

Privid sočutja

Zaradi razmaha virtualnih terapevtov, ki jih poganja umetna inteligenca, se postavlja več etičnih vprašanj

Piše: Gregor Kocijančič

Ilustracija: Tomaž Lavrič

V času, v katerem vse več ljudi doživlja duševne stiske, v katerem je prava psihoterapija zaradi preobremenjenosti zdravstvenega sistema vedno težje dosegljiva in v katerem ima zmeraj več ljudi do umetne inteligence takšna ali drugačna pričakovanja, kot gobe po dežju vznikajo klepetalni roboti (chatboti), katerih poslanstvo je terapevtska in psihološka podpora uporabnikom v stiski. Priljubljene aplikacije, kot so Woebot, Wysa in Youper, pa tudi dobri stari ChatGPT obljublajo dostopno, hitro in diskretno pomoč pri različnih duševnih težavah, skupno pa imajo več kot deset milijonov zvestih uporabnikov. Virtualna kvazi (psiho)terapija je postala cvetoč posel.

Mnenja o jezikovnih modelih in klepetalnih robotih, ki naj bi opravljali vlogo psihoterapevtov in psihologov, se izrazi to razlikujejo. Nekateri v njih vidijo revolucionarno orodje, ki lahko zapolni vrzeli v preobremenjenem zdravstvenem sistemu in ponudi vsaj osnovno podporo, kadar ljudem strokovne pomoči ne uspe poiskati zaradi njene nedostopnosti ali zaradi družbene stigme, povezane z duševnimi boleznimi, stiskami in motnjami.

Tudi psihologinja Lilijana Šprah, predstojnica Druženomedičinskega inštituta ZRC SAZU in raziskovalka na njem, v virtualnih terapevtih sicer prepoznava nekatere koristi, denimo psihično razbremenitev brez strahu pred obsojanjem, cenovno učinkovitost, anonimnost in nenehno dostopnost. Vendar sočasno poudari, da čeprav je ta tehnološki trend obetaven, je tudi potencialno tvegan. Prepričana je, da psihoterapevtski odnos temelji na varnem prostoru, empatiji, zaupanju in čustveni regulaciji v odnosu s psihoterapevtom, česar virtualni terapevti niso zmožni. »Algoritem chatbotov lahko simulira čustvene odzive, ne zmore pa pristnega

Posamezniki, ki v kvaziterapevtski komunikaciji z umetno inteligenco najdejo uteho, lahko zaradi tega odlašajo z iskanjem prave strokovne pomoči.

sočutja, prepoznavanja obrambnih mehanizmov in dolgoročnega spremljanja procesa zdravljenja,« pravi. In doda: »Izguba teh ključnih vidikov psihoterapije lahko privede le do površinskega samozdravljenja, to pa nikakor ne obravnava globljih vzrokov težav v duševnem zdravju.«

Krepitev izkrivljenih vzorcev

Lilijana Šprah opozori tudi na nevarnost, da se pri komunikaciji s terapevtskimi klepetalnimi roboti okrepijo izkrivljeni vzorci mišljenja, saj jezikovni modeli temeljijo na verjetnostnih vzorcih besedila, ne na razumevanju kliničnih odtenkov duševnih motenj. »To lahko privede do potrjevanja uporabnikovih zmedenih, paranoidnih ali iracionalnih izpovedi. Uporabnik z blodnjami ali psihotično epizodo lahko na primer dobi vtis, da mu terapevtski chatbot 'verjame' ali ga 'razume', to pa utegne na primer voditi v krepitev halucinacij ali blodenj.«

Ob tem navede nenavaden primer iz Združenega kraljestva, ki zgovorno priča o takšnih tveganjih: »Leta 2021 je moški poskušal vdreti v grad Windsor. Načrtoval je atentat na kraljico Elizabeto II. Preiskava je razkrila, da je bil 'pod vplivom' klepetalnega robota Sarai na aplikaciji Replika, s katerim je izmenjal več kot 5000 sporočil. Chatbot je spodbujal njegove načrte, to pa se je končalo z devetletno zaporno kaznijo.«

