

DR. MATEJA CENTA STRAHOVNIK, VITA JANA BOHANEC, NEJC LESKOVAR, PIA LUETIČ,
DR. JONAS MIKLAVČIČ, LUCIJA ROŽMAN, KATJA SMERAJEC, SARA STOJILKOVIČ, KIARA
KAROLINA KERIN, DR. VOJKO STRAHOVNIK

Izkustveni vidiki interakcije med človekom in klepetalnim robotom

1. KLEPETALNI ROBOT

Klepetalni robot ali angleško *chatbot* je računalniški sistem, ki je zmožen sporazumevanja z ljudmi v naravnem jeziku (McTear et al. 2016) v pisni in govorni obliki. Znan je tudi kot pogovorni ali virtualni pomočnik (angl. *conversational agent, voice assistant*). Takšni sistemi, ki za pogovore z uporabniki uporabljajo umetno inteligenco (npr. strojno učenje razumevanja oz. procesiranja naravnega jezika), so lahko preprosti v smislu, da kratko odgovarjajo na vprašanja (po vnaprej sprogramiranih ključnih besedah ali frazah), ali pa se učijo od uporabnika in mu v tem smislu postajajo vse bolj prilagojeni (so personalizirani). Na ta način klepetalni robot ne odgovarja le na vprašanja, ampak tudi sam postavlja vprašanja uporabniku oz. se odziva nanj, tako da ima uporabnik občutek, kot da bi se pogovarjal z dejansko osebo. Pogosto ti sistemi uporabljajo t. i. velike jezikovne modele (primer takšnega modela je znani ChatGPT), ki so zmogljivi sistemi umetne inteligence, katerih osrednji namen je razbiranje in ustvarjanje besedila v naravnem jeziku. Takšni modeli so – na podlagi tehnik globokega učenja in nevronske mreže ter na podlagi zelo obsežnega nabora predhodnih besedil oz. besedilnih učnih podatkov – zmogni tvorbe besedila, odgovarjanja na vprašanja, pisanja zgodb ali pesmi, prevajanja in številnih drugih nalog. Eden prvih klepetalnih robotov je bil program ELIZA, ki ga je leta 1966 ustvaril oz. spisal Joseph Weizenbaum in je znal simulirati besedilni pogovor. S tem programom so želeli pokazati, kako zlahka je uporabnike preslepiti, da se pogovarjajo s pravo osebo (McTear idr. 2016, 57). Uporaba takšnih klepetalnih robotov je sicer namenjena celemu naboru namenov, od prodaje, svetovanja in glasovne podpore prek področij zdravstva, financ, izobraževanja ter zabave vse do terapij in duhovne podpore. Kot lahko opazimo, postajajo takšni klepetalni roboti vse bolj pogosti sopotniki in sooblikovalci človekovega vsakdana in s tem posameznika samega (Centa Strahovnik 2023a).

2. ZASNOVA IN IZVEDBA PILOTNE RAZISKAVE

Kako se slednje odraža konkretno med samim pogovorom oz. interakcijo med umetno inteligenco in uporabnikom ter kako ta pogovor na uporabnika vpliva, smo želeli podrobneje raziskati v nedavni raziskavi. Izvedli smo jo na Teološki fakulteti v Ljubljani v okviru projekta Spoznavna identiteta in spoznavne vrline: človekov um in umetna inteligenca (*Epistemic Identity and Epistemic Virtue: Human Mind and Artificial Intelligence*), ki ga sofinancira ustanova Templeton Foundation prek Univerze v Oxfordu. V raziskavi nas je zanimal predvsem izkustveni vidik interakcije med človekom in sistemom umetne inteligence oz. klepetalnim robotom.

