

Etični izzivi umetne inteligence: 47. srečanje Innsbruškega kroga moralnih teologov in socialnih etikov

Innsbruck, 2.–4. januar 2020

Letošnja konferenca moralnih teologov in socialnih etikov v Innsbrucku od 2. do 4. januarja 2020 je bila posvečena etičnim izzivom umetne inteligence. Okrog 40 profesorjev in raziskovalcev iz pretežno nemško govorečih dežel je razpravljalo o trenutnem stanju na področju razvoja umetne inteligence, o etičnih vprašanjih, povezanih z njeno uporabo, in o možnih posledicah uvajanja umetne inteligence za osebno in družbeno življenje v prihodnosti.

Uvodni referat je imel moralni teolog z Univerze v Luzernu, Peter Kirchschräger, ki je opozoril na korenite spremembe na trgu dela v bližnji prihodnosti. Cilj avtomatizacije ni olajšanje delovnega procesa, ampak nadomeščanje delovne sile. Po nekaterih prognozah se bo v prihodnjih 30 letih število plačanih delovnih mest znižalo od 20 do 50 %. Kirchschräger je izpostavil ambivalentno naravo umetne inteligence, ki je istočasno nekaj etično dobrega, saj bistveno olajšuje življenje in pomaga pri reševanju številnih problemov, po drugi strani pa človek postaja čedalje bolj odvisen od umetne inteligence. Tisti, ki obvladuje njeno delovanje, ima lahko nadzor nad številnimi posamezniki in tudi nad družbenim življenjem. Izpostavil je, da ne moremo govoriti o etičnosti umetne inteligence, saj sami računalniški programi nimajo vesti, ne zavesti, ne svobode, ne odgovornosti. Odgovornost ima vedno tisti, ki je ustvaril umetno inteligenco. Zato je nujno, da snovalci umetne inteligence upoštevajo etična načela, kakor so spoštovanje dostojanstva človeške osebe, prepoved diskriminacije in pravičnost, in ne sledijo zgolj načelom ekonomske koristi in učinkovitosti. Kirchschräger je razmišljal o možnosti razvoja tako imenovane močne umetne inteligence, ki bi samo sebe razvijala, in o tem, katera etična načela naj bi takšna avtonomna umetna inteligenca upoštevala. Ob tem se je vprašal, ali ne bi na koncu takšna oblika umetne inteligence tudi na etičnem področju preseгла človeka in bila bolj etična od njega.

Glede realnih možnosti za razvoj »močne« umetne inteligence, ki predpostavlja avtonomno odločanje in delovanje naprav, je bil zelo skeptičen profesor informatike in raziskovalec na področju umetne inteligence, Justus Piater z univerze v Innsbrucku. Za zdaj obstaja samo »šibka« umetna inteligenca. Po njegovem mnenju pri strojih ne moremo govoriti o inteligenci v pravem pomenu besede, saj stroji zgolj predelujejo in povezujejo podatke, niso pa zmožni mišljenja in razumevanja. Res je, da raziskovalci pri razvijanju umetne inteligence posnemajo proces učenja

pri otroku. Razlika med otrokovim učenjem in strojnim učenjem je v tem, da se stroj, potem ko se nekaj »nauči«, nikoli več ne zmoti. Ko robot osvoji tehniko, kako zadeti žogico v koš, nikoli več ne vrže mimo. Vendar tudi pri strojnem učenju ne moremo govoriti o učenju v pravem pomenu besede, saj je to zgolj optimizacija danih možnosti, ko programer predvidi cilj in dá stroju nalogo, naj čim hitreje in čim bolje doseže ta cilj. Prav tako je strojno učenje omejeno na neke določene specifične naloge, ki jih programer predvidi za posamezno napravo, in nikakor ni bistvo v tem, da bi, na primer neki robot, premogel splošno (vseobsegajočo) inteligenco, kakršna je značilna za človeka.

