

Janez Bregant

Filozofska fakulteta Univerze v Mariboru

Umetna inteligenca v praksi (1. del): razvoj, obnašanje in učenje strojev

Ideja o simulaciji inteligentnega obnašanja iz 60. in 70. let 20. stoletja, ki naj bi bila dovolj za govorjenje o pametnih strojih, je še danes aktualna. Sicer je v ospredju t.i. ne-simbolni pristop, ki temelji na urjenju umetnih nevronske mreže, pri čemer se le-te s pomočjo prepoznavanja tipičnih značilnosti in njihovega klasificiranja ter svojo sposobnostjo popravljanja napak učijo, pa vendarle. Iz tega se je že razvil niz izjemno zmogljivih modelov UI, ki postopoma brišejo mejo med naravno in umetno inteligenco ter sprožajo številna moralna vprašanja. V članku najprej analiziramo pojem 'inteligence' v pojmih zaznavanja, spominjanja in učenja, potem na kratko opišemo razvoj UI, pri čemer izpostavimo nekatere ključne dogodke v zgodovini in najbolj znane danes, na koncu pa znotraj simbolnega in ne-simbolnega pristopa k simulaciji inteligentnega obnašanja opišemo strojno učenje, tj. pod katerimi pogoji poteka postopek urjenja (treniranja) umetnega sistema.

Ključne besede: umetna inteligenca, šibka UI, močna UI, strojno učenje, obdelava naravnega jezika.

0. Uvod

Kaj je umetna inteligenca (UI)? Vprašanje, na katerega ni enotnega odgovora, posebej danes, ko so pametni telefoni, takšni in drugačni (osebni) računalniki, pametne ure itd. del našega vsakdanjika. UI je danes tako močno vpeta v naše okolje, da je postala samoumevna, zaradi česar nanjo ne gledamo več kot na nekaj nenaravnega, ampak kot na polnokrven del v osnovi fizične narave sveta, v katerem živimo. Meja med človeško in umetno inteligenco je zaradi tega vedno bolj zabrisana, saj na razliko med človekom, ki misli oziroma je zavestno bitje in strojem, ki to ni, nenačrtno pozabljamo. John McCarthy, eden izmed očetov UI je nekoč dejal: »Kakor hitro deluje, je nihče več ne imenuje UI.« (Dengel, 2019a) Z drugimi besedami, neprestana uporaba npr. spletnih iskalnikov, navigacijskih naprav, mobilnih aplikacij itd. je prerasla v običajno prakso, ki UI omogoča, da se skriva v ozadju in ustvarja vtis, da sploh ni umetna in da je z nami že od nekdaj.

Slednje seveda ne drži, rodila se je po 2. sv. vojni, ko so se pojavili prvi računalniki za splošno uporabo, javnost jih je imenovala »elektronski možgani«, na katerih so tekli programi, ki so bili zelo dobri pri igranju salonskih iger, kot je npr. dama in ki so v tistih časih šteje kot preizkusni kamen inteligentnosti kakršnegakoli sistema. Formalno se je disciplina rodila leta 1956 na slavni Dartmouthski konferenci, njen cilj pa ni bil zgolj posnemati mišljenja, ampak narediti stroj, ki resnično misli. Marvin Minsky jo je opisal kot »znanost o izdelavi strojev, ki so sposobni narediti stvari, za katere je po naših merilih potreben um« (Copeland, 1993: 1) oziroma kot »znanost o tem, kako narediti in/ali programirati računalnike, da bodo sposobni istih stvari kot um«. (Boden, 1990: 1)

Povečanje zanimanja za UI v zadnjem času, ko ne mine dan, da se v javnosti ne bi pojavila kakšna novica v zvezi z njo, pa gre pripisati predvsem naslednjim trem faktorjem, ki so v največji meri odgovorni za njen nagli razvoj:

- a. povečana računska moč strojev (računalnikov) (procesorji zmorejo danes v 1 sek. opraviti 2×10^{15} , tj. 2 bilijardi,¹ elementarnih računskih operacij, kot sta seštevanje in množenje, zapisanih v obliki, ki jih računalnik razume, t.i. operacij s plavajočo vejico;²
- b. dostopnost do ogromnega števila (digitalnih) podatkov kot posledice razvoja znanosti in tehnologije, gospodarske dejavnosti in aktivnosti na družbenih omrežjih (s pomočjo t.i. *spleta stvari*³ je med seboj povezano ogromno število naprav, ki na ta način proizvedejo izjemno količino podatkov, ocenjuje se, da na dan do 2,4 trilijona);⁴
- c. novi izdelki, ki temeljijo na obdelavi omenjenih podatkov in zaradi katerih se pojavljajo nova pravna in etična vprašanja (proizvodnja trotoev,⁵ avtonomnih vozil, virtualnih asistentov itd.). (Dengel, 2019a)

S člankom se pridružujemo omenjenemu trendu: v prvem poglavju predstavimo model inteligentnega ravnanja, ki je na splošni ravni združljiv s tem, kako deluje UI, v drugem skiciramo razvoj UI, ki nas je od premikanja kock v omejenem okolju, npr. na mizi, pripeljal do samodejne vožnje avtomobilov v naravi, v tretjem poglavju pa ponazorimo razliko med t.i. šibko UI, ki temelji na posnemanju inteligentnega obnašanja in močno UI, ki si prizadeva za izdelavo pametnega umetnega bitja po vzoru človeka, opišemo simbolni in ne-simbolni pristop k simulaciji

¹ 1 bilijarda je število s 15 ničlami.

² Angl. *floating-point-operations*.

³ Gre za mrežno povezavo fizičnih in virtualnih predmetov, ki s pomočjo informacijskih in komunikacijskih tehnik omogoča njihovo skupno delovanje.

⁴ 1 trilijon je število z 18 ničlami.

⁵ Angl. *drone*.

inteligentnega obnašanja ter zaključimo s prikazom strojnega učenja, natančneje, pogojev, pod katerimi poteka postopek urjenja (treniranja) umetnega sistema.

