ANαliZA

časopis za kritično misel

Letnik 10, številka 1-2, maj 2006 · ISSN 1408-2969

RAZUMEVANJE IN UMETNA INTELIGENCA

Tomaž Strgar

Stroji, mišljenje in inteligenca

Filozofska fakulteta Univerze v Ljubljani

Citiranje: Tomaž Strgar, »Stroji, mišljenje in inteligenca«, *Analiza: časopis za kritično misel* 10, št. 1-2 (2006): 91–100.

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/deed.sl.

Izdajatelj: Društvo za analitično filozofijo in filozofijo znanosti, Ljubljana.

Elektronska različica je bila leta 2026 rekonstruirana iz izvirnih stavnih datotek in preverjena s tiskano izdajo. Številke strani ustrezajo tiskani izdaji; prelomi strani znotraj prispevka se lahko nekoliko razlikujejo.

Tomaž Strgar
Filozofska fakulteta Univerze v Ljubljani

STROJI, MIŠLJENJE IN INTELIGENCA

V klasičnem tekstu s področja umetne inteligence (UI) “Stroji, ki računajo, in inteligenca”, britanskega matematika in enega pionirjev računalniških znanosti Alana Turinga, objavljenem l. 1950 v reviji *Mind*[1] avtor predloži v obravnavo vprašanje: “Ali so stroji sposobni mišljenja?” Ker je bila v tistem času običajna raba besed “stroj” in “mišljenje” tako rekoč izključujoča, osnovno vprašanje nadomesti s starim Descartesovim vprašanjem o načelni možnosti smiselne besedne komunikacije stroja s človekom. Podobno tudi pojem *inteligentnosti* Turing izrecno povezuje s stroji le v naslovu svojega sestavka. Kljub temu so tovrstna teoretska vprašanja postalo in ostala osrednja tema nove, takrat nastajajoče smeri znanstvenega raziskovanja.

V tej kratki razpravi je moj osnovni namen pokazati, da je sposobnost mišljenja sicer potreben, ne pa tudi zadosten pogoj (umetne) inteligence. Ker sta oba pojma pogosto uporabljana kot enakovredna, še zlasti v teoretskih razpravah o možnostih in omejitvah UI, uvodoma predstavim zgodovinsko ter jezikovno ozadje te problematike. V nadaljevanju svoje poglede najprej osvetlim s pomočjo psiholoških teorij S. Freuda in C. G. Junga, zatem pa kot središčni pojem svoje argumentacije obravnavam pojem razumevanja. V zadnjem poglavju lastno razumevanje razumevanja postavim nasproti stališčem enega bolj znanih nasprotnikov Turingovih stališč, ameriškega filozofa J. Searla, ki možnost razumevanja strojem v celoti odreka.

Ključne besede: umetna inteligenca, Turing, mišljenje, razumevanje, Searle.

1. Babbage, Turing in umetna inteligenca

[1] Del tukaj obravnavanega Turingovega članka “Computing machinery and Intelligence” je v zborniku “Oko duha” (Hofstadter, D. R., Dennett, D., 1990) preveden tudi v slovenščino. Sam se v pričujočem tekstu naslanjam neposredno na izvirnik, tako da tudi citati iz že prevedenega dela omenjenemu prevodu ne sledijo vedno dobesedno. Odstopanja so zgolj subjektivno-estetske narave in tako obstoječemu prevodu v ničemer ne odrekajo pravilnosti in ustreznosti.

Ideja mehaničnega pripomočka za računanje je stara tisočletja, njena najstarejša dobro znana materializacija pa je *abakus*, ki se v nekaterih delih sveta uporablja še dandanes. Zgodovina računanja s pomočjo strojev je precej krajša in sega v Evropo 17. stoletja, ko je Blaise Pascal izdelal prvi stroj za seštevanje, nekaj let kasneje pa je G. W. Leibniz predstavil stroj, ki je poleg seštevanja obvladoval tudi operacije odštevanja, množenja, deljenja in korenjenja.

