je človeški um ne more niti predstavljati.

V vseh teh primerih pa je veljalo eno preprosto pravilo: orodja so razširila naše roke, naše mišice in naše sposobnosti – nikoli pa našega razuma. Kladi lahko udari močnejše kot roka, vendar ne ve, kam udariti. Parni stroj lahko premika tone tovora, vendar ne ve, kam potuje. Računalnik lahko izvede milijarde operacij na sekundo, vendar še vedno potrebuje človeka, ki mu pove, kaj naj počne.

Toda novembra 2022 se je zgodilo nekaj nenavadnega.

Ne v raziskovalnem laboratoriju ali na kakšni konferenci znanstvenikov, temveč na internetu, pred očmi milijonov ljudi. Podjetje OpenAI je javnosti predstavilo orodje z imenom ChatGPT. Na prvi pogled ni bilo videti nič posebnega. Preprosta spletna stran, prazno polje za vnos besedila in utripajoč kazalec. Vsakdo je lahko vpisal vprašanje in v nekaj sekundah dobil odgovor.

Sprva smo mnogi mislili, da gre le za še en pameten pogovorni program.

Toda hitro je postalo jasno, da se dogaja nekaj drugečnega. Sistem ni le odgovarjal – razlagal je, ustvarjal, povezoval in včasih celo presenetil. Odgovori niso bili vedno popolni. Tudi napačni. Toda veliko bolj pogosto so bili presenetljivo dobri. Včasih celo osupljivi.

V prvih petih dneh je ChatGPT dobil več kot milijon uporabnikov. V nekaj mesecih več kot sto milijonov. Nobena tehnologija v zgodovini interneta se ni širila tako hitro.

Ljudje po vsem svetu smo doživljali isti trenutek: sedeli smo pred zaslonom, vpisali vprašanje in dobili odgovor, ki je zvenel, kot da ga je napisal človek. Ne kot seznam povezav, temveč kot razlaga. Kot pogovor. Kot misel.

Mnogi smo prvič začutili nenavaden občutek – kot bi se pogovarjali z nečim, kar razume.

Toda resnica je tišja – in veliko bolj zaskrbljujoča.

Umetna inteligenca ni vstopila v naš svet s ChatGPT-jem. Tam je bila že dolgo prej. Sodeluje pri večini stvari, ki nas vsak dan obkrožajo – od iskalnikov in navigacije do priporočil in filtrov, ki odločajo, kaj sploh vidimo. Povprečen človek se z umetno inteligenco sreča nešteto krat na dan. Večinoma nevede.

In morda najbolj zaskrbljujoče: večine vsebine, ki jo vidimo na svojih zaslonih, ne izberemo sami. Izbere jo umetna inteligenca.

Vprašanje ni več, kaj gledamo. Vprašanje je, kaj nam je sploh dovoljeno videti.

Marca 2016 se je v Seulu odvijala partija starodavne igre Go med človekom in strojem. Na eni strani je sedel Lee Sedol, eden najboljših igralcev na svetu. Na drugi strani ni sedel nihče. Tam je bil sistem AlphaGo.

V četrti partiji je stroj odigral potezo, ki je postala znana kot »Poteza 37«. Sprva se je zdela napačna. Nelogična. Toda nekaj potez pozneje je postalo jasno, da je stroj videl nekaj, česar ljudje nismo. Lee Sedol je dolgo molčal. Pozneje je priznal, da ima občutek, kot da igra proti nečemu, kar razmišlja drugače kot človek.

Na koncu je AlphaGo zmagal.

Ta trenutek je bil prelomnica. Ne zato, ker bi računalnik prvič premagal človeka – to se je zgodilo namreč že prej v šahu. Temveč zato, ker stroj ni samo računal hitreje. Našel je strategije, ki jih človeški um ni poznal.

Prvič smo videli inteligenco, ki ni bila kopija človeškega razmišljanja.

V tem se skriva najgloblja sprememba, ki jo prinaša umetna inteligenca. Dolgo smo verjeli, da bodo računalniki vedno ostali orodja. Izjemno hitra, izjemno uporabna – a še vedno orodja. Danes pa sistemi ne izvajajo le navodil, temveč se učijo iz podatkov, odkrivajo vzorce in včasih najdejo rešitve, ki jih človek ni predvidel.

Ko se to zgodi enkrat, je zanimivo.

Ko se začne dogajati vedno pogosteje, postane jasno, da smo stopili na novo pot. Pot, ki vodi do vprašanja, ki si ga človeštvo še nikoli ni bilo zares prisiljeno zastaviti:

»Kaj se zgodi, ko inteligenca, ki smo jo ustvarili, začne presegati našo?«

Zgodba se začne z vprašanjem, ki ga je postavil Alan Turing. In tam se začne tudi naslednje poglavje.