1. Inteligenca kot spiralni proces

Dejstvo je, da ni soglasja o tem, kaj 'inteligenca' sploh pomeni. Ljudje imamo o tem povsem različne predstave, za enega je inteligenčen nekdo, ki dobro in hitro računa ter tako izkazuje nadarjenost za matematiko, za drugega nekdo, ki si hitro zapomni besedilo, se ga dobro spomni in zna ideje iz njega prenesti v prakso, za tretjega nekdo, ki je sposoben prepoznati sogovornikova čustva, jih razumeti in se ustrezno odzvati itd., za četrtega pa nekdo, ki se je sposoben usklajevati z drugimi, da bi od tega v okviru danih možnosti vsi imeli koristi. Vidimo, da smo v resnici opisali štiri povsem različne vrste sposobnosti (lahko pa bi jih še več), vse pa razumemo kot lastnosti, ki jih v večjem ali manjšem obsegu (ni nujno niti, da vse) imajo bitja, ki so inteligenčna. Lahko bi govorili o neke vrste matematični inteligenci, jezikovni inteligenci, čustveni inteligenci, socialni inteligenci itd. Ker je inteligenca tako širok pojem, ne preseneča niti, da ni njene enotne definicije, niti takrat, ko se išče nekaj splošnejšega, nekaj, kar bi na višjem nivoju vključevalo vse, za kar mislimo, da bi inteligenčna oseba morala imeti. Pomeni lahko, (i) kako se znajdemo v novih situacijah in kako se jim znamo prilagoditi, (ii) kako znamo razmišljati in reševati probleme, (iii) kako se znamo učiti iz knjig in iz izkušenj, (iv) kako znamo opazovati okolje in razumeti dogodke itd. Psihologi kot definicijo inteligence ponujajo naslednje:

- a. sposobnost posameznika, da se obnaša skladno z namenom, razmišlja racionalno in učinkovito obvladuje svoje okolje (Wechsler, 1944);
- b. rezultat procesa pridobivanja, hranjenja, spominjanja, kombiniranja, primerjanja in uporabe informacij ter konceptualnih sposobnosti v novem okolju (Humphreys, 1979);
- c. k cilju usmerjeno prilagojeno obnašanje (Sternberg, Salter: 1982);
- d. sposobnost obvladovanja kognitivne kompleksnosti (Gottfredson, 1998) itd.

Oče preučevanja inteligence William Stern jo je definiral tako: inteligenca je sposobnost posameznika, da svoje mišljenje zavestno prilagodi novih zahtevam; je splošna sposobnost prilagoditve duha na nove naloge in življenjske pogoje. (Stern, 1912) Zakaj ga posebej navajamo? Ker je šel celo tako daleč, da je razvil test za njeno merjenje, pri čemer je kot ključne faktorje inteligentnega obnašanja med drugim poudaril ravno tisto, kar smo na začetku navedli tudi sami: matematične sposobnosti, komunikativnost, jezikovne sposobnosti, zaznava, motorične sposobnosti in socialna razvitost. Še več, omenjeni vidiki so ključni pri še eni tipični lastnosti inteligentnega bitja, *učenju*: ker so v stalni interakciji z vzorci na-

šega obnašanja, ki se nahajajo v možganih, npr. da se ob prihodu pozdravi, da se ob poslušanju državne himne vstane, da se na pogrebu ne smeji itd., nas včasih prisili, da te vzorce ravnanja dopolnimo, spremenimo ali zamenjamo.

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, kar lahko ponazorimo kot *spiralni proces*. Ta se pri človeški inteligenci začne s sprejemanje *podatkov*, ki se vrši preko vida, sluha, tipa, vonja in okusa. Iz podatkov s pomočjo interpretacije dobimo *informacije*, iz njih pa, ko jih postavimo v nek kontekst, *teoretično znanje*.⁶ Npr. številka 1984 je zgolj podatek, ko ugotovimo, da gre za naslov Orwellove knjige, postane informacija, ko pa jo postavimo v kontekst vseprisotnega državnega nadzora, aktualnega tudi danes v obliki izgubljanja zasebnosti pod pretvezo varnosti, postane znanje. Govorimo lahko o transformaciji podatkov v znanje, ki poteka po naslednjih etapah: podatek → informacija → teoretično znanje. To tvori naše *razumevanje*.

To pa še ni dovolj, to znanje je treba prenesti tudi v prakso, tj. uporabiti ga je treba pri odločanju o tem, kako bomo v neki situaciji ravnali. Prvi korak v tej smeri je, da se iz njega s pomočjo naših dejanj najprej razvije *ustrezno obnašanje*, iz njega pa, ko se le-to z vajo utrdi, *sposobnost*. Če jo pogosto uporabljamo, potem v tretjem koraku s pomočjo izkušenj, ki jih tako dobimo, postanemo strokovnjaki na nekem področju, tj. razvijemo *strokovno znanje*.⁷ Npr. ker vemo, da je naša zasebnost ogrožena, bomo pri brskanju po spletu bolj previdni, to obnašanje pa se z vajo spremeni v sposobnost, da se nenačrtovani izgubi zasebnosti vedno izognemo. Ko se to dovolj pogosto ponovi, nabereimo izkušnje, zaradi katerih postanemo strokovnjaki za varovanje zasebnosti na spletu. Govorimo lahko o transformaciji ustreznega obnašanja v znanje, ki poteka po naslednjih etapah: *ustrezno obnašanje* → *sposobnost* → *strokovno znanje*. To nam v novih situacijah omogoča *pravilno ravnanje*. Še več, omogoča nam tudi *predvidevanje*, kateri podatki so na področju, kjer smo strokovnjaki, pomembni, s čimer se mentalni koncepti, ki so naloženi v naših možganih, širijo, kar omogoča človeku, da se v novih situacijah ustrezno odzove. (Dengel, 2019a)

⁶ Nem. *das Geisteswissen*.

⁷ Nem. *das Fachwissen*.

