



ChatGPT o hrvatskome jeziku

Godine 2015. Ivan Pandžić objavio je u *Hrvatskome jeziku* rad *Razumiju li računala hrvatski jezik?* U ovome radu pokazat će se što se u svijetu govornih robota promijenilo u manje od deset godina.

Što je ChatGPT?

ChatGPT (Chat Generative Pre-Trained Transformer) govorni je robot (engl. *chatbot*) koji je razvila tvrtka OpenAI. Prva je inačica objavljena u svibnju 2022. ChatGPT drukčiji je od standardnoga govornog robota jer se može u većoj mjeri prilagoditi korisniku te odgovarati na više standardnih prirodnih jezika (uključujući i hrvatski jezik). On se temelji na GPT-3.5 i GPT-4 arhitekturi. Ta arhitektura koristi se velikim brojem računalnih neurona i parametara, što mu omogućuje da „razumije” i proizvodi složene rečenice te daje iscrpnije odgovore od prethodnih govornih robota.

Standardni govorni roboti često su programirani tako da daju odgovore na određena pitanja te se u praksi koriste u korisničkoj službi kako bi ponudili brz odgovor na česta pitanja, a ako im je upit nejasan, preusmjeravaju ga čovjeku. ChatGPT osim razvijene neuronske mreže ima i pristup velikoj količini podataka prikupljenih s interneta. Stoga program raspolaze informacijama o događajima koji su se dogodili prije posljednjega posuvremenjivanja te uglavnom za svaki upit može bolje prepoznati kontekst i generirati odgovor. Za razliku od drugih govornih robota odgovori nisu unaprijed definirani te se često na isto pitanje mogu dobiti različiti odgovori. Zbog mogućnosti dobivanja odgovora na svako pitanje sve se češće upotrebljavaju kao alat za rješavanje problema pri skupljanju i obradi podataka, programiranju i sastavljanju teksta, ali i za oblikovanje slika i dizajn. ChatGPT integriran je kroz različite platforme te se može besplatno upotrebljavati (npr. na stranicama <https://chat.openai.com/>, <https://chat-gpt.org/> i <https://trychatgpt.info/>).

U ovome će se radu analizirati razgovor s ChatGPT-jem (besplatna inačica) na hrvatskome jeziku i o hrvatskome jeziku. Pri analizi treba uzeti u obzir da je besplatna mrežna inačica ChatGPT-ja posljednji put posuvremenjena u rujnu 2021., te stoga ne može dobro odgovoriti na pitanja koja se odnose na novije događaje. Često je program preopterećen zahtjevima različitih korisnika te treba čekati na odgovor. Postoji naprednija inačica ChatGPT Plus, koja se plaća. Ona brže generira odgovore i ne stavlja korisnika na čekanje ako više korisnika šalje zahtjeve programu. ChatGPT Plus korisnicima daje neke naprednije i novije opcije. Korištenje GPT-4 arhitekturom za pretraživanje moguće je jedino s Plus inačicom te Plus inačica jedina omogućuje instaliranje dodataka

* Josip Mihaljević leksikografski je suradnik u Leksikografskome zavodu Miroslav Krleža.

koji omogućuju povezivanje programa s drugim aplikacijama radi dohvaćanja i obrade novih informacija. Primjerice, može se upotrijebiti ChatGPT Plus sa stranicom *Expedia* za pretraživanje i rezervaciju letova i smještaja, se stranicom *OpenTable* za pronalazak i rezervaciju mjesta u restoranu i povezivanje s *Bing* tražilicom radi dohvaćanja novijih podataka kojih nema u besplatnoj inačici (OpenAI 2023.).

Koliko ChatGPT razumije hrvatski jezik?

