

Izvirni znanstveni članek/Article (1.01)

*Bogoslovni vestnik/Theological Quarterly* 83 (2023) 4, 813—823

Besedilo prejeto/Received:10/2023; sprejeto/Accepted:12/2023

UDK/UDC: 165.2:004.89

DOI: 10.34291/BV2023/04/Zalec

© 2023 Žalec, CC BY 4.0

*Bojan Žalec*

## **Ali je umetna inteligenca inteligenca v pravem pomenu besede?**

### **Vprašanje psihičnih značilnosti in splošnosti *Is Artificial Intelligence an Intelligence in the True Sense of the Word?***

#### *The Issue of Mental Characteristics and Generality*

*Povzetek:* Središčno vprašanje prispevka je formulirano že v naslovu. Avtor zagovarja minimalno definicijo inteligence: inteligenca je samostojno reševanje problemov. Stopnja inteligence je odvisna od kompleksnosti problemov in stopnje samostojnosti njihovega reševanja (Klaus Mainzer). V skladu s to definicijo avtor ugotavlja, da že obstajajo inteligentni stroji. To so sistemi UI,<sup>1</sup> ki so zmožni globokega učenja. Avtor v članku predstavlja kritike trditve, da je UI inteligentna v dobesednem pomenu – in ne zgolj metaforičnem. Te kritike temeljijo na ugovorih, ki jih lahko razdelimo v dve osnovni skupini. Zagovorniki prve linije kritike kot bistvene attribute inteligence navajajo psihične zmožnosti (kot sta fenomenalna zavest in zavedanje) in živost, pri čemer pojem umetnega življenja zavračajo kot nesmisel. Neživi in nepsihični sistemi so zgolj prilagodljivi, ne pa inteligentni sistemi. Predstavniki druge skupine trdijo, da so nepogrešljive značilnosti inteligence splošnost, intuicija, zdravorazumskost in zmožnost abdukcije. Avtor omenjene kritike analizira in ugotavlja, da njihovi zagovorniki dobrih razlogov proti minimalni definiciji niso podali.

*Ključne besede:* definicija inteligence, samostojno reševanje problemov, pripisovanje inteligence umetni inteligenci, psihični atributi, splošnost

*Abstract:* The central question of the article is already formulated in the title. The author advocates a minimal definition of intelligence: intelligence is independent problem solving. The level of intelligence depends on the complexity of the problems and the degree of independence in solving them (Klaus Mainzer). According to this definition, the author notes that intelligent machines already exist. These are AI systems that are capable of deep learning. In the paper, the author presents criticisms of the claim that AI is intelligent in the literal sense

<sup>1</sup> Od tu naprej bomo za izraz umetna inteligenca uporabljali kratico UI.

and not just in the metaphorical sense. These criticisms are based on objections that can be divided into two basic groups. Proponents of the first line of criticism cite mental abilities (such as phenomenal consciousness and awareness) and aliveness as essential attributes of intelligence, rejecting the notion of artificial life as nonsense. Inanimate and non-mental systems are merely adaptive, not intelligent systems. Representatives of the second group claim that the indispensable characteristics of intelligence are generality, intuition, common sense and abduction. The author analyses the aforementioned criticisms and concludes that their proponents have not given good reasons against the minimal definition.

*Keywords:* definition of intelligence, independent problem solving, attribution of intelligence to artificial intelligence, generality, mental attributes

## 1. Pojem inteligence kot take: zagovor minimalne definicije

Menimo, da oblikovanje človeku podobne UI v okvirih trenutnega znanja, možnosti in raziskovalne paradigme UI ni verjetno (Larson 2021; Chollet 2017).<sup>2</sup> Kljub temu, da UI ni verjetna in je omejena, pa se zastavlja vprašanje, ali je v primeru trenutno obstoječe UI upravičeno govoriti o inteligenci v pravem oz. dobesednem pomenu besede ali pa ne (in gre v tem primeru samo za metaforično rabo)? Naš končni sklep v tem prispevku je, da v določenem smislu o inteligentnosti obstoječe UI vendarle lahko govorimo upravičeno – in da gre pri tem za več kot zgolj metaforo.

### 1.1 Človeška inteligenca in UI: samostojnost in odvisnost

UI je kot inteligenca odvisna od človeške inteligence in brez nje ni mogoča. Tu nimamo v mislih zgolj dejstva, da je UI ustvaril človek, ampak močnejšo trditev: človeški um je za obstoj UI kot inteligence trajen pogoj (Müller 2021, 364) – UI brez človeškega uma ne more obstajati. UI ima lastnosti – npr. samostojno reševanje problemov –, zaradi katerih lahko o njej na določen način govorimo kot inteligenci v pravem smislu. Toda tudi če sodobni UI določeno mero samostojnega reševanja problemov lahko pripišemo, pri tem ne smemo pozabiti, da je zgolj orodje človeka in da je delovanje takšne ‚intelligence‘ trajno odvisno od človeškega uma. UI brez človeka namreč sploh nima problemov (Fuchs 2021, 41–42): nima lastnih problemov, nima problemov, ki bi jih reševala – in jih brez človeka, ki jo uporablja, ni zmožna imeti. UI ‚ima‘ probleme, ki jih rešuje, samo posredno, kot orodje človeka.