Izumitelj koncepta modernega digitalnega računalnika pa je bil pravzaprav angleški matematik in izumitelj Charles Babbage. Najpomembnejša lastnost njegovega "analitičnega stroja", ki ga je zasnoval v prvi polovici 19. stoletja in ki je bil povsem mehanski, je njegova univerzalnost. To pomeni, da je isti stroj uporaben za najrazličnejše računske postopke, katerih zaporedje se vnaprej opiše z nizom ukazov – s programom. Kot način podajanja ukazov stroju je Babbage uporabil princip luknjanih kartic, ki so jih leta 1801 začeli uporabljati v tekstilni industriji kot krmilni mehanizem pri strojnem tkanju vzorčastega blaga. Luknjane kartice bi bile hkrati tudi izhod njegovega stroja, ki bi tako lahko, kot se je izrazil Babbage, "pojedel svoj lasten rep".

Analogija s starodavnim simbolom *uroboros*, kačo, ki žre lasten rep in simbolizira vase zaprt razvojni krog, je očitna. Z drugimi besedami, stroj bi bil sposoben po neki začetni zunanji vzpodbudi sam sebi spreminjati način delovanja – to pa je poleg sposobnosti pomnjenja bistven atribut učenja! Babbage svojega stroja sicer ni nikoli dokončal, po njegovih načrtih pa so ga britanski znanstveniki rekonstruirali šele konec 20. stoletja – zadeva deluje.

Sredi dvajsetega stoletja je z razvojem elektronike dobila Babbaggova ideja o univerzalnem računajočem stroju nov zagon. Kot je za zahodno civilizacijo običajno, je bila tudi tu v ozadju uporabnost tehnologije v vojaške namene. Med momente, ki so pomembno vplivali na potek druge svetovne vojne, namreč šteje tudi razbitje tajne nemške komunikacijske šifre Enigma. To je s pomočjo enega prvih elektronskih digitalnih računalnikov uspelo ekipi britanskih znanstvenikov, katere član je bil med drugimi tudi Alan Turing.

Medtem ko je bila za Descartesa sposobnost mišljenja in smiselne besedne komunikacije absolutni privilegij človeka, pa je Turingovo stališče nasprotno (Turing, 1950). Glede same sposobnosti mišljenja pri strojih je Turing sicer previden, saj priznava možnost, da je narava človeškega mišljenja povsem drugačna od postopkov, ki jih za doseganje podobnih ciljev izvaja stroj. Problem zato opiše s pomočjo sklicevanja na neke vrste psihološko "igro oponašanja" s tremi igralci – moškim in žensko v eni ter izpraševalcem v drugi sobi. Komunikacija med njimi poteka v obliki pisnih vprašanj in odgovorov, naloga izpraševalca pa je ugotoviti, kateri od obeh soigralcev je moški in kateri ženska. Pri tem se moški izdaja za žensko, ženska pa poskuša pomagati izpraševalcu k pravilni identifikaciji. Če moškega zdaj zamenjamo s strojem posebne vrste, digitalnim računalnikom, ostale pogoje igre pa pustimo nespremenjene, dobimo varianto igre, danes imenovano Turingov test.

Digitalni računalnik torej opravi Turingov test, če ima izpraševalec približno enako možnost pravilne identifikacije kot pri varianti z moškim in žensko. Turing strne svoja prepričanja o možnostih umetne inteligence z napovedjo, da bo "čez

kakih petdeset let računalnike s kapaciteto spomina okrog 10”[9] možno programirati na način, da bodo igro oponašanja igrali tako dobro, da povprečni izpraševalec po petih minutah izpraševanja ne bo imel več kot 70 odstotkov možnosti za pravilno identifikacijo”. (Turing, 1950: 442) Mnenja o uspešnosti današnjih računalnikov pri Turingovem testu kot tudi o sami vrednosti tega testa kot relevantnega pokazatelja stopnje inteligentnosti pri strojih so sicer močno deljena, prevladuje pa mnenje, da je Turingov test načeloma uporaben in da so današnji računalniki še zelo daleč od tega, da bi ga opravili.

Besedna zveza “umetna inteligenca”, ki danes označuje samostojno smer znanstvenega raziskovanja, je bila skovana šele po Turingovi smrti. Inteligenca je tako postala ključna beseda nove znanosti, ki pa jo še vedno bremenijo ista osnovna teoretska vprašanja kot pred petdesetimi leti. Inteligentnost strojev vse od takrat praktično stagnira, medtem ko je bila rast njihovih zmogljivosti eksponentna. Tudi večina ugovorov Turingovim stališčem ostaja pretežno v polju ugovorov, ki jih je v svojem odličnem tekstu obravnaval že sam.

