

Tine Kolenik

Pedagoška fakulteta Univerze v Ljubljani

Kaj imata skupnega Heidegger in Roomba: utelešena umetna inteligenca kot peskovnik kontinentalne filozofije

Robotski sesalec Roomba ne bi bil mogoč brez bogate zgodovine filozofskega izročila, ki predstavlja temelj klasične umetne inteligence. Načelo slednje je računska metafora za um, kjer je kognicija le manipuliranje arbitrarnih simbolov. A takšna paradigma ima številne težave, kot sta problem okvirja in ozemljitve simbolov. To so razkrile ideje Heideggerja, Merleau-Pontyja idr. kontinentalnih filozofov, ki so skozi robotika Brooksa, izumitelja Roombe, vodile nastanek utelešene umetne inteligence (UUI). UUI temelji na paradigmi nujnosti spoja agenta in okolja ter vznika kognicije iz senzomotorične zanke. A raziskovalci na tem področju ponujajo zelo različne interpretacije teh predpostavk. To ponovno prinaša težave, predvsem v obliki vprašanja avtonomije in nastanka pomena. Namen članka je predstaviti vzajemen odnos filozofija-UUI. Filozofijo vzpostavi kot ključno za razvoj (U)UI ter ponuja pregled trenutnih trendov in praktičnih primerov. Članek na koncu prikaže problematičnost sodobnih smernic ter ponudi morebitne rešitve, ki med drugim izvirajo prav iz filozofije.

Ključne besede: utelešena umetna inteligenca, utelešena kognicija, fenomenologija, klasična umetna inteligenca, enaktivizem

1. Uvod

Roomba, mali robotski sesalnik podjetja iRobot, je skupek nekaj senzorjev, s katerimi zaznava okolico (npr. odbijači, merilniki razdalj), in aktuatorjev, s katerimi deluje v okolju (npr. motorski pogon, sesalni mehanizem). Je tudi posledica velikega dela večstoletnega filozofskega izročila – najsibo to doktrina realizma (Ockham, Galileo, Hobbes, Locke, Hume, Moore, Russell, Kopernik, Descartes, Kepler...), idealizma (Berkeley, Kant, Schopenhauer, Nietzsche, Hegel...) ali

vmesne poti med njima (Heidegger, Merleau-Ponty, Wittgenstein...). Roombo je izumil avstralski robotik Rodney Brooks, oče utelešene umetne inteligence (UUI), ki je na prehodu iz 80. v 90. leta prejšnjega stoletja sprožil revolucijo na področju umetne inteligence (UI). Slogan brooksijanske revolucije je bil preprost: »najboljši model sveta je svet sam« (Brooks, 1997)¹. Brooks je na podlagi te ideje utemeljil svoj pristop k modeliranju inteligentnih robotov, kjer reprezentacije robota niso izhajale od znotraj, temveč so bile svet sam preko neprestanega zaznavanja. To se razlikuje od klasične UI, ki temelji na računski metafori uma kot računalnika in duševnih procesov kot manipuliranja arbitrarnih simbolov, kjer so bile reprezentacije in vse znanje vnaprej podani, programirani v robotih. Eden izmed glavnih vplivov na to revolucijo (in Roombo samo) je bilo sporočilo Heideggerja, da »pomenljivi predmeti, /.../ med katerimi živimo, niso model sveta, ki je spravljen v naših možganih ali umu, temveč svet sam« (Dreyfus, 1972: 265–266).

Heidegger, eden najpomembnejših sodobnih nemških filozofov, je predstavljal, podobno kot Brooks, reakcijo na takratni pogled na svet. Nasprotoval je temu, da je svet mogoče spoznati preko kvantifikacije vsakdanjega izkustva, zbiranja empiričnih podatkov. Trdil je, da tovrstna metodologija ne prinaša spoznavanja obstoječega sveta. Pomembno je znanje *kako* (npr. védenje, kako voziti kolo), ne znanje *da* (npr. védenje o fizičnih silah med vožnjo kolesa). Pomembno je naše delovanje v zvezi s kolesom. Heideggerjeva fenomenološka zapuščina je rodila nove pristope k raziskovanju prvoosebnega izkustva (Stolorow, 2006) in gibanja, ki so prinesla paradigmo utelešene kognicije (UK; Froese, 2007). Nemški fenomenolog je preko Dreyfusovih interpretacij na prehodu iz 80. v 90. leta prejšnjega stoletja močno vplival na Rodneyja Brooksa, velikega kritika dotedanje t.i. stare dobre UI (SDUI; ang. *Good Old-Fashioned Artificial Intelligence* – GOFAI; Dreyfus, 2007). Brooks je po več robotih, ki so utelešali novo, brooksijansko paradigmo, izumil Roombo, robotski sesalnik, ki se z neprestanim zaznavanjem sveta odziva z izredno učinkovitim sesanjem prahu. Ti roboti naj bi Heideggerjeve misli utelešali v tem, da spoznavajo svet z delovanjem, da pridejo preko tega delovanja do znanja *kako* in ne z znanjem *da*, kjer bi imeli že vse vnaprej programirano v sebi.

Za izumitev Roombe je bila torej potrebna dolga in naporna pot številnih mislecev, filozofskih pristopov, empiričnih dognanj in postavljenih paradigem iz najrazličnejših zgodovinskih obdobj. Povezava med filozofijo in UI je še bolj eksplicitna pri zagovornikih stališča, da mora robotika opravljati vlogo orodja za testiranje filozofskih stališč (Harvey, 2000; Froese in Ziemke, 2009). Članek poskuša, predvsem z vidika UI, na kratko opisati prav omenjeno pot, posebna pozornost pa je posvečena zanki med filozofijo in UUI, praktičnim primerom, težavam UUI (vprašanji avtonomije in pomena) in trendom, ki le-te poskušajo rešiti.

¹ Vse citate iz tuje literature je prevedel avtor.

2. Filozofija in stara dobra umetna inteligenca

Rojstvo UI (te, ki jo zdaj nekoliko pokroviteljsko kličemo SDUI, sicer pa klasična oz. simbolna UI) sega v leto 1956, ko se je odvila t. i. dartmouthska konferenca. Osrednji misleci le-te so bili McCarthy, Minsky, Newell, Samuel in Simon. Zaradi dosežkov njihovega dela, ki je navdušilo z navidez zelo človeškimi sposobnostmi (npr. igranje šaha in govorjenje angleščine), so si dartmouthovci zastavili ambiciozen cilj: odkriti algoritem, enačbo, s katero bi opisali človeško oz. splošno inteligenco, ki bi bila zmožna prepričljivo reševati med seboj zelo različne naloge (Crevier, 1993). Ta cilj je odseval močno predpostavko – da naša duševnost resnično operira na podlagi takšnega algoritma. Podobno je razmišljal tudi Turing, eden izmed najpomembnejših znanstvenikov preteklega stoletja. Računska metafora, um kot računalnik, je najmočnejšega in konkretnega utemeljitelja dobila prav v njem, ko je leta 1936 »definiral računanje kot formalno manipulacijo z neinterpretiranimi simboli, ki se izvaja s pomočjo formalnih pravil.« (Markič, 2010: 29). Prav na tem temelji Turingov stroj, ki lahko potencialno izvede prav vsako nalogo, če mu le lahko določimo neko končno število korakov, ki pripeljejo do rešitve. Turing je na podlagi tega umislil stroj, ki bi s takšno formalno strukturo posnemal mišljenje.