<sup>2</sup> Prispevek je nastal v okviru raziskovalnega programa P6-0269 „Religija, etika, edukacija in izzivi sodobne družbe“ ter temeljnega raziskovalnega projekta J6-4626 „Teologija, digitalna kultura in izzivi na človeka osrediščene umetne inteligence“, ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS). Poleg tega je rezultat raziskovanja v sklopu mednarodnega projekta „Teološki, filozofski in etični izzivi umetne inteligence in novih tehnologij (2023–2024)“, ki ga TEOF UL izvaja v okviru razpisa „Nova obzorja za znanost in religijo v Srednji in Vzhodni Evropi (Univerza v Oxfordu; John Templeton Foundation)“.

Tako brez človeka UI potrebnega pogoja za inteligentnost ne izpolnjuje: je namreč zgolj inteligentno orodje. Orodje lahko deluje v skladu s svojo pravo funkcijo le takrat, ko ga kdo uporablja. To velja tudi za UI: svojo funkcijo inteligentnega reševanja problemov lahko opravlja samo, če jo kdo uporablja – ta kdo pa je človek. Bistvena razlika med človeško in umetno inteligenco ni inteligentnost, ampak to, da človeška inteligenca ni golo orodje, temveč sestavni del človeškega uma, ki ni orodje. UI pa ni sestavni del človeškega uma, ampak zgolj njegovo orodje, kar pomeni, da je tudi orodje človeške inteligence. To pomeni, da je UI kot inteligenca eksistenčno odvisna od človeške inteligence, obratno pa ne velja: človeška inteligenca je kot taka od UI neodvisna. To dokazuje že preprosto dejstvo, da je obstajala dolgo pred nastopom UI in – vzporedno neodvisno od nje obstaja še sedaj. Ob tem opozarjamo, da te trditve veljajo za umetno UI kot jo poznamo sedaj – tj. inteligenco, ki nima zmožnosti splošnosti, intuicije, zdravega ali praktičnega razuma, abdukcije, skratka lastnosti, ki so bistvene in ključne značilnosti človeške inteligence (Larson 2021). ‚Samostojnost‘ UI je torej drugačna kot samostojnost človeške inteligence.

## 1.2 Globoko učenje

UI je v zadnjih letih dosegla pomembne dosežke z razmahom globokega učenja (Walsh 2017, 61–65). Globoko učenje je strojno učenje, ki ga izvajajo nevrnske mreže. Nevronske mreže so lahko biološke, kakršne imamo v možganih, ali pa umetne. Sestavljene so iz nevronov. Nevroni so sestavljeni iz telesa in izrastkov oz. žic, ki se stikajo v sinapsah oz. križiščih ali vozliščih. Nevroni procesirajo podatke in med seboj komunicirajo, sprejemajo in pošiljajo signale. Globoko učenje izvajajo mreže, ki so globoke, kar pomeni, da imajo več slojev med seboj povezanih nevronov. Sisteme nevrnskih mrež je mogoče prilagoditi tako, da pravilno prepoznajo vzorce. Lahko jih tudi izurimo, da se sami prilagajajo in vzorce prepoznajo samodejno.

»O globokem učenju govorimo takrat, ko se sistem »učik iz primerov, ko vozlišča vsakokratne nevrnske mreže tako obteži, da lahko po učni fazi posploši naučeno prepoznavanje vzorcev. Tako nastavljen sistem lahko pregleda nove podatkovne zapise in poskuša v njih prepoznati naučene prototipe vzorcev.« (Müller 2021, 353)

Tako so sistemi zmožni prepoznavanja novih struktur, ki s programom ni bilo vnaprej določeno. V tem smislu govorimo o ‚samostojnosti‘ sistema. Tovrstna zmožnost učenja sistemom omogoča prožnost in plastičnost ‚ravnanja‘, zaradi česar lahko mnoge naloge rešujejo uspešneje od togih ekspertnih sistemov iz sedemdesetih in osemdesetih let prejšnjega stoletja. Globoko učenje so omogočili globoke mreže in zelo zmogljivi računalniki, ki lahko hitro izvedejo veliko število operacij in obdelajo velikansko število podatkov. Dosežki globokega učenja so vidni na področju prepoznavanja slik in jezika, ki sta bili šibki točki klasične umetne inteligence. O veliki zmogljivosti, ki so jo tako dosegli sistemi UI, pričajo številni primeri zmag nad človeškimi konkurenti. Sloviti je primer Googlevega računalniškega sistema AlphaGo, ki je premagal južnokorejskega mojstra azijske igre go, za igranje katere mora biti človek inteligenčen, ustvarjalen in zmožen strateškega

razmišljanja (Walsh 2017, 34–6; Müller 2021, 353). Zaradi omenjenih zmožnosti in dosežkov sistemom UI mnogi pripisujejo inteligentnost v smislu samostojnega reševanja problemov. Po drugi strani pa nekateri kljub vsemu tega mnenja ne delijo.