2. Razvoj UI: od *Elize* do *Alexe*

Ne moremo zanikati, da nista zaznava in inteligenca ljudi navduševali že od nekdaj, ko pa so se temu po 2. sv. vojni pridružili še računalniki, ni bilo nobene ovire več, da se ne bi rodila nova disciplina, ki si je leta 1956 v Hanovru v New Hampshiru v ZDA nadela ime umetna inteligenca.⁸

Dartmouthski kolidž je takrat gostil znanstvenike, ki jih je zanimala računalniška inteligenca, med drugim Johna MacCarthyja,⁹ Marvinina Minskyja, Alana Newella, Herberta Simona, Clauda Shannona in Arthura Samuela, ki so razmišljali o tem, kako bi lahko s pomočjo programov proizvedli inteligentne stroje. Res je, da so že takrat razpravljali o npr. računalnikih, nevronskih mrežah, računski teoriji, abstrakciji, kreativnosti itd., nečem, kar je še danes aktualno, kljub temu pa konferenca v številnih drugih predvsem socialnih vidikih ni bila prav uspešna: ljudje so na srečanja prihajali in odhajali sporadično, ni bilo pravega pretoka idej, med znanstveniki je vladalo tekmovanje (Newell in Simon sta s predstavitvijo svojega napredka pri izdelavi inteligentnega programa zasenčila vse ostale) itd. Kljub temu pa je treba omeniti vsaj en pomemben dosežek Dartmouthskega srečanja, prej osamljeni entuziasti so bili naenkrat povezani v znanstveno skupnost z koliko toliko jasnimi raziskovalnimi cilji in močno identiteto (Copeland, 1993: 8–9), ki so pojem 'inteligenca' razumeli kot izvajanje umskih/spoznavnih operacij, pojem 'umetna' pa kot, da jih ne izvaja človek, ampak stroj/sistem (tj. nekaj, kar ni rezultat narave, ampak človekovega ali strojevega dela).^{10 11} Umetno inteligenco bi potem lahko v takšnem splošnem smislu definirali kot *sistem, ki ni plod dela narave, sposoben pa je izvajati kognitivne operacije, kot so zaznavanje, sklepanje, mišljenje, reševanje problemov, načrtovanje, odločanje, učenje, primerjanje, ocenjevanje ipd.*¹²

⁸ Domnevno se je ideja o umetni inteligenci prvič pojavila v McCulloch in Pitts (1943/1990).

⁹ John McCarthy je bil organizator konference, ki jo je naslovil *The Dartmouth Summer Research Project on Artificial Intelligence*; zadnji besedi sta ostali in dali ime na novo rojenemu področju.

¹⁰ Skratka, sklep konference je bil najti načine, kako lahko stroj uporabi jezik, ustvari abstrakcije in koncepte, reši probleme in se sam izboljšuje.

¹¹ Nekateri teoretiki predlagajo drugačno definicijo umetne inteligence, razumejo jo kot »*znanost o inteligenci na sploh /.../*« (Boden, 1990: 1), katere cilj je pojasniti ali celo reproducirati različne zmožnosti duha in ugotoviti, ali lahko inteligenca nastane zgolj v sistemih, ki imajo podobno zgradbo kot možgani ali pa jo je možno realizirati tudi na kakšen drug način.

¹² Kot zanimivost naj omenimo, da je bil McCarthy prepričan, da se lahko glede teh tipično človeških značilnosti pri strojih doseže viden napredek že, če skrbno izbrana skupina strokovnjakov na njih skupaj dela eno poletje, kar je bila tudi njegova motivacija za organizacijo srečanja.

V 50. letih 20. stol. se je utrdilo prepričanje, da je kriterij za pripisovanje inteligence zmožnost logičnega mišljenja, ki se v praksi še danes pogosto preverja z igranjem salonskih iger, kot so dama,¹³ šah ali go. Ideja je bila, da bi moral to sposobnost imeti tudi stroj, če bi želel veljati za pametnega. Temelj za to je leta 1956 postavil John von Neumann s tem, ko je razvil računalnik¹⁴ – hardwer – z imenom *Johnniac*, na katerega so potem Newell, Shaw in Simon namestili program – softwer – z imenom *Logični teoretik*. Ta kombinacija je takrat po mnenju mnogih predstavljala prvega resnega kandidata za to, da bi lahko govorili o inteligentnem stroju. Newell in Simon teh ambicij nista skrivala niti na Dartmouthski konferenci, kamor sta bila povabljeni bolj kot ne drugi violini, kjer pa sta z dosežki *Logičnega teoretika* zasenčila vse ostale udeležence. Ko mu je uspelo dokazati prvih 38 teoremov 2. poglavja iz Whiteheadove in Russellove *Principia Mathematica*,¹⁵ ki velja za temeljno delo s področja logike in matematike, je bila zvezda rojena. Prvič se je zgodilo, da stroj ni zgolj računal, ampak tudi dokazoval, zaradi česar si je *Logični teoretik* upravičeno prislužil naziv prvega inteligentnega programa. (Copeland, 1993: 7, Bregant, 2010: 63–64)

Sledilo je zlato obdobje umetne inteligence, ki je trajalo nekje do leta 1974, v njem pa so raziskovalci postavili temelje (računalniškega) sklepanja, razvili prve nevronske mreže, zasnovane po vzoru bioloških in z vidika simuliranja inteligence dosegli prve uspehe. Verjetno najbolj znan program iz tega obdobja je *Eliza*, ki je nastal v sredini 60 let 20. stoletja na MIT-ju, njegov avtor pa je bil Joseph Weizenbaum. Njena »služba« so bili psihoterapevtski pogovori z ljudmi, ki so bili v takšni ali drugačni duševni stiski. Potekali so preko tipkovnice in ekrana, kar je bil za tisti čas običajen medij komunikacije z zunanjim okoljem, program pa je bil tako prepričljiv, da je z lahkoto prepričal ljudi, da je tudi sam človek in tako so ga »bolniki« tudi obravnavali.

Ko je videl, kaj je ustvaril, se je Weizenbaum zamislil, nikoli si ni namreč predstavljal, da nas lahko razmeroma enostaven računalnik tako prevara. Še več, ljudje so se na *Elizo* celo čustveno navezali, odkrivali so ji svoje najintimnejše skrivnosti in se niso dali prepričati, da gre zgolj za stroj. Celotna njegova tajnica, ki je po-

¹³ Najbolj znan in uspešen program za igranje salonskih iger iz tistega časa je »program za damo« (angl. »the checkers program«), ki ga je že na začetku 50. let 20. stol. razvil Arthur Samuel, deloval pa je na IBM-ovem računalniku 701 iz tistega časa.