Na tem temelji metoda »zgoraj-navzdol« (»*top-down*«) raziskovalcev SDUI. Ti svojim programom podajo vse znanje, ki ga bodo potrebovali za reševanje danih nalog. Ko ti programi rešujejo naloge, se zanašajo na vnaprej vprogramiran repertoar znanja, na knjižnico, katere pot je od zgoraj, iz programa, navzdol, na nalogo iz okolja. Model sveta imajo že v sebi, rešitve za morebitne težave prav tako. Odgovore na vprašanja imajo že vnaprej definirane.

Dreyfus (2007) se spominja, kako so se študenti Newella in Simona v 1960-ih zgrinjali nanj z besedami, da filozofi v 2000 letih razmišljanja v naslanjačih še vedno ne razumejo inteligence, oni pa so v kratkem času v laboratorijih UI ravno v tem uspeli (dokazali naj bi, da inteligenca deluje na trditvah simbolne paradigme). A rojstvo UI je omogočilo prav bogato filozofsko izročilo. Pri tem izstopata dve doktrini: realizem in kartezijska filozofija, ki ju, čeprav se zdita na prvi pogled v nasprotju, SDUI obe predpostavlja. Realizem je doktrina, ki zagovarja stališče, da skozi čutila zaznavamo resnični svet, kakršen je izven nas; ta določa, kako iz njega izvirajoče predmete zaznavamo. Dualizem pa zagovarja, da sta duševnost in telo dve fundamentalno različni snovi, ki ju ne gre reducirati na eno ali drugo. A ravno to počne SDUI – če je misel le manipulacija simbolov, to pomeni, da je ločena od telesa in da le-to ne vpliva na kognicijo, inteligenco (Zebrowski, 2010). Sicer je o mišljenju kot računanju že v 17. stoletju pisal Hobbes, Descartes pa je v istem obdobju razmišljal o reprezentacijah. In to je le začetek – vplivi na UI so prisotni tudi v Leibnizevih univerzalnih karakteristikah, Kantovi univerzalnosti konceptov (koncepti kot pravila), formalni logiki Fregeja in Russella, logičnem atomizmu Wittgensteina itd. SDUI se torej zdi kot nekakšna aplikacija teh

idej (Froese, 2007). Ta aplikacija je še posebej uspešna v različnih ekspertnih sistemih (tj. sistemih, ki posnemajo človeško odločanje, npr. zdravstvena diagnostika; Russell in Norvig, 2003), programih za igranje strateških iger (npr. program za igranje šaha Deep Blue; prav tam) ipd. SDUI je sicer utemeljil tudi klasično informacijsko-procesno kognitivno znanost.

Zdi se torej, da SDUI ne bi bila mogoča brez filozofije. A prav filozofija je bila tista, ki je SDUI zadala tako močne kritike, da je bila paradigmatška sprememba nujna.

3. Filozofija in utelešena umetna inteligenca

Simon in Newell sta v 1950-ih izjavila, da bo računalnik v desetih letih svetovni šahovski prvak in da bo odkril ter dokazal nov matematični izrek (Russell in Norvig, 2003). V 1960-ih je Simon, potem ko računalnik teh dveh ciljev ni dosegel, nekoliko pompozno popravil izjavo in zatrdil, da bo računalnik v dvajsetih letih nadomestil človeka v opravljanju katerekoli naloge (Crevier, 1993). Isto desetletje je Minsky napovedal, da bo UI svoj prvotni cilj dosegla v eni generaciji, v 1970-ih pa, da bodo v nekaj letih zgradili stroj s človeško inteligenco (prav tam). Po takšnem vzorcu se je datum cilja, tj. obstoja stroja s človeško inteligenco, vedno znova prestavljal. A raziskovalci SDUI so še vedno verjeli, da je predvsem računska zmogljivost tista, ki jih loči od splošne UI.

Težav, ki so botrovale prestavljanju datuma splošne inteligence računalnikov, je bilo veliko več, kot so si jih sami raziskovalci upali priznati. SDUI je obtičala na mestu – in to relativno brez napredka ob siceršnjem tehnološkem razvoju. Programi, ki so temeljili na goli računski moči, so imeli velike težave z delovanjem, ki ga ljudje nenehno jemljemo za samoumevnega, predvsem pa ga ne enačimo z inteligenco. To je vključevalo predvsem vidno in slušno zaznavanje ter gibanje. Program, ki je premagal samega Kasparova, ni bil denimo sposoben nečesa, nam tako preprostega, kot je hoja. A to je bila le prva vrsta težav. Druga vrsta je bila bržkone bolj povezana s predpostavkami paradigme v zvezi s človeško inteligenco in kognicijo kot tako. Programi so imeli velike težave z učenjem novih reprezentacij, kategorizacijo, reševanjem nalog, za katere ni eksplicitnih pravil (torej delovanje s pomočjo ozadnjega znanja (*background knowledge*)) ipd. Prvi pogubni udarec je SDUI zadal problem okvirja. Programi so imeli težave z ugotavljanjem, katere informacije so relevantne za njihovo delovanje, saj programer enostavno ni vedel, v kakšni situaciji se bo program znašel. Drugi pogubni udarec je SDUI zadal filozof Searle z miselnim eksperimentom kitajska soba (Searle, 1980), katerega bistvo je, da stroj, ki operira z arbitrarnimi simboli, pač nima pra-

vega razumevanja.² Kitajska soba je vodila v prelomni Harnadov članek »The Symbol Grounding Problem« (1990). Ta prevprašuje ravno predpostavke SDUI. Iz kje vznikne pomen skozi simbole, ki so manipulirani le na podlagi njihove sintaktične, formalne oblike? Kako je lahko njihov pomen ozemljen kjerkoli drugje kot le v nadaljnjih brezpomenskih simbolih? Harnad ponudi potencialne rešitve problema, predvsem pa opozori na to, da se mora problem reševati od »spodaj-navzgor« (»bottom-up«) – simboli morajo biti ozemljeni v senzomotorični interakciji. A vse to je le odmev prvotne Dreyfusove kritike SDUI iz začetka 1970-ih, ki temelji na idejah Heideggerja.

Filozofija je torej, tako se zdi, igrala glavno vlogo pri pokopu SDUI. A takšno stanje v krogih SDUI je ustvarilo možnost za novo skupino mislecev in za novo paradigmo. Ta se je usidrala v kontinentalni filozofiji, predvsem v fenomenološkem izročilu Heideggerja, Husserla, Merleau-Pontyja, Sartra, kasneje pa tudi v idejah različnih sodobnikov paradigme UK, kot so Flores, Lakoff, Varela, Thompson, Noë, Gallagher, Clark, Haugeland, Wheeler idr.