## 2. Kritika mentalističnega in vitalističnega zavračanja inteligentnosti: psihičnost in živost nista potrebna pogoja za inteligentnost

Oglejmo si primer ostre kritike stališča, da lahko UI upravičeno pripišemo inteligentnost. Podal jo je ugledni nemški raziskovalec, psihiater in filozof Thomas Fuchs, ki inteligenco povezuje predvsem z živostjo in psihičnimi značilnostmi, kot so zavedanje, uvid, razumevanje, doživljanje (Fuchs 2021, 35; 41ff). Problem Fuchsove argumentacije je, da te ključne postavke svojega argumenta nikjer ne dokaže. S Fuchsom se lahko celo strinjamo, da je človeška inteligenca taka, da temelji na uvidu, zavedanju itd. Toda to še ne pomeni, da je taka nujno vsaka inteligenca. Zakaj bi morali sprejeti tak pojem inteligence? Strinjamo se, da so živost, fenomenalna izkustva, zavedanje itd. to, kar človeka bistveno ločuje od stroja. Lahko sprejmemo celo, da so te lastnosti bistvene za človeško inteligenco – toda zakaj bi morali sprejeti, da so te stvari potrebne za inteligenco kot tako?

Da to nikakor ni razvidno samo po sebi, dokazuje razumevanje sodobnega nemškega filozofa Tobiasa Müllerja, ki razlikuje med zavestjo, mišljenjem in inteligenco (Müller 2021, 345ff). Sestavine zavesti in mišljenja, kot ta izraz uporablja Müller, opisujemo s pojmi tipa kvalitativno ali fenomenalno izkustvo (npr. takšnosti, *qualia*), poseben način (fenomenalnega) doživljanja (kako je biti ..., *what-is-likeness*),<sup>3</sup> doživljanje (nem. *Erleben*), subjektivnost, zavedanje, razumevanje, uvid, refleksija itd. Nič od tega ne moremo pripisati sistemom UI – s tem se strinja tudi Müller. Pa vendar iz tega ne sklepa (kot Fuchs), da UI zaradi tega ne bi mogli pripisovati inteligentnosti sploh. Müller meni – tako kot mi – da lahko tudi UI »po analogiji s človeško inteligenco« (349–350) inteligentnost v določenem smislu pripišemo, vendar pa se moramo pri tem zavedati vseh razlik med človeško inteligenco in UI. Müller inteligenco definira kot samostojno reševanje problemov,<sup>4</sup> medtem ko zavedanja in drugih psihičnih lastnosti kot potrebnih pogojev zanjo ne navaja – opredelitev inteligence kot samostojnega reševanja problemov pa prevzema od znanstvenika in filozofa Klaus Mainzerja (Mainzer 2016, 3; Müller 2021, 349).

Če sprejmemo, da je samostojno reševanje problemov zadosten pogoj za inteligenco, potem lahko sprejmemo, da stroji niso takšni kot ljudje: da UI ni zmožna stvari, ki so ključne za človeka in njegovo inteligenco, da človeka oz. njegovega

<sup>3</sup> O tem govori slavna in sedaj že ‚klasična‘ razprava Thomasa Nagla *Kako je biti netopir* (Nagel 1990).

<sup>4</sup> Pravzaprav niti ni potrebno, da inteligenco definiramo kot samostojno reševanje problemov, da bi Fuchsovo argumentacijo lahko utemeljeno zavrnil. Dovolj je, da sprejmemo predpostavko, da je samostojno reševanje problemov za pripis inteligentnosti zadosten pogoj.

uma ne moremo naložiti na računalnik itd. – toda iz tega še ne izhaja, da artefakti ne morejo biti inteligentni. Človeška inteligenca je drugačna od inteligence UI, toda to še ne pomeni, da druga sploh ni inteligenca.

Če se strinjamo, da za pripis inteligence zadostuje samostojno reševanje problemov, lahko inteligentnost UI spodbijemo po dveh poteh: da dokažemo, da UI problemov sploh ne rešuje, ali tako, da utemeljeno zanikamo, da jih rešujejo samostojno. Fuchs ne zanika zgolj, da bi UI probleme reševala samostojno, ampak, da jih sploh rešuje. Njegov argument za neinteligentnost strojev je naslednji: stroji problemov sploh ne rešujejo, ker zanje problemi niti ne obstajajo. Probleme rešuje uporabnik UI, ne pa inteligenca sama. Ta argument je spet povezan s postavko, da stroji nimajo zavedanja, razumevanja, vrednotenja, skratka določenih psihičnih lastnosti. Ker se UI ne zaveda in ne vrednoti, po Fuchsovem mnenju o problemu sploh nima predstave – za stroj ni nobenega problema, in ga zato tudi ne rešuje. Probleme z uporabo UI rešuje človek (Fuchs 2021, 42).

Glede pravkar navedenega se lahko s Fuchsom strinjamo, vendar pa moramo takoj dodati, da s tem Fuchs ni še prav nič dokazal, da UI ni inteligentna v dobesednem pomenu, saj iz tega, da x nima (lastnih) problemov, še ne izhaja niti, da x problemov ne more reševati, niti, da jih ne more reševati samostojno. Fuchsova argumentacija je neprepričljiva, saj ponovno že predpostavlja, kar bi moral šele dokazati – in sicer to, da iz umanjkanja psihičnih značilnosti, kot so zavedanje, razumevanje, uvid, tudi vrednotenje, izhaja odsotnost inteligence. Kot že omenjeno, pa zgolj ni razvidna – kar kažeta primera Müllerja in Meinzerja ter vrste drugih uglednih strokovnjakov.