Po moji oceni je eden temeljnih razlogov za to stagnacijo splošno sprejemanje pojmov inteligence ter sposobnosti mišljenja kot enakovrednih. Ta sinonimna uporaba obeh pojmov je še posebej izrazita v razpravah o UI, kaže pa se že v Turingovem mnenju, da je za dobro igranje šaha potreben visok nivo splošne inteligentnosti. To stališče je bilo medtem empirično ovrženo – danes je že malo boljši namizni računalnik s kvalitetnim šahovskim programom sposoben odigrati partijo vrhunskega šaha.

Sposobnost (diskurzivnega) mišljenja je tako kvečjemu potreben, nikakor pa ne tudi zadosten pogoj izkazovanja (umetne) inteligence. V podporo mojim nadaljnjim argumentacijam te trditve sledi kratka predstavitev problematike z jezikovne plati.

2. Jezikovni diskurz

Po SSKJ (2002) *inteligenca* pomeni *nadarjenost za umske dejavnosti*, medtem ko je *mišljenje* opredeljeno kot *najvišja umska dejavnost kot izraz človekove zavesti*. Na prvi pogled je torej stanje stvari ravno obratno, kot ga poskušam prikazati sam. A le na prvi pogled – *mišljenje* je namreč le *ena* izmed dejavnosti uma, pa čeprav najvišja, medtem ko *zavest* zaobjema le *del* človekovega uma.

Pri obravnavi teoretskih problemov UI z jezikovne plati je seveda nujno potrebno upoštevati tudi dejstvo, da je izvorni jezik teh tekstov praviloma angleški. V nadaljevanju bom na kratko obdelal nekaj ključnih pojmov teh razprav.

2.1 Inteligenca [*intelligence*]

Inteligenca je v slovenščino prevzeta iz germanskega jezikovnega področja, in to šele v 19. stoletju (Snoj, 2003). Angleška slovarska literatura razlaga besedo v prvi vrsti kot *sposobnost učenja in razumevanja*, ki jo nekateri viri (npr. *Merriam-Webster’s Collegiate Dictionary*, 1994) povzemajo tudi kot *razum [reason]*. Kot smo že omenili, je *sposobnost učenja* eden bistvenih elementov

koncepta digitalnih računalnikov in je tu ne bom posebej obravnaval, *razum* in *razumevanje* pa obravnavam ločeno v nadaljevanju.

Poleg nekaterih bolj empiričnih opisov, ki so bolj ali manj enakovredni bodisi osnovni ideji Turingovega testa, bodisi t.i. inteligenčnim testom, je teoretsko zanimiva še definicija *inteligence* kot *osnovne in večne lastnosti božanskega uma* v *krščanski znanosti*. Razprava o lastnostih božjega uma bi bila sicer nekoliko izven torišča pričujočega teksta, je pa razlaga pomembna zaradi dimenzij, ki jih vpeljuje v pojem.

2.2 Razum [*reason*]

Beseda *razum* izvira iz praslovanskega jezika in je sestavljena iz predpone *raz-* ter besede *um*, pri čemer je predpona razslužila za poudarjanje razlikovalne sposobnosti mišljenja (Snoj, 2003). Povsem drugačna je tvorba angleškega *reason*. Le-ta se etimološko navezuje na latinske *ration-*, *ratio*, *reti*, ki pomenijo tudi preračunavanje, izračunavanje, računanje.

V novejšem času prihaja do približevanja pomena *razuma* v smeri poudarjanja matematično-logičnih sposobnosti uma. To je razvidno tudi iz besede *racionalen*, ki je bila iz germanskega jezikovnega področja v slovenščino prevzeta v 19. stoletju, nje sopomenka slovanskega izvora pa je *razumski*.

2.3 Um/duh [*mind*]

Za angleški izraz *mind* se je v slovenščini (predvsem v filozofskih tekstih) uveljavilo prevajanje s slovenskim *duh*, kateremu pa v angleščini nekako bolj ustreza beseda *spirit*. Po drugi strani pa pod geslom *um* (Grad, 1990) najdemo vse od *reason, intellect, intelligence* preko *sense, understanding* do *mind* in celo *brain*.

