

Paolo Benanti. *Human in the LOOP: Decisioni umane e intelligenze artificiali*. Mondadori: Milano, 2022. 161 str. ISBN 979-12-206-0047-7.

Paolo Benanti, profesor na Papeški univerzi Gregoriana v Rimu, raziskuje zapleteno razmerje med človeškim odločanjem in UI. Pomenljivi naslov knjige Paola Benantija, ki je izšla leta 2022, *Human in the LOOP: Decisioni umane e intelligenze artificiali*, bi lahko v slovenščino prevedli *Človek v zanki: Človeške odločitve in umetna inteligenca*. V tej knjigi avtor obravnava ključno vprašanje – kako ohraniti človeško delovanje in dostojanstvo v dobi, ko stroji vse bolj vplivajo na naše odločitve? Osnovno vprašanje, ki si ga Benanti postavi na samem začetku in se ponavlja skozi vso knjigo, je: »Če umetna inteligenca (UI) postane nevidna kot voda, kakšne posledice lahko pričakujemo za človeka in družbo?« Kot voda, skozi katero lahko vidimo in je na neki način nevidna, a je z razvojem digitalne tehnike omogočila ponoven razvoj UI. Pri tem razvoju so težnje, ki želijo zakriti njegove negativne posledice za človekovo dostojanstvo. Zato je naloga etike, da pri razvoju nenehno sodeluje: opozarja na probleme oz. zanke, jih dela vidne.

Knjiga je razdeljena na dva dela. V prvem je obravnavan razvoj inteligentnih strojev in njihova sposobnost sprejemanja odločitev na podlagi podatkov in algoritmov. Benanti trdi, da umetna inteligenca sicer lahko obdeluje velike koli-

čine informacij in ugotavlja povezave, vendar pa ji manjkata razumevanje in etično razmišljanje, ki sta značilna za človeško odločanje. To je razlog, zakaj morajo ljudje ostati v te procese vključeni in biti nanje posebej pozorni: zagotoviti morajo, da umetna inteligenca ne bo delovala samostojno, kar bi lahko privedlo do odločitev brez moralnega premisleka (9–86).

V drugem delu so obravnavane različne etične dileme, kot so tiste, ki jih ponazarja eksperiment ‚moralni stroj‘; poudarjena je zapletenost človeških odločitev, ki jih ni mogoče enostavno zreducirati na algoritemske modele. Benanti znova svari pred tem, da bi stroji postali ‚nevidne‘ sile v naših življenjih, podobno kot voda za ribe – vseprisotni, a neopaženi. Treba je zagotoviti, da bodo sistemi UI zasnovani ob upoštevanju človekovega dostojanstva in pravic ter preprečiti zdrs v utilitarizem brez etičnega razmisleka (89–150).

Benanti poudarja, da je človek biološko pomanjkljivo bitje, saj brez pripomočkov ne more preživeti v nobenem naravnem okolju. V tem biološkem stanju je slabši od živali, saj nima posebej prilagojenih organov, ki bi ga neposredno zaščitili, ampak je preživetje odvisno od njegove sposobnosti, kako lahko svoje pomanjkljivosti preseže s pomočjo orodij. Zato človek ‚manjkajoče organe‘ nadomešča, krepi obstoječe in razbremenjuje delovanje organizma (12–15).

Vendar pa človek svojega preseganja ne razvija le na funkcionalni ravni, ampak tudi simbolni. Med temi najstarejšimi sledovi so jamske poslikave in pogrebni obredi, ki kažejo na človekovo iskanje in poskus izražanja človeškega presežka. Človek je torej po svoji identiteti duhovno bitje – in ta njegov konstitutivni presežek se nahaja na križišču njegovega doživljanja sveta in tehnologije: zmožnost delovanja, izražanja, sklicevanja na ‚onostransko‘ predstavlja globoko jedro njegovega eksistencialnega doživljanja. Toda v trenutku, ko je začel ustvarjati sintetično resničnost, je človekov odnos do sveta postal problematičen. Izvor te resničnosti je v presečišču več pojavov: industrializacije, vzpona kapitalističnega gospodarstva konec 19. stoletja, vojn in državnih interesov ter vrste drugih konfliktov 20. stoletja. Avtor je prepričan, da gre danes za revolucijo, ki so jo na področje znanosti in tehnologije prinesli računalniki in računalništvo. Z računalnikom lahko skoraj vsak človeški problem spremenimo v statistične podatke, grafe in enačbe. Zanj je resnično zaskrbljujoče to, da s tem ustvarjamo iluzijo, da so ti problemi rešljivi z računalniki. Danes se za dovolj natančno napovedovanje uporablja korelacija. Ker smo se na stvarnost ne samo naučili gledati kot na niz podatkov, temveč smo se velike količine podatkov naučili tudi zbirati, smo se opremili z novim orodjem za raziskovanje: ne potrebujemo več kamenčkov, kalkulatorjev, temveč nova orodja, ki nam omogočajo vidnost povezav v velikem podatkovju. Vendar pri ugotavljanju morebitnih vzročno-posledičnih povezav (korelacij), na katerih deluje UI, pogosto prihaja do lažnih korelacij. O njih govorimo takrat, ko med dvema