Iz tega, da je x orodje, še ne izhaja, da x ne more biti inteligenčen. Prav tako iz odsotnosti lastnih motivov ne moremo sklepati, da x ni inteligenčen – niti iz orodnosti niti iz odsotnosti motivov namreč ne izhaja, da x ni sposoben samostojnega reševanja problemov. Ta samostojnost je lahko manjša ali večja – toda zanikati, da stroji določeno samostojnost pri reševanju problemov imajo, se zdi ob tem, česar je sodobna UI zmožna, neprepričljivo. Nobena od obeh zgoraj omenjenih poti zanikanja inteligentnosti UI se spričo razvoja in dosežkov sodobne inteligence ne zdi prepričljiva ali obetavna. Nekaj dokazov za to lepo strne že sam Fuchs, ko k svojemu stališču predvideva ugovore. Tako pravi, da

»imamo zdaj opravka z novo generacijo UI, in sicer s stroji, ki se »učijo«. To so umetne nevronske mreže, sposobne simulirati prilagoditvene sposobnosti možganov. Podobno kot biološke sinapse so povezave med nevroni, modeliranimi v računalniku, numerično obtežene in se vnosu signalov v procesu urjenja prilagajajo (globoko učenje). Pogosto uporabljene povezave se okrepijo, redko uporabljene se prekinejo. Tako so sistemu predstavljeni tisoči podobnih vzorcev, na primer različne različice obraza, dokler se ne odzove na razporeditev slikovnih pik, ki se bo najverjetneje ponovila, tj. obraz »prepozna«. Takšni sistemi lahko tudi razlikujejo na primer med psi in mačkami, prepoznavajo glasove na mobilnih telefonih ali prevajajo – in so prisotni že vsepovsod.« (Fuchs 2021, 33)

Neprestan razvoj avtonomne kibernetike oz. robotike bo vodil v razvoj androidov, ki jih bo od ljudi oz. človeških oseb vse težje razlikovati. Katera je fundamentalna inovacija robotike v primerjavi s klasično UI? Operativna mobilnost in kvazitelesna interakcija robotom omogočata, da razvijejo nove oblike povratne informacije in prilagoditev, ki segajo onkraj zmožnosti stacionarnih prilagajajočih se sistemov. Roboti so postali netrivialni stroji, v katerih ne potekajo zgolj preddoločeni programi, ampak lahko regulirajo tudi lasten odnos do okolja. Interaktiven prilagajajoči se stroj se lokalizira v prostoru in ima v tem smislu sam sebe za predmet. Zaznava rezultate svojega vedénja v okolici in lahko také svoj lastni program modificira. Roboti se lahko prilagajajo na podlagi telesnega modela sebe, ki so ga generirali sami. Tako se je štirinožni robot po amputaciji ene noge zmozel prilagoditi tako, da lahko hodi tudi s tremi nogami (35–36).

Prilagoditveni procesi robotov temeljijo na ‚učenju‘ in utelešenosti. Seveda ne gre za pravo utelešeno subjektivnost, ker roboti ne doživljajo zares. A stalna povratna informacija iz resničnega stika s predmeti in prostori jim omogoča, da algoritme optimizirajo tako, kot tega ni bilo mogoče doseči s predhodnim centralnim programiranjem. Ta pristop veliko bolj temelji na vedénju – veliko bolj je decentraliziran, od spodaj navzgor in, če se tako izrazimo, ‚analogen‘, kot je zapisal Fuchs –, učenju skozi delovanje. Ker postopek posnema načela evolucije, lahko na neki način govorimo tudi o evolucionarni robotiki ali umetnem ‚živiljenju‘.<sup>5</sup> Taki postopki so klasični pristop v UI, temelječ na vnaprej določenih programih (od zgoraj navzdol), potisnili v ozadje – vsaj pri funkcijah, kot so nadzor gibanja in orientacija. So pa klasični sistemi še vedno boljši za abstraktne dejavnosti (36).

Fuchsov odgovor na ugovore na podlagi zgoraj navedenega je, da sta v primeru strojev izraza ‚učenje‘ in ‚prepoznavanje‘ uporabljena samo v prenesenem pomenu besede, metaforično. Dobesedno pa stroji nič ne prepoznavajo in se ne učijo, ampak se samo prilagajajo. Zakaj ne? Zato, ker nimajo psihičnih zmožnosti, ki so za učenje in prepoznavanje potrebne. Fuchs učenje opredeljuje kot pridobivanje novih znanj oz. veščin na podlagi doživljanja (nem. *Erlebnis*) ali živega izkustva (ang. *living experience*). Sistemi UI pa nimajo nobenega doživljanja, nobenega živega izkustva (Fuchs 2021, 33). Prepoznavanje temelji na uvidu, zavedanju, razumevanju – sistemom UI pa ne moremo pripisati nič od tega. Človeško učenje temelji na uvidu ali ‚intuiciji‘, ki v različnih kontekstih prepozna isti lik (nem. *Gestalt*) – globoko učenje strojev pa na statistiki, ki je v osnovi induktivno sklepanje na osnovi številnih podatkov; na izredno hitri obdelavi ogromne količine podatkov (Larson 2021, 127; 133ff). Zato se računalnik nauči ‚prepoznati‘ obraz po več milijonih ‚ogledov‘, otroku ali človeku pa so za to lahko dovolj že ogledi, ki jih lahko preštejemo na prste ene roke (Fuchs 2017).