Gre torej za precejšnjo pojmovno zmedo, ki pa je verjetno bolj posledica razsežnosti same vsebine pojma in njegovih različnih filozofsko-psiholoških konceptualizacij kot pa samih jezikovnih razlik. Na tem mestu bi tako izpostavil le moderni psihološki koncept uma, katerega zavestni del je po obsegu neprimerljiv z mnogo obsežnejšim področjem nezavednega. V tem kontekstu je *um* najbrž ustreznejši prevod angleškega pojma *mind*, kot pa bi to bil *duh*.

2.4 Razumevanje [*understanding*]

Etimološka struktura angleškega glagola je očitna (tj. *under* + *standing*; dob. *pod/spodaj* + *stati*), po zgledu nemškega *ver-stehen* pa je tvorjen tudi slovenski nižje pogovorni izraz *za-stopiti* (Snoj, 2003). Sicer pa je slovanska jezikovna skupina pri tvorjenju teh pojmov ubrala povsem drugo pot kot germanska (prim, zgoraj).

V slovenščini se *razumevanje* fonetično sicer veže na *razum*, a z dobesednim prevodom takšnega branja *razumevanja* v angleščino dobimo pojem *reasoning*, ki označuje predvsem racionalni postopek sklepanja oz. dokazovanja.

Razumevanje je torej potrebno brati tudi kot *raz-umevanje*. Če upoštevamo še sinonimno uporabo pojmov um/duh (predvsem glede na ang. *mind),* pridemo do pomena *"raz-duhovljenja"* ali *"materializacije"*. Taka postavitev vsebinsko precej bolj ustreza angleškemu *understanding*, ki predpostavlja nek konkreten subjekt (tudi) Newtonovsko-fizikalnih razsežnosti, ki lahko kot tak tako rekoč dobesedno *stoji pod* samim umskim konceptom kot objektom razumevanja.

Nadaljnje razvijanje teh vprašanj z jezikovne perspektive bi zahtevalo precej več prostora, celo posebno etimološko-filozofsko raziskavo, nek osnoven pregled pojmov pa je bil za pravilno razumevanje razprave, ki sledi, nujno potreben.

3. Ponovni premislek ugovorov Turingovemu stališču

Čeprav Turing v svojem slavnem članku načelno zavrača diskusijo o vprašanju "Ali so stroji sposobni mišljenja?", pa se pri obravnavi stališč, nasprotnih njegovemu, ukvarja predvsem s tem vprašanjem.

Obravnava devet različnih ugovorov, ki jih lahko razdelimo v tri skupine, njihova poimenovanja pa v večini primerov že razkrivajo njihova izhodišča. Prvi trije so načelne narave, to so (1) teološki, (2) "nojevski" oz. "glave v pesku", ter (3) matematični. Druga skupina je bolj psihološko obarvana, gre pa za argumenta glede (4) zavesti in (5) različnih nezmožnosti ter za (6) ugovor Lady Lovelace. Na koncu sledi še obravnava pretežno empirično utemeljenih argumentacij nasprotnih stališč, ki slonijo na (7) zveznosti živčnega sistema, (8) neformalnosti obnašanja ter (9) nadčutni zaznavi.

Menim, da je Turingova zavrnitev teh ugovorov večinoma ustrezna, v kolikor obravnava osnovno vprašanje: "Ali so stroji sposobni mišljenja?" Težave se pojavijo šele, če izkazovanje inteligentnosti *ni enakovredno* sposobnosti mišljenja, sam test pa nam pravzaprav odgovarja na vprašanje oblike: "Ali so stroji sposobni izkazovanja splošne, človekovi podobne inteligence?" Določeno nekonsistentnost lahko pri obravnavi tega novega vprašanja zasledimo predvsem v skupini psihološko utemeljenih ugovorov. Svoje poglede bom v nadaljevanju opisal v dveh točkah – najprej na podlagi vloge *nezavednega*, zatem pa še v povezavi z značajem in pomenom *razumevanja.* Čeprav je prvi od obeh mojih ugovorov pravzaprav zaobsežen v drugem, ju obravnavam ločeno, kot sledi.

