

Dijakinja: Zala Breclj

Gimnazija Nova Gorica, Nova Gorica

NATEČAJ ODDELKA ZA FILOZOFIJO FILOZOFISKE FAKULTETE V MARIBORU:

»ALI LAHKO STROJ FILOZOFIRA: UMETNA INTELIGENCA IN FILOZOFIJA«

Umetna inteligenca je postala prek raznovrstnih tako imenovanih »pametnih naprav«, kar so na primer pametni telefon, robotski sesalnik in podobno, vse bolj del našega življenja. Na nek način nam lajšajo marsikatero neljubo opravilo, saj je veliko enostavneje s pomočjo telefonske aplikacije sprogramirati robotski sesalnik, da ti vsak teden ali vsak dan ob vnaprej določenem času očedi bivalne prostore, kakor da to opravilo opravljaš sam. Življenja brez pametnih telefonov si velika večina prebivalcev razvitega sveta ne more predstavljati. Skoraj vsak učenec ali dijak si ne predstavlja ur matematike, fizike ali kemije brez pomoči žepnega računalnika. Takšne in drugačne naprave so, in glede na trende tudi bodo, pomemben del naših življenj, kar terja razmislek o njihovem statusu v odnosu do drugih entitet.

Pri odgovarjanju na vprašanje iz naslova moramo najprej premisliti pojma »stroji« in »filozofiranje« oziroma »mišljenje«, saj je konsenz o pomenu pojmov nujen za uspešen razmislek o katerikoli temi. Če si pod besedo »stroj« predstavljamo parno lokomotivo, stružnico ali kateri drugi stroj iz časa industrijske revolucije, ali pa računalnik, pametni telefon, računalnik ali otroškega robota, ki se lahko premika naokoli, antropomorfnega robotskega »dečka« Davida iz Spielbergovega filma A.I., njegovega robotskega medvedka Teddya ali računalnik Hal 9000 iz Kubrickove 2001: Vesoljske odiseje, se bomo težje sporazumeli o kriterijih za »mislečnost« strojev. David je namreč po zunanem izgledu deček, ne pa hladen stroj, zaradi česar bi mu po logiki antropomorfiziranja z lahkoto dali status »nekoga«-entitete z duhom, medtem ko bi za življenja znanstvenikov, posajenih v stanje nekakšne hibernacije, odgovornega Hala veliko težje uvrstili med misleče entitete, saj ni na videz niti malo podoben ljudem.

Zaradi pasti antropomorfizmov bi superračunalnik lažje uvrstili med entitete s statusom »nečesa«-entitete brez duha, ki se sama sebe ne zaveda, kar bi vplivalo tudi na naš odnos do njega. Tudi Hala sta Dave in Fran želela izključiti, ko sta ocenila, da ni tako zanesljiv, kot naj bi bil. Pri definiranju stroja ne moremo mimo dejstva, da smo zaradi togosti in zanesljivosti delovanja DNA, encimov in drugih osnovnih, s človeškim telesom povezanih procesov, tudi mi precej podobni strojem, kajti osmoza transportira vodo in druge snovi skozi fosfolipidni dvosloj oz. celično membrano enako togo in zanesljivo, kakor deluje računalnikovo električno vezje. Ta podobnost nakaže, da bi bili ob primernih kriterijih lahko stroji tudi ljudje. Vsakomur, ki zaradi življenja v razvitem svetu še predobro pozna avtomobile, se bo porodila asociacija na servisiranje in menjavo avtodelov ter na »smrt« odsluženega prevoznega sredstva.

Pomen glagola »misliti« oz. »filozofirati« bi lahko določili kot sočasno sposobnost mišljenja, sposobnost čutenja in zavest, ki so samo različne fasete istega fenomena in nobena ne more obstajati brez drugih dveh. Lepa ilustracija je primer (intelektualne) debate, saj mora vse vpletene tema zanimati, morajo jo razumeti in želijo si, da bi jih sogovorci ali poslušalci poslušali, razumeli in spoštovali zaradi tistega, kar so izustili. Da debaterje tema zanima, je dokaz debata sama, saj jo poganjajo skrite strasti in interesi, ki so samo drugo ime za cel spekter nezavednih naravnosti. Te namreč odražajo njihov slog in osebnost, ki izhaja iz neznanskega števila drobnih naklonjenosti, nagnjenj in naravnosti. Pri obravnavani temi del problema leži tudi v tem, da zaenkrat še ne znamo ločiti med fenomenom (npr. mišljenjem, znanjem kitajščine...) in njegovo simulacijo, saj imamo v vseh primerih opravka zgolj z nekakšnim izdelkom, »outputom«, ki nam ga posreduje telo oz. »hardver« in na podlagi katerega naj bi sklepali na »softver« -npr. duh. Na kakšne težave bomo pri risanju ločnice naleteli, nam nazorno prikaže miselni eksperiment »kitajska soba« ameriškega filozofa Johna Searla, saj nam pokaže, da lahko ustrezno logaritmiran program ustvarja tako kakovostne, prepričljive in pristne izdelke oz. »outpute«, da so lahko druge entitete (npr. ljudje) prepričani, da imajo opravka z osebo, ki kitajščino obvlada do potankosti, četudi ta oseba ne razume niti besedice v tem jeziku.