Brooksijanska revolucija je neposredna posledica kritik SDUI, ki so izšle iz filozofije. Zdi se, da vse glavne Brooksove ideje nosijo misli Heideggerja in njegovih sodobnikov (Dreyfus, 2007). Heidegger je dokončno zavrnil dihotomijo med svetom in nami samimi, s tem pa postavil naš obstoj ali *Dasein* (tu-bit, biti-v-svetu) kot dejavnost znotraj te domene. To pomeni, da svet ne obstaja zunaj nas. Mi smo v njem in ne ločeni od njega. Iz tega izhajajo pomembne implikacije: kar zaznavamo, je odvisno od nas samih oz. tega, kar smo. Igramo netrivialno vlogo v oblikovanju resničnosti, predvsem dela, ki vpliva na nas same, na našo izkušnjo resničnosti. Svet seveda obstaja kot tak, a naše zaznavanje in konstrukcijo le-tega kujemo (tudi) mi. Posebno vlogo ima tu naše aktivno delovanje v svetu, preko katerega spoznavamo svet, pridobimo znanje *kako*. Primer tega je naš odnos do orodij. Heidegger (1953) pravi, da naša dejanja z različnimi orodji spremenijo zaznavno naravo samega orodja in našega odnosa z njim. Ko pišemo, smo tako zavzeti v tem dejanju pisanja, da preidemo v doživljajsko stanje nekakšnega zlitja s pisalom samim; pisalo postane utelešeno, ločnica med spoznavajočim subjektom in spoznavanim objektom se zabriše. Posledica tega je nekakšno izginotje materialnosti. Vse to vključuje preprosti slogan brooksijanske revolucije: »najboljši model sveta je svet sam«, ki razblinja dihotomijo med svetom, reprezentiranim v umu in svetom izven uma. A Brooksove ideje nosijo širšo zapuščino. Omeniti gre še Merleau-Pontyja, ki nadaljuje Heideggerjeve ideje. Vplivi so vidni v Brooksovi ideji inteligence brez reprezentacij in v njegovem grajenju robotov. Brooks nikoli ne gradi duha znotraj stroja, temveč utelešena bitja, ki delujejo v dialogu s svetom in nimajo potrebe po nekem ločenem umu za inteligenco, saj ta ni potreben za to, da utelešeno bitje izvaja kompleksne naloge v okolju.

² Tega sicer ne gre zamešati s še eno kritiko SDUI, ki jo izraža Blockov miselni eksperiment kitajski možgani (Bringsjord in Arkoudas, 2014).

Seveda brooksijanska revolucija ni potekala le znotraj področja UI, ampak tudi znotraj kognitivne znanosti, ki se je prav tako oddaljevala od simbolne paradigme in začela raziskovati predpostavke UK. Pomembnih mejnikov je znotraj kognitivne znanosti več, zato je težko govoriti o le enem poglavitnem akterju, kakršen je bil Brooks za UUI. Tako Lakoff (1987) pokaže, da jezik na vseh nivojih poteka skozi metafore, ki so ubesedene z dejanji (npr. prepir kot vojna, kjer braniš svoje mnenje, napadaš nasprotnikove argumente itd.). Maturana in Varela (1984) vpe-ljeta koncept strukturnega spoja (nujno medsebojno vplivanje dveh sistemov, npr. agent-okolje ali agent-agent, za katera je bistvena zgodovina interakcij). Tega kasneje Varela, Thompson in Rosch (1991) razširijo v močnejšo tezo utelešenosti in enaktivizma. Noč (2005) trdi, da zaznavanje ni nekaj, kar se nam zgodi, temveč je nekaj, kar mi počnemo. Razmah takšnih razmišljanj in raziskav, ki naj bi potrjevale teze UK,³ je precejšnji. Trenutni prevladujoči trend med raziskovalci UK (še posebej v nevroznanosti in UI) je, da »mišljenje ni manipuliranje simbolov, temveč notranje simuliranje s pomočjo senzomotorične kognicije«⁴ (Farkaš et al., 2012). Kognicijo obravnavajo kot namenjeno delovanju, ne računanju, ozemljeno v senzomotorični zanki in nerazdružljivo od telesa. Takšna stališča podpira kopica empiričnih dokazov. Nevrokomputacijski modeli kažejo, da se komunikacijske vrline razvijejo le, ko senzomotorično delovanje in razumevanje dejanj drugih delujeta skladno (Rizzolatti in Arbib, 1998); rezultati s področja nevroslikanja prikazujejo, da se ob misli na brco zasveti homunkularni del motoričnega korteksa za nogo (Pulvermüller, 2005); kognitivna neurofarmakologija pa izvaja izredno zanimive eksperimente, kjer omejitve telesa na nek način vpliva na naše čustveni odziv (na slednje je vplivalo vbrizganje botoksa v ustnice), presojanje in razumevanje (Havas et al., 2010).⁵

Omeniti velja, da je tovrstno razumevanje UK le najmočnejši trenutni trend, ki se kaže v kognitivni znanosti. UK kot raziskovalni krog nikakor ni enoten v predpostavkah in trditvah. Zdi se, da so se začetki, na katere je tako zelo vplivala fenomenologija in so nosili idejo močnejše utelešenosti, nekoliko omilili, če ne celo strivializirali (Froese, 2007). Kljub temu da vplivov fenomenologije ne gre zani-kati, se zdi, da je vodilni trendi ne upoštevajo (Dreyfus, 2007). To je jasno tako v UK kot v UUI, kjer so vsi raziskovalci združeni le v imenu in v njihovi drži proti simbolni paradigmi. Gre torej bolj za to, proti čemu UK nastopa, in ne to, kaj UK

³ Za namen tega članka približno enačim utelešeno kognicijo (*embodied cognition*), ozemljeno kognicijo (*grounded cognition*), umeščeno kognicijo (*situated cognition*) in ostale vzporedne »paradigme«, ki med seboj niso strogo ločene. Razlike seveda obstajajo, a niso bistvene za ta članek.

⁴ Mišljenje torej poteka na podlagi našega zaznavanja in delovanja.

⁵ Pomembno je dodati, da tovrstne raziskave pritrjujejo le prevladujočemu trendu znotraj UK. Polje le-te ni enotno, kar pomeni, da takšne raziskave ne podpirajo vseh tez (npr. Varelove teze, da telo ne le vpliva na kognicijo, temveč jo konstituira). Polje predvsem družiti kritika simbolne paradigme.

zagovarja (Ziemke, 2004). Razlike se kažejo na več nivojih. Prvi primer bi bil odnos do reprezentacij – od tega, da jih imamo in da so ozemljene v našem senzomotoričnem sistemu, do radikalnejših smeri, ki trdijo, da reprezentacij nimamo. Drugi primer bi bil odnos do tega, za kaj sploh je kognicija – od informacijsko-procesne paradigme, ki trdi, da smo pasivni agenti, ki sprejemamo informacije iz okolja za reševanje nalog, do enaktivizma Varele, Thompsona in Rosch (1991). Slednji trdijo, da celotno bistvo organizma (njegov svet, kognicija, adaptacija, če jih lahko ločimo) vznikne v interakcijskem trikotniku uma, telesa in okolja. Različne vidike, ki se pojavljajo na področju UK, je združila Margaret Wilson: »1) kognicija je umeščena; 2) kognicija deluje pod časovnim pritiskom; 3) kognitivno breme prelagamo v okolje; 4) okolje je del kognitivnega sistema; 5) kognicija je za akcijo; 6) telo je osnova odklopljene kognicije« (Wilson, 2002: 625). Temu primerno je tudi raznoliko, razpršeno in veliko manj poenoteno polje UUI.