Ne glede na bogastvo in zanimivost Fuchsovih besedil pa moramo vendarle ugotoviti, da se tudi v zavračanju pomislekov ali ugovorov na podlagi zgoraj navedenega Fuchs zateka k argumentiranju, ki smo ga že omenili kot njegov stalni oz.

<sup>5</sup> O umetnem živiljenju lahko po Fuchsu govorimo samo v prenesenem smislu, dobesedno pa je pojem umetnega živiljenja po njegovem mnenju nesmisel, saj je protisloven (Fuchs 2021, 38).

ponavljajoči se način zavračanja inteligentnosti UI: preprosto zatrjuje, da je za specifično človeške stvari – in torej tudi za inteligenco – potrebno zavedanje, um, ki interpretira, razume itd. Tega pa računalnik nima. Temelj njegove interpretacije je torej, da računalnik nima potreb in nobenih psihičnih značilnosti, kot so intencionalnost, prepričanje, želje, čustva, občutki ugodja ali neugodja, vrednotenje, nameni, motivi, odločanje itd. O vseh teh stvareh lahko pri računalniku govorimo samo metaforično, ne pa dobesedno – nič od tega računalnik nima. Zakaj ne? Zato, ker nima človeškega telesa in ker ni živ. Ni zavesti brez življenja in ni psihičnih lastnosti brez živosti. Če pa entiteta nima psihičnih značilnosti, ki jih ima človek, potem tudi ne more izvajati opravil, ki jih izvaja človek. Ne more se učiti, ne more imeti zdravega razuma, intuicije itd., kar pa pomeni, da ne more biti inteligentna v pravem pomenu besede. Navedene trditve v Fuchsovi argumentaciji delujejo kot temeljne premise, ki jih ne dokazuje, ampak jih predpostavlja kot resnične. S tem pa predpostavlja – kot smo že omenili –, kar bi moral šele dokazati.

Bolj zmerno, trezno, stvarno in znanstveno previdno (in zato bolj sprejemljivo od Fuchsovega) je stališče že omenjenega Tobiasa Müllerja. Kot rečeno, Müller meni, da je sistem inteligenten, če lahko probleme rešuje učinkovito in samostojno. Stopnja inteligence pa je odvisna od stopnje samostojnosti, stopnje kompleksnosti problema in stopnje učinkovitosti postopka reševanja problema (Mainzer 2016, 3; Müller 2021, 349). Kot opozarja Müller, že uporaba izrazov ‚samostojno‘ in ‚problem‘ inteligenco postavlja v smotrnostno-racionalni sklop, kajti v procesu reševanja problema gre za doseg cilja in premagovanje ovir. To pomeni, da mora biti določen cilj dosežen s primernimi in kar se da učinkovitimi sredstvi ter metodami. S tem je reševanje problemov izvorno postavljeno v zavestni življenjski proces nekega ciljno orientiranega bitja, ki samo od sebe določeno rešitev išče takó, da si predstavlja in predvideva prihodnje situacije (Fuchs 2020, 45ff). Čeprav je pri tem v mnogih primerih rešitev mogoče najti tudi s pomočjo računalnika, je računalnik le orodje, s katerim lahko pri izvornem problemu računamo določene vidike, ki se dajo kvantificirati (Müller 2021, 349). Vendar pa ta povratna vezava na človeški kontekst problema po mnenju Müllerja ne izključuje možnosti, da lahko inteligenco analogno pripišemo tudi računalniku – kolikor se to nanaša na njegove zmožnosti, da učinkovito izvede določene izračune, ki zadevajo kakšen delni vidik človeškega problema (349–350). Pri tem analognem prenosu pa moramo upoštevati, da se pomen besede ‚samostojno‘ v kontekstu UI spremeni. ‚Samostojno‘ tu pomeni iskanje novih vzorcev v podatkih, ki so jih vnaprej določili inženirji in programerji, medtem ko gre pri živem bitju dejansko za intrinzično dejavnost, ki je vpeta, vključena, integrirana v njegov življenjski proces (350).

Fuchsova napaka je, da pojem inteligence – ki pa kot tak, po svoji vsebini ali bistvu, ni niti vitalističen niti psihičen, ampak funkcionalen – vitalizira in psihologizira.<sup>6</sup> Iz dejstva, da je bila edina inteligenca, poznana do pojava računalnikov, inteligenca živih in psihičnih bitij, Fuchs neutemeljeno posplošuje, da je to edina

<sup>6</sup> Kot se izrazi Müller, pojem inteligenca v kontekstu reševanja problemov označuje praktično zmožnost (2021, 349).

možna inteligenca. Poleg tega inteligenco kot tako brez utemeljitve počloveči. Iz dejstva, da človeško inteligenco spremljajo zavedanje, uvid, razumevanje itd., sklepa, da to velja za inteligenco kot tako. Vendar za ta sklep ne navaja argumentov, sklicevanje na samorazvidnost pa, kot že rečeno, nima prave teže in ni prepričljivo.

