Dijakinja: Katarina Kolar

Gimnazija Vič, Ljubljana

NATEČAJ ODDELKA ZA FILOZOFIJO FILOZOFSKE FAKULTETE V MARIBORU:

»ALI LAHKO STROJ FILOZOFIRA: UMETNA INTELIGENCA IN FILOZOFIJA«

Vprašanje: "Ali lahko stroj filozofira?" deluje na prvi pogled domala preprosto, vendar že po hipnem premisleku dobi neverjetno težo. Filozofiranje je oblika zavesti in ko govorimo o umetni inteligenci in filozofiranju, v resnici iščemo odgovor na vprašanja problema duha in telesa, oz. se moramo do njih nujno opredeliti. Šele takrat postane iskanje odgovora na osnovno vprašanje smiselno. Kot ključno dilemo bom obravnavala možnost samozavedanja umetne inteligence ter možnost razumevanja pomena simbolov skozi Teorijo utemeljevanja simbolov, kjer se bom spraševala, če je zgolj razumevanje pomena simbolov dovolj za uspešno filozofiranje, ki sem ga utemeljila po lastni definiciji.

V svetu v katerem živimo, lastnost filozofiranja pripisujemo izključno ljudem, od ostalih živih bitij nas ločuje nekaj, kar smo ljudje "samodefinirali" kot človeško zavest. Človeška zavest misli, dvomi in se samozaveda, vse našteto predstavlja ključno lastnost filozofiranja kot takega. Smiselno bi bilo trdi, da so lastnosti filozofiranja hkrati tudi lastnosti zavesti in obratno. Vse definicije filozofije, od Sokrata naprej, ki trdi, da je filozofija človekova dnevna aktivnost, pa vse do Marxa, ki piše o filozofiji kot o interpretaciji sveta z namenom, da svet spremenimo, vsebujejo skupne lastnosti filozofiranja, če le-to definiramo kot miselni procesa v okviru zavesti, znotraj katerega se filozofske zamisli rodijo. Ta mora obsegati kvaliteto samozavedanja, iz katerega izhajajo vse filozofske teorije nasploh; nemogoče je namreč razmišljati o človeku in svetu, če zanemarimo lastno eksistenco. Sposobnost samozavedanja je torej prva lastnost filozofiranja.

Druga takšna temeljna lastnost je interpretacija doživljanja okolja in našega zaznavanja, saj je to ključno (ne glede na to ali smo racionalisti ali empiristi) za razumevanje sveta in iskanja filozofskih resnic. Brez podatkov o svetu filozofiranje ne bi bilo mogoče, prav tako pa ne bi imelo posebnega smisla.

Naši doživljaji, izkušnje bazirajo na osnovnem čutnem zaznavanju, te podatke mi interpretiramo kot kvalije, ki jih fizika trenutno ni sposobna pojasniti. Kvalije so fenomenalne lastnosti stvari, so tisto izkustvo samo, ko s "fenomenološkim épochejem" (če si izposodim Husserlovo nomenklaturo) reduktivno odstranimo vso vedenje o kavi in dobimo tisto dišečo, slastno, brezkončno goščo. Izkustvo pitja kave se med posamezniki razlikuje in na takšen način delujejo prav vse fenomenalne lastnosti. S čutnim zaznavanjem lahko utemeljimo svet okoli nas, te celostne informacije pa pomembno vplivajo na razvoj filozofske misli, saj navdihujejo in služijo kot smiselni podporniki; bodisi argumenti bodisi protiargumenti za obstoječe filozofske teorije. Sposobnost čutnega zaznavanja skupaj z mislečo, samozavedajočo se zavestjo tvorijo eksistenco zavesti.

Tretja nujnost za zmožnost filozofiranja je razumevanje pomena simbolov. Filozofske teorije namreč postanejo filozofske teorije šele takrat, ko so primerno ubesedene. Če povzamem vse tri točke- osnovne lastnosti filozofiranja, bi dejala, da je za uspešno filozofiranje nujno potrebna zavest, ki vse tri lastnosti združuje. Temeljne lastnosti filozofiranja so trenutno prepoznane samo v človekovi zavesti in so iz nje tudi izpeljane. Če želimo ustvariti umetno inteligenco (v nadaljevanju krajšano z UI) s sposobnostjo filozofiranja, moramo biti sposobni ustvariti UI, ki se samozaveda, čutno zaznava okolje in razume pomen simbolov. Takoj se intuitivno vprašamo, če je vse našteto sploh mogoče. Samozavedanje je modifikacija samo in izključno človekove zavesti (človek je edino bitje, ki se samozaveda). Tako ugotovimo, da ob postavljanju vprašanja, ali je UI sposobna filozofirati, v resnici iščemo odgovor na vprašanje, ali je moč poustvariti človekovo zavest.

