OD DESCARTESA DO ALEXE: FILOZOFSKI POGLED NA RAZVOJ UMETNE INTELIGENCE

JANEZ BREGANT
Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
janez.bregant@um.si

Sinopsis Mehanicističen odgovor na vprašanje o tem, kaj je mišljenje, sega v 17. stol., ko je Thomas Hobbes zapisal, da ni nič drugega kot računanje. Čeprav se je mogoče sam zdel naklonjen takšni materialistični ideji, ki lastnost duha formulira v pojmih mehanskih procesov, pa ni bilo jasno, kako naj bi to potekalo. Začetki simulacije inteligentnega obnašanja segajo v 60. leta 20. stol., ko je bil v modi t. i. simbolni pristop k modeliranju inteligence, ki je temeljil na deduktivnem sklepanju. Kmalu se je izkazalo, da vsega znanja, ki bi ga stroji rabili za to, da bi uspešno opravljali tudi vsakodnevne kompleksnejše naloge, kljub vsemu ne moremo predstaviti v strukturah, podobnim stavkom. Tako se je razvila nesimbolna umetna inteligenca (UI), ki pri simulaciji inteligentnega obnašanja stavi na induktivno sklepanje, verjetnost in prepoznavanje vzorcev. Danes se zdi, da je pri razvoju različnih aplikacij meja samo še nebo, vprašanje pa je za kakšno ceno. V članku najprej predstavimo vizionarje, ki so predvideli nastanek strojev in se spraševali, ali bodo ti kdaj mislili, potem opišemo prvi val razvoja UI, ki razume mišljenje kot simbolno manipulacijo, nato pa še drugega, ki inteligenco modelira s pomočjo nevronskih mrež, ki se urijo z učnimi primerki.

Ključne besede:
umetna inteligenca,
simbolna UI,
nesimbolna UI,
nevronske mreže,
strojno učenje,
rudarjenje
podatkov

Univerzitetna založba
Univerze v Mariboru

DOI https://doi.org/10.18690/um.ff.11.2022.1
ISBN 978-961-286-675-4

FROM DESCARTES TO ALEXA: A PHILOSOPHICAL VIEW ON THE EVOLUTION OF ARTIFICIAL INTELLIGENCE

JANEZ BREGANT
University of Maribor, Faculty of Arts, Maribor, Slovenia
janez.bregant@um.si

Abstract The mechanicist answer to the question of what is thinking goes back to the 17th century when Thomas Hobbes wrote that it is nothing but calculating. Despite being fond of such a materialist idea, which formulates the mind in terms of mechanical processes, it was not clear how this was supposed to work. The beginning of the intelligent behaviour's simulation dates to the 1960s when the symbolic approach based on deductive reasoning was at work. It soon turned out that not all knowledge needed by the machines to complete the complex tasks typical for our everyday life could be represented by structures that resemble sentences. This gave birth to the non-symbolic artificial intelligence (AI), which simulates intelligent behaviour by using inductive reasoning, probability, and pattern recognition. The article first introduces visionaries who predicted the arrival of machines and wondered if they could ever possibly think. It then describes the first wave of the AI's evolution, which understands thinking as symbol manipulation. This is followed by the outline of the second wave of AI's evolution, which models intelligence with the help of neural networks trained by learning examples.

Keywords: artificial intelligence, symbolic AI, non-symbolic AI, neural networks, machine learning, data mining

https://doi.org/10.18690/um.ff.11.2022.1 DOI
978-961-286-675-4 ISBN
University of Maribor Press

1 Uvod

Kaj je umetna inteligenca (UI)? Google Maps, Instagram, Twiter, Facebook, Dropbox, YouTube, Netflix, Google Translate, Siri, Alexa, iPhone, iPad, Mac itd., z drugimi besedami, spletni brskalniki, aplikacije za družbena omrežja in shranjevanje v oblakih, za predvajanje glasbe ali filmov, za pomoč v obliki virtualnih prevajalnikov in asistentov, pametni telefoni, tablice, računalniki itd. Meja med inteligenco, ki je naravna (človeška), in tisto, ki je umetna (strojna), je danes zabrisana, pogosta uporaba takšnih in drugačnih pametnih orodij, ki je prerasla v običajno prakso, pa ustvarja površinski vtis, da med njima ni razlike. Takšno pozabljanje na to, da je človeška inteligenca zavestna, strojna pa zgolj mehanska, če vse skupaj poenostavimo, omogoča UI, da se v družbi »skrije« in vzbuja občutek, da je z nami že od nekdaj.

Začetki tega, kar danes imenujemo UI, segajo v obdobje po drugi sv. vojni, ko so nastali prvi računalniki s programi, ki so bili zelo uspešni pri igranju npr. dame, ko je za inteligentno štelo obnašanje, ki je vključevalo opravljanje logičnih operacij. Najbolj znan program za igranje dame, ki se je že bil sposoben učiti, kar mu je omogočalo napredovanje in zmago nad svojim izumiteljem, je že na začetku 50. let 20. stol. razvil Arthur Samuel, deloval pa je na IBM-ovem računalniku 701. Za rojstvo UI kot discipline pa štejemo leto 1956, ko so se na Darthmouthski konferenci zbrali vsi raziskovalci strojne inteligence, ki so takrat kaj pomenili in se povezali v formalno družbeno skupnost s ciljem narediti stroj, ki je resnično inteligenten. Običajno se opisuje kot »znanost o izdelavi strojev, ki so sposobni narediti stvari, za katere je po naših merilih potreben um« (Copeland 1993: 1), ali kot »znanost o tem, kako narediti in/ali programirati računalnike, da bodo sposobni istih stvari kot um«. (Boden 1990: 1) V bistvu je šlo za izdelavo stroja (računalnika), ki je z namenom doseganja ciljev zmožen opravljati zapletene kognitivne operacije, povezane z obnašanjem, načrtovanjem, reševanjem problemov, napovedovanjem, sklepanjem itd., in operacije, ki vključujejo matematične sposobnosti, komunikativnost, jezikovne sposobnosti, zaznavo, motorične sposobnosti, socialno razvitost in učenje.[1]

[1] V splošnem bi lahko rekli, da gre za zaznavanje, spominjanje in učenje, da lahko to, kar smo doživeli, posplošimo in uporabimo za reševanje konkretnih življenjskih problemov, tj. da se znamo v družbenem okolju smiselno obnašati.

V tem obdobju bi lahko govorili o prvem valu razvoja UI,[2] raziskovalci so se ukvarjali z modeliranjem človeških psihičnih sposobnosti, kot so logično mišljenje, ustvarjalnost, uporaba jezika, učenje, reševanje problemov itd., kar bi strojem omogočalo avtonomno delovanje v negotovem svetu. To, včasih se imenuje zlato obdobje UI, je trajalo vse do sredine 70. let 20. stol., v njem pa so bili postavljeni temelji računalniškega sklepanja, razvite so bile prve nevronske mreže in doseženi prvi uspehi pri posnemanju človeške inteligence. Temu začetnemu optimizmu je sledilo obdobje streznitve, programi, ki so bili napisani za reševanje preprostih problemov, kot je npr. premikanje kock v prostoru in so temeljili na logičnem sklepanju v strogo določenem okolju, se niso obnesli pri reševanju zapletenih življenjskih problemov v spreminjajočem se svetu, ki je temeljilo na statističnih zakonitostih (verjetnosti) in zahtevalo veliko računsko moč za obdelavo velike količine podatkov.

Ponovni vzpon UI se je začel spet v 80. in 90. letih 20. stol., ko so bile zaradi vedno večje računske moči računalnikov, vedno večjega števila podatkov, ki so bili na voljo in vedno boljšega dostopa do njih, omenjene tegobe počasi odpravljene. Govorimo lahko o drugem valu razvoja UI, ki pa ne temelji več na dedukciji in pisanju navodil za manipuliranje s simboli kot prvi, ampak na indukciji in pisanju algoritmov za strojno učenje. (Markič 2019) Raziskovalci so svojo pozornost preusmerili na razvijanje pametnih orodij, ki delujejo po principu učenja na osnovi predhodnih izkušenj in interakcije z okoljem, svoj navdih pa črpajo iz povezovanja različnih disciplin, kot so kibernetika, teorija verjetnosti in statistika. Uporabljajo na podlagi učnih primerkov modelirane nevronske mreže, ki pa so še vedno specializirane zgolj za opravljanje nalog na nekem ozkem področju, npr. za prepoznavanje obrazov, prevajanje ali avtonomno vožnjo, tako da v tem trenutku še ne moremo govoriti o neki splošni UI, ki bi bila primerljiva s človeško, kjer gre za obvladovanje večih, med seboj tudi precej različnih področij in sposobnost reševanja nalog znotraj njih.

V članku se najprej zazremo v zgodovino in prikažemo avtorje, ki so že pred davnim časom mišljenje na takšen ali drugačen primerjali z računanjem, vizionarje, ki so predvideli nastanek strojev in se spraševali, ali bodo ti kdaj mislili, ter načrtovalce abstraktnih računalnikov, ki niso nikoli zagledali luči sveta. Potem podrobneje predstavimo prvi val razvoja UI, nekaj njegovih dosežkov in filozofsko vprašanje, ali nismo tudi ljudje zgolj računalniki, ki predstavlja vrh razmišljanja o podobnosti

[2] Razdelitev na dva vala razvoja UI predlaga Cantwell Smith (2019) in uvaja npr. Markič (2019).

med stroji in ljudmi. Temu sledi opis drugega vala razvoja UI z nekaterimi najbolj znanimi primeri umetnih sistemov danes s poudarkom na izgubi zasebnosti kot ceni, ki jo moramo z moralnega vidika plačati za tako hiter tehnološki razvoj. Končamo s kratkim povzetkom in navedbo virov ter literature.

2 Izlet v zgodovino

»Mišljenje ni nič drugega kot računanje,« pravi Hobbes leta 1651 v *Leviathanu* (Hobbes 1651/2006: 32). Verjetno gre za prvi opis določene mentalne operacije v pojmih računanja, ki vključuje simbolno manipulacijo, ki se dogaja v naših možganih. Primerki misli so tako v bistvu primerki možganov, ali kot jih imenuje Hobbes, 'fantazme', mišljenje pa neke vrste mentalni pogovor.