### 3. Splošnost

Druga linija argumentiranja proti minimalni definiciji temelji na trditvi, da je bistvena značilnost prave inteligence oz. inteligence kot take splošnost. To argumentacijo lahko zasledimo v knjigi *Mit o umetni inteligenci* Erika J. Larsona (2021). Larson se tu sicer osredotoča na dokazovanje, zakaj ni verjetno, da bi v okviru sedanje raziskovalne paradigme pri raziskovanju UI prišlo do oblikovanja človeku podobne UI in eksplozije inteligence (Irwing John Good (Good 1965; Chollet 2017)<sup>7</sup>), vendar pa v knjigi najdemo izjave, ki kažejo na to, da meni, da je splošnost bistvena značilnost ne le človeške ali človeški podobne inteligence, ampak tudi inteligence kot take. Da bi to trditev bolje razumeli, si na kratko oglejmo, kaj je s splošnostjo inteligence mišljeno.

Da bi dosegel zadane cilje, se mora vsak sistem strojnega učenja naučiti nekaj specifičnega. Raziskovalci temu pravijo, da stroj usmerimo pristransko. Pristranskost v strojnem učenju torej pomeni, da je sistem oblikovan in nastavljen tako, da se nauči nekaj določenega (Larson 2021, 28). To pa seveda pomeni ustvarjanje aplikacij za reševanje ozkih problemov (28–29). Zato se denimo sistemi globokega učenja, ki jih npr. Facebook uporablja za prepoznavanje človeških obrazov, niso hkrati naučili, kako izračunati davke (29). Stanje je dejansko še slabše: raziskovalci so ugotovili, da če sistem strojnega učenja usmerijo pristransko oz. specializirajo, da se nauči posebne aplikacije ali naloge, to pomeni, da bo slabše deloval pri drugih nalogah. Obstaja obratno sorazmerje med uspehom sistema, da se nauči neke določene stvari, in zmožnostjo, da se nauči neke druge naloge (29) – to velja celo za zelo podobne naloge. Računalniški sistem, ki se nauči vrhunškega igranja šaha, se ne bo naučil vrhunškega igranja goja – in obratno. Sistem za go je bil oblikovan specifično, tj. s posebno pristranskostjo za učenje (pravil) goja. Težava je,

<sup>7</sup> Cholletov članek o neverjetnosti t. i. eksplozije inteligence je sprožil veliko zanimanje. Do sedaj ima že več kot osemnajst milijonov ogledov. Izraz ‚eksplozija inteligence‘ je v 60. letih 20. st. uporabil britanski matematik – Turingov sodelavec – Irwing John Good za trditev o nastopu superinteligence, ki se bo razvila do nepredstavljive stopnje. Ta razvoj bo potekal s silovitim naraščanjem. Chollet v članku dokazuje, da taka eksplozija ni verjetna. Njegovo dokazovanje je relevantno za današnji čas, ko se ponovno pojavljajo trditve o eksploziji inteligence, formulirane kot napovedi nastopa singularnosti (Kurzweil 2005), superinteligence (Bostrom 2020) in podobno. »Leta 1965 je I. J. Good prvič opisal pojem ‚eksplozije inteligence‘, ki se nanaša na umetno inteligenco (UI), z naslednjimi besedami: »Naj ultrainteligentni stroj definiramo kot stroj, ki lahko daleč preseže vse intelektualne dejavnosti katerega koli človeka – še tako pametnega. Ker je načrtovanje strojev ena od teh intelektualnih dejavnosti, bi ultrainteligentni stroj lahko oblikoval še boljše stroje; potem bi nedvomno prišlo do ‚eksplozije inteligence‘ in človeška inteligenca bi ostala daleč zadaj. Tako je prvi ultrainteligentni stroj zadnji izum, ki ga mora človek kdaj narediti – pod pogojem, da je stroj dovolj ubogljiv, da nam pove, kako ga obdržati pod nadzorom.« (Chollet 2017)



da se pristranskosti ne moremo znebiti, ker je sestavni del strojnega učenja. Vsak UI sistem učenja, ki ga ne usmerimo pristransko, se pri uporabi za poljubne probleme ne obnese nič bolje od naključnega uspeha. Resnično nepristranski sistem (sistem, ki ga raziskovalci ne usmerijo pristransko) je neuporaben. Strojni sistem se torej nauči tistega, kar želijo njegovi oblikovalci. Pristransko usmerjanje dejansko pomeni, da sistem naredimo ozkega v smislu, da njegove uporabe ni mogoče posplošiti na ostala področja. Ozkost je od uspešnosti sistema neločljiva – ozkost in uspešnost sta pri strojnih sistemih plati istega kovanca (30).

Larson v knjigi nazorno pokaže, da je za splošnost inteligence potrebna abdukcija. Abdukcija je posebna vrsta sklepanja, ki ni ne dedukcija ne indukcija. Gre za neke vrste ugibanje, ki pa ni golo ugibanje – vključuje intuicijo, analizo in zdravi razum (157–190). Za oblikovanje splošne UI bi potrebovali takšno teorijo abdukcije, ki bi bila dovolj uporabna za inženirje in programerje, da bi lahko oblikovali splošno UI. Take teorije pa nimamo in po mnenju mnogih vrhunskih inženirjev in programerjev je tudi še zdaleč ni na vidiku. Za oblikovanje splošne UI nam torej manjka temeljna teorija (190).