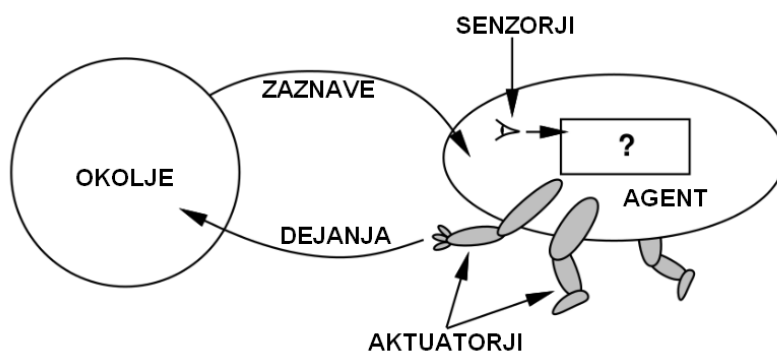
4. Utelesena umetna inteligenca v akciji

Znotraj UUI velja še močnejša težnja po postavljanju enotnih temeljev kot v kognitivni znanosti. Poskus slednjega predstavlja že Brooks v enem od dveh svojih prelomnih člankov, ki sta sprožila revolucijo. V »Intelligence without reason« Brooks (1995) našteje ključne značilnosti utelešenega sistema (str. 27):

1. [Umeščenost] Roboti so umeščeni v svet – ne ukvarjajo se z abstraktnimi opisi, ampak s *tukaj in zdaj* pričujočega sveta, kar neposredno vpliva na njihovo vedenje.
2. [Utelesenost] Roboti imajo telesa in dostopajo do sveta neposredno – njihova dejanja so del dinamike s svetom in imajo takojšnji zankasti učinek na njihovo zaznavanje.
3. [Inteligenca] Roboti izgledajo inteligentno – a izvor te inteligence ni le njihova koda. Inteligenca izhaja tudi iz njihove umeščenosti v svet, iz njihove pretvorbe dražljajev znotraj senzorjev in iz njihovega fizičnega spoja s svetom.
4. [Emergenca] Inteligenca sistema vznikne iz sistemovih interakcij s svetom in iz posrednih interakcij med njegovimi komponentami – včasih je težko pokazati na točen dogodek ali mesto znotraj sistema in zatrditi, da je to vzrok za neko dejanje.

Ko pride do modeliranja višjih kognitivnih procesov, se našteje značilnosti poskušajo realizirati z rabo umetnih nevronske mreže (UNM). UNM so se v disciplini ustalile zato, ker so osnovane na nevronske mreže človeških možganov, ker se lahko učijo, ker delujejo od »spodaj-navzgor«, ker so robustne, ker znanje ni diskretno shranjeno itd. UNM, dane v robotsko telo, tako za raziskovalce predstavljajo rešitev težav in potešitev kritik SDUI (Pfeifer et al., 2007). A večkrat kot ne je uporabljen hibridni pristop (s simbolnimi metodami) – »spodaj-navzgor« in

»zgoraj-navzdol«, nekaj podanega eksplicitnega znanja (sploh v primerih, ko gre za modeliranje zelo specifičnih kognitivnih procesov), diskretne vrednosti nevronov⁶ ipd. Raziskovalci skozi modeliranje nižjih kognitivnih procesov iščejo predvsem vznik obnašanja skozi spoj preprostih robotov in okolja. To poglavje bo zato razdeljeno v dva dela: prvega bodo predstavljali primeri modelov nižjih kognitivnih procesov (npr. senzomotorika), drugega pa primeri modelov višjih kognitivnih procesov (npr. abstrakcija). Tovrstno razdeljevanje izvira iz tega, da prevladujoči trendi na področju kognicijo obravnavajo kot kontinuum med nižjimi in višjimi kognitivnimi procesi; raziskovanje (in uspeh) z modeliranjem prvih je veliko pogostejše (Mirolli in Parisi, 2011). Ne glede na predmet modeliranja so utelešeni sistemi (oz. agenti), abstraktno gledano, funkcija iz zgodovine zaznav v dejanja ($f: P^* \rightarrow A$), izgledajo pa takole:



Slika 1: Prikaz senzomotorične povratne zanke med agentom in okoljem (Russell in Norvig, 2003: 32)

4.1 Primeri modeliranja nižjih kognitivnih procesov

Prvi sklop UII v praksi predstavljajo primeri modelov nižjih kognitivnih procesov. Pred UII se je na tovrstne procese, predvsem na motoriko in gibanje, gledalo kot na problem nadzora – telo je bilo le orodje, ki mora od »zgoraj-navzdol« prejeti pravilne ukaze. UII prikaže, da nižja kognicija (npr. gibanje) vznikne skozi dinamičen odnos interakcije telesa in okolja, ki zahteva zelo malo, če sploh kaj, ukazov od »zgoraj«.

Tega se je seveda zavedal že Brooks. Zaradi neuspeha SDUI pri izdelavi sistemov, ki bi bili sposobni opravljati preproste naloge v okolju, je že v zgodnjih 1980-ih v svojem laboratoriju na MIT začel z delom na robotu, ki bi se bil sposoben prilagajati in spreminjati z okoljem. Nastal je Herbert. Herbertova naloga je

⁶ To pomeni, da en nevron predstavlja točno eno stvar (npr. modro barvo).

bila navidez preprosta: tavati je moral po laboratoriju in metati prazne pločevinke pijač, na katere je naletel, v koš za smeti. Slednje mu je omogočila Brooksova posebna subsumpcijska arhitektura (*subsumption architecture*), ki so jo sestavljale med seboj povezane mreže, vsaka odgovorna za svojo nalogo (razlikovanje med polnimi in praznimi pločevinkami, izogibanje mizam itd.). Herbertova dejanja so vzniknila iz različnih aktivacij in inhibicij – tako je recimo miza aktivirala mrežo za izogibanje oviram, ta pa je posledično inhibirala mrežo za premikanje naprej. To naj bi bila fizična realizacija Brooksove ideje, da je svet sam najboljša reprezentacija. Herbert ni posedoval vnaprej sprogramiranega notranjega zemljevida prostora, ni bil opremljen z diskretnim znanjem o vseh možnih pločevinkah, ki jih laboranti ne bodo vrgli v smeti, njegovo obnašanje je vzniknilo od »spodaj-navzgor«, ne od »zgoraj-navzdol«. S tem je bil rešen problem okvirja. Vseskozi se je neposredno odzival na okolje, tukaj in zdaj (Clark, 1997).