3.1 Razsežnosti nezavednega

Sredi dvajsetega stoletja je bil behaviorizem prevladujoča paradigma psiholoških teorij in raziskav, posebno še v Združenih državah. Tudi sam Turingov test je v bistvu behavioristične test. Seveda pa koncept nezavednega ni bil neznan – pri zavračanju *ugovorov iz različnih nezmožnosti* tudi Turing sam pripiše izvor le-teh *nezavednemu* pospoševanju takrat aktualnih lastnosti strojev na njihove potencialne zmožnosti.

Ta uporaba pojma nezavednega sicer verjetno ne temelji na Freudovem tridelnem konceptu človeškega uma *id-ego-superego,* kjer je nezavedno (t.j. *id* in *superego)* osnovni izvor večine naših motivacij. Po Freudovski terminologiji bi bilo ta

psihološki mehanizem moč opredeliti celo kot *predzave- den*, kar pomeni sicer nezavedno vsebino uma, ki pa jo je mogoče brez večjih težav ozavestiti.

Na tem mestu lahko postavimo še *ad hoc* Freudovsko analogijo Turingovega stroja: *predzavedno – spominska enota, zavest ~ centralna procesna enota, nezavedno – kontrolna enota.* S tem smo sicer dobili izredno močno centralno procesno enoto, kljub temu pa se zdi, da bi Freud podprl Turingova stališča, vsaj kar se tiče zavrnitve t.i. *ugovora Lady Lovelace.* Omenjena Lady je bila sicer Babbagova prijateljica in ena redkih sodobnikov, ki je njegovo idejo o računajočem stroju sploh razumela, v zgodovini računalništva pa je danes poznana kot prvi programer. Bistvo tega ugovora je njeno stališče, da lahko stroj naredi izključno tisto, kar mu zna ukazati programer, medtem ko je človek ustvarjalno bitje *po sebi.*

Če zdaj zamenjamo Freudov koncept človeškega uma s konceptom naslednjega izmed pionirjev moderne psihologije, Carla Gustava Junga, dobi Turingov izziv povsem nove razsežnosti. Medtem ko je po Freudu nezavedno ogromno, temačno, kaotično področje, ki zaobsega posameznikovo duševnost skoraj v celoti, je le-to še vedno neznatnih dimenzij v primerjavi z Jungovim konceptom *kolektivnega nezavednega.* Po Jungu je to neke vrste historični rezervoar, ki vsebuje vse izkušnje človeštva kot vrste.

Kolektivno nezavedno je glede na osnovni Turingov namen pravzaprav preširok koncept, je pa Jung razvil še neko drugo teorijo, ki je za naše vprašanje precej bolj zanimiva. Poleg dobro poznane dvodelne tipologije osebnosti, tj. introvertiranosti ter ekstrovertiranosti, je švicarski psiholog opredelil tudi štiri osnovne načine naše komunikacije s svetom – zunanjim in notranjim. Ti načini, ali funkcije, so po Jungu štirje: *čutnost, mišljenje, intuicija* in *čustvovanje.*

Tu je *intuicija* pretežno nezaveden proces, ki ne deluje algoritmično oziroma njen način delovanja ni znan, medtem ko imata *čutnost* in *čustvovanje* pomembno vlogo v strukturi razumevanja.

3.2 Struktura razumevanja

V originalnem Turingovem tekstu se pojem *razumevanja [understanding]* sicer pojavi, a bolj mimogrede. Prvi ali pa vsaj najbolj znan kritik Turingovih stališč, za katerega je *razumevanje* osrednji pojem njegovih argumentacij, je ameriški filozof John Searle s svojim dobro znanim miselnim eksperimentom *kitajske sobe* (Searle, 1980). Njegove poglede in izpeljave teh vprašanj nekoliko podrobneje obravnavam v naslednjem poglavju.

Moje razumevanje *razumevanja* je precej različno od Searlovega, ki samega izvora tega fenomena ter njegove vloge pri inteligenci v svojih izpeljavah niti ne problematizira. *Razumevanje* kot *raz-umevanje [understanding]* namreč v pojem inteligence eksplicitno vključuje novo kvaliteto, ki je v pojmu mišljenja popolnoma odsotna. Gre seveda za *raz-mislitev,* za neke vrste *konkretizacijo* oziroma *"ozemljitev"* abstraktnega značaja mišljenja kot manipulacije s formalnimi simboli. Tu pride torej do izraza celotna kompleksnost človeškega bitja, ne pa le zavestni del njegovega uma oziroma razum [reason].