¹⁴ Prvi sprogramiran računalnik je sicer v Nemčiji leta 1941 izdelal že Konrad Zuse. Bil je sposoben 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. Malo pozneje, leta 1943, so tudi Britanci razvili svoj prvi računalnik, ki se je imenoval *Kolos* (velikan), namenjen pa je bil dekodiranju nemških sporočil, program zanj – pravila, po katerih je informacije, tj. črke in besede, na smiselni način pretvoril v vsakdanji jezik – pa je priskrbel takrat v Bletchley Parku delujoča skupina dešifradorjev. Po 2. sv. vojni so Britanci v Manchesteru leta 1948 na noge postavili sprogramiran računalnik z imenom *Mark I*, ki za svoje delovanje ni uporabljal več relejev, ampak elektrone, danes bi rekli tranzistorje – polprevodnike – in je bil prvi elektronski računalnik, ki je kadarkoli ugledal luč sveta. (Copeland, 1993: 1–5, Bregant, 2010: 63)

¹⁵ Whitehead, Russell (1910–1913/1997).

znala cel projekt, je vztrajala, naj vsi zapustijo sobo, da se bo lahko z njo na samem pogovarjala. Weizenbaumova groza pa je še kar rasla, saj so jo nekateri psihoterapevti bili celo pripravljeni testirati na svojih pacientih. Prepričani so bili, da bo program, ko bo zrel za klinično uporabo, predstavljal terapevtsko orodje, ki bo lahko nadomestilo psihologe tam, kjer jih primanjkuje. Z bridkostjo je ugotovil, da je naša družba brez predsodkov pripravljena zaupati skrb za blagostanje ljudi računalniku, zaradi česar je postal nasprotnik UI. Zanj vprašanje ni bilo več, ali lahko naredimo stroje, podobne človeku, ampak, ali to smemo, menil je namreč, da UI ne bo nikoli povsem sposobna razumeti in pravilno ovrednotiti položaja, v katerem se lahko znajde človek.

Kakorkoli, če ocenjujemo *Elizo* z vidika njenih kognitivnih sposobnosti vidimo, da je znala v bistvu samo eno stvar: sogovornikove odgovore je zgolj ponovila ali pa jih spremenila v vprašanja (kar je, mimogrede, klasičen način /posrednega/ psihoterapevtskega pogovora). Imela ni nobenih zmožnosti, ki se običajno povezujejo z UI: ni poznala svojega okolja, ni bila sposobna razmišljati in načrtovati dejanj, ni razumela motivov zanje in ni se mogla ničesar naučiti. Kljub temu je dolgo časa stala na hodniku MIT-jevega laboratorija za UI in neobremenjeno klepetala z gostujočimi raziskovalci in novinarji o njihovih srčnih težavah. (Copeland, 1993: 13–15)

Naprednejši je program, ki se je imenoval *Shrdlu*. Nastal je v 70. letih 20. stol. prav tako na MIT-ju, razvil pa ga je Terry Winograd. S pomočjo robotske roke je na mizi premikal predmete različnih oblik, barv in velikosti in je bil sploh prvi program, ki mu lahko pripišemo omejeno opravljanje kognitivnimi operacij. Ko je dobil navodila, naj na določen način uredi predmete na mizi, je program za to razvil in izpeljal lasten načrt. 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, kar očitno kaže na njegove sklepalne zmožnosti. Tako je bil prvi program, ki je bil zmožen izpolniti določene pogoje za pripis inteligence: razumel je navodila, s pomočjo sklepanja je našel odgovore na zapletena vprašanja, razumel pa je celo, vsaj deloma, motive zanje. (Copeland, 1993: 15–18)

Omeniti velja še program z imenom *Hacker*, ki je v sredini 70. let 20. stoletja prav tako nastal na MIT-ju, njegov avtor pa je bil Gerald Sussman. Njegov cilj je bil razbiti mit o računalnikih, ki nikoli ne bodo mogli napisati programa. Šlo je za podobno okolje kot pri *Shrdlu*, tudi tukaj je robotska roka na mizi premikala predmete. Toda, *Hacker* je imel do sedaj še ne videno lastnost, pisal je programe za računalnike, na katerih je bil sam nameščen. Ključnega pomena za to, da je ta program lahko napisal program, je bil njegova knjižnica tehnik programiranja, ki ni bila nič drugega kot shramba dejstev o tem, kako napišemo program, ki jo je dobil od svojega programerja. Poleg tega se je podobno kot človeški hekerji, učil

iz izkušenj, vsak neuspešen program je bil vir informacij o tem, česar se ne dela, zaradi česar se je s prakso izboljševal, prav tako pa je s tem povečeval svojo knjižnico tehnik programiranja. Drži, da je bil kontekst, v katerem je *Hacker* deloval, izmišljen in omejen, ter da je bil sposoben napisati zgolj najbolj enostavne programe, drži pa tudi, da je bil s tem razbit mit o računalnikih, ki sami tega ne morejo narediti. In naj je bilo do zapletenih in naprednih programov, ki so nastali na takšen način, še tako daleč, ideja o tem, da bodo stroji programiranje nekoč vzeli v svoje roke, že takrat ni bila iz trte izvita. (Copeland, 1993: 18–21)

Sledilo je obdobje zatona, saj prvotni algoritmi, ki so delovali na enostavnih problemih in se nanašali na premikanje kock, iskanje poti v labirintu ali orientacijo v prostoru, niso bili uporabni za reševanje problemov, s katerimi se srečujemo v vsakdanjem življenju. Metode njihovega delovanja, ki so temeljile zgolj na sklepanju, niso bile primerne za reševanje kompleksnejših problemov, metode, ki so temeljile na statističnih zakonitostih, pa so zahtevale veliko podatkov in računske moči, česar pa takrat še ni bilo na voljo. Ponovni vzpon UI se je začel v začetku 90. let 20. stoletja, ko so se pojavili šahovski programi, ki so bili kmalu sposobni premagovati šahovske vele mojstre. Da je šahovski program prvič premagal svetovnega prvaka se je zgodilo malo kasneje, ko je v sredini desetletja s 3,5:2,5 v šestih partijah premoč *Globoki modrini* moral priznati Gari Kasparov. Sposoben je bil preračunati 200 milijonov potez v sekundi, vnaprej pa jih je videl več kot 20.¹⁶ Danes sicer pri razvoju šahovskih programov ni več poudarek na hardveru (sposobnosti računanja), najmočnejši šahovski programi na svetu (*Rybka* ali *Fritz*) tečejo na navadnih prenosnih računalnikih z Intelovimi dvojedrnimi procesorji, ampak na softveru (sposobnosti razumevanja).