Podobni poti so sledili nadaljnji modeli nižjih kognitivnih procesov. Nekateri so bili še bolj preprosti od Herberta – ne le, da niso imeli vprogramiranega kontrolnega centra »tam zgoraj«, od koder bi prejeli ukaze, imeli niso niti aktuatorjev. S tem so raziskovalci želeli še radikalneje prikazati, kako je lahko svet sam najboljša reprezentacija. Tovrstni pristop temelji na različnih eksperimentih, ki raziskujejo dinamiko med telesom in okoljem, iz katere vznikne določeno (vsaj navidezno) inteligentno obnašanje. Primer tega je raziskava, kjer so mrtvo ribo položili v umetno generirane vodne tokove, telo ribe pa se je nemudoma prilagodilo okolju ter se obnašalo in obvladovalo tokove (telo se je samodejno premikalo proti toku), kot da bi bila riba živa (Liao et al., 2003). Tovrstne raziskave so že v začetku 1990-ih med drugim navdihnile t.i. pasivne dinamične sprehajalce (*passive dynamic walkers*). Slednji so sposobni hoje po različnih vzdolžnih klančinah brez kakršnihkoli aktuatorjev in vgrajenega nadzora. Nimajo ne motorjev ne senzorjev ne mikroprocesorjev. Uspešna hoja navzdol, tudi ob spreminjajočem se, nepoznanem in nejasnem terenu, vznikne preprosto iz dinamike med klancem in telesom robota (McGeer, 1990).

Emergentno vedenje, ki ni bilo predvideno in vprogramirano v robota samega, ki ni eksplicitno podano v sistemu in ne poteka od »zgoraj-navzdol«, je bilo eno glavnih vodil raziskovalcev UUI. Preprost primer takšne emergence iz 1990-ih je kvadrasti robot, ki je imel na vsaki strani po en infrardeči senzor (Wilson in Golonka, 2013). Preko njega je robot ob zaznavi predmeta spremenil smer – robot se je torej izogibal oviram. Dan je bil v okolje s kockami, lažjimi od sebe, razmetanimi po prostoru. Ko so mu raziskovalci odklopili sprednji senzor, se je začel zaletavati v kocke in jih potiskati naprej. Ko so med potiskanjem neprednji senzori zaznali oviro (zid ali kocko), je robot nehal potiskati kocko pred seboj in se obrnil. To je povzročilo, da je robot praktično pospravljal kocke, ki so končale zložene v kupih. Pospravljanje je torej vzniknilo skozi zelo preproste elemente prostora in samega robota.

Raziskovalci UII se pri modeliranju večkrat še neposredneje zatečejo v naravo. Primer tega je robotski črček (Webb, 1995). Črčki se reproducirajo tako, da samice iščejo samce, ki se najglasneje oglašajo. Med seboj povezane bobenčke imajo na sprednjih nožicah, ki so povezane z vmesnimi nevroni za obračanje. Celoten sistem črčka se odziva le na specifične frekvence. Samice se torej vedno obrnejo proti samcu, ki je najglasnejši. Ravno to so raziskovalci reproducirali v robotskem črčku z majhnim številom zelo osnovnih mehanizmov.

Kompleksnost modelov s takšnimi pristopi seveda variira. Tako poskušajo raziskovalci odkriti različne rezultate emergence s kompleksnim oblikovanjem človeških mišic (emergenca sistema tetiv, ki uravnavajo kopico dejanj, za katera se ne potrebuje možganov *per se*), oči (emergenca paralakse gibanja), koordinacijskih vzorcev (emergenca preživetja v Sahari) itn. (Pfeifer et al., 2007).

Modeliranje nižje kognicije v UII je prek koncepta emergence pokazalo, da lahko kompleksno obnašanje vznikne iz le nekaj preprostih komponent robota, ki je postavljen v pravo okolje (prav tam). To je doseženo brez eksplicitnih reprezentacij ali neposrednega programiranja takšnega vedenja. Modeli torej kažejo na to, da lahko koncepta utelešenosti in umeščenosti nadomestita kompleksne notranje algoritme in vnaprej podano znanje.

4.2 Primeri modeliranja višjih kognitivnih procesov

Področje modeliranja višjih kognitivnih procesov je precej drugačno kot prejšnje. Raziskovanja je veliko manj, čemur botrujejo tudi težave pri modeliranju takšnih procesov (na kar opozarjajo tudi kritiki, večinoma pristaši SDUI) – enega izmed razlogov za to navaja Clark (1997). Takšni procesi naj bi zahtevali več reprezentacij (*representation-hungry problems*). Emergenca ni tako očitno v ospredju, velikokrat je prisotna kompleksna kognitivna arhitektura (iz UNM), vedenje modelov je manj očitno, manj vidno. Zaradi specifičnosti vseh najrazličnejših procesov je modeliranje izredno osredotočeno, razpršeno in raznoliko.

Med modeli višjih kognitivnih procesov se najde nekaj takšnih, ki poskušajo svoj cilj doseči s pristopi modelov nižjih kognitivnih procesov (emergenca skozi zelo osnovno mehansko delovanje robota v povezavi z okoljem). Primer tega je bil ambiciozen projekt Brooksa, ki je skupaj s filozofom Dennettom ustvaril robota Coga (Dennett, 1997). Njegov cilj je bil nivo človeške inteligence (predvsem govor, manipulacija predmetov, raziskovanje, učenje itd.), do katere bi prišlo popolnoma od »spodaj-navzgor«. Na odpadu je končal leta 2003, brez pretiranega pompa okoli neuspeha. Vznikanje kompleksnega obnašanja, ki bi se približalo višjim kognitivnim procesom, najdemo sicer že v 1980-ih, in sicer v poznanih Braitenbergovih vozilih (Braitenberg, 1984). Braitenberg je bil psiholog, ki je trdil, da lahko z roboti simuliramo miselni eksperiment, kjer čustveno (psihološko) obnašanje vznikne skozi interakcijo med zelo preprostimi telesi več agentov (ali med agentom in okoljem) z le enim senzorjem in enim aktuatorjem. Njegovo na-

čelo je bilo, da je sinteza lažja kot analiza – ustvariti nekaj, kar izgleda kompleksno, je lažje, kot analiza nečesa, kar izgleda kompleksno. Tako je robota s senzorjem, ki ob zaznavi svetlobe sorazmerno požene kolesa, dokler svetlobe ne zaznava več, strah svetlobe. V njej je viden, kar ogroža njegov obstoj. Zelo preprost sistem simulira torej strah. Če bi senzor in motor robota povezali tako, da bi se robot premikal proti svetlobi in v njegovo okolje postavili še enega robota, ki je enako zgrajen, bi dobili ljubezen. Obema bi dodali le še en aktuator – lučko. S tem bi se privlačila. Tako nekako je razmišljal Braitenberg.