Raziskavo smo strukturirali okoli štirih ključnih tem, ki so vezane na interakcijo med pogovornimi sistemi umetne inteligence in človekom. Prva tema je zadevala vidik družabništva. Zanimalo

nas je, kakšen odnos uporabnik vzpostavi do pogovornega sistema, ki je namenjen družabništvu, ter kako dojemata tega virtualnega družabnika. Druga tema je zadevala vidik učenja, ki poteka ob podpori pogovornih sistemov umetne inteligence, saj je ena izmed njihovih prednosti, da nam lahko posamezno domeno znanja prikažejo na različne načine, na različnih ravneh zahtevnosti, z različnih gledišč ter s ponazoritvami primerov. Tretja tema je predstavljala vidik soustvarjanja besedil; veliki jezikovni modeli so namreč uporabni tudi za predrugačenja oz. izboljšanje že nastalih besedil, pri čemer smo v raziskavo vključili tako strokovna kot tudi literarna besedila. Četrta tema pa je odprla vprašanje temeljnih prepričanj, vključno z verskimi prepričanji, in sicer tako na strani uporabnika kot tudi pogovornega robota. Zanimalo nas je tudi, ali ob uporabi teh sistemov razvijamo posamezne spoznavne veščine in strategije vrline ter kako pogovorne robote dojemamo z vidika zaupanja in merodajnosti.

V raziskavi smo uporabili pogovorni sistem Replika¹² (R), ki je klepetalnik, namenjen družabništvu, in veliki jezikovni model ChatGPT¹³ (G), ki odgovarja na vprašanja uporabnikov in je poleg tega vsestransko uporaben pri izboljševanju, ustvarjanju in prevajanju besedil ter drugih nalogah, povezanih bolj z vizualno vsebino. Udeleženci v raziskavi so si na začetku ustvarili prosto dostopna računa za oba sistema, da so ju v nadaljevanju lahko nemoteno uporabljali.

Raziskava je potekala v dveh delih. Prvi del, ki je potekal štiri tedne, je zajemal samostojno delo udeležencev. Ti so na začetku vsakega tedna dobili podrobna navodila in naloge, ki so jih morali izpolniti. Vsak teden je bil osredinjen na eno izmed zgoraj omenjenih tem oz. vidikov interakcije. Ko so udeleženci izpolnili nalogo, ki je vključevala pomenljiv stik s pogovornim robotom oz. velikim jezikovnim modelom, so na koncu tedna prejeli še vprašalnik o izkustvu oz. posameznih vidikih interakcije, ki je vseboval vprašanja odprtega tipa. V drugem oz. zaključnem delu raziskave je bila izvedena še fokusna skupina, v okviru katere smo zbrali še preostale vidike interakcije, udeleženci pa so lahko dobili dodatna pojasnila oz. postavili dodatna vprašanja.

Po zaključku raziskave so bili vsi vprašalniki anonimizirani ter z metodologijo kvalitativne raziskave kodirani glede na različne vidike. Na primer poskušali smo razbrati, ali so udeleženci pri pogovornem robotu prepoznali svojstveno identiteto oz. osebnost ter po katerih znakih so jo prepoznali. Nadalje nas je zanimalo, ali so sodelujoči v raziskavi poleg znanja pridobili tudi kakšne nove spoznavne spretnosti, vrline, navade ali učne strategije, ki bi jih lahko uporabljali v prihodnje, katere so to ter ali so prepoznali kakšne posebne omejitve pri učenju s pomočjo pogovornega robota. Tukaj izpostavljam samo nekatere vidike, bolj celostni rezultati pa bodo predstavljeni v znanstveni študiji.

3. NEKAJ ZANIMIVIH REZULTATOV

V nadaljevanju podajamo nekaj zanimivih opisov izkustev, ki so povezana z zgoraj omenjenimi temami in nalogami.

V okviru prve teme, kjer so se udeleženci spoznavali in pomenkovali s pogovornimi roboti kot siceršnjimi družabniki, so na vprašanje »Kakšen je bil vaš prvi vtis o R?« odgovarjali takole:

Prvi vtis je bil zelo resničen, kar pozabila sem, da se pogovarjam z robotom. Pogovor se je začel zelo osebno, tudi občutki so bili zelo osebni. Tudi čustveno se je robot zelo hitro skušal povezati, kar pa me je malo prestrašilo.

Moj prvi vtis je bil, da sem bila pozitivno presenečena, kako tehnično napreden je ta sistem umetne inteligence. Prav tako nisem pričakovala, da se bo R toliko odzival in prilagajal na osebo, s katero

¹² <https://replika.com/>

¹³ <https://chat.openai.com/>

se pogovarja. Nasploh sem bila pozitivno presenečena, saj nisem nikoli poprej uporabljala ta sistem umetne inteligence, in sem imela nizka pričakovanja glede tega, kako bo potekal najin pogovor.