Razumevanje seveda *je* notranje stanje (zavestnega) uma, pri čemer pa le-to za lastno vzpostavitev terja *reflektiran odnos* z zunanjim svetom kot virom (povratne) informacije. Šele preko tega odnosa abstraktni jezikovnomišljenjski pojmi pridobivajo svoj pomen kot 'neke vrste metapodatek oziroma *zunanjo* referenco. Pri tem je *zunanji svet* potrebno razumeti kot zunanji glede na bitnost, ki jo lahko opišemo kot *zavestni del uma*, *miselni proces* ali *tok misli, trenutno zavedanje* ipd.

Ko Turing na koncu svoje razprave obravnava možne strategije razvoja umetne inteligence, ugotavlja, da bi obetavna smer razvoja lahko bila "opremiti stroj z najboljšimi čutnimi organi, kar jih je moč kupiti za denar, ter ga potem naučiti razumeti in govoriti angleško". Pri obravnavi *argumentov glede različnih nezmožnosti* pa kot nekoliko neresno izpostavi nezmožnost stroja, da bi *se sladkal z jagodami s smetano.* S Turingom se sicer lahko strinjam, da bi bil vsak poskus konstruiranja tovrstnega stroja sicer "idiotizem" (Turing, 1950: 448), vendar pa so po drugi strani tovrstne *zmožnosti* (skupaj s posledičnimi *izkušnjami)* temelj vsakršnega razumevanja. Turing je morda upal, da bo ob zadostnih zmogljivostih stroja možna simulacija resničnega razumevanja do stopnje, ki bo le-temu omogočila zadostne sposobnosti za prepričljivo sodelovanje v igri oponašanja, a zaenkrat kaže, da temu ni tako.

Da bi torej stroj usposobili za *razumevanje* stvari na način, primerljiv s človekovim, bi ga morali oskrbeti z umetnim telesom, ki bi bilo čimbolj podobno človeškemu – navzven *in* navznoter! Za dodatno osvetlitev te problematike se bom naslonil na dobro poznan primer gdč. Hellen Keller, ki je bila slepa in gluha, a je kljub temu izkazovala visoko stopnjo inteligentnosti.

Znano je namreč dejstvo, da ljudje z okvarami posameznih čutnih organov te svoje pomanjkljivosti navadno kompenzirajo z dodatno izostritvijo ostalih čutilnih sposobnosti, ki so polno delujoče. Prvi poudarek bi torej bil, da je pomembna predvsem *kvaliteta* oziroma *potencialne zmožnosti* posameznih čutov, ne pa le njihov goli obstoj.

Naslednji vidik obravnavane teme je problem procesiranja čutnih podatkov. Ko je tok čutnih podatkov iz enega ali več čutnih organov oviran ali zaustavljen (npr. kot posledica fizične okvare posameznega čutila), so s tem (raz)umu na voljo dodatne "kapacitete" za obdelavo podatkov oziroma pridobivanje informacij iz drugih virov. Eden od virov je tu lahko tudi *nezavedno*, obravnavano v prejšnjem podpoglavju – nekatere azijske filozofske tradicije, npr. budistična filozofija duha, dejansko postulirajo dodatni, notranji čut na v osnovi enak način kot ostalih pet čutov, s čimer postane koncipiranje nezavednega kot posebnega dela posameznikovega uma metodološko nepotrebno.

V tej zvezi bi lahko ponovno proučili tudi argument glede nadčutne zaznave, ki ga je precej resno obravnaval že Turing, a za splošno strukturo razumevanja le-ta pravzaprav niti ni pomembna, pa naj jo sprejmemo kot možno, kot gotovo ali pa jo povsem zavrnemo.

4. Umetna inteligenca in razumevanje

Vrnimo se zdaj k Searlu, za katerega je vprašanje *razumevanja [understanding]* ključni moment teoretskih razprav o umetni inteligenci. Searle (Searle, 1980) najprej vzpostavi razliko med "močno" in "šibko" UI. Pri tem gre v bistvu za vprašanje "narave mišljenja" pri človeku na eni in pri stroju na drugi strani. To vprašanje je izpostavil že Turing (Turing, 1950) (prim, zgoraj), a se je njegovi neposredni obravnavi izognil. Po prepričanju zagovornikov "šibke" UI, med katere se uvrsti tudi Searle sam, so strojni postopki le *simulacija* človeškega uma, medtem ko se za pristaše "močne" UI v ustrezno programiranem računalniku um *dejansko* vzpostavi.