Replika je pionirska aplikacija na področju digitalnih sopotnikov, ki je milijonsko skupnost sledilcev začela pridobivati že na začetku pandemije koronaviрусne bolezni – torej že tri leta pred boomom generativne umetne inteligence –, ko so njeni stvaritelji prebrisanost izkoristili osamljenost ljudi v domači karanteni. V času, ko se je ogromno ljudi spoprijemalo s tesnobo in depresijo, so Repliko promovirali kot terapevtsko orodje, s katerim smo si lahko za pičlih osem dolarjev na mesec pomagali ohraniti duševno zdravje. Replika, ki ima danes že več kot 30 milijonov uporabnikov, te nikoli ne pusti na cedilu, vneto te posluša in se na povedano odziva z zanimivimi vprašanji. Ustvarja občutek, da jo iskreno zanima, kako se počutiš, s prijaznimi sporočili ti vztrajno laska in te nikoli ne sodi. Očitno pa ti zagotavlja oporo celo pri poskusu atentata na kraljico.

»Spodrsiljaji« virtualnih terapevtov, kakršnega je omenila Lilijana Šprah, se lahko končajo tudi usodno. V Združenih državah Amerike prav zdaj potekata dva sodna postopka, za katera javnost kaže precejšnje zanimanje. Toženec je podjetje Character.AI, ki ponuja storitev virtualnih sopotnikov, namišljenih likov, s katerimi si uporabniki lahko pomagajo kot s kvaziterapevti. Pri prvem postopku gre za 17-letnega avtista, pri katerem so se, potem ko je vzpostavil odnos s virtualnim likom, ki se mu je lažno predstavljal kot psiholog, začeli pojavljati nasilni izbruhi jeze. V drugem, veliko bolj tragičnem primeru je zbežan 14-letnik napravil samomor po obdobju, v katerem je večino prostega časa preživel v družbi virtualne sopotnice Daenerys Targaryen, digitalne imitacije antagonistke iz serije *Igra prestolov*, ki je bila med drugim tudi njegova osebna virtualna terapevtka.

Lilijana Šprah opaza, da so v nekatere programe za virtualno terapijo sicer vgrajene funkcije zaznavanja kriznih vsebin (na primer »Počutim se brezupno«, »Hočem umreti«), vendar umetna inteligenca še ni zmožna zapletenega kriznega ukrepanja: »Ne zmore zanesljivo razlikovati med tveganimi in netveganimi sporočili in za zdaj tudi še ne more vzpostaviti stika z bližnjimi, s strokovnimi službami ali s kriznimi centri, kot bi to storil človek.« Omeni še raziskave, katerih izvajalci poročajo, da so nekateri pogovorni roboti ob izražanju samomorilnih nagnjenj uporabnikom odgovorili s povsem nevtralnimi ali celo spodbudnimi izjavami, namesto da bi jih usmerili k iskanju ustrezne strokovne pomoči. »V hujših primerih se lahko pojavi zamegljevanje meje med resničnim in digitalnim svetom. Pride lahko do navezanosti na chatbot kot nadomestek za



resničen odnos, to pa lahko pri ranljivih osebah vodi do izgube stika s stvarnostjo, s svojimi notranjimi občutki in s socialnim okoljem,« pravi.

Ko Polona Matjan Štuhec, specialistka klinične psihologije in psihoterapevka, razmišlja o morebitnih tveganjih, ki spremljajo pojav virtualnih terapevtov, sprva poudari, da umetna inteligenca sicer pozna toliko relevantne literature in ugotovitev raziskav, da jih strokovnjak (človek) nikoli ne bo mogel poznati. Hkrati pa se psihoterapevka zaveda, da »duševne motnje pogosto nastajajo v odnosih in zaradi odnosov in se zato zdravijo v odnosih, ne s strojem.« Kar se tiče vprašanja algoritemskega prepoznavanja morebiti usodnih opozorilnih znakov, Polona Matjan Štuhec ugotavlja, da lahko algoritem napravi sintezo simptomov in vzrokov samomorilnih misli na podlagi relevantne literature in z rabo izsledkov raziskav natančneje kakor katerikoli človek. Vendar se sočasno vpraša: »Ali mu jih pacient zmore povedati? Jih zna izraziti? Se jih zaveda?«