pojavoroma, ki statistično sicer sta v korelaciji, ne obstaja nobena dejanska neposredna vzročno-posledična povezava (16–50.)

Računalnik je postal medij par excellence 21. stoletja prav zato, ker obdeluje jezik vseh drugih medijev v digitalni obliki. Zlasti je orodje za pisanje za vse: novinarje, pisatelje, znanstvenike, inženirje, pesnike in umetnike. V veliki meri je spremenil tradicionalne tehnike pisanja, enako pa velja za urejanje, fotomontažo in sam tisk (51–52).

Cilj UI – ki je mlada disciplina, stara komaj šestdeset let in izhaja iz sklopa znanosti, teorij in tehnik – je posnemanje kognitivnih sposobnosti človeka. Pri opisu UI se uporabljajo pojmi, ki so izposojeni iz matematične logike, statistike, verjetnosti, računalniške nevrobiologije in računalništva. Problem, ki ga avtor vedno znova izpostavlja, je transparentnost delovanja: velike težave so pri razločljivosti procesov odločanja, ki jih UI izvaja (52–59).

Računalništvo postane sredstvo, s katerim računalnik lahko odredi končno zaporedje osnovnih operacij (tistih, ki jih procesor izvede v določenem času in jih ni mogoče dodatno zmanjšati ali razčleniti v druge operacije za pretvorbo vhoda v koherenten izhod).

Med drugim velja poudariti, da avtor omenja temeljno razliko med računalniškimi algoritmi in programsko kodo. Prvi so logični računski postopki za pridobitev rezultata z združevanjem osnovnih operacij; koda je izražena s programskim jezikom: formalnim jezikom, zasnovanim za realizacijo sklopov navodil (vhod), s katerimi se ustvarijo izhodni podatki (izhod). Razvojni jeziki abstraktne človeške pojme v bistvu pre-

tvarjajo v strojno kodo, tj. jezik, s katerim deluje računalnik. Pojma sta povezana, vendar na različnih ravneh. Računalniški algoritem je mogoče pretvoriti v različne programske kode. Te kode se lahko razlikujejo ne le po različnih sintaksah (formalnih jezikih, ki se danes uporabljajo), temveč tudi po načinu, kako programer logični postopek za optimizacijo izvede jezikovno (programski slog). Algoritem je končna zaporedna vrsta osnovnih pravil ali navodil, urejenih po natančnem vzorcu, ki omogoča njihovo izvajanje korak za korakom za reševanje določene vrste problema in ustvarjanje niza končnih podatkov na podlagi niza začetnih podatkov. Algoritemska navodila so lahko ali linearna ali ciklična. V sodobnih programskih jezikih lahko algoritem s pomočjo tako imenovanih knjižnic prikliče druge algoritme zunaj kode in tako uporablja algoritme, ki jih izvajajo drugi. Sodobna koda je še posebej zapletena – ne le zaradi kompleksnosti problemov, temveč tudi zato, ker noben programer programa dejansko ne napiše popolnoma sam: ustvari ga z vključitvijo in izvajanjem algoritmov, napisanih in razdeljenih v tako imenovanih 'knjižnicah' (60–71).

Danes razvoj ni usmerjen toliko v strojno učenje kot v globoko učenje. To temelji na različnih ravneh predstavitve, ki ustrezajo hierarhijam značilnosti dejavnikov ali pojmov, kjer so pojmi višje ravni opredeljeni na podlagi pojmov nižje ravni. Z drugimi besedami, stroj je tisti, ki sam najde vzorce in podatke poveže tako, da se med njimi pojavi pomen – nepovezane podatke zavrže ali pa določi kategorije podatkov, za katere se je zdelo, da za odgovor na vprašanje niso pomembni. To lahko povzamemo v treh sposobnostih stroja. Prva sposob-

nost, ki jo avtor opredeli kot videnje ali opazovanje, pomeni, da je stroj v podatkih sposoben zaznati elemente pravilnosti. Druga sposobnost vključuje zmožnost predvidevanja učinkov namernih sprememb v okolju. Od tod izhaja sposobnost predvidevanja, katera od več možnih sprememb bi lahko privedla do želenega rezultata. Da bi dosegli to zavedanje, je potrebna tretja raven: domišljija. Cilj tako imenovane močne umetne inteligence je ustvariti stroje s človeku podobno inteligenco – stroje, ki so sposobni vzročnega sklepanja in pogovora z ljudmi (72–79).

