

Nedovršena basna o vrapcima

Beše sezona građenja gnezda, i nakon višednevnog teškog rada, vrapci sedošu u smiraj dana, opuštajući se i cvrkućuci.

„Tako smo mali i slabašni. Zamislite kako bi nam život bio lak kad bismo imali sovu koja bi nam mogla pomoći da izgradimo naša gnezda!”

„Da!”, reče drugi. „I mogli bismo je uposliti da se brine o našim starima i o mladuncima.”

„Mogla bi nam davati savete i držati komšijsku mačku na oku”, dodade treći.

Onda Hraniša, ptica starešina, reče: „Pošaljimo izviđače u svim pravcima da negde pokušaju naći napušteno sovče ili možda jaje. Poslužilo bi i mladunče vrane ili mladunče lasice. Ovo bi mogla da bude najbolja stvar koja nam se ikad desila, bar od otvaranja Paviljona besko-načnog zrnevlja u onom dvorištu preko plotu.”

Jato beše uzbuđeno i vrapci posvuda počesu da cvrkuću iz petnih žila.

Jedino Drndisad, jednooki vrabac razdražljivog temperamenta, ostade neubeđen u mudrost tog poduhvata. I reče: „Ovo će nam sigurno doći glave. Zar ne treba prvo da savladamo veštinu pripitomljavanja i kroćenja sova, pa tek onda da takvog stvora dovedemo k sebi?”

Hraniša odgovori: „Kroćenje sove je, po svemu sudeći, izrazito težak posao. Biće nam teško da nađemo i sovino jaje. Počnimo odatle. Kad uspemo da uzgojimo sovu, možemo početi da razmišljamo o suočavanju s tim drugim zadatkom.”

„Taj plan je manjkav!” – ciknu Drndisad; ali njegovo negodovanje beše uzalud, jer jato je već bilo poletelo da izvrši uputstva što ih je izneo Hraniša.

Samo dva ili tri vrapca ostadoše. Zajedno su pokušavali da utvrde kako se sove mogu ukrotiti ili pripitomiti. Ubrzo shvatiše da je Hraniša bio u pravu: beše to izrazito težak izazov, pogotovo u odsustvu stvarne sove na kojoj bi se mogli oprobati. Ipak, oni nastaviše kako su znali i umeli, stalno strahujući od toga da će se jato vratiti sa sovinim jajetom pre nego što pronađu rešenje za problem kontrole.

Ne zna se kako se priča završava, ali autor posvećuje ovu knjigu Drndisadu i njegovim sledbenicima.

PREDGOVOR

U vašoj lobanji je stvar koja čita. Ta stvar, ljudski mozak, poseduje izvesne sposobnosti koje nedostaju mozgovima drugih životinja. Tim osobenim sposobnostima dugujemo naše dominantno mesto na Zemlji. Druge životinje imaju snažnije mišiće i oštrije kandže, ali mi imamo pametnije mozgove. Naša skromna prednost u opštoj inteligenciji dovela je do toga da razvijemo jezik, tehnologiju i složeno društveno uređenje. I kako se svaka generacija nadograđivala na postignuća prethodnikā, tako je prednost rasla.

Ako jednog dana napravimo mašinske mozgove koji nadmašuju ljudske mozgove u opštoj inteligenciji, onda ova superinteligencija može postati veoma moćna. I kao što sudbina gorila sada zavisi više od nas ljudi nego od samih gorila, tako bi i sudbina naše vrste mogla zavisi od delanja mašinske superinteligencije.

Mi ipak imamo jednu prednost: mi smo ti koji treba da napravimo tu stvar. U principu, mogli bismo napraviti takvu vrstu superinteligencije koja bi čuvala ljudske vrednosti. Sigurno bismo imali dobar razlog da to učinimo. U praksi, problem kontrole – problem toga kako kontrolisati ono što bi superinteligencija činila – izgleda vrlo težak. Takođe se čini da ćemo imati samo jednu šansu. Jednom kad neprijateljska superinteligencija nastane, ona će nas sprečiti da je zamenimo ili da izmenimo njene preference. Naša sudbina bila bi zapečaćena.

U ovoj knjizi pokušavam da razumem izazove koje pruža očekivani dolazak superinteligencije, te kako bismo na njih najbolje mogli odgovoriti. Sasvim je moguće da je ovo najvažniji i u najvećoj meri zastrašujući izazov s kojim se čovečanstvo ikada suočilo. I – bilo da uspemo ili ne – verovatno je poslednji izazov s kojim ćemo se ikad suočiti.