Trenutno pa so najbolj razvpiti modeli UI verjetno *avtonomna vozila*. Nobena skrivnost ni, da so že nekaj časa na cesti avtomobili, ki imajo določeno stopnjo samostojnosti: npr. voznika ne samo opozorijo na nepredvideno menjavo voznega pasa, ampak avtomobil tudi usmerijo nazaj na pravega (nadzor menjava/zapustitve voznega pasu), pri nizkih hitrostih v mestu sami zavirajo, da preprečijo nalet, če ocenijo, da je/bo voznik to spregledal (sistem pomoči za zaviranje v sili), samodejno nadzirajo in prilagajajo hitrost glede na pred njimi vozeče avtomobile (prilagodljiv tempomat) itd. Opravljajo pa lahko tudi bolj kompleksne operacije, kot npr. da se sami odločijo, kam zaviti, s kakšno hitrostjo peljati in kako sploh priti do cilja. 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 brez človeške pomoči izvrševati vse potrebne operacije. Od potnika se ne zahteva niti, da prevzame nadzor nad vozilom niti, da je v njem prisoten. Že danes takšni avtomobili delujejo z visoko stopnjo zanesljivosti, kljub temu

¹⁶ S tem je bil na lestvici prvih 500 superračunalnikov tistega časa na 258. mestu.

pa prihaja do občasnih napak, ki lahko ogrozijo človeško življenje. Najbolj znan takšen primer se je zgodil podjetju *Uber*, ki je testiralo samovozeči avtomobil, ta pa je pri tem s hitrostjo 70 km 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. 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 z ovinkom na pločnik ali bankino. Preiskava je pozneje odkrila, da avtonomna vozila podjetja *Uber* sploh niso bila sprogramirana tako, da bi reagirala na nepravilna prečkanja ceste, kljub temu, da so bila že pred tem večkrat udeležena v nesrečah in da je do zbitja kolesarke prišlo, ker je nadzorni sistem ni prepoznal kot človeka. Na srečo jih v masovni proizvodnji še ni, opisane nezgode pa pričetek njihove prodaje zamikajo v prihodnost.

Naslednji znani modeli UI, ki jih lahko v takšni ali drugačni izvedbi že kupimo na trgu, kar pomeni, da so povsem funkcionalni in pripravljeni za uporabo v vsakdanjem življenju, so *osebni asistenti*. Najbolj znani so *Siri* (Apple), *Googlov asistent*¹⁷ in *Alexa* (Amazon). Namesto nas lahko preko zvočnih ukazov opravijo go-ro opravil, 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 itd. V bistvu gre za virtualne pomočnike, locirane v »obla-ku«, s katerimi komuniciramo v naravnem jeziku: ko postavimo vprašanje, ti zvočne valove spremenijo v besedilo, kar jim omogoča, da zberejo potrebne informacije iz tistih virov, ki so za izvršitev zahtevane naloge relevantni.

Nekateri virtualni pomočniki so tako pristni, da iz izseka telefonskega pogovora med njimi in človekom, ne moremo ugotoviti niti, da je eden izmed sogovornikov UI niti, kdo to je. Govorimo o sistemu *Google Duplex*, ki predstavlja nadaljevanje razvoja Googlovih virtualnih asistentov, s to razliko, da je ta izurjen za obdelavo naravnega jezika. Trenutno so njegove jezikovne naloge omejene zgolj na rezervacije mize v restavracijah ali terminov pri frizerju ter posredovanje odpiralnih časov, v prihodnosti pa naj bi bil sposoben opravljati tudi bolj kompleksna opravila, ki ne temeljijo zgolj na telefonskih pogovorih, so pa z njimi povezana in jih vključujejo. Npr. če bomo od njega zahtevali, da naj v nedeljo ob 13.00 rezervira mizo za dve osebi v restavraciji *Darjeeling*, bo sistem poiskal njeno telefonsko številko, tja poklical, se s človekom na drugi strani vse potrebno dogovoril ter na koncu vnesel termin kosila v uporabnikov koledar. (Kremp, 2018) Očitno ni več tako daleč čas, ko nas bo UI poznala tako dobro, da jo bomo lahko vprašali, kam naj gremo na dopust in kaj naj tam počnemo, njen odgovor pa bo takšen, da se bo v celoti skladal z našimi potrebami, interesi in željami.

¹⁷ Angl. *Google Assistant*.

Na koncu omenimo še verjetno trenutno najbolj uporaben model UI, *Googlov prevajalnik*.¹⁸ Na trgu je že več kot 10 let, pred časom pa je bil deležen pomembnih popravkov, presedlal je na t.i. *mrežno prevajanje*, katerega bistvena prednost je, da prevaja cele stavke, ne izmenično posameznih besed, ki jih tvorijo. Z uporabo širšega konteksta, takšen sistem lažje najde najbolj relevanten prevod, ki ga potem preuredi in prilagodi do te mere, da je podoben človeškemu. Odkriti skuša semantiko stavkov, pri čemer upošteva okoliščine, v katerih besede nastopajo, zaradi česar njegovi prevodi postajajo vse bolj naravni. Program omogoča tudi neposredno prevajanje v drug jezik, kar končno uporabniško izkušnjo, ker se tako izognemo vmesni postaji, ki povečuje verjetnost napak in nesmislov v končnem izdelku, še izboljša. Uporaba širšega konteksta, upoštevanje semantike stavkov in neposredno prevajanje, pa so vplivali na izboljšanje vrstnega reda besed v prevodu, kar je povečalo njegovo razumljivost. Tega stari program ni bil vedno zmožen zagotoviti, zaradi česar je bil pogosto tarča posmeha.