> Ko človek razmišlja ne dela ničesar drugega kot, da si zamišlja končno vsoto, ki jo dobi s seštevanjem delov, ali ostanek, ki ga dobi z odštevanjem ene vsote od druge. /…/ Te operacije se ne dogajajo samo pri številih, ampak pri vseh načinih stvari, ki jih lahko združujemo ali odvzemamo. (Hobbes 1651/2006: 32)

Ko razmišljamo, podobno kot pri računanju s pomočjo svinčnika in papirja, uporabljamo simbolne operacije, le da te niso izražene z govorjenimi ali pisanimi simboli, temveč v posebnem nevralnem zapisu/kodi. (Markič 2019) Haugeland (1986) pravi, da je pri Hobbsu mišljenje mehanski proces, podoben upravljanju mentalnega abaka: fantazme premetavamo sem in tja skladno s pravili razuma, podobno kot na pravem abaku skladno z računskimi pravili sem in tja premikamo kroglice. Hobbsova izvirnost se kaže tudi v tem, da je predpostavil obstoj umetnih sistemov, ker njegov materializem, po katerem življenje ni nič drugega kot gibanje telesnih organov, ki je podobno gibanju zobnikov, vzmeti in koles pri stroju, obstoj takšnih *avtomatov*, kot jih sam imenuje, tudi dopušča. »Kaj pa je srce drugega kot vzmet, živci strune in sklepi kolesa, kar telesu skupaj omogoča gibanje /…/?« (Hobbes 1651/2006: 9)

Tudi Leibniz je menil, da je mišljenje računanje, in predlagal izoblikovanje univerzalnega pojmovnega jezika (lat. *characteristica universalis*), neke vrste abecede človeškega mišljenja, s katero bi lahko izrazili matematične, znanstvene in metafizične resnice. »/…/ nihče še ne poskušal ustvariti jezika /…/, v katerem znaki in simboli služijo [misli] na isti način kot matematični znaki služijo številom ali

algebraični znaki količinam.« (Leibniz 1679/1969: 222) Njegov univerzalni jezik je bil v resnici del večjega projekta, in sicer izoblikovanja univerzalne znanosti (lat. *scientia universalis*), tj. transformacija celotnega človeškega znanja v eno sistematično celoto, v kateri bi mišljenje potekalo kot računanje. Da bi bil ta program uspešen, je Leibniz predlagal še univerzalni mehanizem sklepanja (lat. *calculus ratiocinator*), zbirko načinov manipulacije znanja, zapisanega v računalniški obliki, z namenom odkrivanja logičnih relacij med idejami in njihovimi posledicami.

> Naj bodo pojmi, iz katerih je sestavljeno vse drugo, kar obstaja, prvič določeni z znaki, ki bodo neke vrste abeceda. Bilo bi priročno, če bi bili karseda naravni, npr. pike za števila, črte za relacije med entitetami /…/ Če bodo pravilno in domiselno izbrani, bo univerzalni jezik enostaven in običajen ter za njegovo razumevanje ne bomo rabili slovarja. Poleg tega si bomo na tak način zagotovili znanje o vsem, kar obstaja. (Leibniz 1666/1966: 10–11)

Descartes na drugi strani pa primerke misli nikakor ne vidi kot primerke možganov. To mu onemogoča njegov pogled na t. i. *problem duha in telesa*,[3] tj. vprašanje, kakšen je odnos med mentalnim in fizičnim oziroma med našimi psihičnimi in možganskimi stanji. Njegov odgovor je splošno znan *dualizem substanc*, stališče, ki predpostavlja obstoj dveh različnih in ločenih substanc, misleče/mentalne – *res cogitans* in razsežne/fizične – *res extensa*. V *Razpravi o metodi* leta 1637 zapiše takole: »Potemtakem je ta jaz, namreč duša, po kateri sem, kar sem, popolnoma različna od telesa in jo je celo lažje spoznati kakor pa telo; in četudi telesa ne bi bilo, ne bi prenehala biti vse tisto, kar je.« (Descartes 1637/2007: 51, 53)[4] Njegovo prvo načelo filozofije je sicer *mislim, torej sem* (lat. *cogito ergo sum*), saj odkrije, da on sam, ki misli, da je vse zmotno, nujno nekaj je. Vprašanje je, kaj ta *jaz*, v katerega ne more dvomiti (če v vse drugo že lahko), je. Na to odgovori z miselnim eksperimentom, v katerem si predstavlja sebe brez telesa, brez kakršnihkoli organov (svoje telo povsem odmisli), in ugotovi, da kljub temu še vedno nekaj je, je tisto, kar je odmislilo telo, tj. misleča stvar. »Iz tega sem spoznal, da sem substanca, katere celotno bistvo ali narava

[3] Zanj glej npr. Kim (2001).

[4] V *Meditacijah* to izrazi takole: »In čeprav morda (ali bolje: zagotovo …) imam telo, ki je zelo tesno združeno z mano, je vendar – ker imam na eni strani jasno in razločno idejo samega sebe, kolikor sem samo misleča, nerazsežna stvar, in na drugi strani jasno idejo telesa, kolikor je telo zgolj razsežna, nemisleča stvar – je torej vendar gotovo, da sem v resnici različen od svojega telesa in da bi lahko bival brez njega.« (Descartes 1641/1988: 107) (Zadnji del citata, tj. od vezaja do pike, je zaradi napake prevajalca spremenjen; za primerjavo glej Descartes 1641/1985a: VI meditacija.)

je zgolj mišljenje in ki za bivanje ne potrebuje nobenega kraja in ni odvisna od nobene materialne stvari.« (Descartes 1637/2007: 51)

Ker so lastnosti, ki jih pripisuje duhu, npr. misli, občutki, čustva, lastnosti, ki jih pripisuje telesu, pa npr. razsežnost, velikost, oblika, zaradi česar je duh stvar, ki misli in čuti, telo pa stvar z maso in lokacijo (psihični pojavi, kot so čustva, so po svoji naravi drugačni in ločeni od fizičnih pojavov, kot je proženje nevronov), Descartes mišljenja nikakor ne more opisati v pojmih (mehanskega) računanja, kot je to storil Hobbes. Res je, da med duhom in telesom poteka interakcija, ni pa jasno, kako je to mogoče, glede na to, da mentalna substanca nima lokacije v prostoru, fizična pa jo ima. Ta temeljni problem dualizma substanc, tj. kako lahko človeški duh vpliva na aktivnosti telesa (npr. krčenje in raztezanje mišic), je Descartesov sodobnik Gassendi izpostavil z besedami »Pojasniti moraš, kako je ta interakcija, če si breztelesen, se ne raztezaš v prostoru in si nedeljiv, možna. /… / Kako lahko, če nimaš delov, vplivaš na dele? /… / In če si nekaj ločenega, kako lahko skupaj s snovjo tvoriš celoto?« (Descartes 1985b: 238)

Kakorkoli, Descartesovi opisi delovanja telesa (za razliko od duha) pa so bili povsem mehanicistični. Ne samo da je v *Razpravi o metodi* za tisti čas podal podroben prikaz anatomije našega telesa, razmerja med organi, njihovih funkcij in delovanja, dotaknil se je tudi vprašanja podobnosti med človekom in strojem, ki ga prav tako imenuje avtomat.

Dopušča možnost, da bi obstajali stroji, ki bi bili sposobni posnemati naša dejanja, vendar se kljub temu nikoli ne bi moglo zgoditi, da bi jih zamenjali za ljudi. »Kot prvo ne bi nikoli znali uporabljati besed in drugih znakov ter jih sestavljati, kot to počnemo mi, da drugim razodenemo svoje misli. Lahko si sicer zamislimo, da bi bil kakšen stroj narejen tako, da bi izgovarjal besede, in celo, da bi izrekel kakšno besedo o telesnih dejanjih, ki bi povzročila spremembe v njegovih telesnih organih /…/ ne moremo pa si zamisliti, da bi te besede združeval v različne tvorbe in z njimi smiselno odgovarjal na vse, kar bi kdo rekel vpričo njega, kot to zmorejo celo najbolj omejeni ljudje.« (Descartes 1637/2007: 83) Razlika, na katero opozarja Descartes, je v tem, da stroji ne delujejo po spoznanju duha, ampak zgolj naravnanosti (sprogramiranosti) njihovih organov. Ta jim sicer omogoča mehansko manipulacijo s simboli, če uporabimo sodobnejšo terminologijo, ki pa ne vključuje racionalne spoznave.

> Razum je namreč univerzalno orodje, ki ga lahko uporabljamo v vseh okoliščinah, nasprotno pa morajo biti ti organi za vsako posebno dejanje posebej naravnani. Zato je praktično nemogoče, da bi bilo v kakšnem stanju toliko različnih organov, da bi mu to omogočalo delovati v vseh življenjskih slučajih, kakor to nam omogoča razum. (Descartes 1637/2007: 83, 85)

Zaključimo lahko, da Descartes nasprotuje strojnemu mišljenju, saj zanj imeti duha pomeni biti sposoben delovati na različnih med seboj tudi precej različnih področij, kar pa nujno vključuje uporabo razuma, ki pa ga avtomati, ker so specializirani zgolj za opravljanje nalog z nekega ozkega specifičnega področja, ne premorejo.[5]

Pomemben mejnik v zgodovini UI predstavlja »izdelava«, bolje rečeno razvoj, prvega računalnika, ki se je pojavil okrog leta 1833, njegov avtor pa je bil (zgodovinsko gledano prvi računalničar) Charles Babbage. Imenoval se je *Analitični stroj*,[6] [7] njegov ustroj pa je vključeval dve izjemni ideji, ki skupaj predstavljata temelj računalništva: a. operacije je bilo mogoče v celoti programirati in b. programi so lahko vsebovali pogojne izjave tipa *če – potem – drugače* (*če* je odgovor »Y«, potem zapiši »pravilno«, *drugače* zapiši »narobe«).[8] (Haugeland 1986: 126) Sestavljen je bil iz treh delov: *mlina* (aritmetične enote), *shrambe* (spominske enote) in *kontrolne enote*. Mlin lahko izvaja štiri osnovne računske operacije, kontrolna enota pa je »sodnik«, ki ne manipulira s števili, ampak glede na navodila zgolj ukazuje, s katerimi števili se sešteva, odšteva, množi ali deli ter kam se zapisujejo odgovori. Znotraj svojih mej lahko Analitični stroj opravi katerokoli nalogo, ki mu jo naložimo, tj. ko enkrat na ustrezen način določimo pravila igre, kontrolna enota poskrbi za to, da se jih igralci (aritmetična in spominska enota) držijo.

[5] Iz *Razprave o metodi* ni jasno, ali stroji razuma nimajo zato, ker je ta del mentalne substance, ki je ni v prostoru, stroji pa so telesni, ali pa zato, ker pri izvrševanju nalog niso sposobni pokrivati več različnih področij. Glede na to, da je Descartes trdil, da med mislečo in razsežno substanco obstaja vzajemno delovanje in to s pomočjo »češerike«, sicer neuspešno, tudi dokazoval, se zdi bolj verjetna druga možnost. Če to drži, potem nasprotuje le pripisovanju inteligence aktualnim modelovm UI, ki so specializirani samo za opravljanje določenih nalog z enega ozkega področja, ne pa tudi obstoju splošne UI.

[6] *Analytical Engine*.

[7] Babbage je na papirju razvil dva računalnika, prvi, ki se je v angl. imenoval *Difference Machine*, je bil sicer izviren, za nadaljnji razvoj pa ne tako pomemben izdelek.

[8] Angl. *conditional branches*.