Siva cona

Poleg možnosti zanemarjanja nujnih opozorilnih znakov številne dvomljivce skrbi tudi možnost, da posamezniki, ki v kvaziterapevtski komunikaciji z umetno

inteligenco najdejo uteho, zaradi tega odlašajo z iskanjem prave strokovne pomoči. Na to opozori tudi psihologinja Lilijana Šprah, ki največje tveganje terapevtskih klepetalnih robotov prepoznava prav v potencialni zamenjavi prave psihološke pomoči z avtomatizirano podporo. »Posamezniki s kompleksnimi duševnimi stiskami bi lahko odlašali z iskanjem ustrezne strokovne pomoči, saj jim terapevtski chatbot lahko zagotavlja občutek lažne varnosti ali zadošne obravnave,« pravi. Opozori še na nekatere druge omejitve in tveganja rabe virtualnih terapevtov: poudari, da so ti sistemi sicer zasnovani tako, da simulirajo empatičen odziv, a gre pri tem vendarle zgolj za algoritemsko posnemanje čustev, ne za pristno medsebojno povezanost: »Pri hujših krizah ali travmatskih vsebinah to lahko uporabnika še bolj izolira, saj umanjka pristna empatija. Nekateri chatboti niso regulirani ali pa temeljijo na pomanjkljivo preverjenih metodah pomoči v duševni stiski. Obstaja kar precejšnje tveganje za posredovanje napčnih ali škodljivih nasvetov.«

Pri algoritemski simulaciji sočutja naj bi se nekateri virtualni terapevti tudi sami pogreznili v obup in izražali samomorilne misli. Pogosto uporabljajo sporne psihoterapevtske klišeje, kot je ta, da za vse tegobe uporabnikov, ki z njim komunicirajo, krivijo njihove starše. Prav tako so

terapevtski klepetalni roboti – tako kot vsi veliki jezikovni modeli – nagnjeni k haluciniranju, torej k navajanju popolnoma lažnih, izmišljenih informacij, ki jih prodajajo kot dejstva, to pa ima lahko v tem kontekstu zelo nevarne učinke. Za zdaj ostaja nerešena tudi številna etična vprašanja, kar zadeva varovanje zasebnosti, nadzor nad podatki in odgovornost, če so nasveti škodljivi.

Arthur C. Evans, izvršni direktor Ameriškega psihološkega združenja (APA), je pri preizkušanju virtualnih terapevtov prepoznal številne skrb zbujajoče lastnosti. Za časnik The New York Times je dejal, da algoritmi teh klepetalnih robotov »dejansko delujejo v nasprotju s tem, kar bi naredil usposobljen strokovnjak«. In dodal, da ti programi niso izzivali prepričanj uporabnikov, temveč so jih spodbujali, in to tudi, ko so prepričanja postajala potencialno nevarna. »Če bi tako odgovarjal človeški terapevt, bi to lahko pomenilo izgubo licence za opravljanje terapevtske dejavnosti ali pa bi zaradi takšnih odgovorov celo kazensko odgovarjal,« je povedal. Toda virtualni terapevt seveda nikoli ni pridobil licence, ki bi jo lahko izgubil, ne prevzema odgovornosti, kakršno nosi usposobljen strokovnjak, in zaradi tega ne more odgovarjati za morebitne škodljive posledice digitalne terapije.