Trenutno podpira 105 jezikov, sposoben pa je celo prevajati v jezik, ki ga ne pozna, tj. ni bil del njegovega urjenja,¹⁹ 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. Kljub temu, da izkazuje doslednost pri uporabi pomenov besed, pa se sistem pri besedah, ki imajo v jeziku, v katerega se prevaja, več pomenov, pogosto zmoti. Razen tega je zelo občutljiv na slovnične napake, kar drastično zmanjšuje njegovo zanesljivost. Zanimariti pa ne gre niti dejstva, da je slednja odvisna tudi od kompleksnosti jezika, zaradi česar so nekateri prevodi boljši kot drugi. V kakovosti končnega izdelka tipično prednjačijo evropski jeziki, pa še to ne vsi (v prednosti so vplivni jeziki kot angleščina, nemščina, francoščina itd.), močno pa zaostajajo afriški.

Verjetno se niti ne zavedamo, da razumevanje jezika, ki se nam zdi tako samo po sebi umevno, zahteva veliko mere inteligence. To se kaže v tem, da ne zahteva zgolj poznavanja dejstev, ampak tudi konteksta, v katerega je pogovor umeščen in splošnega znanja, prepoznati pa je treba tudi način, kako je nekaj povedano. Zato ni čudno, da se stroji že skoraj 50 let trudijo to doseči, pa jim do končnega uspeha še vedno veliko manjka. V luči tega so pravkar opisani primeri UI še posebej pomembni, saj kažejo na napredek na področju znotraj UI, ki ga imenujemo *obdelava naravnega jezika*.^{20 21}

¹⁸ Angl. *Google Translate*.

¹⁹ Uspešno reševanje problemov ali odgovarjanje na vprašanja, ki jih UI v fazi učenja še ni srečala, v angl. imenujemo *Zero-Shot-Learning*.

²⁰ Angl. *natural language processing*.

²¹ Napredek na področju obdelave naravnega jezika se kaže v naslednjem: (i) programi za prevajanje so danes že tako dobri, da se jih poslužujejo že poklicni prevajalci za izdelavo surovih osnutkov, (ii) spletni iskalniki so sposobni na naša vprašanja ponuditi že zelo natančne odgovore, ne pa zgolj narediti seznam

Ker očitno velja, da je učenje bistvena zmožnost inteligentnega bitja, ali umetnega ali naravnega, si v nadaljevanju oglejmo, kako pri stroju poteka.

3. Strojno učenje

Šibka UI in močna UI

Dartmouthska konferenca je temeljila na domnevi, da lahko vsak vidik učenja ali katerakoli drugo značilnost inteligence opišemo tako natančno, da jo lahko stroji simulirajo. Razprave o tem, ali je to dovolj za pripisovanje inteligence umetnim sistemom, trajajo še danes. Avtorji, ki to zagovarjajo, trdijo, da je tisto, kar šteje, rezultat in če je rešitev neke kognitivne naloge ustrezna 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*. Na drugi strani pa so raziskovalci, ki menijo, da je treba umetno inteligenco izenačiti s človeško v smislu proizvodnje umetnega bitja, ki bi imelo zavest, čustva in občutke ter bilo človeku z duhovnega vidika enakopravno. Samo v tem primeru bi potem lahko upravičeno govorili o inteligentnem stroju v pravem pomenu besede. To stališče imenujemo *močna umetna inteligenca*.²² (Dengel, 2019a)

Metodi strojnega učenja

(i) simbolna UI

Vzor za to, kako simulacija inteligentnega obnašanja poteka, najdemo že pri starih Grkih, zlasti pri Aristotelu, ki je s svojimi kategorijami, tj. različnimi vrstami ali načini bivanja, postavil temelje simbolnega razmišljanja. Prvo vprašanje, na katerega je moral odgovoriti, je bilo, »kaj obstaja«. Da bi to ugotovil je Aristotel ustvaril 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 (npr. gori, seka, tepe), j. trpeti/pasivnost (npr. nadlegovan). (Aristotel, 2004) Drugo vprašanje se je zastavilo samo od sebe, »kakšne lastnosti imajo stvari in v kakšnem odnosu so med seboj«.

Njegov cilj je bil zgraditi model sveta, ki bi vključeval njegove različne vidike, njihove značilnosti in odnose med njimi. Zanimivo je, da uporabljajo danes

največkrat obiskanih strani, ki ustrezajo vnesenemu geslu in (iii) računalnikom lahko ukaze tudi narekujemo, kar pomeni, da so sposobni govorjeno besedo pretvoriti v njim razumljiv jezik. (Socher, 2019)

²² Ne bomo se spuščali v to, kateri pristop je pravilen, šibkejši ali močnejši, dejstvo je, da lahko kognitivne sposobnosti z različnimi metodami na računalnikih uspešno simuliramo že danes.

ustvarjalci UI pri razvijanju računalniškega jezika, s katerim znanje o svetu formalno reprezentirajo, podobne modele. Primer tega so t.i. *semantične mreže*, ki so kot pri Aristotelu sestavljene iz različnih kategorij, npr. kraja, oseb, dogodkov, časa, imajo različne lastnosti in so med seboj v določenih odnosih. Če bi želeli na takšen način prikazati predavanje, bi to vključevalo predavalnico (kraj), z učiteljem in študenti (osebe), ki ob določeni uri (čas) razpravljajo o določeni temi na način, da učitelj podaja snov, študenti pa poslušajo in sprašujejo (odnos). Grafično bi to lahko ponazorili kot 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. (Dengel, 2019a)

Simulacija inteligentnega obnašanja tako poteka preko semantičnih mrež, ki so na simbolni način predstavljeno znanje, ki omogoča, da s pomočjo deduktivnega in induktivnega sklepanja, ki poleg aplikacije podatkov vključuje tudi uporabo pravil, sistem izpeljuje nove zaključke o svetu in se tako *uči*. Umetno inteligenco, ki uporablja takšno metodo učenja, imenujemo *simbolna UI*. V njenem ozadju je nekaj, kar ima že zelo dolgo brado in sega nazaj vse do Hobbesa, ki je trdil »Mišljenje ni nič drugega kot računanje«. (Hobbes, 1651/2006: 32) Njegove besede vsebujejo dve domnevi: (i) mišljenje je mentalni proces, ki obsega manipulacijo s simboli, ki se dogaja v možganih, zaradi česar so primerki misli v bistvu primerki možganov, ali kot jih imenuje Hobbes, »fantazme«. ²³ In (ii) mišljenje je najbolj čisto, ko zvesto sledi metodičnim pravilom, zaradi česar lahko rečemo, da je mišljenje mehanski proces, je nekaj takega kot upravljanje 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. (Haugeland, 1986, Bregant, 2010, 2016) Govorimo o t.i. *simbolni hipotezi*, ki 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.