To početje, ki iz specifičnega avtomatskega sistema zgolj s primernim opisovanjem in spreminjanjem navodil naredi avtomatski splošni/formalni sistem (računalnik), pa že imenujemo programiranje.[9] Tudi navadni kalkulator lahko sicer prav tako izvaja iste štiri operacije, a to razliko, da moramo posamezne ukaze vnašati ročno. Če želimo izračunati 3(x +y) – 5 za dana x in y, moramo najprej vstaviti dana x in y, potem njuno vsoto pomnožiti s 3 in na koncu od zmnožka odšteti 5. Analitični stroj pa je bil na drugi strani že takrat (kot danes računalniki) sposoben na osnovi specifikacij, ki, kot smo že omenili, vključujejo pogojne izjave tipa *če – potem – drugače*, za izračun poljubnega zaporedja osnovnih računskih operacij za poljubne spremenljivke, izražene v njemu razumljivem »jeziku«, avtomatsko izračunati katerokoli dano formulo. »Tako je bil Babbage prvi, ki je »naredil« sprogramiran računalnik, pri čemer je treba »naredil« vzeti z rezervo: ker je bil njegov stroj prevelik in prezapleten, zaradi česar ga ni skoraj nihče razumel, kaj šele bil sposoben zgraditi, ni v praksi nikoli zaživel.« (Bregant 2010: 62)

Nič manj pomembna ni iznajdba še enega takega stroja, ki ga je doletela ista usoda, tj. nikoli ni bil realiziran v praksi, čeprav iz drugega razloga – že v osnovi je bil mišljen le kot enostavni abstraktni sistem za dokazovanje teoretičnih predpostavk – je leta 1936 razvil Alan Turing.[10] Potem, ko je računanje definiral kot formalno manipulacijo z neinterpretiranimi simboli, ki se izvaja z uporabo formalnih pravil (Turing 1936; Markič 2019), je opisal še napravo, ki je tako računanje sposobna izvesti. Znana je pod imenom *Turingov stroj* in sestavljena iz *glave* in *traku*. Trak, ki je shramba podatkov, je razdeljen na neskončno število kvadratov, izmed katerih so eni vedno zasedeni s simbolom, tj. figurami, s katerimi se manipulira, iz predpisane abecede, drugi pa prazni. Glava, ki je izvrševalec operacij, se pomika preko kvadratov in skenira njihove simbole. Naenkrat obdeluje samo en kvadrat in to je edini kvadrat, s katerega takrat bere ali na katerega zapisuje, odvisno pač od tega, v katerem notranjem stanju je, tj. kaj je predpisano delo, ki ga mora na njem opraviti. Ko konča, se pomakne naprej ali nazaj. Na katerem kvadratu se ustavi in kaj z njim stori, je vnaprej določeno s pravili, po katerih deluje. Ta vsebujejo opise tega, (i) kateri simbol je treba zapisati na dani kvadrat (ali s katerim novim je treba nadomestiti starega), (ii) kateri kvadrat je treba skenirati naslednji in (iii) kaj je treba z naslednjim kvadratom storiti. Tudi Turingov stroj je tako preprost avtomatski formalni sistem, podoben Analitičnemu stroju, s katerim lahko izvršimo vsako nalogo, za katero smo

[9] Turing pozneje pravi, da lahko takšne računalnike, ki imajo »posebno lastnost /…/, da lahko posnemajo kateri koli diskretni stroj, opišemo [kot] /…/ *univerzalne* stroje.« (Turing 1950/1990: 64)
[10] Zanj glej Turing (1936), Haugeland (1986).

jasno specificirali korake, ki so potrebni za njeno izpolnitev. Čas za prihod računalnikov, ki ne bodo le imaginarne tvorbe na listu papirja, je napočil.

3 1. val razvoja UI: simbolno modeliranje

Računalniki

Razvoj UI v neki točki sovpada z razvojem računalnikov. Beseda 'računalnik' naj bi se prvič pojavila leta 1613, ko jo je Richard Brathwaite uporabil za opis osebe, ki je računala, kasneje pa tudi za naprave, ki so bile sposobne izvajati računske operacije. Sodoben pojem elektronske računske naprave pa se v bistvu ni pojavil vse do leta 1945, ko ga je von Neumann uporabil v svojem znanem poročilu o EDVAC-u,[11] ki je bilo prvi javno objavljeni opis logične zgradbe računalnika, v katerem so bili podatki o programu in podatki o navodilih shranjeni v spominski enoti, kar je dobilo ime von Neumannov model (ali arhitektura). (Berkeley 2018) Prvi računalnik, ki ni bil zgolj neuresničen teoretični eksperiment naj bi leta 1941 izdelal Konrad Zuse. Sposoben je bil opraviti katerokoli računsko operacijo, na kasnejši razvoj računalnikov pa ni imel nobenega vpliva, saj zanj takrat, ko je nastal, zaradi začetka 2. sv. vojne ni vedel skoraj nihče. Leta 1943 so v Bletchley Parku, kjer je bil sedež zavezniške skupine za odkrivanje kod, v katerih so bila zapisana nemška sporočila, tudi Britanci razvili svoj prvi računalnik, ki se je imenoval *Kolos* (velikan). Razvozlaval je na prvi pogled nesmiselno nemško komunikacijo o npr. premikih njihovih čet in pomembno prispeval k zmagi zaveznikov nad Nemci v 2. sv. vojni. Leta 1948 je v Manchestru nastal prvi povsem elektronski računalnik z imenom *Mark I*, ki je bil tudi prvi, narejen za »masovno« proizvodnjo in komercialno prodajo. Izdelali so jih 9 in jih 9 tudi prodali. Kasneje so prevladujoč položaj pri proizvodnji računalnikov prevzele ZDA, ki so svoj *ENIAC*[12] sicer predstavile že leta 1945, vendar je bil ta v primerjavi z *Markom I* z vidika tega, kako ga je bilo treba programirati za vsako novo nalogo, nočna mora vsakega operaterja. Šlo je za velikansko operacijo ročnega pretikanja kablov iz enih vtičnic v druge, ker je programerjem vzelo dva dni, da so ga pripravili za novo delo. Na drugi strani pa je bil *Mark I* sposoben z vstavljenega preluknjanega papirnega traku, na katerem so bila zapisana navodila, te preprosto skopirati v svoj spomin in takoj začeti z naslednjim opravilom. (Copeland 1993; Bregant 2010)

[11] Electronic Discrete Variable Automatic Computer.
[12] Electronic Numerical Integrator and Computer.

Kaj je torej računalnik? Računalnik je *avtomatski* formalni sistem,[13] ki je sposoben *interpretirati* simbole.[14] Kaj to pomeni? To pomeni, da lahko prepozna »figure«/simbole, s katerimi se igra in na njih skladno z navodili izvaja manipulacije (to vključuje tudi »sodnika«/kontrolorja, ki skrbi za to, da igra poteka po pravilih – določa, kdo je na potezi, s katero figuro naj igra in razglasi rezultat) ter da lahko ugotovi, kaj figure/simboli pomenijo, tj. kaj izražajo ali predstavljajo v vsakdanjem jeziku. Naloga interpretacije, kjer računalnik najprej ugotovi, kaj pomeni vsak enostaven simbol (npr. besede), potem pa, kaj pomeni vsak sestavljen (npr. stavki), je, da simbole iz enega sistema, ki ni razumljiv, »prevede« v drugega, ki je. Velja pravilo, da če računalnik poskrbi za sintakso, tj. da se drži specificiranih formalnih pravil, s tem hkrati poskrbi tudi za semantiko, tj. iznosi/outputi dobijo v standardnem jeziku smisel. »Če stroj sledi znotraj sistema definiranim sintaktičnim pravilom pri tvorjenju novih formul, potem bodo interpretirane formule ohranjale svojo semantično vrednost.« (Markič 1997: 43–44)

Poenostavljeno lahko rečemo, da je računalnik *stroj, ki manipulira s simboli*[15] in je s tega vidika podoben človeku. Tudi ljudje za prikaz stvarnosti uporabljamo besede in stavke, ki so simbolni opis predmetov in dogodkov v našem svetu. To preprosto imenujemo *simbolna hipoteza*, izhaja pa iz tega, da tudi naš duh ni nič več kot univerzalni sistem simbolov, zaradi česar je tudi človeška spoznava zgolj operiranje z njimi. V splošnem pravi, da je mišljenje računanje, ki vključuje manipulacijo s simboli, ta pa je lahko realizirana v sistemih, ki jo izvajajo preko preklapljanja med 0 in 1 kot elementoma binarnega sistema. Pri človeku sta to v svojem prelomnem članku z naslovom »A Logical Calculus of the Ideas Immanent in Nervous Activity«, ki pomeni začetek pristopa k simbolnemu modeliranju mišljenja,[16] ki je zaznamoval prvi val razvoja UI, leta 1943 pokazala nevrofiziolog McCulloch in matematik Pitts. Dokazovala sta, da je edino, kar je pri nevronih pomembno to, da so sproženi ali nesproženi. Prvo lahko označimo z npr. 1 ali »da«, drugo pa z npr. 0 ali »ne«. To, ali so sproženi ali nesproženi, pa je odvisno od tega, ali je dosežen oziroma presežen njihov aktivacijski prag. (Bregant 2016) Naši možgani tako delujejo kot neke vrste preprosti manipulatorji s simboli, saj ni nevron nič drugega kot naprava, ki lahko

[13] Formalni sistem je funkcionalna celota med seboj načrtno povezanih odvisnih elementov, ki deluje na predpisan in ustaljen način.
[14] Obstajajo tudi ročni formalni sistemi, šah je eden izmed njih, kjer nekdo od zunaj – igralec – premika figure in ustvarja nove položaje.
[15] Te simbole imenujemo *biti* – 0 in 1 – in označujejo besede ter števila, zapisani pa so v vrstah, tj. registrih.
[16] Zanimivo je, da pomeni članek hkrati tudi začetek modeliranja nevronskih mrež, ki se običajno bolj povezuje z drugim valom razvoja UI. (Russell in Norvig 2010; Markič 2019)

fizično realizira eno izmed obeh stanj. Ker pa je »zakon 'vse ali nič', kateremu je podvržena živčna aktivnost, zadosten za to, da je lahko aktivnost kateregakoli nevrona predstavljena kot propozicija« (McCulloch in Pitts 1943/1990: 23–24), si lahko nevrone zamislimo kot propozicije, ki so resnične (1) ali neresnične (0), obtežene povezave med njimi pa kot logične veznike in tako v pojmih logike izračunamo vse, kar lahko jezikovno opišemo.

Naenkrat se je zdelo, da je Hobbsova ideja o tem, da je mišljenje računanje, dobila sprejemljivo fizično razlago in odkrila način za izdelavo umetnih sistemov, ki so sposobni posnemati katerikoli vidik inteligence. In ko so se v Hanovru v New Hampshiru v ZDA leta 1956 na tamkajšnjem Dartmouthskem kolidžu med drugim zbrali John MacCarthy,[17] Marvin Minsky, Alan Newell, Herbert Simon, Claud Shannon in Arthur Samuel, ki sta jih zanimala razvoj in izdelava inteligentnega stroja, se je rodila nova disciplina, ki si je nadela ime umetna inteligenca.[18] Srečanje je temeljilo na domnevi, da lahko katerakoli značilnost inteligence v jeziku logike opišemo tako natančno, da jo lahko stroji posnemajo oziroma podvojijo, včasih se ta namera izraža s preprostim vprašanjem »Ali lahko stroj misli?«,[19] zaradi česar jim inteligenco tudi moramo pripisati.

Programi

V tistem času se je utrdilo prepričanje, da je kriterij za pripisovanje inteligence zmožnost logičnega mišljenja, sposobnost, ki bi jo moral imeti tudi stroj, da bi ga lahko imeli za inteligentnega. Najresnejši kandidat za to je bil po prevladujočem mnenju takrat program z imenom *Logični teoretik*, ki so ga leta 1956 razvili Newell, Shaw in Simon, tekel pa je na računalniku z imenom *Johnniac*, ki ga je izdelal von Neumann. Da bi to dosegel, bi moral biti program sposoben dokazati logične teoreme tipa $p \rightarrow (p \vee q)$. (Newell, Shaw, Simon, 1963) Pri tem lahko to storimo s

[17] John McCarthy je bil organizator konference, ki jo je naslovil *The Darthmouth Summer Research Project on Artificial Intelligence*; zadnji besedi sta ostali in dali ime na novo rojenemu področju.