Prema trenutno dostupnim podacima ChatGPT može razumjeti više od 50 jezika, što uključuje i hrvatski jezik (Botpress Community 2023.). ChatGPT-ju može se postaviti upit na hrvatskome jeziku te će on odgovoriti na hrvatskome jeziku. Moguće je napisati i upit na engleskome jeziku te u upitu tražiti da napiše odgovor na hrvatskome jeziku. Nažalost, teško je analizirati koji je bolji način komunikacije s programom pri sastavljanju teksta na hrvatskome jeziku jer program na jednako postavljena pitanja gotovo svaki put nanovo generira odgovor te se često koristi istim izvorima pri generiranju odgovora (podacima s interneta napisanim na hrvatskome jeziku do rujna 2021.). Ipak, katkad je lakše dobiti pouzdan odgovor ako se upit postavi na engleskome jeziku te traži prijevod na hrvatski jezik. Razlog je za to vjerojatno taj da ChatGPT može pretraživati i tekstove napisane na engleskome te ih prevesti na hrvatski, dok kod upita na hrvatskome jeziku vjerojatno uglavnom pretražuje izvore na hrvatskome jeziku. Primjerice, na upit na hrvatskome *nabroji informatičke rječnike na hrvatskome jeziku* odgovorio je: *Nažalost, kao jezični model bez pristupa internetu, nemam mogućnost trenutnog pretraživanja interneta niti mogućnost ažuriranja s popisom svih informatičkih rječnika na hrvatskom jeziku koji postoje*. Kad je upit postavljen na engleskome *can you mention computer science dictionaries written in Croatian (please answer in Croatian)* te se tražio odgovor na hrvatskome, ChatGPT nabrojio je ove rječnike: *Računalni leksikon, Rječnik informacijske tehnologije, Rječnik informatike i računarstva, Rječnik računalstva i informatike, Informatički rječnik*. Za većinu njih navodi da su izdavači Mate d.o.o. i Element d.o.o., koji nisu objavili spomenute rječnike (ne navode ih na svojim mrežnim stranicama), ali zato objavljuju druge informatičke i računalne udžbenike i priručnike.

Na 1. slici može se vidjeti te usporediti odgovore dobivene na istoznačan upit postavljen na hrvatskome i engleskome jeziku. Korisnik je tražio da računalo napiše nešto o hrvatskome jeziku



Hrvatski jezik je južnoslavenski jezik s bogatom poviješću i dubokim ukorijenjenjem u kulturi Hrvatske. Kao službeni jezik u Republici Hrvatskoj i jedan od službenih jezika u Bosni i Hercegovini, hrvatski se govori i u dijelovima susjednih zemalja te u dijaspori. Koristi se latinica s dodatkom dijakritičkih znakova (ččžđ) koji mu daju prepoznatljivu fonetičku raznolikost. Hrvatski jezik obiluje različitim dijalektima, a standardni oblik temelji se na štokavskom dijalektu, posebice na njegovim ikavskim i ijekavskim varijantama. Sa svojom bogatom književnom tradicijom koja seže od srednjeg vijeka do suvremenih pisaca, hrvatski jezik odražava duh i identitet hrvatskog naroda te ima ključnu ulogu u očuvanju i promicanju hrvatske kulture.





Hrvatski jezik je južnoslavenski jezik koji se koristi kao službeni jezik u Republici Hrvatskoj, a također je jedan od službenih jezika u Bosni i Hercegovini. S obzirom na svoje duboko ukorijenjenje u hrvatskoj kulturi, hrvatski je ključni element nacionalnog identiteta. Koristi se latinica s dodatkom dijakritičkih znakova (čćžšđ) koji mu daju specifičan zvuk i omogućuju precizno izgovaranje riječi. Hrvatski jezik ima raznolik sustav padaža i bogatstvo glagolskih oblika, što ga čini gramatički složenim i izražajnim. Kroz svoju dugu povijest, hrvatski jezik oblikovan je utjecajima drugih jezika, ali je i sam doprinio stvaranju vrijednih književnih i kulturnih djela. Hrvatski jezik nije samo sredstvo komunikacije, već je i temelj hrvatske kulturne baštine i nacionalne svijesti.