A takšni modeli delujejo naivno in ne oponašajo procesov, ki se dogajajo v ljudeh. Večina raziskovalcev vseeno sledi temu, da višji kognitivni procesi izvirajo iz nižjih. Velikokrat je le vprašanje, kaj je tisto, kar premosti ta most med obema vrstama procesov. Kaže se močan trend, ki trdi, da je jezik tisti, ki nižjo kognicijo potisne do višje (Mirolli in Parisi, 2011). Jezik naj bi nižjo kognicijo vodil do učenja, kategorizacije, abstrakcije, spomina, samokontrole in bogatega mentalnega življenja. Veliko raziskovalcev višjih kognitivnih procesov v UUI prav tako trdi, da so rešili problem ozemljitve simbolov (*symbol grounding problem*; Steels, 2008) – nekateri pri tem povzdigujejo prav vlogo jezika kot kognitivnega orodja, ki pripelje do vznika pomena s senzomotoričnem delovanjem v okolju (Farkaš et al., 2012). Gre za modele, ki v senzomotorični interakciji z okoljem, predmeti v njem in jezikovnimi signali, uspejo povezati jezikovni signal z nekim dejanjem. Do tega pride z učenjem (konkretno gre za spreminjanje uteži (umetnih sinaps) v UNM) od »spodaj-navzgor«.

Druge raziskovalne skupine jeziku ne pripisujejo takšne pomembnosti in modelirajo agente, ki naj bi konstruirali pomen brez jezika. Pomembno jim je, da agent skonstruirajo pomene, ki so pomembni zanj – do tega pride s pomočjo kognitivnih procesov, preteklih izkušenj in njihovih interakcij z okoljem. S tem naj bi prišlo do utelešenega znanja iz zaznavanja in delovanja v okolju. Primer takšnega raziskovanja je model, ki je postavljen v okolje z igračami, sadjem in pohištvom (Takáč, 2008). Agent lahko med raziskovanjem v vsakem koraku naključno izbere predmet in ga ali dvigne ali pa položi, to pa povzroči spremembe različnih parametrov (sila, stopinje premikanja agentovih aktuatorjev). V primeru dvigovanja se predmetu poveča vertikalna komponenta, če je sila agenta večja od teže predmeta, sicer dejanje nima efekta. V primeru spuščanja se predmetu zniža vertikalna komponenta glede na spremembo stopinje premikanja, a le, če predmet ni na tleh. Po enem dejanju agent zapiše povzročene spremembe v obliki kavzalnih razmerij med dejanji, predmeti in dinamiko med obema. Tako ustvari različne kategorije, ki so med sabo različno povezane, pomen pa tvori predmet, povezano dejanje, kavzalno razmerje med njima itd. Gre torej za neke vrste »danosti« (*affordances*) delovanja v prostoru, zgodovine tega delovanja, odnos med agentom in predmeti ... torej širokim diapazonom odnosov, ki se skujejo med agentom in okoljem, in ti za agenta predstavljajo pomen nekega predmeta.

Nekateri krogi raziskovalcev poskušajo napraviti še korak dlje, kar se tiče abstraktnosti, in modelirati doživljanje (Schillaci et al., 2016). Ti prav tako trdijo, da doživljanje izhaja iz senzomotoričnih interakcij z okoljem. Agent se uči uporabljanja orodij s poskušanjem ali opazovanjem človeka, izkušnje učenja pa je sposoben tudi podoživljati. Z uporabo nižjih kognitivnih procesov rekreira nekakšen svet, po katerem vznikne doživljanje. To rekreiranje služi tudi učinkovitejšemu delovanju v svetu, npr. uporabljanju orodij.

Področje UUI, ki se ukvarja z modeliranjem višjih kognitivnih procesov, je torej izredno raznoliko in razkropljeno. Skupna točka je nenehen pritisk kritikov, ki so večinoma zagovorniki SDUI. Težko je namreč modelirati višje kognitivne procese brez privzemanja nekega vnaprej podanega znanja; redki so modeli, ki bi bili izključno »spodaj-navzgor«. To krha predpostavke UUI, da kognicija izhaja od »spodaj-navzgor«. S tem se ponovno pojavi vprašanje pomena, ker do tega ne pride od »spodaj-navzgor«, temveč ga, kot v SDUI, določijo programerji. Vprašanje pomena vodi do vprašanja avtonomije, ki je zaradi istih razlogov modelom ne gre pripisovati – ti se namreč ne vedejo glede na notranje cilje, ki bi bili resnično pomembni za njih same, in ne za njihove kreatorje. Tovrstne težave vodijo do kritike, da UUI preprosto ni dovolj utelešena – ali vsaj ne pravilno (Froese, 2007).

5. Filozofija in potreba po novih paradigmah

UUI, ki le težka pojasnjuje višje kognitivne procese le na podlagi svoje paradigme, je zato vplivala na filozofske prakse, ki se ukvarjajo s tovrstnimi vprašanji. Nekateri filozofi so raziskave UUI pripeljale do trditev, da um ne služi za razmišljanje, kot je tradicionalno okarakteriziran, temveč za dejanja, za to, da stori- mo tisto, kar v resničnem svetu pod časovnim pritiskom pač moramo (Clark, 1997), da je senzomotorika fizično vpletena v konceptualno procesiranje (Rupert, 2009) ali pa to, da je vidno zaznavanje nekaj, kar počnemo (Hickerson, 2007) oz. da sestavlja našo zavest (Gangopadhyay et al., 2010). V splošnem je prišlo do pojava močnih kritik v prid simbolni paradigmi ali pa do radikalizacije pogleda na utelešenost (Chemero, 2009). Nastali sta dve šoli misli: kompatibilistična in puristična (Calvo in Gomila, 2008). Prva trdi, da lahko za razlago različnih kognitivnih procesov uporabljamo različne metode – nekateri procesi bodo razloženi s simbolnimi, drugi pa z utelešenimi. Druga pa popolnoma zavrača simbolno paradigmo in se zavzema za to, da bi bile za raziskovanje kognicije rabljene izključno utelešene metode, brez kognitivističnih teorij (npr. teorija dinamičnih sistemov, konceptualna integracija idr.).

Z začetkom 21. stoletja se zdi, da se je raziskovanje na področju UUI nekoliko ustavilo. Razlogov za to je več. Prvič, glavni akterji so prenehali z udejstvovanjem znotraj področja (Ziemke je nehal s pisanjem člankov, Brooks je ustanovil podjetje iRobot in izdeluje Roombe, Pfeifer, ki je konec 90. prinesel področje v Evropo, se je upokojil itd.). Drugič, samo področje se je močno usmerilo, tj. izoblikovala sta se dva tabora: tehnični (Bongard, Lipson, Floreano) in filozofski (Di Paolo, Froese). Tretjič, področje še vedno ni paradigmatsko poenoteno. In četrtič, pojavile so se močne kritike (in ne s strani zagovornikov SDUI) prevladujočih trendov UUI in UK.