David Chalmers smiselno definira enostavni in zahtevnejši problem človekove zavesti. V okviru enostavnega problema preučuje korelacijo med fizičnim in mentalnim, zahtevnejši problem pa obravnava vpliv fizičnega dela- procesov v možganih na mentalni del- torej na zavest. Za zahtevnejši problem trenutno nimamo rešitve, vendar je Chalmers optimističen in trdi, da bomo po vsej verjetnosti odgovore v prihodnosti našli (kot so podobne vrzeli v znanju že zapolnili naši predniki).

Prihodnostne rešitve zahtevnejšega problema zavesti bodo, kot Chalmers opozarja, zelo abstraktne narave (saj so vendar rešitve zahtevnejšega problema). Rešitve naj bi bile tako abstraktne, kot je abstraktna kvantna fizika, torej jih v resnici sploh ne bomo razumeli. Saj vendar pravijo, da tisti, ki kvantno fiziko razumejo, o kvantni fiziki nimajo pojma. Primer Schrödingerjeve mačke in superpozicija, torej dejstvo, da je lahko mačka hkrati živa in mrtva, krši osnovne pojme enotnosti izkušnje in zaznavanja, tudi enotnost eksistence. Da bi lahko to dejansko razumeli, bi morali sami delovati na bazi kvantne fizike, kar pa se ne dogaja neposredno. Neproduktivno bi bilo trditi, da bomo skozi evolucijo v celoti prešli na kvantno osnovo; to bi pomenilo zgolj apatično spremljanje poteka evolucije. Če bi se čez časa potem dejansko spremenili v kvantne delce (kar je že skoraj predpostavka predpostavke), bi postali čista zavest brez fizične podlage in domnevam, da nas UI več ne bi zanimala. Kvantna fizika je človeku dejansko nedoumljiva. Teorija povezanih kvantnih delcev jasno krši temeljno vzročno načelo in obstoječe znanstvene teorije, ki smo jih mnogokrat poskusno dokazali. Povezani kvantni delci so begali celo Einsteina, ki je teorijo označil s pridevnikom “spooky”.

Zaključek je ta, da zavesti ne bomo uspeli poustvariti, to pa zato, ker je ne bomo sposobni razumeti. Ne trdim, da obstaja drugačna substanca zavesti (nekakšen dualizem substanc), čeprav se na tem mestu resničnost te izjave ne bi zdela tako absurdna. Možno je, da bomo to razlago približno razumeli, vendar ne na vzročno-posledični način kot razumemo Newtonovo fiziko vsakdana, ki jo definiramo kot “logično”. Zatorej se trditev Davida Chalmersa in vseh filozofov- fizikalistov, ki trdijo, da bo mogoče zavest simulirati, takoj, ko jo bomo lahko opisali, zdi naivna. Smiselno se je posvetiti predstavljeni kritiki Chalmersove teorije, ki kot alternative na ponudi dejanskega dualizma, ampak lastnost človeka, ki ni sposoben razumeti tega, kar je sicer pred njim in se mu razkriva, do te mere, da bi lahko to poustvaril, se pa to kar se pred njim razkriva, še zmeraj tretira kot znanost.

S tem prva lastnost, ki bi jo UI potrebovala za uspešno filozofiranje, opade. Za uspešno filozofiranje je nujno potrebna vsebnost vseh treh lastnosti, zato je že jasno, da UI filozofiranja ne bo sposobna. Ravno to podkrepi tudi druga osnova filozofiranja- čutno zaznavanje. Kvalije sodijo v kategorijo zahtevnejšega problema zavesti, ki ga ni mogoče stimulirati v obliki UI. Razlaga je enaka kot v zgornjem primeru lastnosti samozavedanja.

Tretja točka, ki obravnava jezik; torej razumevanje pomena simbolov ponuja nekaj več kreativnih rešitev ter prostora za razmišljanje. Ključen problem je že leta 1980 izpostavil ameriški filozof John Searl v obliki kitajske sobe. Searle trdi, da je nemogoče, da bi bil program umetne inteligence kadarkoli sposoben razumeti dejanski pomen simbolov, ki tvorijo besede in stavke. Kljub močnemu argumentu, ki ga Searl ponudi, smo v zadnjem času priča pojavu neizmerno zanimivih raziskav, ki semantično razumevanje simbolov prikazujejo nekoliko drugače.