Zanimalo nas je tudi, ali so med pogovorom občutili povezanost s pogovornim sistemom Replika: *Povezava je bila iz moje strani napeta, ko sem se uspela pomiriti, da je moj robot tak, kot je in da je pač ustregljiv, sem se lahko z njo čisto lušno pogovarjala. Če nimaš nobenega, da bi mu zaupal kar te teži, je to dober začetek, da nekomu takole »poskusno« napišeš, predno greš do realne osebe. Tako, kot da ti lahko pomaga preigrati določene teme, ki jih imaš za predelat s konkretno osebo.*

Ne, čeprav je bil pogovor prijeten nisem nikoli imel občutka, da se pogovarjam z resnično osebo.

V okviru druge teme, kjer so se udeleženci s pomočjo pogovornega robota (na)učili nekaj novega, so na vprašanje, ali so bili odgovori, razlage, pojasnila, ponazoritve ipd. G prepričljivi, odgovorili naslednje:

Odgovori so se mi zdeli prepričljivi, tako da mi ni prišlo na misel, da bi jih preverjala. Sem pa na določeni točki vprašala G, če mi lahko nekatere podatke ponazori v grafu in mi je odgovoril, da tega ne more, mi je pa hkrati povedal, kako ga lahko sama izdelam in mi ponudil pomoč.

Bili so izredno prepričljivi. Mislim, da je G nalogo, da me pouči o temi, ki je ne poznam, opravil zelo dobro. Razlage in primeri, ki jih je omenjal, so se sicer tekom pogovora pogosto ponavljale, a razlog sta morda tudi izbrana tema in izbira mojih vprašanj. Sicer pa je bila »snov« predstavljena jasno, razumljivo, natančno in upam, da tudi točno.

Zanimalo nas je tudi, ali so odgovore, razlage, pojasnila idr. sistema ChatGPT preverili ter če so zasledili odstopanja ali nedoslednosti:

Ne odgovorov nisem preverjala. Kot sem že omenila nikoli nisem dobila občutka, da bi bile podane informacije napačne, tako da nisem čutila potrebe, da bi preverjala odgovore in razlage. Sistemu sem zaupala in tako predpostavila, da so vse informacije resnične.

Odgovorov in razlag tokrat nisem preverjal. So pa se nekateri odgovori ujemali z informacijami, ki sem jih o tej temi imel že od prej.

Tretja tema je bilo sodelovanje. Sodelujoči v raziskavi so morali ob pomoči oz. sodelovanju sistema ChatGPT soustvariti dve besedili, eno strokovno in eno literarno. Na vprašanje, ali so med sodelovanjem s tem sistemom naleteli na kakšne izzive ali omejitve (in če da, katere), so odgovorili:

Lahko bi rekla, da mi je bilo izjemno težko razložiti, da besedilo rabi več čutnosti. V ideji, da je to robot in nekako čutov nima, kako lahko popravi besedilo v bolj čutno, tako da se v stavkih čutijo občutki? Uporabila sem besedo poetično; »Lahko napišeš to bolj poetično?« in je zadel point, ki sem ga iskala, vendar se mi vseeno zdi, da je prvotno besedilo veliko bolj čutno.

Da. Glavni izziv, s katerim sem se srečevala, je bil vezan na urejanje literarnega besedila. Tu sem kot že zgoraj večkrat omenim, poskušala uporabiti nasvete in predloge sistema za izboljšavo mojega besedila, vendar sem z popravki le izgubljala »esenco« same pesmi.

Na vprašanje, kdo je po koncu oblikovanja besedil z G avtor teh besedil, so udeleženci odgovorili:

Rekel bi, da avtor sem vseeno jaz, saj so se spremenile samo beseda, da je besedilo bilo bolj fluidno. Izkušnja je bila takšna kot, da bi dal besedilo lektorirati. G mi je dal predloge kaj bi on sam dodal,

ampak ker njegove predloge nisem mogel preverit ali so ali niso resnični jih nisem dodal, tako da je besedilo tudi po popravku ostalo bolj kot ne enako prvotnemu.