[18] Domnevno se je ideja o umetni inteligenci prvič pojavila v McCulloch in Pitts (1943/1990).

[19] Turing pa ni znan samo kot izumitelj enega izmed abstraktnih računalnikov, ampak tudi kot avtor posebnega preizkusa, ki naj bi pokazal, ali stroj lahko misli. Imenuje se *Turingov test* zasnovan pa je na t. i. igri oponašanja, v kateri v *standardni verziji* preizkusa nastopajo trije akterji, dva človeka in računalnik. Eden izmed ljudi je spraševalec, ki mora ugotoviti, kateri izmed preostalih dveh igralcev, ki odgovarjata na njegova vprašanja, je računalnik. Vsi trije so v ločenih sobah, komunikacija pa poteka preko ekrana in tipkovnice. Računalnik lahko stori karkoli, da bi s svojimi odgovori spraševalca prevaral in izsilil napačno identifikacijo, človek pa mora na vprašanja odgovarjati po resnici. Test se ponovi večkrat, pri čemer se človeški igralci vedno menjajo. Stroj ga opravi, če na koncu spraševalec v več kot 50 % primerov računalnik zamenja za človeka. V tem primeru moramo tudi stroju, če želimo biti dosledni, pripisati mišljenje. (Turing 1950/1990; Bregant, 2016)

pogojnim dokazom, pri katerem je pogojnik dokazan takrat, ko nam uspe izpeljati njegov konsekvent. V tem postopku vedno najprej predpostavimo antecedent pogojnika, v našem slučaju je to *p*, potem pa nadaljujemo z najbolj primerno logično operacijo, da bi čimprej dobili sklep. V našem primeru je to adicija, s katero takoj dobimo *p V q*, s čimer je gornji logični izrek že dokazan. To je *Logičnemu teoretiku* s tem, ko je dokazal prvih 38 teoremov drugega poglavja iz Whiteheadove in Russellove *Principia Mathematica*,[20] ki velja za temeljno delo s področja logike in matematike, tudi uspelo. To je bilo prvič, da stroj ni samo izvajal računskih operacij, ampak dokazoval, zaradi česar gre verjetno za prvi praktični izkaz strojne inteligence nasploh. (Copeland 1993; Bregant 2010)

Po tem preboju na področju razvoja programov, ki so zmožni interpretirati simbole, tj. jih tolmačiti tako, da imajo v vsakdanjem jeziku smisel in na ta način posnemati človeško obnašanje do te mere, da jim lahko pripišemo inteligenco, je v 60. in 70. letih 20. stol. luč sveta ogledalo kar nekaj programov, ki so v večji ali manjši meri izpolnili takratne zahteve za njen pripis. Prvi, ki ga je vredno omeniti, je tudi najbolj znan, in sicer je to *Eliza*, ki je nastal v sredini 60 let 20. stol. na MIT-ju, njegov avtor pa je bil Joseph Weizenbaum. Njena naloga so bili psihoterapevtski pogovori z ljudmi, ki so bili v takšni ali drugačni duševni stiski, ki so potekali preko tipkovnice in ekrana. Z vidika kognitivnih sposobnosti je znala *Eliza* samo eno stvar: sogovornikove odgovore je zgolj ponovila in čakala na odziv ali pa jih spremenila v vprašanja (kar je, mimogrede, klasičen način psihoterapevtskega pogovora). Imela ni nobenih lastnosti, ki se običajno povezujejo s pripisovanjem UI, npr. ni poznala svojega okolja, ni bila sposobna razmišljati in načrtovati dejanj, ni razumela motivov zanje in ni se mogla ničesar naučiti, zaradi česar je ne moremo imeti za inteligenten program v pravem pomenu besede.

Kljub temu je bila tako prepričljiva, da je brez težav prevarala »bolnike«, da je človek. Weizenbaum je bil zgrožen, nikoli si ni predstavljal, da nas lahko tako razmeroma enostaven računalnik tako preslepi. Ljudje so se na *Elizo* celo čustveno navezali, tudi njegova tajnica, ki je bila seznanjena s celotnim projektom, je zahtevala, naj vsi zapustijo sobo, da se bo lahko z njo nemoteno pogovarjala. Zaupali so ji svoje najintimnejše skrivnosti in se niso dali prepričati, da gre zgolj za stroj. »Nisem si mislil, da lahko relativno kratek čas, ki ga normalni ljudje preživijo z relativno preprostim računalnikom, v njih povzroči tako močne fantazije.« (Weizenbaum

[20] Whitehead, Russell (1910–1913/1997).

1976: 7) Še več, nekateri psihiatri so jo bili pripravljeni testirati na svojih pacientih, v *Journal of Nervous and Mental Disease* pa so bili prepričani, da bo program, ko bo zrel za klinično uporabo, predstavljal terapevtsko orodje, s katerim bi lahko v eni uri obravnavali več sto pacientov in tako nadomestili psihiatre v ustanovah za duševno bolne. Z žalostjo je ugotovil, da je naša družba brez predsodkov pripravljena zaupati skrb za blagostanje ljudi računalniku, zaradi česar je postal nasprotnik UI. Menil je, da ni »[njen] cilj ni nič manj kot to, da izdela stroj po vzoru človeka, robota, ki bo odraščal, se učil jezika na isti način kot otrok, spoznaval svet s pomočjo čutil in na koncu razmišljal o vsem znanju, ki ga človeštvo premore.« (Weizenbaum 1976: 202–203) Zato zanj vprašanje ni bilo več, ali lahko naredimo stroj, podoben človeku, ampak, ali to smemo, saj je bil prepričan, da sistemi UI ne bodo nikoli povsem sposobni razumeti in pravilno ovrednotiti položaja, v katerem se lahko znajde človek, zaradi česar niso primerni za opravljanje našega dela. (Copeland 1993; Bregant, 2019)

Naprednejši program, ki naj bi imel vse tiste spoznavne zmožnosti, ki so *Elizi* manjkale, je v 70. letih 20. stol. na MIT-ju razvil Terry Winograd, imenoval pa se je *Shrdlu*. S pomočjo »roke« in ukazov, kaj naj naredi, je na mizi premikal predmete različnih oblik, barv in velikosti, pri čemer je za opravljanje zahtevane naloge razvil in izpeljal lasten načrt. Še več, v ozkem specifičnem okolju, ki je bilo natančno določeno, je znal logično sklepati. Npr. če določimo, da so predmeti, ki so naši, piramide in tisti, ki niso beli, potem pa ga vprašamo, ali je v belem kvadru, ki vsebuje belo piramido in modro kocko, kakšen predmet, ki je naš, *Shrdlu* pravilno odgovori, da je to modra kocka. Ker mu tako lahko pripišemo opravljanje kognitivnih operacij (sicer zgolj v nespreminjajočem se okolju), je bil prvi program, ki je bil sposoben izpolniti omejen pogoj za pripis inteligence: »razumel« je navodila, s pomočjo sklepanja je našel odgovore na kompleksna vprašanja, »razumel« pa je tudi, vsaj deloma, motive zanje. (Copeland 1993; Bregant 2019)

Na koncu naj omenimo še mogoče najbolj napreden program iz tistega časa, ki je nastal v sredini 70. let 20. stol. na MIT-ju, z imenom *Hacker*. Njegov avtor je bil Gerald Sussman, cilj pa razbiti mit o tem, da se računalniki nikoli ne bodo mogli sami programirati. *Hacker* ima to, do sedaj še nevideno zmožnost, in sicer sposoben je razviti programe za računalnik, na katerem teče tudi sam. Programi, ki jih piše, v dobro določenem in ozko omejenem okolju nadzirajo »roko«, ki na mizi premika s črkami označene kocke. Recimo (Copeland 1993), da mora *Hacker* razviti program, tj. napisati postopek izvedbe, ne pa to zgolj izvesti, kar je delal npr. *Shrdlu*, ki bo

omogočal, da kocko A, ki leži pod kocko B, postavimo nad kocko C. Najprej pogleda v svojo knjižnico, da bi našel karkoli relevantnega za to nalogo, ali kakšen program, ki ga je dobil od svojega avtorja, ali kakšen program, ki ga je že sam napisal. Edino, kar najde, je npr. vzorec *postaviti na*, kar »roki« omogoča, da en predmet postavi ali na drugega ali na mizo. Dalje vidi, da bi lahko ta vzorec uporabil za to, da bi položil kocko A na kocko C, če kocka A ne bi imela ničesar nad sabo. Potem v svoji knjižnici išče dalje in najde npr. vzorec *spustiti na*, kar »roki« omogoča, da en predmet spusti ali na drugega ali na mizo. Sedaj lahko zahtevano nalogo opravi: najprej kocko B spusti na mizo, potem pa kocko A postavi na kocko C.

Bistvenega pomena za to, da je pri programiranju uspešen, je njegova knjižnica tehnik programiranja, ki je v resnici shramba dejstev in receptov, pa tudi trikov in ukan, znanih hekerjem, o tem, kako napišemo program. Poleg tega se je sposoben učiti iz izkušenj, vsak neuspešen poskus programiranja je vir informacij o tem, česa se ne dela. Na ta način se s prakso izboljšuje, vse, kar se novega nauči, pa shrani v svojo knjižnico tehnik programiranja, ki jo je dobil od svojega avtorja, in jo tako povečuje. Res je, da je bil kontekst, v katerem je *Hacker* deloval, izmišljen in zanesljiv, ter da je bil sposoben napisati zgolj najbolj enostavne programe, res pa je tudi, da je bil s tem razbit mit o računalnikih, ki sami tega niso zmožni narediti. In naj je bilo do zapletenih in naprednih programov še tako daleč, ideja o tem, da bodo stroji programiranje nekoč vzeli v svoje roke, ni bila več neutemeljena. (Copeland 1993; Bregant 2019)

Ali smo računalniki?

Eno izmed bolj provokativnih filozofskih vprašanj, ki izhaja iz opisa računalnika kot fizičnega sistema, ki s pomočjo programa, implementiranega v strojnem jeziku, realizira operacije s simboli, je, ali nismo tudi ljudje zgolj (neke vrste) računalniki. O tem, da smo ljudje stroji, je pisal že La Mettrie v svojem *Strojnem človeku* leta 1748: »Človek je stroj, zgrajen na takšen način, da si je to nemogoče predstavljati in ga zato tudi nemogoče definirati.«[21] (La Metrrie 1748/1996: 5) Danes imamo na voljo boljšo predstavo o tem, kakšen stroj naj bi človek bil, vse skupaj pa se je začelo v sredini 19. stol., ko je bilo odkrito, da je delovanje živčnega sistema v resnici sprejemanje in prevajanje električnih impulzov, kasneje pa do podrobnosti raziskane kemične

[21] Pri tem ne pozabi omeniti, da je o tem govoril že Descartes v *Razpravi o metodi* leta 1637: »Nihče ne zanika, da je ta slavni filozof naredil mnogo napak, je pa razumel živalsko naravo in bil prvi, ki je dovršeno pokazal, da so živali stroji. (La Mettrie 1748/1996: 35)

značilnosti nevronov in njihov električni potencial. Že La Mettrie je predpostavljal, da je »posebna lastnost našega stroja, da vsako njegovo vlakno, tudi najmanjše, oscilira. To naravno nihanje je kot ura, ki včasih zamre in se mora potem obnoviti. Če postane šibko, se mora okrepiti, če pa premočno, oslabiti.« (La Mettrie 1748/1996: 31) Vendar je minilo še kar nekaj časa, da smo ugotovili, kaj je namen vzajemnega proženja nevronov, kar nam je pomagalo razumeti delovanje naših možganov. To se je zgodilo z omenjenim člankom McCullocha in Pittsa približno sto let pozneje, ki je pokazal »kako lahko operacije nevronov in njihove povezave z ostalimi nevroni modeliramo s pojmi logike.« (Markič 2010: 30). To pomeni, da si lahko nevrone zamislimo kot propozicije, ki so resnične (1) ali neresnične (0), obtežene povezave med njimi pa kot logične veznike in tako v pojmih logike izračunamo vse, kar lahko v jeziku opišemo.