Prav korelacija je danes ključ, s pomočjo katerega je mogoče iz podatkov pridobiti informacije. Korelacija v statistiki pomeni razmerje med dvema naključnima spremenljivkama, tako da vsaka vrednost prve spremenljivke z določeno pravilnostjo ustreza vrednosti druge. Ni nujno, da gre za dejansko vzročno-posledično povezavo, ampak preprosto za tendenco spreminjanja ene spremenljivke v odvisnosti od druge. Veliko podatkovje razkriva, da je ideal ugotavljanja vzročnih mehanizmov iluzija: priča smo nastanku nove epistemološke paradigme, ki prehaja od vzročnosti h korelaciji. Vendar pa obstaja težava v načinu zbiranja in označevanja podatkov (80–83).

V nadaljevanju avtor v povezavi z zbiranjem podatkov izpostavlja nekaj primerov vprašanj: Zakaj je bil nabor podatkov ustvarjen? Kdo je oblikovanje nabora podatkov financiral? Kakšna predhodna obdelava/čiščenje je bilo opravljeno? Če se sklicujete na ljudi – ali so jim povedali, za kaj bo nabor podatkov uporabljen, in ali so se s tem strinjali? Če da, na kakšen način? Ali so jim bili na voljo mehanizmi za preklic

privolitve v prihodnosti ali za določene namene? Ali se bo zbirka podatkov posodabljala? Če da, kako pogosto in kdo jo bo posodabljal?

Sklicuje se na Isaaca Asimova, ki je pred več kot petdesetimi leti predvidel potrebo po etičnih pravilih, ki bi usmerjala vedênje robotov. Ko obravnavamo zmožnost delovanja strojev, zato najprej pomislimo na tri zakone robotike Isaaca Asimova: 1. robot človeka ne sme poškodovati ali z neukrepanjem dopustiti, da bi bil človek poškodovan; 2. robot mora ubogati ukaze ljudi, razen kadar so ti ukazi v nasprotju s prvim zakonom; 3. robot mora varovati svoj obstoj, če to varovanje ni v nasprotju s prvim ali drugim zakonom (84–85).

Benanti je prepričan, da bi se brez izrecnega etičnega upravljanja UI, algoritmov in njihovega povezovanja z ljudmi lahko znašli pred ustvarjanjem tretjega (umetnega) kognitivnega sistema. Človeški razvoj je treba razumeti kot cilj in ne kot sredstvo. Govoriti o razvoju torej pomeni, da v središče pozornosti ne smemo postavljati tehničnih zmogljivosti, temveč mora tam ostati človek. Etična uporaba tehnologije danes pomeni, da si inovacije prizadevamo spremeniti v razvoj; pomeni usmerjanje tehnologije v razvoj in za razvoj – ne pa zgolj iskanje napredka kot cilja samega po sebi (86–113).

Avtor svoje razmišljanje zaključuje s trditvijo, da so potrebni različni pristopi – vključno s humanističnim pristopom in tudi prispevkom različnih religij. Razvoj, ki je za soočanje z izzivi spreminjajoče se dobe sicer potreben, bi moral biti globalen, pluralen, ploden in prijazen. Z izrazom globalen nakazuje, da bi moral biti dostopen vsem ljudem brez izjem. S pluralnostjo želi poudariti pozornost na družbeni kontekst, v katerem živimo, in spoštljiv odnos do različnih kultur. Usmerjenost v prihodnost pa izraža z besedo ploden – tj. sposoben postaviti temelje za prihodnje generacije. Naposled izpostavlja tudi prijaznost oz. spoštljivi odnos do zemlje, ki nas gosti, do virov in do vseh živih bitij (122–151).

Ob koncu lahko poudarimo, da je knjiga odličen komentar na aktualno dogajanje, ki poteka v presečišču tehnologije in etike. Avtor zagovarja model, pri katerem človek ostaja sestavni del procesov odločanja, ki vključujejo UI. Edino na tak način lahko naš moralni kompas v vse bolj avtomatiziranem svetu ohranimo. Knjiga je poziv oblikovalcem politik, tehnologom in širši družbi, naj razmislijo, kako lahko UI odgovorno izkoristimo – obenem pa zaščitimo človeške vrednote.

David Kraner