(ii) *ne-simbolna UI*

Obstaja pa tudi simulacija inteligentnega obnašanja, ki ne vključuje na simbolni način predstavljenega znanja, njen vzor pa je delovanje dejanskega nevrona. S tega vidika je zanj bistveno naslednje: (a) da je povezan z drugimi nevroni, (b) da so te povezave različno obtežene (močne), (c) da obstaja prag, ki določa, ali je aktiviran ali ne in (d) da je tako »izključen« (0) ali »vključen« (1). Skratka, biološki nevroni tvorijo prepleteno mrežo povezav, ki so različno močne, slednje je deloma odvisno od premera vlaken, ki so povezana, deloma pa od kemične zgradbe

²³ »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 odvezujemo.« (Hobbes, 1651/2006: 32)

stika (sinaps). Ko vsota vhodnih signalov določenega nevrona doseže ali preseže njegov aktivacijski prag, nevron »zagori« (se sproži) ter posreduje izhodni signal naprej svojemu najbližjemu sosedu. In podobno so danes zgrajene tudi umetne nevronske mreže, ki predstavljajo orodje ne-simbolno predstavljenega znanja, pri čemer so njihovi osnovni gradniki t.i. idealizirani nevroni, 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)

Simulacija inteligentnega obnašanja tukaj temelji na prepoznavanju, upoštevanju in sortiranju tipičnih značilnosti predmetov, 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, iz česar sistem izpeljuje nove zaključke o svetu in se tako *uči*. UI, ki uporablja takšno metodo učenja, imenujemo *ne-simbolna UI*. Na takšen način je algoritem sposoben ponovno prepoznati npr. limono in jo ločiti od drugih predmetov, ki so prav tako rumeni, pri čemer upošteva tudi npr. njeno obliko in še kaj, kar je za uspešno prepoznavanje tega, katero sadje je pred nami, pomembno, v fazi urjenja dobil od svojega programerja in shranil v svoji bazi podatkov. Problem je, da je stvarnost veliko bolj kompleksna kot prepoznavanje sadja, zaradi česar je določanje ustreznih značilnosti razlikovanja, če želimo, da takšen sistem dobro deluje tudi v smislu robustnosti, tj. da je odporen na motnje (popačena, meglena ali temna slika), zahtevno početje, ki pogosto ne obrodi sadov.

Toda rešitev je že na vidiku in sicer v obliki metode, ki se imenuje *globoko učenje*.²⁴ V splošnem gre za to, da se je sistem sposoben brez predhodne človekove pomoči povsem samostojno naučiti, katere so tiste značilnosti, ki pridejo v poštev pri prepoznavanju, tj. v fazi učenja mu ne rabimo več povedati, katere so tipične značilnosti predmetov na slikah, ampak mu njihove slike zgolj pokažemo in rečemo npr. »To je limona«, »To je marelica«, »To je jagoda«. Sistem potem sam ugotovi, katere so tiste značilnosti, ki so za posamezno sadje tipične in tako bistvene za njihovo kasnejše ustrezno ločevanje. (Dengel, 2019a)

Postopki strojnega učenja

Strojno učenje temelji na tipično človeški sposobnosti učenja iz primerov: na začetku dobiva sistem vedno takšne primere z oznakami, ki spadajo v isto skupino, npr. same paradižnike. Na osnovi tega je sposoben izpeljati posplošitev, ki mu potem pomaga tudi pri razvrščanju takšnih predmetov, ki jih ne pozna oziroma ki nimajo nobene oznake, bodisi paradižnike ali kaj drugega. Da bi bilo takšno uče-

²⁴ Angl. *deep learning*.

nje uspešno, morajo biti izpolnjeni naslednji trije pogoji: (i) količina podatkov mora biti zelo velika, (ii) iz primerov, ki rabijo za urjenje, mora biti razviden vzorec, na osnovi katerega lahko sistem predmete grupira in (iii) ne sme še obstajati matematična formula, ki glede na dane značilnosti že uspešno razlikuje eno skupino od druge.

Ko je temu zadoščeno, tj. ko obstaja takšna podatkovna baza, ostane vprašanje, kateri postopek učenja izbrati. V osnovi poznamo tri: a. *klasifikacija*, b. *grupiranje*²⁵ in c. *regresijo*.²⁶ Klasifikacija pomeni sistematično umestitev primerov v že znane razrede, ki so med seboj jasno ločeni in rabijo kot orodje, s katerim podatke uredimo. Umestitev v razrede temelji na določenih značilnostih, razredi pa imajo svoja imena, t.i. *oznake*.²⁷ V podatkovnem nizu slik sadja je npr. ena oznaka limona, druga marelica, tretja jagoda, z njihovo pomočjo pa sistem sadje razvršča. Pri grupiranju imena razredov pred začetkom postopka še niso znana, tukaj prevzame algoritem to nalogo, da za določeno količino primerov, ki jih sortira po podobnih ali skladnih značilnostih ustvari oznako. Rezultat so skupine podobnih elementov, ki jih lahko razumemo tudi kot kategorije, ki se med seboj razlikujejo glede na tipične značilnosti. Regresija pa išče matematično zvezo med dvema značilnostma. Cilj je ugotoviti, ali ena značilnost vpliva na drugo (ciljno značilnost) in ali lahko potem s pomočjo prve napovemo vrednost druge, npr. ali zaposlitev na fakulteti vpliva na plače njenih uslužbencev in ali lahko iz tega, da nekdo dela na fakulteti napovemo, koliko zasluži. Ciljne značilnosti ne želimo v celoti pojasniti, ampak zgolj ugotoviti, ali obstaja vpliv in kako velik je. (Dengel, 2019b)