Tudi računalniku t. i. »logična vrata« omogočajo, da podobno kot nevron po istem principu preklaplja med 1 in 0. Recimo, če v računalnik preko tipkovnice vnesem 'mačka' (vhod ali vnos), računalnik preklaplja med 0 in 1, odvisno od kode, ki jo vsaka črka ima, dokler ne pride do binarnega zapisa, npr. 1000011 1000001 1000111 1001111 1000001, ki pomeni izpis 'mačka' na mojem ekranu (izhod ali iznos). Ker tako tudi računalnik ni nič drugega kot naprava, ki lahko fizično realizira eno od obeh stanj – manipulator s simboli – do njegovega enačenja z našimi možgani na osnovi opisanega delovanja nevrona ni bilo več daleč. Kljub temu, da so pri računalnikih manipulacije s simboli realizirane z bistveno drugačnimi fizični operacijami kot pri nas, to ni pomembno, saj enačenje računalnika in možganov temelji na ideji o *večvrstni realizaciji simbolov*. Vsebuje domnevo, da so lahko simboli realizirani na različne načine, in sicer z barvo, ogljem, svinčnikom, elektromagnetnimi stikali (releji), elektronkami, silikonskimi polprevodniki (tranzistorji) itd. (Bregant 2016), in se skriva v naslednjem Turingovem citatu: »Pogosto se misli, da je pomembno, da so sodobni digitalni računalniki električni in da je tudi živčni sistem električen. /…/ ker so vsi digitalni računalniki v nekem smislu enakovredni, vidimo, da uporaba elektrike z vidika teorije ni pomembna.« [22] (Turing 1950: 439)

[22] Za izdelavo *Analitičnega stroja* je Babbage predvidel ključno mehanske dela, tj. kolesa in zobnike.

Eden izmed zagovornikov enačenja računalnikov in možganov je Pylyshyn:

> Semantična vsebina [prepričanj in namer] je kodirana s strani možganskih lastnosti na isti splošni način, na katerega so semantične vsebine računalniških predstav kodirane s strani fizično uprimerjanih simbolnih struktur. (Pylyshyn 1984: 258)

Naše mentalne operacije so v bistvu manipulacije s simbolnimi reprezentacijami, ki so zgolj podobne običajnim stavkom kateregakoli govorjenega jezika: »/…/ [simbolne reprezentacije] so simbolni izrazi v notranjem, fizično uprimerjanem simbolnem sistemu, ki se včasih imenuje 'mentalese' ali 'jezik misli'.«[23] (Pylyshyn 1984: 194) Ali je manipulacija s simboli kot podobnost med računalniki in možgani dovolj za enačenje ljudi s stroji oziroma ali ne obstaja dovolj razlik, ki to postavljajo pod vprašaj, pa je treba še raziskati.

Vnet nasprotnik enačenja računalnikov z nami pa je Searle, ki je leta 1980 v svojem članku *Duhovi, možgani in programi* predstavil znani miselni eksperiment z imenom »kitajska soba«, s katerim je hotel za vse večne čase že v kali zatreti tudi načelni poskus pripisovanja inteligence kakršnemukoli stroju. Predstavljajte si Janeza, ki ne obvlada nobenega drugega jezika razen slovenščine, zaprtega v sobi z odprtino, skozi katero dobi tri zvezke besedila v kitajščini, *skripte* (scenarije za različne situacije), *zgodbo* in *vprašanja*, poleg tega pa še v slovenščini formalna pravila za povezovanje kitajskih pismenk, ki mu omogočajo, da jih lahko spozna izključno po njihovih oblikah. Zunaj sobe se od njega pričakuje, da odgovori na vprašanja, odgovore pa posreduje skozi za to pripravljeno odprtino. Zamislite si, da se Searle v rokovanju s simboli sčasoma tako izuri, da se njegovi odgovori v kitajščini ne razlikujejo od odgovorov rojenih Kitajcev. Vprašanje je sedaj, ali Janez, ki deluje kot računalniški program, tj. izvaja operacije na formalno opredeljenih elementih oziroma proizvaja kitajske odgovore tako, da uporablja neraztolmačene formalne simbole, obvlada kitajsko? Searlov odgovor je kategoričen *ne*, saj je ključna razlika med stroji in ljudmi v tem, da ljudje *razumemo* (v tem primeru zgodbo), stroji pa ne.

[23] Za jezik misli glej tudi Fodor (1975), Markič (2010).

Skratka, Searlov miselni eksperiment pokaže, da obvladovanje sintakse, tj. poznavanje niza pravil za simbolno manipulacijo, še ne pomeni obvladovanja semantike, tj. poznavanje tega, kar simboli dejansko pomenijo. To pa je v nasprotju z enačenjem računalnikov in možganov, saj bi morali biti stroji, da bi bili podobni ljudem (in obratno), sposobni razumeti formalne simbole, s katerimi manipulirajo, kot to velja za nas, kar pa očitno ni mogoče. »/…/ v dobesednem pomenu programirani računalnik razume toliko, kot razumeta avto in seštevalni stroj, to pa je natanko nič. Računalnikovo razumevanje ni niti delno niti nepopolno (kot je moje razumevanje nemščine); je povsem nično.« (Searle 1980/1990: 366)

To, da Searle »kitajske sobe« ni nikoli zapisal v obliki argumenta, je še najmanj, da pa na naslednji ugovor (če malo poenostavimo) zgolj zamahne z roko, pa je že zaskrbljujoče. Namreč, kako vemo, da ljudje to, o čemer se z njimi pogovarjamo, sploh razumejo? Zdi se, da razen njihovih smiselnih odgovorov na vprašanja, ki se pojavijo, ni na voljo nobenega drugega dokaza za to. Searle sicer ta ugovor povzame takole: »Kako pa veste, da drugi ljudje zares razumejo kitajsko ali karkoli drugega? Samo po njihovem vedenju. Računalnik lahko ravno tako uspešno opravi vedenjske teste kot oni (v načelu); če torej pripisujemo drugim ljudem spoznavnost, jo morate načeloma pripisati tudi računalnikom.« (Searle 1980/1990: 373) Če pa to drži, bi morali tudi računalnik, ki se je v danem okolju sposoben obnašati na ustrezen način, bodisi verbalno bodisi fizično obravnavati kot bitje ali sistem, ki to, kar je vzrok za njegovo takšno ali drugačno delovanje, razume. V nasprotnem primeru pri pripisovanju duševnosti uporabljamo različna merila, zaradi česar smo lahko upravičeno obtoženi t. i. *šovinizma vrst*, ki (v tem primeru) ljudi obravnava drugače kot stroje, pa čeprav za to ni nobenega razloga.

Kakorkoli, po tem, lahko bi rekli zlatem obdobju UI, ko je, kot smo videli, raziskovalcem s pomočjo simulacije človeškega obnašanja, ki je temeljilo na transparentni simbolni manipulaciji, uspelo razviti modele UI, ki so bili (do neke mere in po nekih standardih) inteligentni, kar je vse navdajalo z optimističnimi pričakovanji in napovedmi, pa je sledilo obdobje zatona. Programi, kot sta bila *Shrdlu* in *Hacker*, ki so temeljili na deduktivni formalni logiki in delovanju v dobro definiranem nespreminjajočem se okolju, niso bili uporabni za reševanje kompleksnih nalog iz vsakdanjega negotovega sveta. To bi moralo temeljiti na induktivnem (statističnem) sklepanju in obdelavi velike količine podatkov, kar pa od sistemov UI zahteva veliko računsko moč, česar pa takrat še ni bilo na voljo.

4 2. val razvoja UI: nesimbolno modeliranje

Posnemanje inteligence

O ponovnem vzponu UI lahko govorimo na začetku 80., o njegovem zagonu pa v 90. letih 20. stol.,[24] ko se je začela povečevati računska moč strojev, količina informacij, ki je bila s prihodom spleta v obtoku in uporaba računalnikov, ki je omogočala relativno enostaven dostop do njih. Gre za drugi val razvoja UI, ki pa po novem temelji na drugačni predstavitvi podatkov, in sicer na indukciji (posplošitev, sklepanje po analogiji, vzročno sklepanje ali sklepanje na najboljšo pojasnitev), ki stavi na predhodne izkušnje in interakcijo z okoljem ter dogodke napoveduje z višjo ali nižjo verjetnostjo, pogosto s pomočjo statistike, pri čemer dopušča (manjšo) možnost, da se motimo. Programi za manipuliranje s simboli, ki so obvladovali prvi val, so v drugem postali algoritmi za strojno učenje, ki pa se je najprej zgledovalo po dejanskih nevronskih mrežah (prvi model sta leta 1943 izdelala že omenjena McCulloch in Pitts, ki tako veljata za začetnika obeh valov), tj. po možganskih procesih, ki so odgovorni za učenje pri človeku in se od tega (s prihodom konekcionizma) oddaljilo šele kasneje.[25]

Kakorkoli, omenjena Dartmouthska konferenca (in danes področje UI nasploh) je slonela na prepričanju udeležencev, da lahko katerokoli značilnost človeške inteligence opišemo tako natančno, da jo lahko stroj posnema. Vzor za to, kako simulacija inteligentnega obnašanja poteka, najdemo sicer že pri Aristotelu, ki je s svojimi *kategorijami*, tj. različnimi vrstami ali načini bivanja, postavil temelje. Njegov cilj je bil sicer zgraditi model sveta, ki bi vključeval njegove različne vidike, njihove značilnosti in odnose med njimi. Vprašanja, na katera je moral odgovoriti, da bi pri tem uspel, so bila, »kaj obstaja«, »kakšne lastnosti imajo stvari« in »v kakšnem odnosu so med seboj«. Za odgovor na njih je razvil sistem 10 kategorij, ki upodabljajo svet: a. substanca/bitnost (npr. človek, miza, računalnik), b. kvantiteta/kolikost (npr. štiri noge), c. kvaliteta/kakšnost (npr. bel, visok, debel) d. relacija/odnos (npr. večji,), e. mesto v prostoru/kje (npr. v šoli, na trgu, v avtu), f. časovnost/kdaj (npr. včeraj), g. imeti (npr. 7 let, brata, 3 avte), h. pozicija/položaj (npr. sedi), i. delovanje/akcija

[24] Iz tega obdobja so verjetno najbolj znani šahovski programi, ki so bili kmalu sposobni premagovati šahovske velemojstre. Šahovski program je prvič premagal svetovnega prvaka v sredini 90. let 20. stol. Takrat je *Globoki modrini* (*Deep Blue*) s 3,5:2,5 v šestih partijah premoč moral priznati Gari Kasparov.