Kritikam Dreyfusa (2007), ki se od 1970-ih niso pretirano spremenile in še vedno opozarjajo, da je UI premalo heideggerska (slednje velja tako za UUI kot za trende UK), so se pridružili predvsem zagovorniki iz filozofskega tabora UUI (Di Paolo, Froese). Gre za kritike, da roboti v UUI še vedno ne ustvarjajo resničnega pomena, pomena, ki bi bil bistveno pomemben za njih same, ki bi izhajal iz njih samih glede na to, kaj potrebujejo za preživetje (Dreyfus, 2007). Heideggerjeve ideje torej niso dovolj realizirane – kljub temu da znanje *da*, vnaprej podano znanje, ni več osrednje, se zdi, da morebitno znanje *kako*, ki naj bi ga zdaj roboti imeli, ni znanje njih samih (temveč prej programerjev). Kot stvari stojijo, je pomen, ki naj bi vzniknil v robotih, pomemben za programerje, ki robotu zadajo to nalogo – s tem tudi same ozemljitve simbolov ni v robotu. Cilje postavijo programerji in ne izvirajo iz robotov.

Ta težava pomena je povezana še z enim konceptom, ki je ključnega pomena: avtonomijo.⁷ Avtonomija je koncept, ki se dandanes v širši UI pojavlja zelo trivialno – če ima robot ločeno telo od svojega kreatorja in če lahko sam izpolnjuje določene naloge v prostoru (kot je seveda vnaprej določen), potem je avtonomen. A avtonomen je lahko le, če cilji izhajajo iz njega samega in so le-ti pomembni zanj – ne za njegovega kreatorja –, šele takrat pa seveda lahko vznikne pomen, ki bo imanenten robotu samemu (Kolenik, 2016). Trenutni trendi UUI tega ne omogočajo niti ne raziskujejo v pravi smeri, saj se raziskovalci ne zavedajo svoje zmote v razmišljanju, da je namreč ustvarjanje pomena lahko ukazano, vsiljeno od njih (Froese, 2007). Zdi se torej, da filozofija še drugič v zgodovini UI zadaja tako močne kritike, da so paradigmatske spremembe nujne (Froese in Ziemke, 2009). Tako Dreyfusova heideggerska kritika iz 1970-ih. še vedno velja: modeli trenutno ne »najdejo pomena v svetu in se ne redigirajo v prepoznavanju za njih relevantnih delov sveta glede na njihovo obliko telesa, način premikanja, osebno in družbeno identiteto itd.« (Dreyfus, 2007: 265).

⁷ Avtonomijo definiram kot določanje lastnih zakonitosti.

6. Zaključek

Morda smo vsi zgrešili neko organizacijsko načelo bioloških sistemov ali pa vsaj neko splošno resnico o njih. Morda obstaja način, ki nam bo preko opazovanja bioloških sistemov omogočil vpogled v to inherentno nujnost interakcij med deli le-teh, ki je do sedaj popolnoma umanjkal iz sistemov umetne inteligence. /.../ Mislim, da trenutno preprosto ne razumemo in da je potrebna bistvena sprememba v našem mišljenju /.../. (Brooks, 1997: 300)

S temi treznimi besedami je Brooks pospremil kritike in očitne težave, s katerimi se bori UUI. Slednja je sicer več kot le močno stopila na prste simbolni paradigmi in SDUI ter pokazala, da delovanje v resničnem svetu ne mora potekati od »zgoraj-navzdol« s predprogramirano knjižnico znanja, ker duševnost pač ni manipulacija z arbitrarnimi simboli, ki ima že vnaprej postavljena pravila. A pristopi brez reprezentacijskega odzivanja na dražljaje od »spodaj-navzgor« ali obravnavanja uma kot simulacije nižjih kognitivnih procesov ali katerega podobnega tretjega trenda, ki se podpisuje pod termin UK, težka razlagajo višje kognitivne procese, vsekakor pa ne rešujejo globljih vprašanj avtonomije in pomena. Vračanje k SDUI ali nekakšna radikalizacija UUI se zdita le še več istega, saj se še vedno ne pripoznava pomembnost netrivialne avtonomije, iz katere šele lahko izhaja pomen. Brooks je, kljub svojim zmotam, imel prav: potrebna je bistvena sprememba v razmišljanju. In ravno to ponujajo nove paradigme (ali »resnične«, »prave« paradigme UK), ki se razvijajo v krogih filozofije in UI. Takšni sta, med drugim, enaktivizem (Varela idr., 1991) in umetno življenje (*artificial life*).

Enaktivizem bi lahko šteli med te »prave« utelešene paradigme. Med prvimi je namreč močno poudaril pomembnost telesa in njegovega odnosa z okoljem, a so se njegove ideje nekoliko izgubile, omilile ali nekako strivializirale, še posebej v UUI (Rohde, 2010). Enaktivizem, ki briše ločnico med spoznavajočim subjektom in spoznavanim objektom, poskuša rešiti težavo pomena prav preko razreševanja vprašanja avtonomije. Vzrok avtonomije je po enaktivizmu *autopoiesis* (Maturana in Varela, 1984). *Autopoiesis* označuje sistemov (biološki) proces samoustvarjanja in samoorganizacije preko neprestanega proizvajanja sposobnosti proizvajanja samega sebe. To povzroči, da je sistem avtonomen. In le avtonomen sistem lahko ustvarja pomen, ki je ozemljen v doživljanju, v *Daseinu*. Umetno življenje kot področje UI veliko pozornost posveča adaptaciji, ki je nekako enačena s kognicijo. Če je torej sistem zmožen avtonomne adaptacije, gre za kognitiven sistem. Ta avtonomija mora biti posledica okoljskih pogojev in ne vnaprej pripravljenega robota (Froese in Ziemke, 2009). Obe paradigmi sicer predstavljata manjšino znotraj svojih področij, zaradi relativne mladosti pa tudi nista tako dobro razdelani kot npr. simbolna paradigma. Kljub temu predstavljata možnost razrešitve nekaterih problemov, ki pestijo filozofijo, UI in kognitivno znanost že dolgo časa (Rohde, 2010).

What Do Heidegger and Roomba Have in Common? Embodied Artificial Intelligence as a Playground for Continental Philosophy

Roomba, the robotic vacuum cleaner, would not exist today without the rich history of philosophy, which presents the groundwork for traditional artificial intelligence. The latter sees cognition as an arbitrary symbol manipulation in its “mind as computer” metaphor. This view led to many problems, such as the “frame” and the symbol grounding problem, which sparked the AI revolution. The ideas of Heidegger, Merleau-Ponty and of other continental philosophers, which were realized by the roboticist Brooks, Roomba’s creator, were among the driving forces of embodied artificial intelligence (EAI). Its basis is the emergence of cognition from the sensorimotor loop between an agent and the environment. However, the immensely dispersed field of EAI is without common paradigmatic grounds. Aware of its problems regarding autonomy and meaning, this paper focuses on the philosophy-EAI relationship. It identifies philosophy as crucial for (E)AI, and offers a review of current trends and real-life examples. The paper concludes with a critique of current trends, and suggests some solutions originating from philosophy.

Keywords: embodied artificial intelligence, embodied cognition, phenomenology, traditional artificial intelligence, enactivism

Literatura

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, Mass.: MIT Press.