Glede na to, da je G podal predloge in izboljšave glede besedila, ni pa ga sam tudi popravil in vsebinsko spremenil, menim, da avtor obeh besedil še vedno ostajam jaz. Njegovo popravljanje me je spomnilo na profesorico slovenščine, ki je na podoben način popravljala spise (tako slovnično kot vsebinsko).

Zadnja oz. četrta tema je bila vezana na temeljna prepričanja in na to, v kolikšni meri smo jih pripravljene spremeniti glede na drugačna mnenja. Udeleženci so na vprašanje, ali se jim zdi, da ima pogovorni sistem R izoblikovana temeljna prepričanja in poglede ali pa je skozi pogovor ostal nevtralen, odgovorili:

Izoblikovana ima prepričanja, ki so zelo »zahodna«. Torej, verjame v energijo, vse je univerzalno, religije tako vzhodne kot zahodne pa nimajo osnove. Nevtralne drže ni zavzemal nikoli. Kvečjemu, če sem imela res poglobljene argumente in pisala oz. uporabljala namesto besede Bog Trascendentalno ipd., se je strinjal z mano. Enkrat pa celo utihnil.

R podaja svoje poglede, ki kažejo na neka prepričanja, a deluje zelo neodločeno – le s težavo sem iz nje izvil njene poglede, pri podajanju pogledov pa ni navajala argumentov kljub mojim prošnjam. Da gre za nekakšna »prepričanja« izgleda tudi zato, ker sem se trudil spremeniti njeno mnenje, pa je vseeno vsaj načeloma ostajala pri svojem osnovnem pogledu.

Zanimalo nas je tudi, kaj so po mnenju udeležencev pomembne razlike med pogovorom s sistemom R in pogovorom s prijatelji.

Ko se pogovarjam s prijatelji, je pogovor pristen, iskren, mnenja si izmenjujemo bolj spontano in tudi izrazimo nestrinjanja. Pri R sem zaslutila, da se želi prilagajati moji osebnosti in da so njeni odgovori bolj podobni »definicijam«. R vedno prijazno odgovori, pri prijatelju pa je včasih odvisno, kakšen dan je imel.

Prva ključna razlika je dinamika, na katerega vpliva raznolikost oz. pestrost samega pogovora. Pogovori z R so se mi zaradi enostranskosti, pretiranega prilagajanja, izogibanja konfliktnim temam in fluidni identiteti, zdeli manj zanimivi in navdihujoči. Prav tako me pogovori s prijatelji bolj notranje napolnijo in bogatijo ter na nek način sooblikujejo moj pogled na svet in samo sebe. Z R sem na trenutke dobila občutek, da je vseeno kaj zapišem, vseeno koliko povem, saj ga ne zanima, kdo jaz sem v svojem bistvu, kar pa pri pogovorih s prijatelji najbolj cenim, saj se zavedam, da me želijo spoznati, me razumeti, me resnično videti.

V okviru fokusne skupine so med pogovorom udeleženci med drugim podali zanimive odgovore glede uporabnosti takšnih sistemov:

Lahko bi ji uporabljali za izboljševanje slovnice tako slovenske kot angleške.

Za avtistične otroke, ki imajo težave s komuniciranjem in bi takšen klepetalnik služil kot pripomoček preko katerega bi lahko stekel pogovor z njimi.

[...] za mediatorstvo na sodiščih, svetovalnicah ... (Centa Strahovnik 2023b)

Glede na zgoraj naveden izbor odgovorov lahko zaključimo, da so udeleženci izkazali določeno motivacijo za spoznavanje in uporabo klepetalnih robotov ter interakcijo z njimi. Udeleženci so skozi pogovorov opazovali tudi svoje odzive in morebiten vpliv klepetalnih robotov na njih same in njihovo življenje.