Kljub hitremu napredku – po prepričanju raziskovalca Piaterja – se v zadnjih desetih letih ni zgodil kvalitativni skok na področju razvoja umetne inteligence. Naprave so resda zmožne obdelati velikanske količine podatkov, niso pa zmožne proizvesti nečesa novega; to pomeni, da ne moremo govoriti o močni umetni inteligenci. Dejansko se naprave ne učijo konceptov, ne »razumejo«, kaj se dogaja, ampak zgolj zaznavajo in obdelujejo podatke. Ker digitalni sistemi ne razumejo vsebin in jih niso zmožni interpretirati, posredujejo kot relevantne tiste podatke, ki so najbolj množično obiskani oziroma za katere je nekdo plačal, da se lansirajo kot najbolj pomembni. Neutemeljena stališča imajo enako težo kakor argumentirana stališča, podkrepljena z dejstvi, saj stroji niso zmožni avtonomnega etičnega razločevanja. Znotraj internetne platforme, v kateri so razširjena številna socialna omrežja, obstaja velika nevarnost, da se bo zrušilo profesionalno novinarstvo, ki pomeni enega od stebrov sodobne demokratične ureditve. Glede trga dela je bil informatik Piater mnogo bolj optimističen od teologa Kirchschrägerja. Po njegovem prepričanju ni grožnje, da bi umetna inteligenca uničila večino delovnih mest, saj dosedanja zgodovina človeštva kaže, da človek vedno najde kreativne rešitve. Posamezna delovana mesta bodo res nadomestili avtomatizirani stroji, vendar se bodo oblikovali novi poklici.

Velik izziv za snovalce umetne inteligence je: pridobiti družbeno naklonjenost za uvajanje novih naprav, kakor so na primer samovozeča vozila in roboti na področju nege starostnikov, v vsakdanje življenje. Avtomatizirana vozila bi bila v marsičem lahko rešitev za prometne zamaške v številnih mestih, vendar ima večina prebivalcev še vedno strah pred takšnimi vozili, ker ne zaupa delovanju vozil brez človeškega nadzora. Tudi v negi bi bili roboti kot pomočniki pri čiščenju prostorov, pri dviganju s postelje in pri spremljanju starejših ljudi na sprehodih zelo dobrodošli, seveda pa podoba robotizacije pri negi starejših v nas spodbuja zelo negativne podobe, saj vključuje tudi grožnjo umanjkanja človeške bližine v zadnjem obdobju življenja.

Aaron Butler, znanstveni sodelavec in raziskovalec na univerzi v Luzernu, je predstavil nekaj izhodišč pri oblikovanju etike za umetno inteligenco. Bistvo je predvsem v programiranju strojev, da bi pri odločanju izbirali v skladu z etičnimi merili. Poleg že znanih dilem iz zgodovine etike, kakor je na primer tramvajsko dilema (*Trolley problem*), ko se je treba odločiti, ali bo tramvaj v neizogibni situaciji povozil eno ali pet oseb, je pri programiranju samovozečih vozil še mnogo podobnih zapletenih situacij, za katere si strokovnjaki prizadevajo, da bi izdelali

čimbolj popoln etični sistem. Uporaba umetne inteligence v bančništvu ali na sodiščih v marsičem nadomešča človeško odločanje in vnaprej pripravi nekatere odločitve, zato je pomembno, da bi ta proces vodila zelo izostrena etična merila. Pri tem se seveda postavlja vprašanje o končni odgovornosti pri oblikovanju nekaterih odločitev, predvsem zaradi nepreglednosti samega strojnega delovanja v procesu odločanja. V razpravi po Butlerjevi predstavitvi so številni diskutanti poudarili, da ne moremo govoriti o moralni odgovornosti strojev in tudi ne o etiki umetne inteligence. Prav tako je po mnenju razpravljavcev o avtonomiji mogoče govoriti samo pri človeku, ni pa možno govoriti o »elektronski osebi«, ki bi ji pripisovali neko določeno odgovornost pri podpisovanju pogodb ali dogovorov. V razpravi se je tudi razvila misel: želja po implementaciji etičnih načel v umetno inteligenco temelji na predpostavki, da je mogoče etiko prevesti v jezik matematike. Nekatere principe je resda mogoče prevesti v matematično govorico, večino moralnih načel pa izražamo v naravni govorici, ki je ni mogoče prevesti v matematični govor.