[25] Danes sicer za izdelovanje učinkovitih algoritmov še vedno uporabljamo izraz nevronske mreže, ampak sodobni modeli UI, kot je npr. varovanje pred nezaželeno elektronsko pošto, ne delujejo več po principih delovanja našega živčnega sistema. (Markič 2019)

(npr. gori, seka, tepe), j. trpeti/pasivnost (npr. nadlegovan). (Aristotel 2004; Bregant, 2019) V grafičnem smislu gre v bistvu za mrežo z vozlišči in povezavami, semantična pa se imenuje zato, ker je sestavljena iz pojmov ter relacij med njimi in ker prikazuje v kakšni medsebojni zvezi so stvari in dejanja ter kako vzajemno delujejo. Ker so semantične mreže v resnici na simbolni način predstavljeno znanje, ki sistemu omogoča, da s pomočjo deduktivnega sklepanja, ki poleg aplikacije podatkov vključuje tudi uporabo pravil, izpeljuje nove zaključke o svetu (in se tako *uči*), v tem primeru govorimo o simbolnem modeliranju inteligence oziroma strojnem učenju s pomočjo simbolne manipulacije, tipičnem za prvi val razvoja UI.

V drugem valu razvoja UI pa se izkaže, da to ni edina možnost in da obstaja posnemanje inteligentnega obnašanja, ki ne vključuje na simbolni način predstavljenih informacij, njen vzor pa je delovanje dejanskega nevrona. V tem primeru govorimo o t. i. *nevronskih mrežah*, za njihovo učinkovito obratovanje pa je bistveno naslednje: (a) da je vsak nevron povezan z drugimi, (b) da so te povezave različno obtežene (močne), (c) da obstaja prag, ki določa, ali je nevron aktiviran ali ne in (d) da je tako »izključen« (0) ali »vključen« (1). Skratka, tudi biološki nevroni tvorijo prepleteno mrežo povezav, ki so različno močne, kar je deloma odvisno od premera vlaken, ki so povezana, deloma pa od kemične zgradbe stika (sinaps). Ko vsota vhodnih signalov določenega nevrona doseže ali preseže njegov aktivacijski prag, se nevron sproži in posreduje izhodni signal naprej svojemu najbližjemu sosedu. Podobno so danes zgrajene tudi umetne nevronske mreže, ki predstavljajo orodje nesimbolno predstavljenega znanja, pri čemer so njihovi osnovni gradniki t. i. idealizirani nevroni, tj. preproste, neinteligentne enote, ki so lahko vklopljene ali izklopljene. Takšni umetni nevroni so med seboj povezani, vsak izmed njih pa ima določeno aktivacijsko vrednost, ki jo preko vezi posreduje drugim enotam in s tem pripomore k povečanju (ekscitiranje) ali zmanjšanju (inhibiranje) njihove aktivacijske vrednosti, kar vpliva na to, ali se sprožijo ali ne. (Bregant 2016: 97–98) Ker so nevronske mreže v bistvu na nesimbolni način predstavljeno znanje, ki sistemu dopušča, da iz velike količine podatkov s pomočjo statistične verjetnosti, induktivnega sklepanja in predhodnih izkušenj izlušči ponavljajoče se vzorce (in se tako *uči*), tj. usvoji znanje, ki mu zagotavlja bolj ali manj uspešno napovedovanje dogodkov v našem stalno spreminjajočem se svetu, v tem primeru govorimo o nesimbolnem modeliranju inteligence oziroma strojnem učenju s pomočjo učnih primerkov.

Razprava o tem, ali je simulacija inteligentnega obnašanja dovolj za to, da umetnemu sistemu, ki ga je sposoben realizirati, pripišemo inteligenco, še vedno teče. Zagovorniki menijo, da je pomemben zgolj rezultat in če je rešitev nekega kognitivnega problema pravilna in merljiva ter tako prepričljiva kot pri človeku, ni nobenega razloga za to, da bi stroju inteligenco odrekli. To zelo razširjeno stališče imenujemo *šibka umetna inteligenca*. Nasprotniki pa mislijo, da je ključna izdelava in da lahko govorimo o inteligentnem stroju v pravem pomenu besede samo takrat, ko ima ta enako kot človek tudi fenomenalno zavest,[26] npr. čustva, vizualno izkustvo ali občutke, s čimer bi mu bil z duhovnega vidika enakopraven. To bolj skrajno stališče pa imenujemo *močna umetna inteligenca*.[27]

Strojno učenje

Strojno učenje temelji na prepoznavanju, upoštevanju in sortiranju bistvenih značilnosti predmetov (običajno se pravi, da gre za značilnosti, ki imajo največjo pojasnjevalno moč), tj. tistih lastnosti, ki jih ločijo od drugih stvari istega roda (v definiciji podane v vrstni razliki) oziroma tiste, zaradi katerih nekaj je to, kar je. V resnici gre za nekaj, kar nam je dobro znano, in sicer učenje iz primerov. Njegov cilj je izpeljava posplošitev o predmetih (iz njihovih znanih primerkov), s pomočjo katerih je sistem te predmete kasneje v svetu sposoben prepoznati in razvrstiti brez kakršnekoli tuje pomoči. Da bi bilo takšno urjenje uspešno, mora vključevati veliko količino podatkov oziroma učnih primerkov, iz katerih je razviden vzorec in za katere še ne obstaja nobena formula, po kateri bi sistem predmete glede na dano značilnost med seboj že lahko uspešno razlikoval.

Ko takšno podatkovno bazo zagotovimo, pa imamo na voljo tri osnovne postopke strojnega učenja, *klasifikacijo*, *grupiranje*[28] in *regresijo*.[29] Pri klasifikaciji gre za sistematično umeščanje primerkov v že znane razrede. Slednji so med seboj jasno ločeni in rabijo kot orodje, s katerim podatke uredimo. Razvrščanje v razrede poteka glede na dane značilnosti, razredi pa imajo svoja imena, t. i. *oznake*.[30] V podatkovnem nizu slik sadja je npr. ena oznaka limona, druga marelica in tretja jagoda, sistem pa z njihovo pomočjo glede na določene lastnosti, ki jih predmeti imajo, sadje razvršča.

[26] Zavest, ki vključuje mentalna stanja, individualizirana na osnovi tega, »kako je biti« (angl. *what it's like*).
[27] Ne bomo se spuščali v to, kateri pristop je pravilen, šibkejši ali močnejši, dejstvo je, da lahko kognitivne sposobnosti oziroma inteligentno obnašanje z različnimi metodami in modeli uspešno posnemamo že danes.
[28] Angl. *clustering*.
[29] Angl. *regression*.
[30] Angl. *labels*.

Tudi pri grupiranju gre za sortiranje elementov v razrede, ampak s to razliko, da njihov imena pred začetkom postopka še niso znana. Tukaj algoritem sam opravi to nalogo, tj. najprej določeno količino primerkov razporedi po podobnih ali skladnih značilnostih, potem pa za vsako takšno množico ustvari oznako. Rezultat so skupine podobnih elementov, ki jih lahko razumemo tudi kot kategorije, ki se med seboj razlikujejo glede na svoje tipične lastnosti. Pri regresiji pa gre za iskanje matematične zveze med dvema značilnostma. Ugotoviti želimo, ali ena lastnost vpliva na drugo (ciljno lastnost) in ali lahko potem s pomočjo prve napovemo vrednost druge, npr. ali zaposlitev na banki vpliva na plače njenih zaposlenih (ciljna značilnost) in ali lahko iz tega, da nekdo dela na banki napovemo, koliko zasluži. Ciljne značilnosti ne želimo v celoti pojasniti, ampak zgolj ugotoviti, kaj (če sploh) nanjo vpliva in kako močno.« (Dengel 2019a; Bregant 2019)

Glede na to, čemu je sistem namenjen, pa je dalje odvisno, katero od treh vrst strojnega učenja bomo izbrali. Prvo vrsto imenujemo *nadzorovano učenje*.[31] Zanj je značilno, da so v fazi urjenja vsi podatki, ki jih sistem dobi, opremljeni tudi s pravilnimi odgovori, tj. oznakami, npr. to je mačka, to je pes, to je konj. To mu omogoča, da popravlja napake in da lahko na koncu iz vseh dobljenih informacij izpelje splošni model, ki ga potem uporablja za npr. razvrščanje živali. Druga vrsta nosi naziv *nenadzorovano učenje*.[32] Tukaj podatki, ki jih sistem dobi v procesu urjenja, nimajo nobenih dodatnih oznak, iz podobnih ali istih značilnosti primerkov mora sam ustvariti skupine, ki jih imenujemo *gruče*.[33] Z drugimi besedami, algoritem vhodne podatke razdeli v več kategorij s tipičnimi značilnostmi, njihovo število in vrste pa iz dobljenih informacij izlušči sam brez nadzora učitelja. Tretji vrsti pa pravimo *vzpodbujevalno učenje*.[34] Zanj je značilno, da dobi sistem v fazi urjenja le občasno povratno informacijo o tem, kako uspešen je pri opravljanju svoje naloge. V bistvu sam razvija strategije za reševanje problemov ali opravljanje nalog, nagrada oziroma vzpodbuda v smislu pozitivne ali negativne povratne informacije pa mu omogoča, da se v bodoče izogne napakam. Tako je sposoben bolje oceniti, ali njegovo ravnanje v neki situaciji vodi k uspehu ali neuspehu, kar mu olajša izdelavo učinkovitejših načrtov za spopad s težavami, hkrati pa predstavlja tudi motivacijo za še boljše dosežke. (Dengel 2019a; Bregant 2019)

[31] Angl. *supervised learning*.
[32] Angl. *unsupervised learning*.
[33] Angl. *clusters*.
[34] Angl. *reinforcement learning*.

Danes je verjetno najbolj znan in uporabljen model UI, ki temelji na nesimbolnem predstavljanju znanja *Googlov prevajalnik*.[35] Na voljo je že več kot 15 let, pred časom pa je presedlal na t. i. *mrežno prevajanje*, katerega ključna prednost je, da ne prevaja posameznih besed, ampak cele stavke. Ker s tem do neke mere upošteva tudi kontekst, takšen sistem prevod lažje preuredi in prilagodi tako, da je podoben človeškemu. Pri odkrivanju pomena stavkov si pomaga z upoštevanjem okoliščin, v katerih so zapisani ali izrečeni, zaradi česar njegovi prevodi postajajo vse bolj naravni. Program prevaja neposredno iz enega jezika v drugega, s čimer se izogne vmesni postaji, ki povečuje verjetnost napak v končnem izdelku, kar uporabniško izkušnjo še izboljša. Upoštevanje širšega konteksta, spoštovanje semantike stavkov in neposredno prevajanje so vplivali na izboljšanje vrstnega reda besed v prevodu, kar je povečalo njegovo razumljivost. Tega stari program običajno ni bil zmožen zagotoviti, zaradi česar je bil pogosto deležen posmeha.