Bringsjord, S. in Arkoudas, K. (2014). »Philosophical Foundations«. V Ramsey, B. in Frankish, K. (ur.), *The Cambridge Handbook of Artificial Intelligence*, New York: Cambridge University Press, str. 34–63.

Brooks, R. A. (1995). »Intelligence without reason«. V Steels, L. in Brooks, R. A. (ur.), *The artificial life route to artificial intelligence: Building embodied, situated agents*, Hillsdale: Lawrence Erlbaum Associates, Inc., str. 25–81.

Brooks, R. A. (1997). »From earwigs to humans«. *Robotics and Autonomous Systems*, 20(2-4), str. 291–304.

Calvo, P. in Gomila, T. (2008). *Handbook of Cognitive Science: An Embodied Approach*. Oxford: Elsevier Ltd.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, Mass.: MIT Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass.: MIT Press.

Crevier, D. (1993). *AI: The Tumultuous Search for Artificial Intelligence*. New York: Basic Books.

- Dennett, D. (1997). »Consciousness in human and robot minds.« V Ito, M., Miyashita, Y. in Rolls, E. T. (ur.), *Cognition, computation and consciousness*, Oxford: Oxford University Press, str. 17–30.
- Dreyfus, H. L. (1972). *What Computers Can't Do*. New York: MIT Press.
- Dreyfus, H. L. (2007). »Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian«. *Philosophical Psychology*, 20(2), str. 247–268.
- Farkaš, I., Malík, T. in Rebrová, K. (2012). »Grounding the meanings in sensorimotor behavior using reinforcement learning«. *Frontiers in Neurobotics*, 6(1).
- Froese, T. (2007). »On the Role of AI in the Ongoing Paradigm Shift within the Cognitive Sciences«. V Lungarella, M., Lida, F., Bongard, J. in Pfeifer, R. (ur.), *50 Years of Artificial Intelligence*, Berlin: Springer-Verlag, str. 63–75.
- Froese, T. in Ziemke, T. (2009). »Enactive artificial intelligence: Investigating the systemic organization of life and mind«. *Artificial Intelligence*, 173(3-4), str. 466–500.
- Gangopadhyay, N., Madary, M. in Spicer, F. (ur.). (2010). *Perception, Action and Consciousness*. New York: Oxford University Press.
- Harnad, S. (1990). »The symbol grounding problem«. *Physica D*, 42, str. 335–346.
- Harvey, I. (2000). »Robotics: Philosophy of Mind Using a Screwdriver«. V Gomi, T. (ur.), *Evolutionary Robotics: From Intelligent Robots to Artificial Life*, Ontario: AAI Books, str. 207–230.
- Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J. in Davidson, R. J. (2010). »Cosmetic Use of Botulinum Toxin-A Affects Processing of Emotional Language«. *Psychological Science*, 21(7), str. 895–900.
- Heidegger, M. (1953). *Being and Time*. Tübingen: Max Niemeyer Verlag.
- Hickerson, R. (2007). »Perception as Knowing How to Act: Alva Noë's Action in Perception«. *Philosophical Psychology*, 20(4), str. 505–517.
- Kolenik, T. (2016). »Embodied Cognitive Robotics, the Question of Meaning and the Necessity of Non-Trivial Autonomy«. V Markič, O., Gams, M., Strle, T. in Kordeš, U. (ur.), *Cognitive science: proceedings of the 19th International Multiconference Information Society – IS 2016*, Ljubljana: Institut Jožef Stefan, str. 24–27.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

- Liao, J., Beal, D. N. Lauder, G. V. in Triantafyllou, M. S. (2003). »Fish exploiting vortices decrease muscle activity«. *Science*, 302, str. 1566–1569.
- Markič, O. (2010). *Kognitivna znanost: filozofska vprašanja*. Maribor: Aristej.
- Maturana, H. R. in Varela, F. J. (1984). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, Mass.: Shambhala.
- McGeer, T. (1990). »Passive Dynamic Walking«. *The International Journal of Robotics Research*, 9(2), str. 62–82.
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. New York: Humanities Press.
- Merleau-Ponty, M. (1963). *The structure of behaviour*. London: Methuen.
- Mirolli, M. in Parisi, D. (2011). »Towards a Vygotskian cognitive robotics: The role of language as a cognitive tool«. *New Ideas in Psychology*, 29, str. 298–311.
- Noë, A. (2005). *Action in Perception*. Cambridge, Mass.: MIT Press.
- Pfeifer, R., Bongard, J. in Grand, S. (2007). *How the body shapes the way we think: A new view of intelligence*. Cambridge, Mass.: MIT Press.
- Rizzolatti, G. in Arbib, M. (1998). »Language within our grasp«. *Trends in Neurosciences*, 21, str. 188–194.
- Pulvermüller, F. (2005). »Brain mechanisms linking language and action«. *Nature Reviews Neuroscience*, 6, str. 576–582.
- Rohde, M. (2010). *Enaction, Embodiment, Evolutionary Robotics*. Berlin: Springer.
- Russell, S. in Norwig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. New York: Oxford University Press.
- Schillaci, G., Hafner, V. V. in Lara, B. (2016). »Exploration Behaviors, Body Representations, and Simulation Processes for the Development of Cognition in Artificial Agents«. *Frontiers in Robotics and AI*, 3.
- Searle, J. (1980). »Minds, brains, and programs«. *Behavioral and Brain Sciences*, 3, str. 417–424.
- Steels, L. (2008). »The symbol grounding problem has been solved, so what's next?«. V de Vega, M., Glenberg, A. in Graesser, A. (ur.), *Symbols and Embodiment: Debates on Meaning and Cognition*, Oxford Scholarship Online, str. 223–244.
- Stolorow, R. D. (2006). »Heidegger's investigative method in *Being and Time*«. *Psychoanalytic Psychology*, 23(3), str. 594–602.

Takáč, M. (2008). »Autonomous construction of ecologically and socially relevant semantics«. *Cognitive Systems Research*, 9, str. 293–311.

Varela, F. J., Thompson, E. in Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, Mass.: MIT Press.

Webb, B. (1995). »Using robots to model animals: a cricket test«. *Robotics and Autonomous Systems*, 16, str. 117–134.

Wilson, A. D. in Golonka, S. (2013). »Embodied cognition is not what you think it is«. *Frontiers in Psychology*, 4(58).

Wilson, M. (2002). »Six views of embodied cognition«. *Psychonomic Bulletin & Review*, 9(4), str. 625–636.

Zebrowski, R. (2010). »In dialogue with the world: Merleau-Ponty, Rodney Brooks and embodied artificial intelligence«. *Journal of Consciousness Studies*, 17(7-8), str. 156–172.

Ziemke, T. (2004). »Embodied AI as science: Models of embodied cognition, embodied models of cognition, or both?«. V Iida, F., Pfeifer, R., Steels, L. in Kuniyoshi, Y. (ur.), *Embodied Artificial Intelligence*, Heidelberg: Springer-Verlag, str. 27–36.