Iste dileme -ali ima obravnavana entiteta duha ali ne - se pojavijo pri temi tega eseja, saj ne poznamo instrumenta ali načina, s pomočjo katerega bi se lahko neposredno prepričali o njegovem obstoju. Tudi tukaj imamo namreč opravka zgolj z govorico telesa, mimiko obraza, glasovi, dotiki, vizualnimi dražljaji ..., ki jih proizvaja telo in so potemtakem nekakšen »output« ali izdelek, na podlagi katerega naj bi sklepali na »softver« oz. duha. Posredno sklepanje na duha bi bilo mogoče izpeljati s pomočjo Turingovega testa, saj tudi ta - tako kot detektor laži - meri odzive telesa, s temi pa nato ugotavlja, ali ima pred seboj lažnivca oz. entiteto z dušo. Njegova slabost je, da lahko z gotovostjo trdimo, da je vsako bitje ali stroj, ki ga je sposobno opraviti, entiteta z dušo; medtem ko za osebkke, ki ga ne opravijo, ne moremo ničesar zagotovo trditi. To pomeni, da bi lahko vsakemu bitju ali stvari ali stroju, ki bi prestalo takšno testiranje, dali nekakšno licenco, da zna misliti. Kar bi bilo sila podobno opravljanju vozniškega izpita, kjer ti, ko ga uspeš opraviti, priznajo, da znaš voziti motorno vozilo določene kategorije. Tako bi morale obstajati več različnih stopenj Turingovega testa, ki bi se razlikovale po času trajanja, zahtevnosti, globini vprašanj in izpraševalcu, saj že danes marsikatero bitje takšnega preizkusa ne bi opravilo. Za primer si pogledjmo psa ali mačka, ki naj bi sicer imela nekakšno zavest, vendar Turingovega testa ne bi opravila.

Najzahtevnejša raven preizkusa bi trajala najdlje, vprašanja bi bila najgloblja (npr. zadevala bi čustva, humor...), pa še izpraševalci bi bili kar se da kvalificirani. To bi pomenilo, da bi bili specializirani v najmanj opaznih segmentih govorice telesa, gibov, pomagali bi si z raznorodnimi napravami, ki bi merile najrazličnejše telesne odzive, spraševali pa bi po najljubših knjigah, različnih filmskih in glasbenih delih, pomenu tega ali onega velikega romana ali kakšnega drugega književnega dela, po rešitvah zapletenih romantičnih, prijateljskih ali družinskih situacijah, po smešnih prigodah, vicih vključno z njihovim prepoznavanjem med

pogovorom...skratka po vsem, kjer je potrebno razumeti čustva. Če bi se obravnavanec dosledno in vztrajno izmikal odgovorom na tovrstna vprašanja, bi lahko izpraševalec upravičeno posumil, da nima opravka z mislečim sogovornikom. Pri tem bi moral biti izkušen in pozoren, kajti mogoče bi bilo, da bi se entiteta odgovorom izmikala na zelo prefinjen način, npr. z odgovori v slogu »film me ne zanima«, »te pesmi ne poznam«, »nimam časa za branje, poslušanje ali gledanje«... Indikator, da izpraševanec ni sposoben mišljenja, bi lahko bili tudi neprimerno (premili ali pretirano intenzivni) odzivi na vprašanja, vezana na emocije, kakor je to mogoče videti pri Leonu v Iztrebljevalcu. Slednji se je namreč pri vprašanju o njegovi mami odzval pretirano agresivno, saj je ustrelil iztrebljevalca Holdna, s čimer se izda.

In vendar mislim, da osrednje vprašanje teme umetne inteligence in filozofije ni, ali lahko stroj filozofira, temveč kdaj oziroma na kateri točki smo določeni entiteti (npr. stroju) pripravljeni priznati, da filozofira. Vprašanje, ali stroj lahko filozofira, je po bistroumnosti (ali debilnosti) primerljivo z vprašanjem, ali stroj lahko šahira. Ko je leta 1997 v New Yorku IBM-ov računalnik Deep Blue premagal takratnega svetovnega šahovskega prvaka Garija Kasparova, je ves svet videl, da je ustrezno programiran stroj sposoben odigrati vrhunsko partijo, vendar so ostala vprašanja, kot so: »Je bil Deep Blue vesel svoje zmage?«, »Se je sploh zavedal, da šahira?« in enako bo z mislečimi stroji, ki bodo opravili Turingov test. Na vprašanje, ali stroj lahko filozofira, ni mogoče odgovoriti ontološko, temveč le tako, da dorečemo pojme »stroj«, »filozofirati« in mejo, pri kateri določeni entiteti priznamo, da lahko misli.

Poleg tega se lahko jutri zgodi, da bo nek stroj, ki smo ga že ali ga še bomo izdelali, napisal najboljšo filozofsko knjigo doslej ali pa bo kot Sokrat hodil po Ljubljani ali San Franciscu in sodeloval v izjemno kakovostnih debatah, vendar kaj se dogaja v njegovem duhu oz. ali kaj takega, kot je njegov duh, sploh obstaja pa ne bomo mogli nikoli neposredno preveriti, kajti opravka bomo vedno imeli zgolj z govorico njegovega telesa oz. njegovimi telesnimi reakcijami, torej z nekakšnim »outputom«. Sokratska filozofija je govorica telesa, saj imamo opravka s telesnimi odzivi (mimiko obraza, kretnjami rok, telesno držo, glasom, vidnimi in čutnimi dražljaji...), ne moremo pa videti v duha posameznika. Lahko se s tem sprijaznimo in na podlagi »outputa« kot kriterija določimo minimalne standarde za filozofirajoči stroj, ali pa smo večno frustrirani.