S takim razumevanjem pomena splošnosti se vsekakor lahko strinjamo. Splošnost je bistvena značilnost človeške inteligence – in oblikovanja splošne UI v sklopu sedanje paradigme ni na vidiku.

Zato ni razumno, da bi pričakovali nastop človeku podobne inteligence, kaj šele superinteligence, ki bi človeško v vseh pogledih neizmerno prekašala. S čim Larson utemeljuje trditev, da je splošnost bistvena značilnost inteligence kot take? V celotni knjigi z ničemer! To pa pomeni, da značilnost, ki je bistven atribut človeške inteligence, brez dobrega razloga kot bistveno lastnost razširja na inteligenco kot tako. Larson zagreši podobno napako kot Fuchs, čeprav se ne osredotoča na iste vidike človeške inteligence kot Fuchs: brez tehtnega razloga počloveči pojem inteligence kot take.

#### 4. Zaključek

Dokazovanje, da UI problemov ne rešuje – ali da jih do določene mere ne rešuje samostojno –, ni prepričljivo, pa tudi ne trditev, da problemi, ki UI jih rešuje, niso kompleksni. Torej ostane le še možnost, da UI kljub samostojnemu reševanju problemov inteligentnost odrekamo. To že v luči zgoraj navedenih razlogov ni razumna opcija. Zato ugotavljamo, da je najbolj razumna pojmovna shema, ki sprejema minimalno definicijo inteligence. To definicijo vsekakor lahko obdržimo, kajti njeni kritiki nič zares ne dokazujejo – ker že kar predpostavljajo, kar naj bi šele dokazali. Minimalna definicija je v skladu z razširjeno uporabo izraza inteligenca in s prepričanjem mnogih strokovnjakov za umetno inteligenco. Vse navedeno kaže na to, da je breme dokaza, zakaj je ta definicija neprimerna, na njenih kritikih. Kritiki pri njenem zavračanju zagrešijo *petitio principii*, pa tudi njihovi drugi razlogi so za zavrnitev minimalne definicije neprepričljivi, čeprav upravičeno opozarjajo

na velike nevarnosti, ki jih napačno razumevanje inteligence prinaša. Vendar pa te nevarnosti ne izhajajo iz sprejemanja minimalne definicije, ampak iz napačnih trditev o verjetnosti, da se lahko oblikuje oz. nastopi človekovi podobna inteligenca ali celo superinteligenca, ki bi človeško inteligenco močno prekašala – in to ne samo pri ozko določenih nalogah.

Sprejemanje minimalne definicije prispeva k jasnejši pojmovni shemi s posebnim pojmom za zmožnost samostojnega reševanja problemov – ne glede na vse ostale lastnosti. To je pomembna zmožnost, ki označuje kvalitativno razliko med entitetami, ki jo posedujejo, in tistimi, ki je ne. Ta kvalitativna razlika je vsekakor drugačna od tiste, ki jo označujejo pojmi, kot so psihično, fenomenalna zavest, zavedanje, živo, splošnost, intuicija itd. Če inteligentnost pojmuje kot tako, da že implicira omenjene lastnosti, te kvalitativne razlike ne zamejimo jasno. Če pa se strinjamo, da je za samostojno reševanje problemov koristno in celo potrebno uporabljati poseben pojem – kateri pojem je za to oznako boljši kandidat kot inteligentnost? Zato ugotavljamo, da je pojmovna shema, ki pojma inteligentnosti v pomenu minimalne definicije ne vsebuje, v resnici nejasna in nenaravna. Take pojmovne sheme so dejansko lahko škodljive, omogočajo zlorabe in podcenjujejo intuitivnost človeškega mišljenja, ki se odraža tudi v tem, da določeno uporabo pojma občuti kot naravno. Zato je opredelitev inteligence pomembno vprašanje – ne gre le za prazno in brezplodno pričkanje o uporabi besede, za t. i. golo besedno vojno. Brezkompromisno zavračanje naravne in jasne pojmovne sheme prav lahko kaže na prisotnost predsodkov. Nevarnostim, izzivom in problemom, ki jih razvoj umetne inteligence prinaša, pa ne bomo kos tako, da bomo gojili predsodke in se odpovedovali stvarni, naravni in razumni uporabi pojmov,<sup>8</sup> kar bi pomenilo svojevrstno goljufanje – ampak tako, da bo naša kritična drža filozofije in javne teologije temeljila na sprejemanju stvarnosti, ki med drugim prinaša izziv umetne inteligence v dobesednem pomenu besede. Kritična ost mora biti usmerjena v izrabo širjenja trditev o (inteligentni) umetni inteligenci za ideologije in ravnanje, ki ogrožajo človeka, človečnost in spoštovanje njegovega dostojanstva.<sup>9</sup> Težava torej ni v sprejemanju umetne inteligence kot inteligence, prav nasprotno, temveč v interpretaciji implikacij tega sprejemanja – in njihovi zlorabi.

<sup>8</sup> Ta je vedno že intuitivna. Intuicijo pa seveda lahko poskusimo zajeti v eksplicitni definiciji pojma.