Od letos podpira nekaj čez 130 jezikov, sposoben pa je celo prevajati v jezik, ki ga ne pozna, tj. ni bil del njegovega urjenja,[36] če sta si oba jezika, tisti, iz katerega se prevaja in tisti, v katerega se prevaja, dovolj blizu. Njegovi prevodi so človeškim presenetljivo podobni z vidika smiselnosti, dolžine in strukture stavkov. Ni pa tako zanesljiv pri razumevanju besed, ki imajo več pomenov, kar v prevodih pogosto vodi do nesmislov, občutljiv je na slovnične napake, kakovost prevodov pa je odvisna tudi od kompleksnosti in razširjenosti jezika. Tako so s tega vidika v prednosti vplivni evropski jeziki (angleščina, nemščina, francoščina itd.), močno pa zaostajajo afriški. (Bregant 2019)

S stališča razširjenosti pa prednjačijo sistemi UI, ki nam olajšajo opravljanje tistega vsakodnevnega dela, ki za marsikoga pomeni tratenje časa. Gre za *osebne asistente*, izmed katerih so najbolj znani *Siri* (Apple), *Googlov asistent*[37] in *Alexa* (Amazon). Namesto nas lahko preko zvočnih ukazov opravijo goro nalog, če so povezani z napravami, ki njihovo delovanje podpirajo: igrajo željeno glasbo, naročajo hrano iz restavracije ali izdelke po spletu, nam berejo naša elektronska sporočila, načrtujejo urnike itd. V resnici gre za virtualne pomočnike, ki se nahajajo v »oblaku«, s katerimi se pogovarjamo v naravnem jeziku: ko postavimo vprašanje, ti zvočne valove

[35] *Google Translate*.
[36] Uspešno reševanje problemov ali odgovarjanje na vprašanja, ki jih UI v fazi učenja še ni srečala, v angl. imenujemo *Zero-Shot-Learning*.
[37] *Google Assistant*.

spremenijo v besedilo, kar jim omogoča, da zberejo potrebne informacije iz tistih virov, ki so za izvršitev zahtevane naloge relevantni.

Še več, komunikacija, ki jo obvladajo, je dvosmerna, na naše zahteve so sposobni tudi odgovoriti in to ne samo z nepristnim robotskim glasom, ampak glasom, za katerega bi lahko dali roko v ogenj, da je »človeški«. *Google Duplex*, ki je dodatek h Googlovemu asistentu in je izurjen za obdelavo naravnega jezika, je program, katerega glas je tako pristen, da iz izseka telefonskega pogovora med njimi in človekom, ne moremo ugotoviti niti, da je eden izmed sogovornikov UI, niti, kdo to je. Trenutno so njegove jezikovne naloge, če malo ironiziramo, omejene zgolj na rezervacijo mize v restavraciji ali termina pri frizerju ter posredovanje odpiralnih časov, ni pa daleč čas, ko bo dovolj napreden, da bo sposoben opravljati tudi bolj kompleksna dela.[38] (Kremp 2018; Bregant 2019)

Omenimo še mogoče najbolj razvpite modele UI, ki zadnjih nekaj let polnijo časopisne stolpce, in sicer *avtonomna vozila*. Nobena skrivnost ni, da so že nekaj časa na cesti avtomobili, ki imajo določeno stopnjo samostojnosti: npr. voznika opozorijo na nepričakovano menjavo voznega pasu (nadzor menjave/zapustitve voznega pasu), pri nizkih hitrostih v mestu sami zavirajo, da preprečijo nalet, če ocenijo, da bo voznik reagiral prepozno (sistem pomoči za zaviranje v sili), samodejno nadzirajo in prilagajajo hitrost glede na pred njimi vozeče avtomobile in celo potek cest (prilagodljiv/predvidljiv tempomat) itd. Vse to omogoča množica algoritmov, ki so se sposobni učiti, pa čeprav vožnje, okolja ali avtomobila ne razumejo tako kot mi.

V primeru povsem avtonomnega vozila pa gre za avtomobil, ki je sposoben zaznavati okolje in nas brez naše pomoči pripeljati s točke A na točko B, pri čemer se npr. sam odloči, kam zaviti, s kakšno hitrostjo peljati in kako sploh priti do cilja. Od potnika se ne v nobeni situaciji ne zahteva, da prevzame nadzor nad vozilom oz. da je v njem sploh prisoten. Takšni avtomobili danes že delujejo z visoko stopnjo zanesljivosti, kljub temu pa prihaja do občasnih napak, ki lahko ogrozijo človeško življenje. Verjetno najbolj znan takšen primer se je pred časom zgodil podjetju *Uber*, ki je testiralo samovozeči avtomobil, ta pa je pri tem s hitrostjo 70 km/h zbil kolesarko, ki je nepravilno prečkala cesto. Čeprav je vozilo možnost trka zaznalo že slabih 6 sekund pred njim, se algoritem za nadzor vožnje ni odločil za zaviranje.

[38] Program je sicer deležen utemeljenih očitkov, da s tem, ko snema naše vzorce obnašanja in preference, ki jih imamo, z namenom analize in priprave odgovora na naše zahteve, kar mu omogoča učenje in prehod na višjo, bolj inteligentno stopnjo, preveč posega v našo zasebnost.

Dejstvo je, da bi se v takšni situaciji človek odzval drugače: ali bi npr. zaviral in kolesarko pustil, da prečka cesto, ali pa se ji izognil na pločnik ali bankino. Preiskava je pozneje odkrila, da je do zbitja kolesarke prišlo, ker je nadzorni sistem sploh ni prepoznal kot človeka in da avtonomna vozila podjetja *Uber* sploh niso bila sprogramirana tako, da bi reagirala na nepravilno prečkanje ceste. (Bregant 2019)

Samo vprašanje časa je, kdaj nas bo UI poznala tako dobro, da bo na naše vprašanje, »Kam naj gremo na dopust?«, izbrala destinacijo, rezervirala nastanitev, organizirala aktivnosti in to vse v skladu z našimi zahtevami, preferencami in nagnjenji, medtem ko bomo mi po prihodu iz službe utrujeni počivali na kavču. Ali smo na takšno prihodnost pripravljeni in kakšna je sploh cena, ki bi jo morali za to plačati?

Izguba zasebnosti

Omenili bomo zgolj odpoved nečemu,[39] kar se ponuja samo od sebe in kar do določene mere, ne da bi se tega sploh prav zavedali ali da bi nas to posebno motilo, tako ali tako že počnemo. Gre za izgubo zasebnosti, nekaj, kar naj bi nam bilo z evropsko *Splošno uredbo o varovanju osebnih podatkov* (SUVP)[40] na formalni ravni zagotovljeno.

Kakorkoli, danes so algoritmi, ki pod pretvezo večje varnosti pomagajo represivnim organom pri preprečevanju kriminala in iskanju krivcev, del našega vsakdanjika ne glede na to, da namestitev nadzornih kamer, prepoznavanje obrazov s slik, ki jih naredijo, in hitra identifikacija oseb, ki so na njih, mobilni telefoni, ki stalno sporočajo lokacijo imetnika, spletni iskalniki, ki beležijo zgodovino obiskanih strani, ponudniki spletnih nakupov, ki shranjujejo naša naročila z namenom ugotavljanja naših preferenc in ponujanja podobnega blaga v prihodnosti itd., bistveno zmanjšujejo našo zasebnost in odpirajo Pandorino skrinjico zlorab. Ker tako nekdo vedno ve, kaj delamo, lahko iz tega rekonstruira našo dnevno rutino in odstopanje od nje: kje smo doma, kje smo v službi, kdaj gremo v službo in kdaj se vrnemo iz nje, kateri so naši hobiji, kaj so naši interesi itd. »Zdi se, da bi se lahko v bližnji prihodnosti uresničila nočna mora vseh tistih, ki že dolgo opozarjajo na to, da bo razvoj novih tehnologij v slogu 1984, kjer te veliki brat opazuje, ljudi povsem prikrajšal za svobodo.« (Bregant 2019: 10)

[39] Nekateri izmed ostalih moralnih problemov, ki izhajajo iz nepreračunljive rabe modelov UI, bodo analizirani v preostalih člankih tega zbornika.
[40] Angl. *General Data Protection Regulation* (GDPR).

Uporabniki na družbenih omrežjih na različne načine delijo svoje izkušnje z drugimi, tj. z besedami, glasbo, sliko, filmi, všečki itd. izražajo svoja mnenja o dogodkih v svetu, vsak izmed njih pa predstavlja primer odpovedi zasebnosti. S kritiziranjem, zavračanjem, sprejemanjem, polemiziranjem ipd., puščajo na spletu sledi in prostodušno kažejo, kaj so njihove preference, pričakovanja in prioritete, s čimer vplivajo na svoje zasebno in javno življenje. Kajti pet velikih tehnoloških podjetij *Google*, *Apple*, *Facebook*, *Amazon* in *Microsoft* (GAFAM) nikoli ne spi. Mnenja ljudi, izražena na takšne načine, UI odkriva, povezuje in interpretira, kar v splošnem imenujemo *rudarjenje podatkov*.[41] Tukaj gre v bistvu za pomoč pri identificiranju skritih vzorcev in odstopanj v njih ter ocenjevanje tega, ali so med seboj povezani: npr. združevanje kupcev z istim okusom, opozarjanje na anomalije v proizvodnem postopku, ki kažejo na napake v delovanju sistema, iskanje zveze med vremenom in količino pridelka ali industrializacijo ter podnebnimi spremembami itd.

Vse skupaj se spremeni, ko se rudarjenje podatkov,[42] izrodi v njihovo zbiranje informacij o posamezniku, uporabnik sodobnih modelov UI pa postane izdelek. Pri tem vlada *Google*, ki je s svojimi brezplačnimi storitvami od nas želel le eno, predajo naših osebnih podatkov. Ker se pri tem nismo obotavljali, s čimer smo mu omogočili neprestano črpanje informacij (*Gmail*, *Chrome*, *YouTube*, *Maps* brez predaha polnijo njegove podatkovne zbirke), danes ve, kaj zanima večino uporabnikov spleta.[43] Težava je v tem, da takšna skoncentriranost podatkov tehnološkim družbam omogoča, da s selekcijo podatkov o svetu vplivajo na to, kaj se v njem dogaja in kako ga dojemamo, zaradi česar pride do upravljanja družbe s pomočjo UI: kam bo družba zavila postane odvisno od potreb, interesov, prioritet in preferenc upravljalca, tj. *GAFAMA*. V dobi digitalnih tehnologij smo tako postali surovina, iz katere tehnološki giganti naredijo izdelek: kaj bomo kupili, koga bomo volili, kaj bomo oblekli, kam bomo šli, kaj bomo delali ipd. Ti podatki se prodajajo naprej tudi z namenom spreminjanja naših navad, običajev in želja, s čimer izgubimo pravico do

[41] Angl. *Data-Mining*.