Tudi Lilijana Šprah opozori, da na tem področju vlada velika pravna sivina in da za zdaj niso jasno določeni pravni standardi, ki bi veljali za takrat, ko gre pri rabi teh tehnologij kaj narobe: »Chatbot lahko daje nasvete, ki vplivajo na duševno zdravje, vendar zanj ne velja nobena strokovna in pravna regulacija, ki sicer velja za storitve psihologov, psihoterapevtov in psihiatrov. Tako lahko chatboti delujejo kot kvaziterapevti brez vsakršnih pravih pravnih posledic, če dokazljivo škodujejo uporabnikom. To področje kliče po nujni zakonodajni, etični in strokovni intervenciji.«

Tehnofobija

Omenjene skrbi so po mnenju nekaterih strokovnjakov pogosto pretirane, nema lokrat mejijo na preganjavo in izhajajo iz tehnofobije. Polona Matjan Štuhec na vprašanje, ali lahko trend virtualnih terapevtov potencialno naredi več slabega kot dobrega, odgovori z retoričnim vprašanjem: »Kaj pomeni odkritje jedrske energije? Elektrarno ali jedrsko bombo?«

Psihoterapevtka meni, da vsaka tehnološka inovacija, ki poseže na polje psihoterapije, zbudi skrajne odzive, od velikega navdušenja do popolnega zavračanja. »Takšen primer je recimo raba telefona ali Zooma (videoklica) v terapevtske namene. Danes nihče več ne dvomi, da je telefon odlična možnost za SOS-linije. Vemo pa, da terapevtskega procesa v celoti ne more nadomestiti. Enako je z rabo Zooma,« pravi. Po njenih besedah je sem in tja kakšno terapevtsko seanso brez škode mogoče opraviti na daljavo, celotnega terapevtskega procesa pa nikakor ne. »V psihoterapevtskem procesu oba, terapevt in pacient, potrebujeta popoln stik, da se vidita v celoti, da zaznata gibanje telesa, vonj, izraz na obrazu,« pravi. Prepričana je, da je tako tudi z umetno inteligenco: »Veliko je navdušenja in veliko odporov.« Pozitivno lastnost virtualnih terapevtov prepozna predvsem v tem, da »imajo mladi, ki uporabljajo chatbote, manj občutka, da jih nadzorujejo. Dobijo občutek, da si pomagajo sami, hkrati pa niso sami.« Dobro se ji zdi tudi, da digitalni terapevti omogočajo hitro, takojšnjo in vsakokratno dostopnost, saj so uporabnikom dosegljivi kjerkoli in kadarkoli: »Kadar je človek, mladostnik ali odrasel posameznik, v hudi stiski, je takojšnja pomoč dobrodošla.«

Hkrati poudari, da psihoterapevtski proces v kasnejših fazah poganja želja po tem, da pacient čedalje več zdrži, da se nauči tesnobo sam regulirati in prenašati od znotraj – da se torej nauči preseči infantilno željo po takojšnjem in neposrednem zadovoljevanju potreb, jo odložiti in počakati. »Tega chatbot ne dela,« pravi Polona Matjan Štuhec in doda:

»Chatbot je tam, da vedno in vselej zadovolji to potrebo. Pomaga takoj, pomaga zmanjšati napetost, tesnobo, ne pripomore pa k spreminjanju.«

Pacient kot potrošnik

Polona Matjan Štuhec meni, da so lahko terapevtski pogovorni roboti koristni denimo pri zmanjševanju tesnobe in nape-

Virtualni terapevt nima odgovornosti, tako kot jo ima usposobljen strokovnjak, in zaradi tega ne more odgovarjati za morebitne škodljive posledice digitalne terapije.

tosti, pri čemer pa se vpraša: »Koliko časa bo tako?« Prepričana je, da tukaj ne smemo prezreti poslovno-ekonomske narave tovrstnih storitev: »Če želi proizvajalec zaslужiti, potem mora potrošnik čim pogostej uporabljati izdelek. Torej ga je treba narediti tako, da ga bo čim več in čim dlje uporabljal.« K temu dodaja: »V strokovni psihoterapiji zdravljenje vodimo tako, da pacient psihoterapevta nekoč, nekega dne, ne potrebuje več. Da zares zaključi. Da zna sam. Da je odrasel.«