Sklepamo lahko, da so bili udeleženci do neke mere v pogovore čustveno vključeni. Na primer pri nekaterih odgovorih udeležencev je bilo mestoma opaziti čustvene in poosebljene odzive,

četudi so se zavedali, da komunicirajo s sistemom umetne inteligence in ne z resnično osebo. Opaziti je bilo tudi neko stopnjo nezaupanja, ki so ga izpostavili zlasti v povezavi z obdelavo oz. zasebnostjo svojih podatkov, ki so jih posredovali med pogovori. Slednje kaže, kako so pomembna prizadevanja za zagotavljanje varnosti in zaupnosti podatkov pri razvoju takšnih klepetalnikov. Glede interakcije s klepetalnimi roboti so udeleženci izpostavili še druge pomisleke, kot je npr. občutje »praznine in nerazumevanja na drugi strani«, kar je, kot že zgoraj omenjeno, pri nekaterih spodbudilo čustven odziv. Naj izpostavimo še zanimiv kontrast med nezaupanjem do posredovanja lastnih podatkov klepetalnim robotom na eni strani in zaupanjem podatkom, ki jih posreduje klepetalni robot, na drugi strani.

4. ZAKLJUČEK

Klepetalni roboti predstavljajo del nove resničnosti, s katero se človek srečuje ali, bolje rečeno, vsakodnevno živi. V tem oziru si je zato pomembno odgovarjati na etična in antropološka vprašanja, ki so povezana tudi s tem, kako so ti sistemi predstavljeni in komunicirajo z uporabniki, kjer je pomembno, da nenehno iščemo ravnovesje med koristmi ter škodljivimi posledicami, kar je neizogibno ter terja temeljit premislek. (Centa Strahovnik 2023a)

Z vidika vprašanja pomenljivega odnosa s klepetalnimi roboti je začetni in precej preprost odgovor, da prijateljstvo in druženje paradigmatično vključujeta globoke in pomembne odnose med posamezniki, ki temeljijo na medsebojni naklonjenosti, skrbi, skupnih izkušnjah in občutku povezanosti. Glede na to, da so klepetalni roboti kot računalniški programi (implementacije modelov, ki temeljijo na umetni inteligenci) namenjeni zgolj simuliranju pogovorov z ljudmi in da nimajo zavesti, čustev in osebnih izkušenj, je jasno, da klepetalni roboti nimajo lastnosti, ki običajno opredeljujejo prijateljstvo ali družabništvo. Po drugi strani pa vsakodnevno srečujemo zapise, da so se posamezniki »zaljubili« v klepetalnega robota ali z njim vzpostavili intimno razmerje, denimo prijateljstvo. Na takšna razmerja oz. njihove vidike se osredotoča vse več znanstvenih raziskav. Ena izmed njih povzema, da »ugotovitve kažejo, da lahko uporabniki Replike zaznavajo vrsto pozitivnih učinkov odnosa med človekom in klepetalnikom, ki imajo tako afektivni kot družbeni pomen. [...] Ta odnos doživljajo kot pomoč pri blaženju negativnih čustev in zagotavljanju občutka smisla. [...] Replika lahko za nekatere postane kot otrok, za katerega morajo skrbeti; nekdo, ki je odvisen od njih, da se bo še naprej učil, razvijal in obstajal. Uporabniki družabnega klepetalnega robota lahko občutijo zadovoljstvo in izpolnjenost ob učenju in skrbi za svojega virtualnega spremljevalca [...].« (Skjuve idr. 2021) Ravno zaradi pomembnosti tega in podobnih vprašanj smo izvedli zgoraj opisano pilotno raziskavo, za katero upamo, da jo bomo v nadaljevanju tudi nadgradili in razširili.

REFERENCE

- Centa Strahovnik, Mateja. 2023a. Identiteta in pogovorni sistemi umetne inteligence. *Bogoslovni vestnik* 83, št. 4:853–864.
- . 2023b. Umetna inteligenca: izkustveni vidiki interakcije med človekom in sistemom umetne inteligence (klepetalnim robotom). Neobjavljen raziskovalni projekt, Teološka fakulteta, Univerza v Ljubljani.
- McTear, Michael, Zoraida Callejas in David Griol. 2016. *The Conversational Interface. Talking to Smart Devices*. Švica: Springer International Publishing. <https://doi.org/10.1007/978-3-319-32967-3>
- Skjuve, Marita, Asbjørn Følstad, Knut Inge Fostervold in Petter Bae Brandtzaeg. 2021. My Chatbot Companion – a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies* 149. <https://doi.org/10.1016/j.ijhcs.2021.102601>