[42] Za boljši občutek glede tega, o kakšni količini podatkov, ki je na voljo omenjenim družbam, govorimo, si oglejmo, koliko informacij je bilo na svetu znotraj različnih omrežij leta 2019 v obtoku v 1 minuti: 18 milijonov poslanih SMS sporočil, 4.3 milijona ogledov video vsebin na *YouTubu*, 481.000 poslanih čivkov, 187 milijonov poslanih elektronskih sporočil, 3.7 milijona iskanj z *Googlom*, 973.000 vpisov v *Facebook*, za 862.823 ameriških dolarjev opravljenih nakupov, 375.000 naloženih aplikacij, 67 nameščenih virtualnih asistentov itd. (Dengel 2019c; Bregant, 2020)

[43] »Tudi ostali tehnološki velikani niso nobena izjema: *Microsoft* se ukvarja s ciljnimi oglasi, tehnologijo prepoznavanja obraza, virtualno resničnostjo itd., *Facebook* preko oglasov in všečkov zbira informacije o imenih in naslovih ljudi, odnosih med njimi, njihovih družinah, lokaciji in pogovorih itd. ter je tako lastnik največje podatkovne baze o nas, *Amazon* pa preko spletne trgovine, v kateri je mogoče kupiti tako rekoč vse in *Alexi*, ki govori iz zvočnika *Echo*, podatke pa pošilja v oblak, pozna naše želje, potrebe in interese, da o naslovu dostave, kreditnih karticah in mestu nakupa niti ne govorimo.« (Bregant 2020: 8)

prihodnosti, svobode in zasebnosti. Takšno usmerjanje naših življenjskih navad, skladno z interesi tehnoloških družb, pa imenujemo *nadzorovalni kapitalizem*.[44] [45] (Grobelnik 2018; Masten 2019; Zuboff 2019)

Kaj se nam torej obeta v prihodnosti? Zaenkrat so modeli UI izdelani tako, da lahko opravljajo le eno nalogo, pa še to zgolj na nekem ozkem področju, zaradi česar še ne moremo govoriti o neki splošni UI, primerljivi s človeško. Pričakujemo lahko, da bo šel nadaljnji razvoj UI v smeri sistemov, ki bodo zmožni hkrati izvrševati različne operacije z različnih specializiranih področij (kot človek), na kakšen način se bodo do takšnega znanja dokopali, pa bomo videli.

5 Zaključek

Dartmouthski posvet je temeljil na prepričanju, da je možno katerokoli značilnost človeške inteligence opisati tako natančno, da jo lahko stroj posnema. V duhu te ideje je zakoličil pristop, ki je postal sinonim za prvi val razvoja UI, in sicer simbolno manipulacijo. Ta temelji na deduktivni formalni logiki in delovanju v dobro definiranem nespreminjajočem se okolju ter na ta način UI zagotavlja varen kontekst, znotraj katerega do neke mere in po nekih standardih realizira inteligentno obnašanje. V tem primeru govorimo o na simbolni način predstavljenem znanju, ki sistemu omogoča, da s pomočjo deduktivnega sklepanja, ki vključuje aplikacijo podatkov in uporabo pravil, izpeljuje nove sklepe o svetu in se tako uči. Kmalu se je izkazalo, da takšni programi, niso uporabni za reševanje kompleksnih nalog iz vsakdanjega negotovega sveta, ki zahteva več od izvrševanja enostavnih operacij znotraj ozko določenega specializiranega okolja. To je omogočil šele drugi val razvoja UI, ki pa v nasprotju s prvim temelji na drugačni predstavitvi znanja, ki se zgleduje po delovanju biološkega nevrona in omogoča nesimbolni pristop k modeliranju duha tudi v spreminjajočem se okolju. Vključuje induktivno sklepanje in interakcijo z okoljem ter upošteva pretekle izkušnje in statistične zakonitosti, iz česar s pomočjo verjetnosti in učnih primerkov napoveduje prihodnje dogodke. Govorimo o nevronskih mrežah, kjer gre za na nesimbolni način predstavljeno

[44] Angl. *surveillence capitalism*.

[45] »V praksi se je to pokazalo pri zlorabi osebnih podatkov podjetja *Cambridge Analytica*, ki je z aplikacijo, ponujeno *Facebooku*, s prošnjo sodelovanja v akademski raziskavi, nezakonito pridobilo osebne podatke 50 milijonov uporabnikov Facebooka (njegov dizajn je omogočal tudi to, ne samo zbiranje odgovorov) in iz njih izdelalo njihove psihografične profile, s katerimi so ugotavljali, kakšno propagando je treba uporabiti, da bodo glasovali po naročnikovih željah.« (Bregant 2020: 9)

znanje, ki sistemu dopušča, da realizira inteligentno obnašanje tako, da iz obdelave velike količine podatkov izlušči ponavljajoče se vzorce in se tako uči.

Kakorkoli, če si izposodimo Searlovo misel o tem, da je za obvladovanje jezika bistveno njegovo razumevanje, se zdi, da smo od govorjenja o inteligentnih strojih ne glede na to, da so nekatere specifične naloge ti že danes sposobni opraviti bolje od nas, še vedno precej oddaljeni. Kljub napredku na področju *obdelave naravnega jezika*,[46] ki se kaže v tem, da novodobne aplikacije v svoji tudi glasovni dvosmerni komunikaciji poleg poznavanja dejstev, do neke mere upoštevajo tudi kontekst, stroji temu, kako si inteligentno obnašanje predstavljamo, niti slučajno še niso blizu. To namreč vključuje tudi prepoznavanje uporabljenega pomena in načina, kako je nekaj povedano. Npr. ali je izjava »ta stavek vsebuje eno napako« resnična ali neresnična? Na prvi pogled se zdi, da je neresnična, saj ne vsebuje nobene pravopisne napake, lahko pa je tudi resnična, saj se napaka skriva v pomenu, ki predpostavlja nekaj, česar ni. Zdi se torej, da bi morali znati računalniki, če bi hoteli biti inteligentni v pravem pomenu besede, obdelovati naravni jezik na opisan način, da o zahtevi po sposobnosti izvrševati več opravil z različnih tudi bistveno drugačnih področij, kar smo že večkrat omenili, niti ne govorimo. Ker tako daleč še nismo, tudi ideja o prihodu *superinteligence*, umetnega sistema z inteligenco, ki bistveno prekaša inteligenco najbolj pametnih in najbolj nadarjenih ljudi, ostaja bolj kot ne futurističen konstrukt.

Viri in literatura

Aristotel (2004). *Kategorije*. Ljubljana: ZRC SAZU.

Berkeley, I. S. N. (2018). »A Computational Conundrum: »What is a Computer? A Historical Overview«. *Minds & Machines*, 28, str. 375–383.

Bregant, J. (2010). »Ali lahko stroj misli?«. *Analiza*, 4, str. 55–72.

Bregant, J. (2016). »Možgani v primežu računalnikov«. *Analiza*, 1, str. 87–114.

Bregant, J. (2019). »Umetna inteligenca v praksi (1. del): razvoj, obnašanje in učenje strojev«. *Analiza*, 2, str. 39–55.

Bregant, J. (2020). »Umetna inteligenca v praksi (2. del): nekaj etičnih pomislekov«. *Analiza*, 1, str. 5–20.

Cantwell Smith, B. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge, MA, London: The MIT Press.

Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.

Dengel, A. (2019a). »Maschinelles Lernen – das Gehirn als Vorbild für künstliche Neuronale Netze«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S. (urd.), *Künstliche Intelligenz: Die Zukunft von Mensch und Maschine*. Hamburg: ZEIT Akademie Gmbh, str. 23–32.

[46] Angl. *natural language processing*.

Dengel, A. (2019b). »Künstliche Intelligenz – Eine Einführung«. V Dengel, A., Socher, R., Kirchner E., A., Ogolla, S. (urd.), *Künstliche Intelligenz: Die Zukunft von Mensch und Maschine*. Hamburg: ZEIT Akademie Gmbh, str. 13–22.

Dengel, A. (2019c). »Multimedia-Data-Mining: Trends und Emotionen in Big Data erkennen«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S., Künstliche Intelligenz: Die Zukunft von Mensch und Maschine. Hamburg: ZEIT Akademie Gmbh, str. 74–84.

Descartes, R. (1637/2007). *Razprava o metodi*. Ljubljana: Slovenska matica.

Descartes, R. (1641/1988). *Meditacije*. Ljubljana: Slovenska Matica.

Descartes, R. (1641/1985a). *Meditations*. V Cottingham, J., Stoothoff, R. in Murdoch, D. (urd.), *The Philosophical Writtings of Descartes*, Cambridge: Cambridge University Press.

Descartes, R. (1649/1985b). *The Passions of the Soul*. V Cottingham, J., Stoothoff, R., Murdoch, D. (urd.), *The Philosophical Writtings of Descartes*, Cambridge: Cambridge University Press.

Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Grobelnik, M. (2018). »Podatki so nova nafta. Kdor ima dostop do podatkov, lahko rešuje probleme«. *Mladina*, 39.

Haugeland, J. (1986). *Artificial Intelligence: The Very Idea*. Cambridge: The MIT Press.

Hobbes, T. (1651/2006). *Leviathan (Revised Student Edition)*. Cambridge: Cambridge University Press.

Kim, J. (2001). *Mind in a Physical World*. Cambridge: MIT Press.

Kremp, M. (2018). »Google Duplex ist gruselig gut«. *Spiegel Online* (27. november 2019). URL= https://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html.

Leibniz, G. W. (1679/1969). »On the General Characteristic«. V Loemker, L. E. (urd.), *Philosophical Papers and Letters*. Dordrecht: Springer.

Leibniz, G. W. (1666/1966). »On The Art of Combination«. V Parkinson, G. H. R, (urd.), *Leibniz: Logical Papers*. Oxford: Oxford University Press.

Markič, O. (1997). »Klasična kognitivna znanost in simbolni modeli«. *Analiza 1*, 1999, str. 38–52.

Markič, O. (2010). *Kognitivna znanost*. Maribor: Aristej.

Markič, O. (2019). »Prvi in drugi val umetne intelgience«. V Malec, M., Markič, O. (urd.*) Misli svetlobe in senc: razprave o filozofskem delu Marka Uršiča*. Lubljana: UL, str. 201–211.

Masten, A. (2019). »Kaj vse pomeni klik na 'Strinjam se': O ekonomiji podatkov in monopolih«. *MMC RTV SLO* (2. december 2019). URL = https://www.rtvslo.si/mmc-podrobno/na-pragu-digitalne-diktature-brez-zasebnosti-in-brez-svobode/489378.

McCulloch, W., Pitts, W. (1943/1990). »A Logical Calculus of the Ideas Immanent in Nervous Activity«. V Boden, M. A. (ur.), *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press, str. 22–39.

Mettrie, Julien Offray de (1748/1996). *Machine Man and Other Writings*. Cambridge: Cambridge University Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens democracy*. Crown: New York.

Russell, S., Norvig, P. (2010). *Artificial Intelligence: A Modern Approach (3rd. ed.)*. Upper Saddle River: Prentice Hall.

Searle, J. (1980/1990). »Duhovi, možgani in programi«. V Hofstadter, D. R. in Dennet, D. C. (urd.), *Oko duha*, Ljubljana: Mladinska knjiga, str. 361–379.

Turing, A. (1950/1990). »Stroji, ki računajo, in inteligenca«. V Hofstadter, D. R., Dennet, D. C. (ur.), *Oko duha*, Ljubljana: Mladinska knjiga.

Turing, A. (1936). »On Computable Numbers, with an Application to the *Entscheidungsproblem*«. *Proceedings of the London Mathematical Society*, 42, str. 230–265.

Weizenbaum, J. (1976). *Computer Power and Human Reason*. San Francisco: W. H. Freeman.

Whitehead, A. N., Russell, B. (1910–1913/1997). *Principia Mathematica*. Cambridge: Cambridge University Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. New York: PublicAffairs.