Oblike strojnega učenja

Obstajajo tri različne oblike strojnega učenja, katero bomo uporabili pa je odvisno od tega, čemu je sistem namenjen. Prvo obliko imenujemo *nadzorovano učenje*.²⁸ Pri njem moramo v fazi urjenja vse podatke, ki rabijo kot primeri, opremiti tudi s pravilnimi odgovori, tj. oznakami, npr. to je mačka, to je pes, to je konj. Z njimi sistem popravlja napake, tako da na koncu iz vseh primerov izpelje splošni model, ki ga potem uporablja za npr. razvrščanje živali. Druga oblika nosi naziv *nenadzorovano učenje*.²⁹ Tukaj dobi sistem vse podatke v fazi urjenja brez dodatnih informacij in iz njih na osnovi podobnih ali istih značilnosti sam ustvari skupine ali *gruče*,³⁰ z drugimi besedami, algoritem uči stroj tako, da vhodne podatke razdeli v

²⁵ Angl. *clustering*.

²⁶ Angl. *regression*.

²⁷ Angl. *labels*.

²⁸ Angl. *supervised learning*.

²⁹ Angl. *unsupervised learning*.

³⁰ Angl. *clusters*.

več kategorij s svojimi značilnostmi, njihovo število in katere to so, pa iz vhodnih podatkov izlušči sam brez nadzora učitelja. Tretji obliki pravimo *vzpodbujevalno učenje*.³¹ V tem primeru dobi sistem v fazi urjenja le občasno povratno informacijo iz okolja in se skuša na takšen način naučiti, kaj je prav in kaj ne, da bi se v bodoče izognil napakam ter bil sposoben bolje oceniti, ali bo njegovo ravnanje v neki situaciji vodilo k uspehu ali neuspehu. Povedano drugače, stroj sam razvije ustrezne strategije za rešitve določenih problemov, in sicer s pomočjo nagrade v smislu pozitivne ali negativne povratne informacije. (Dengel, 2019b)

In kako izgleda v luči tega bližnja prihodnost? Trenutni modeli UI, so izdelani tako, da so sposobni opravljati zgolj eno nalogo: npr. prvi algoritem prepozna oblike, drugi vozi avtomobil, tretji igra šah ipd. Tako ne bo ostalo. Nadaljnji razvoj UI bo šel v smeri združevanja opravil, tj. razvoja strojev, ki bodo lahko hkrati opravljali različne naloge. Za to pa bodo potrebni sistemi, ki se bodo lahko učili tudi brez urjenja. (Socher, 2019)

4. Sklep

Zapuščina pionirjev UI iz 60. in 70. let 20. stoletja, imenovane tudi zlata doba UI, je še danes aktualna. Njihova ideja o simulaciji inteligentnega obnašanja, ki naj bi bila dovolj za govorjenje o pametnih strojih, je še danes živa, prav tako pa tudi njihov pristop, ki temelji na simbolni predstavitvi sveta s pomočjo semantičnih mrež, za katerega je sicer temelje postavil že Aristotel s svojimi kategorijami. Močnejši računalniki, velika količina podatkov, ki so na voljo in razvoj umetnih nevronske mreže, ki jih je mogoče uriti, pa so danes omogočili še en drug način simulacije inteligentnega ravnanja, ki je ne-simbolen, temelji pa na prepoznavanju tipičnih značilnosti, njihovem sortiranju in uvrščanju npr. predmetov v razrede z njihovo pomočjo. Iz tega se je (ali se kmalu bo) razvil niz izjemno zmogljivih modelov UI, avtonomna vozila, osebni asistenti, mrežni prevajalniki, nadzorni sistemi, delovni roboti, avtomatizirano orožje itd., katerih uporaba sproža številna etična vprašanja, ki si jih bomo ogledali v 2. delu članka v naslednji številki *Analize*.

³¹ Angl. *reinforcement learning*.

Artificial Intelligence at Work: Evolution, Behaviour and Learning of Machines

The idea of simulating the intelligent behaviour from 1960s and 1970s being enough to speak about smart machines is still alive. It is the so called non-symbolic approach based on the training of artificial neural networks capable of learning by identifying typical features, classifying them and correcting their own errors, that is in the foreground now. They gave rise to sets of extremely powerful AI models which are gradually dissolving borders between the natural and artificial intelligence. The paper analyses the concept of intelligence in terms of perception, memory and learning, then briefly outlines the evolution of AI stating some crucial historical and well known nowadays events, and finally within the symbolic and non-symbolic approach to the simulation of smart behaviour introduces machine learning, more specifically, under what conditions the training of a given artificial system takes place.

Keywords: artificial intelligence, weak AI, strong AI, machine learning, natural language processing.

Literatura:

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

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

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

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

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

Gottfredson, L., (1998). »The General Intelligence Factor«. *Scientific American Presents*, 9, str. 24–29.

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

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

Humphreys, L. G. (1979). »The construct of general intelligence. *Intelligence*«, 3, str. 105–120.

Kremp, M. (2018). »Google Duplex ist gruselig gut«. Spiegel Online. Dostopno na: <https://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html> [27.11.2019].

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.

Socher, R. (2019). »Sprache und KI: Computer, die den Menschen verstehen«. V Dengel, A., Socher, R., Kirchner E. A., Ogolla, S., *Künstliche Intelligenz: Die Zukunft von Mensch und Maschine*. Hamburg: ZEIT Akademie GmbH, str. 60–65.

Sternberg, R. J., Salter W (1982). *Handbook of human intelligence*. Cambridge, UK: Cambridge University Press.

Stern, William (1912) Die psychologischen Methoden der Intelligenzprüfung: und deren Anwendung an Schulkindern. Leipzig: J. A. Barth.

Wechsler, D (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

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.