Ova knjiga ne raspravlja o tome da li smo na pragu velikog proboja na polju veštačke inteligencije, ni da li s bilo kakvom preciznošću možemo predvideti kada bi se takav razvitak mogao desiti. Izgleda nekoliko verovatno da će do toga doći tokom ovog veka, ali to ne znamo pouzdano. U prvih nekoliko glava raspravlja se o mogućim putanjama i kazuje ponešto o vremenu. Veći deo knjige, međutim, tiče se onoga što se dešava nakon toga. U njemu ispitujemo kinetiku eksplozije inteligencije, oblike i moći superinteligencije, te strateške izbore što stoje na raspolaganju superinteligentnom delatniku* koji je ostvario presudnu prednost. Potom izmeštamo pažnju na problem kontrole i pitamo šta se može učiniti da se oblikuju početni uslovi tako da postignemo ishod koji je povoljan i koji čovek može preživeti. Pred kraj knjige, odmičemo se i razmatramo širu sliku koja izranja iz naših istraživanja. Iznose se predlozi o tome šta treba učiniti sada kako bismo povećali svoje šanse da izbegnemo egzistencijalnu katastrofu u vremenu koje će doći.

Ovu knjigu nije bilo lako napisati. Nadam se da će putanja koju sam utro omogućiti drugim istraživačima da brže i podesnije dosegnu nove granice kako bi tamo stigli odmorni i spremni da se priključe daljem širenju dosega našeg shvatanja. (A ako je put koji je napravljen malo kvrgav i vijugav, nadam se da kritičari pri proceni rezultata neće potceniti nepristupačnost terena *ex ante!*)

Ovu knjigu nije bilo lako napisati: pokušao sam da je učinim lakom za čitanje, ali ne mislim da sam u tome sasvim uspeo. Pišući je, na umu sam, kao ciljanu publiku, imao raniji vremenski isečak sopstvenog bića, i pokušao sam da napišem takvu knjigu kakvu bih ja sa uživanjem čitao. To se može pokazati kao mala ciljna grupa. Ipak, mislim da bi

* Prim. prev: često se zatiče slučaj da se engleska reč „agent” prevodi domaćim paronimom „agent”, iako je engleska reč znatno šireg obima značenja, te pored „zastupnika” označava i „ono što izvršava radnju”; iz tog razloga odlučeno je da se reč „agent” prevodi kao „delatnik”, izuzev u nekoliko slučajeva u kojima je smisleno zadržan tehnički izraz u programerskom kontekstu; sledstveno tome, u Glavi 9, engleski termin „principal-agent problem”, koji inače ima mnogostruke prevode („vlasničko-zastupnički problem” ili „problem nalagodavca i nalagoprimsca”), doslednosti radi, preveden je kao „problem nalagodavac-delatnik”.

sadržaj mogao biti dostupan mnogim ljudima samo ako malo razmisle o tome i odupru se iskušenju da svaku novu zamisao odmah pogrešno razumeju time što će je utopiti u najpribližniji kliše koji imaju na svojoj kulturološkoj polici. Laici koji čitaju ovu knjigu ne treba da se obeshrabre pred povremenim primesama matematike ili specijalističkog žargona, jer je glavnu poentu uvek moguće prozreti iz okolnih objašnjenja. (I obratno, za one čitaoce koji žele više detalja, podosta može da se nađe u beleškama.**)

Mnoge tvrdnje iznete u ovoj knjizi verovatno su pogrešne.*** Takođe je verovatno da postoje razmatranja od kritične važnosti koja nisam uzeo u obzir, čime sam neke ili sve svoje zaključke učinio nevaljanim. Potrudio sam se da ukažem na nijanse i stepene neizvesnosti unutar teksta – time što sam ga opteretio mrljom neodređenih izraza „moguće”, „možda”, „može”, „moglo bi”, „čini se”, „verovatno”, „vrlo izvesno”, „sasvim izvesno”. Svaki kvalifikator stavljen je tamo gde je – pažljivo i namerno. Ta mestimična upotreba epistemičke skromnosti ipak nije dovoljna; ona se mora propratiti sistemskim priznanjem neizvesnosti i pogrešivosti. To nije lažna skromnost: dok ja verujem kako je moja knjiga verovatno ozbiljno pogrešna ili zavaravajuća, mislim da su alternativna gledišta pružena u literaturi znatno gora – uključujući i podrazumevano gledište ili „nultu hipotezu” prema kojoj za sada bezbedno i smisleno možemo ignorisati izgled superinteligencije.

** Ne sadrže, međutim, sve beleške korisne informacije.

*** Ne znam koje.