Zadnje predavanje je imel filozof Georg Gasser z univerze v Innsbrucku, ki je spregovoril o možnosti in etičnih izzivih uvajanja robotov v socialno življenje. Predstavil je izkušnjo iz dunajskega doma starejših, kjer imajo za »pomočnika« tudi robota negovalca, ki so ga poimenovali Henry. Henry, na primer, spremlja skupino starejših dementnih ljudi pri nordijski hoji na fizioterapijo. Fizioterapevti ugotavljajo pozitivne posledice robota Henryja na skupino, saj pomeni orientacijsko točko, počaka počasnejše, jih spodbuja, predvaja zeleno glasbo. Skupina in osebje sprejemajo Henryjeva rutinska opravila (čiščenje, pobiranje stvari), medtem ko imajo pri socialnih stikih (higiena) rajši nego človeške osebe. Henry ima težave pri interakciji z ljudmi, to pa starejše občasno frustrira, saj se Henry ne odziva v skladu z njihovimi pričakovanji.

Ugotovljeno je, da sam videz robota vpliva na sprejetost pri ljudeh. Če je videti podoben človeku, mu ljudje pripisujejo več možnosti, da bo imel tudi socialne kompetence. Vendar pa je nekaj zelo zanimivo: če je robot preveč podoben človeku, pri ljudeh upade želja po sodelovanju in se v njih sprožajo negativna čustva do robota (*uncanny valley effect* oz. učinek grozljive doline). Ljudje si želimo videti razliko in ne pristajamo na zavajanje, zato tudi ni verjetno, da bi bili lahko roboti kadarkoli socialni in čustveni partnerji ljudi. Iz tega Gasser sklepa: ni nevarnosti, da bi roboti nadomestili medčloveške odnose, saj niso sposobni dati socialnega priznanja človeku, ampak lahko zgolj izpolnijo neko določene želje. Roboti so vedno na voljo in jih je možno kontrolirati, nimajo osebnih potreb in želja; vendar pa ne zmorejo sprejeti drugega kot samostojno osebo in niso sposobni ustvarjati odnose. Nimajo sposobnosti sočustvovanja in prepoznavanja drugega v njegovi nemoči. Uvajanje robotov kot pomoč pri oskrbi in negi je po Gasserjevem mnenju dobrodošlo, vendar je nujno potrebno, da ljudi predtem ustrezno poučimo; podpišejo naj – podobno kakor pri posegih v zdravstvu – zavestni pristanek, da sprejemajo takšno obliko tehnologije v svoje življenje. Predvsem je pomembno, da se prepreči vsaka oblika zavajanja, saj roboti niso odnosna bitja, niso zmožni empatije in zato ne morejo nadomestiti človeške bližine.

V plenarni razpravi smo udeleženci ugotavljali, da etike ni mogoče prevesti v jezik matematike. Etika je več kakor samo upoštevanje nekih določenih pravil. Govorimo o sposobnosti presojanja in vrednotenja zaradi zmožnosti uživanja v neko konkretno situacijo. Tega pa nobena digitalna naprava ni zmožna. Etika je preveč kompleksna, da bi jo lahko prevedli v računalniške programe.

Človek, ki je ustvarjen po božji podobi, je ranljivo in nepopolno bitje in kot takšno je odprt za odnose z drugimi. Poleg matematične inteligentnosti njegovo razumevanje resničnosti sestavljajo še druge razsežnosti, kakor so čustvena, socialna, umetniška, duhovna inteligentnost, tega pa naprave z »umetno« inteligenco niso zmožne. Predvidevamo, da nam bo nadaljnji razvoj na področju umetne inteligence v marsičem olajšal življenje, nikakor pa nam ne bo mogel nadomestiti tistega najglobljega, kar je položeno v središče človeške osebnosti: želje po odnosu, po sprejetosti, po ljubezni, po tem, da bi ljubili in bili ljubljene. Tega ni mogoče živeti v digitalni obliki, nujno je potrebna dejanska telesna navzočnost drugega.

Udeleženci smo za srečanje prihodnje leto določili naslov: Skupno dobro.

Roman Globokar