Desetletje po objavi že omenjenega miselnega eksperimenta *kitajske sobe* Searle (Searle, 1990) strne svoje poglede v formalizirano obliko. Kot izhodišče postavi tri osnovne aksiome (Searle, 1990: 27): "(A|) Programi so formalni (sintaktični)", "(A,) Um vsebuje pomen (semantiko)" in "(As) Sintaksa sama pomena niti ne vzpostavlja niti zanj ne zadostuje". Iz teh predpostavk izpelje naslednji sklep (Searle, 1990: 29): "(C|) Sami programi uma niti ne vzpostavljajo niti zanj ne zadostujejo."

Z navedenimi predpostavkami se sicer lahko strinjam, ravno tako tudi z zaključkom, vendar pa Searle potem doda še naslednji aksiom: "(As) Možgani povzročijo pojavitev uma." Šele po tem dodatku se mu zdi zavrnitev "močne" UI dokončno utemeljena. Ta četrti aksiom je, poleg tega da pravzaprav niti ni bistven za podporo njegovih stališč, najmanj zelo vprašljiv.

Vzpostavitev pomena oziroma intencionalne vsebine uma (v A_3) bi težko upravičili kot posledico *notranjih* lastnosti možganov, kakor nam to sugerira Searle. Abstraktnim simbolom, nad katerimi deluje mišljenje oziroma um, podeljuje *semantično vrednost* šele njihovo *razumevanje*, ki je posledica interakcije med *zavestjo* in *"Celoto"*, tj. vsoto *vseh* možnih pojavov, vključno z zavestjo samo. Pri tem sta mentalno *in* fizično telo subjekta temelj oziroma posrednik, ki to interakcijo omogoča. *Pomen* oziroma *razumevanje* formalnih simbolov se tu sicer res vzpostavi kot *notranje* stanje zavesti, utemeljeno pa je na *izkušnjah* oziroma v *odnosih* subjekta z zunanjim svetom. Možgani kot fiziološka podlaga človeškega uma pri tem igrajo le vlogo neke vrste "pretvornika signalov". Možgani so seveda izjemno kompleksna struktura, kljub temu pa bi jim namesto *vzročne* vloge precej lažje pripisali *omogočanje* pojavitve uma.

Ugovori Searlovim izpeljavam pa, podobno kot to velja za kritike Turingovih stališč, večinoma niso preveč prepričljivi. Obakrat je namreč problem v samih temeljnih predpostavkah, ne pa v nadaljnji obravnavi le-teh. Pri Turingu je to enačenje inteligence s sposobnostjo mišljenja, ki se kaže v njegovi napačni opredelitvi sposobnosti igranja šaha, medtem ko se Searle v strahu pred (kartezijanskim) dualizmom zateče v "biološki naturalizem" in prenese "odgovornost" za inicialno proženje vzročnih nizov na fiziologijo možganov.

Preden nadaljujem, dodajam še pripombo, da naj bi razumevanje teh mojih stališč ne slonelo na kakih določenih metafizičnih ali ontoloških predpostavkah – tu gre bolj za fenomenološko analizo obravnavane tematike. Enako velja tudi za preostali del tega teksta.

Pojem razumevanja v svoji angleški obliki *"understanding"* ne vključuje nobenih določil subjekta, kateremu se pripisuje – na človeka oziroma višje živali ga veže predvsem njegova običajna uporaba. V angleščini je torej nekako lažje pripisati sposobnost razumevanja stroju, medtem ko bi mu morali v slovanskih jezikih v tem primeru posledično pripisati tudi posedovanje uma. V ta namen bi bilo tako potrebno koncept uma razširiti na vsakršne bitnosti z *lastno notranjo sposobnostjo zadrževanja reprezentacij objektov ter manipulacije z njimi.*

V primeru človeka so to "mentalne podobe" in sposobnost (diskurzivnega) mišljenja, pri stroju pa se to običajno označuje kot *formalni simboli* oziroma *programska oprema [software]*. Tisto, kar v največji meri omogoča *vzpostavitev pomena* teh "mentalnih podob" je pri človeku njegov *fiziološki aspekt*, pri stroju pa t.i. *strojna oprema [hardware]*.