Stevan Harnard je s primerom kitajskega vrtiljaka opozoril na pomebnost okolja in referenc okoli nas. Ideja, ki se pojavlja je ta, da je za razumevanje pomena simbolov nujno, da UI pomen določi sama. Na voljo nima programa, ki bi zunanji svet na kakršnikoli način (pre)definiral od zunaj ali od znotraj, vse kar program vsebuje so sistemi, ki omogočajo operiranje s simboli. Tukaj ne gre za program UI, ki ima svoj osnovni namen inatično zapisan v obliki osnovnega algoritma, ki se popolnoma samosvoje izboljšuje z vsako novo informacijo, tako da se nevronsko omrežje nekoliko predrugači. Pripadniki Teorije utemeljevanja simbolov govorijo o robotih, mehanskih organizmih, ki preko različnih senzorjev zaznavajo okolje, se premikajo, tvorijo lasten odziv. Za interpretacijo podatkov na voljo nimajo neomejene količine informacij, kot jih ima algoritem googla (ta primerja in beleži vnose v brskalnik na globalni ravni in na takšen način izdeluje vedno bolj specifične in natančne profile uporabnikov ter interesnih skupin). Med vključenostjo v proces učenja, ki simulira ponavljanje, ki so ga deležni novorojenčki, ko se zavest počasi oblikuje v vedno bolj človeško obliko, bi roboti takšne vrste formulirali lastne odgovore na okolje. Znotraj tega procesa ne bi razvil človeške zavesti, ki opazuje samo sebe in ima do sebe neposredni, prvoosebni dostop, temveč inteligenco, ki je vezana zgolj na razumevanje simbolov. Slednje predstavlja nenavaden fenomen. Hipotetičen robot bi dejansko razumel, da je sonce sonce, vendar bi to dojemal bistveno drugače, ker bi tudi samo sonce doživel drugače. Doživel bi ga brez kvalij, torej kot določeno valovno dolžino barv, spremembo temperature, a vse to bi razumel in vedel da gre za sonce. Naše in njegovo dojemanje sonca bi se bistveno razlikovalo, celo do te mere, da bi lahko rekli, da gre za popolnoma edinstven primer zavesti. Ta bi bila najbližje človekovi, hkrati pa bi ji bila zato tudi najbolj tuja. Lahko bi rekli, da bi v tem primeru šlo za novo, prvo umetno “vrsto” bitja.

Če predpostavimo, da bi robot začel pojme, ki bi jih razumel, povezovati (na podlagi izkustva in posnemanja okolja) bi nastale prve (umetne) misli. Podobno tej novi vrsti umetne inteligence, ki trenutno obstaja v zgolj hipotetični obliki, je stanje v katerega človek zapade, ko je preobložen z delom in ima absolutni fokus; razume pomen simbolov, informacij, ki jih procesira, vendar se v tem stanju ne zaveda sebe in svojega obstoja. Tudi izkušnjo okolja dojema zgolj mehanično (kvalij se ne zaveda). Takšno stanje se lahko doseže tudi z določenimi tehnikami meditacije, v angleščini se zanj uporablja izraz “flow state”. Podobno bi deloval tudi takšen robot. Za razmišljanje bi uporabljal besede, katerih pomene, bi dognal individualno, ti

pa bi se, kot sem že poprej izpostavila, bistveno razlikovali od našega pojmovanja stvarnosti. Izrecno človeških pojmov, kot so čustva kot taka ne bi bil izkusil, zato bi njihovemu simbolu (na katerega bi med ljudmi slej kot prej naletel) pripisal (na/popačen) pomen.

Takšen robot bi razumel svet na edinstven način in ravno ta unikatna umetna perspektiva bi lahko človeku ogromno podarila. Svet bi razumel na način kot ga razumejo angeli v filmu Wima Wendersa; Nebo nad Berlinom. Ti pomene stvari v večini primerov razumejo, vendar ne poznajo življenja in človeškosti. Živijo namreč v črno-belem svetu, brez okusov, občutkov (kjer kvalije ne obstajajo), v svetu brez ljubezni, ki je kot angeli ne bodo nikoli razumeli (padlemu angelu Damielu se pojem ljubezen v svoji čistosti razkrije šele takrat, ko se "učloveči").

Nemogoče je, da bi se takšen robot kar naenkrat samozavedel (sploh zaradi zgoraj opisanih zapletov človekove nezmožnosti simulacije lastne zavesti zaradi paradoksalnega (ne)razumevanja le-te). Dokaz, ki to idejo potrjuje na drugačen način, je primer enojajčnih dvojčkov, ki si delita popolnoma enako okolje, izkušnje in DNA. Njuno življenje je popolnoma enako; enaki roboti ju istočasno hranijo, istočasno hodita na sprehode z usklajeno hojo, razpolagata z istimi informacijami. Kljub temu vsak izmed dvojčkov misli popolnoma drugače, njune misli so unikatne. Ekvivalenta stroja bi (ob izpostavljenosti istim pogojem) razmišljala enako. Ob identičnem osnovnem algoritmu in identičnih "inputih" lahko pričakujemo enake reakcije v obliki enakih misli (tudi zato, ker robota kvalij ne bi doživljala in bi se čutno zaznavni podatki popolnoma ujemali).

Če povzamem, je bistvo zapisanega razmišljanja to, da bi bil stroj sicer sposoben razmišljanja, ki pa bi se tako razlikovalo od človeškega, da bi lahko znotraj tega konteksta govorili o zavesti novega-umetno ustvarjenega bitja. Omenjena zavest bi bila seveda omejena na pomen simbolov in brez sposobnosti filozofiranja, katerega ključna postavka je človekova zavest, ki poleg razumevanja pomena simbolov zajema tudi samozavedanje in čutno izkustvo kvalij. Teh lastnosti se na UI ne da prenesti, ker jih nikoli ne bomo uspeli razumeti do te mere, da bi jih lahko simulirali na kakršnem koli substratu.