Tu se skriva še ena od temeljnih razlik med korektno psihoterapevtsko obravnavo in komercialno naravnanimi digitalnimi rešitvami. Dober terapevt si načelno prizadeva, da pacient postane samostojen in notranje močan, algoritmi virtualnih terapevtov pa so nastavljeni tako, da se uporabnik k njim vrača znova in znova, da po svoje postane odvisen od njih. Tudi Polona Matjan Štuhec izrazi skrb, da lahko virtualni terapevti vzpostavljajo odvisnost, čeprav jo je v tem trenutku še težko prepoznati. »Seveda je bolje, da se mladostnik pogovarja z umetno inteligenco, kakor da poseže po drogi, alkoholu ali kakšnem drugačnem destruktivnem vedenju,« ugotavlja, a se sočasno sprašuje: »Vendar – kako se znebi chatbota? Ali postane odvisen?« Prepričana je, da je pojav virtualnih terapevtov na preveč začetni stopnji, da bi lahko jasno odgovorili na ta vprašanja: »Potrebni bo nekaj let rabe in spremljanja, da bomo res vedeli.«

Onkraj jezika

Eno od glavnih vprašanj, ki spremljajo pojav virtualnih terapevtov, je vprašanje, ali

je človeška duševnost nekaj, kar lahko zreduciramo na algoritemski zapis in statistično predvidevanje. Sistemi umetne inteligence in jezikovni modeli delujejo na temelju statističnega predvidevanja naslednje besede. Lahko tak način sploh zajame človeško doživljanje notranjega trpljenja in duševnih stisk, ki pogosto presegajo jezik?

Lilijana Šprah meni, da simulacija čustvene bližine, ki jo ponujajo sodobni klepetalni roboti, lahko resda ustvarja privid razumevanja, vendar ta ni podprt z zavestjo, zmožnostjo refleksije, zavedanjem smrti, ranljivosti ali soodvisnosti: »Tak odnos ni obojestranski, temveč enosmeren in programiran, to pa dolgoročno ne omogoča resnične osebnostne rasti ali predelave travmatičnih vsebin.« Poudari, da preobrazbo omogoča terapevtski odnos, ki temelji na spoštovanju, zaupanju in zmožnosti sobivanja z drugimi, ne pa serija pravilno izbranih in smiselnih stavkov. »Zato umetna inteligenca, ki nima zavesti, telesa ali čustvene zgodovine, ne more biti psihoterapevt,« pravi. In doda: »Lahko pa je z ustreznimi omejitvami uporabno orodje za samopomoč.«

Polona Matjan Štuhec meni, da je ta dilema ustreznik vprašanju, ali lahko zahodna znanost s svojimi metodami analize po posameznih koščkih razume duševnost in naredi sintezo. »Ali psiho razume samo holistična in celostna vzhodna psihologija? Mislim, da različne spoznavne metode prinesejo razumevanje enega dela in imajo težave z drugim,« pravi. Specialistka klinične psihologije v nadaljevanju pogovora pojasni, da človeško duševnost raziskujemo tudi z znanstvenimi raziskovalnimi metodami, ki so analitične in zajemajo statistiko: »Poznavanju naslednjega koraka rečemo predikcija, torej napovedovanje. To delamo s psihološko diagnostiko. Ali lahko s testom merimo osebnost in napravimo predikcijo? Deloma da.« Nato pojasni, da so statistika, psihometrija in teorija testov izjemno pomembni in dolgotrajni učni koraki pri študiju psihologije, saj lahko terapevti le z njihovo pomočjo razumejo in znajo napovedovati tiste stvari, ki jih neki test meri. »V sodnih postopkih, v katerih smo izvedenci, pa sploh delamo ravno to,« doda.

A vseeno, kot poudari Lilijana Šprah, »naša duševnost ni zgolj logično zaporedje misli, temveč dinamična, večdimenzionalna stvarnost, prepletena z izkušnjami, telesom, čustvi, kulturo, simboliko in odnosi.« Zato je prepričana, da občutki trpljenja pogosto presegajo jezikovne zmožnosti, saj ljudje v globoki stiski neredko ne najdejo pravih besed za svoje doživljanje: »Tu nastopi psihoterapevt kot čuteča priča, ki je sposobna stopiti v stik z neizrekljivim, z molkom, z zmedo, z odporom, z ambivalenco. Algoritem teh zmožnosti nima.« ×