<sup>9</sup> Primerki takšne ideologije so »teorije« nekaterih transhumanistov. Za kritično razpravo o njih prim. Pouliquen 2022; Petkovšek in Žalec, ur. 2021; Klun 2019; Strahovnik 2019; Globokar 2019; Osredkar 2019; Platovnjak in Svetelj 2019; Stegu 2019; Vodičar 2019; Pohar 2019; Štivić 2021; Sturm 2021; Dittrich 2021.

## Reference

- Bostrom, Nick.** 2020. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Chollet, François.** 2017. The implausibility of intelligence explosion. *Medium*. 27. 11. <https://medium.com/@francois.chollet/the-possibility-of-intelligence-explosion-5be4a9eda6ec> (pridobljeno 22. 8. 2023).
- Dittrich, Tristan Samuel.** 2021. Transhumanistisches Glückstreben und christliche Heilshoffnung: Ein Vergleich. *Zeitschrift für Theologie und Philosophie* 143, št. 3:452–474.
- Fuchs, Thomas.** 2017. Verkörperertes Wissen – verkörperertes Gedächtnis. V: Gregor Eitzelmüller, Thomas Fuchs in Christian Tewes, ur. *Verkörperung – Eine neue interdisziplinäre Anthropologie*, 57-77. Berlin: Walter de Gruyter.
- . 2020. *Die Verteidigung des Menschen: Grundfragen einer verkörpererten Anthropologie*. Berlin: Suhrkamp.
- . 2021. *In Defense of Human Being: Foundational Questions of an Embodied Anthropologie*. Oxford: Oxford University Press.
- Globokar, Roman.** 2019. Normativnost človeške narave v času biotehnoškega izpopolnjevanja človeka. *Bogoslovni vestnik* 79, št. 3:611–628. <https://doi.org/10.34291/bv2019/03/globokar>
- Good, Irving John.** 1965. Speculations Concerning the First Ultra-intelligent Machine. *Advances in Computers* 6:31–88. [https://doi.org/10.1016/s0065-2458\(08\)60418-0](https://doi.org/10.1016/s0065-2458(08)60418-0)
- Klun, Branko.** 2019. Transhumanizem in transcendenca človeka. *Bogoslovni vestnik* 79, št. 3:589–600. <https://doi.org/10.34291/bv2019/03/klun>
- Kurzweil, Ray.** 2005. *The Singularity is Near: When Humans Transcend Biology*. London: Duckworth Overlook.
- Larson, Erik J.** 2021. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Mainzer, Klaus.** 2016. *Künstliche Intelligenz – Wann übernehmen die Maschinen?*. Berlin: Springer.
- Müller, Tobias.** 2021. Künstliche Intelligenz und menschliches Selbstverständnis: Zu anthropologischen Herausforderungen der Digitalisierung. *Zeitschrift für Theologie und Philosophie* 143, št. 3:341–364.
- Nagel, Thomas.** 1990. Kako je biti netopir. V: Douglas R. Hofstadter in Daniel C. Dennett, ur. *Oko duha*, 399–410. Ljubljana: Mladinska knjiga.
- Osredkar, Mari Jože.** 2019. Religija kot izziv za transhumanizem. *Bogoslovni vestnik* 79, št. 3:657–668. <https://doi.org/10.34291/bv2019/03/osredkar>
- Petkovšek, Robert, in Bojan Žalec, ur.** 2021. *Transhumanism as a Challenge for Ethics and Religion*. Zürich: Lit.
- Platovnjak, Ivan, in Tone Svetelj.** 2019. To Live a Life in Christ's Way: the Answer to a Truncated View of Transhumanism on Human Life. *Bogoslovni vestnik* 79, št. 3:669–682. <https://doi.org/10.34291/bv2019/03/platovnjak>
- Pohar, Borut.** 2019. Transhumanizem v službi človekove odgovornosti do stvarstva. *Bogoslovni vestnik* 79, št. 3:643–656. <https://doi.org/10.34291/bv2019/03/pohar>
- Pouliquen, Tanguy Marie.** 2022. *Čar novih tehnologij in transhumanizem: 115 vprašanj*. Ljubljana: Družina.
- Stegu, Tadej.** 2019. Transhumanizem in krščanska antropologija. *Bogoslovni vestnik* 79, št. 3:683–692. <https://doi.org/10.34291/bv2019/03/stegu>
- Strahovnik, Vojko.** 2019. Vrline in transhumanistična nadgradnja človeka. *Bogoslovni vestnik* 79, št. 3:601–610. <https://doi.org/10.34291/bv2019/03/strahovnik>
- Sturm, Wilfried.** 2021. Transhumanismus und Digitalisierung: Theologisch-anthropologische Perspektiven. *Zeitschrift für Theologie und Philosophie* 143, št. 3:425–451.
- Štivič, Stjepan.** 2021. Upanje v krščanstvu in transhumanizem. *Bogoslovni vestnik* 81, št. 4:849–856. <https://doi.org/10.34291/bv2021/04/stivic>
- Vodičar, Janez.** 2019. Transhumanizem in katoliška vzgoja. *Bogoslovni vestnik* 79, št. 3:693–704. <https://doi.org/10.34291/bv2019/03/vodicar>
- Walsh, Toby.** 2017. *Android Dreams: The Past, Present, and Future of Artificial Intelligence*. London: C Hurst & Co Publishers Ltd.