Na prvi pogled ta analogija morda deluje precej domače, a so v njej pravzaprav vsebovani ključni poudarki mojega stališča. Tu je namreč v igri *celotna kompleksnost človeškega' organizma* skupaj s *celotno kompleksnostjo človeškega uma.* V popularni različici te analogije vlogo strojne opreme prevzemajo izključno možgani, medtem ko programska oprema pokriva predvsem sposobnosti diskurzivnega mišljenja.

Ne nazadnje je tu še dejstvo, da so vse te človeške vsebine in funkcionalnosti koncentrirane v dokaj kompaktni masi velikostnega reda nekaj deset kilogramov, ki je poleg tega sposobna tudi dokaj učinkovite interakcije z okoljem – bodisi po sebi, bodisi s pomočjo "dodatne opreme", zasnovane v umu ter izdelane in uporabljane s pomočjo telesa ... Umetna izdelava nečesa vsaj nekoliko podobnih lastnosti je z ozirom na današnjo stopnjo razvoja tehnologije in računalništva, pa tudi glede na stopnjo znanstvenih spoznanj o delovanju človekovega telesa in uma, povsem nepredstavljiva.

5. Zaključek

Možnosti razvoja umetne inteligence so gotovo še daleč od tega, da bi bile izčrpane, pa čeprav le ob uporabi digitalnih računalnikov. Na drugi strani pa je verjetno še neprimerno večje polje lastnosti in zmožnosti človeške inteligence, ki je za sodobno znanost še vedno uganka. Ostaja nam torej le upanje, da bosta obe področji napredovali na način, ki bo v dobrobit vseh čutečih bitij ...

THINKING MECHANISMS AND INTELLIGENCE

In the classical text on the possibilities and limitations of artificial intelligence (AI) "Computing Machinery and Intelligence" (1950), A. M. Turing considered the question "Can machines think?" dismissed it as being rather ambiguous, and introduced a kind of game which he called the "imitation game", instead. Similarly,

the term *intelligence*^* explicitly related to machines solely in the title of his poper. Still, these theoretical issues (i.e. thought and intelligence in machines) turned out to become the fundamental topic of the emerged scientific field.

In this dispute, my basic objective is to point out that utilizing (or constructing) thinking mechanisms is a necessary, but insufficient condition for exhibiting (artificial) intelligence. As the two are often regarded as synonymous, especially in the field of Al, I begin my discussion by presenting the historical and linguistic background of the matter. As next I relate the subject in hand to well known psychological theories of S. Freud and C G. Jung. Finally, the importance of understanding turns out to be the most important topic of the subject in hand. I confront my understanding of understanding to the views of American philosopher J. Searle, who totally denies the possibility of understanding in machines.

Keywords: Artificial Intelligence, Turing, thought, understanding, Searle.

Literatura

Grad, Leeming (1990). *Slovensko angleški slovar.* Ljubljana: DZS.

Hofstadter, D. R., Dennett, D. (1990). *Oko duha.* Ljubljana: Mladinska knjiga.

Jacker, C. (1966). *Človek, spomin in stroji.* Ljubljana: Mladinska knjiga. Prevedel F. Jerman.

Merriam-Webster's Collegiate Dictionary. 10th ed. (1994). Springfield, Massachusetts, U.S.A.: Merriam-Webster, Incorporated.

Searle, J. (1980). "Minds, brains, and programs". *Behavioral and Brain Sciences,* III, str. 417-424.

Searle, J. (1990). "Is the Brain's Mind a Computer Program?". *Scientific American,* 262, str. 26-31.

Slovar slovenskega knjižnega jezika (2002). Ljubljana: DZS.

Snoj, M. (2003). *Slovenski etimološki slovar.* Druga izdaja. Ljubljana: Modrijan.

The Oxford Dictionary of Modern English. 2nd ed. (1984). Ljubljana: Cankarjeva založba.

Turing, A. M. (1950). "Computing Machinery and Intelligence". *Mind,* 59, str. 433-460.