1. slika: Primjer teksta o hrvatskome jeziku za istoznačan upit postavljen na hrvatskome i engleskome jeziku (21. srpnja 2023.)

Odgovor koji daje ChatGPT katkad ima pravopisnih, gramatičkih i posebno stilskih pogrešaka. Stoga ga korisnik, osim provjere točnosti podataka, treba i lektorirati. Program ovako objašnjava svoje znanje hrvatskoga jezika (vidi 2. sliku):



Treba li tebi lektor?



Kao umjetna inteligencija, ne posjedujem osobnost, stvarne svijesti ili svijesti o sebi pa samim time nemam osjećaj za potrebu za lektorom. Međutim, kao jezični model, ja sam vrlo napredan i razvijen za obradu jezika te pokušavam osigurati što veću točnost i jasnoću u svojim odgovorima.



Unatoč tome, kao i svaki sustav, moguće su pogreške u prepoznavanju ili interpretaciji riječi, izraza ili konteksta. Ako primijetite bilo kakve greške ili netočnosti u mojim odgovorima, svakako slobodno istaknite, kako bih mogao poboljšati svoje performanse. U takvim situacijama uvijek je korisno koristiti ljudsku provjeru i intervenciju kako bi se osigurala najveća preciznost i pouzdanost.

2. slika: ChatGPT odgovara na pitanje treba li mu lektor (24. srpnja 2023.)

Za svaki odgovor koji program ponudi može se poslati povratna informaciju kako bi se točnost odgovora u programu poboljšala.

Na 3. slici nalaze se odgovori koje ChatGPT daje na pitanje *Koji normativni priručnici postoje za hrvatski jezik?*. Naveo je normativne priručnike, ali je posve pogrešno naveo bibliografske podatke.



Koji normativni priručnici postoje za hrvatski jezik?



Za hrvatski jezik postoje nekoliko normativnih priručnika koji služe kao referentne knjige za pravopis, gramatiku i pravila jezičnog izražavanja. Ovdje su neki od najvažnijih normativnih priručnika za hrvatski jezik:

1. "Hrvatski pravopis" (The Croatian Orthography): Ovo je ključni priručnik koji regulira pravila pisanja i pravopis hrvatskog jezika. Izdan je od strane Hrvatske akademije znanosti i umjetnosti (HAZU) i pruža smjernice o pravilnom pisanju riječi, interpunkciji, sklanjanju i sl.
2. "Hrvatski jezični savjetnik" (The Croatian Language Advisor): Ovaj priručnik također izdaje Hrvatska akademija znanosti i umjetnosti. Nudi savjete i smjernice o jezičnom izražavanju, stilu pisanja, upotrebi riječi i gramatičkim konstrukcijama.
3. "Rječnik hrvatskoga jezika" (The Dictionary of the Croatian Language): To je temeljni priručnik za standardni rječnik hrvatskog jezika. Izdan je također od strane HAZU i sadrži široki raspon riječi i njihovih značenja.
4. "Pravopisni priručnik hrvatskoga jezika" (The Croatian Language Spelling Handbook): Ovaj priručnik, izdan od strane Instituta za hrvatski jezik i jezikoslovlje, pruža smjernice o pravilnom pisanju i pravopisu hrvatskog jezika.
5. "Gramatika hrvatskoga jezika" (The Grammar of the Croatian Language): To je opsežan priručnik za gramatiku hrvatskog jezika. Postoje različita izdanja i autori, a neka izdanja također su objavljena od strane HAZU.
6. "Rječnik stranih riječi" (The Dictionary of Foreign Words): Ovaj priručnik, također objavljen od strane HAZU, sadrži objašnjenja stranih riječi koje su ušle u hrvatski jezik.

Ovi normativni priručnici redovito se ažuriraju kako bi se pratila promjena jezika i jezičnih potreba suvremenog društva. Oni su od velike važnosti za očuvanje standarda hrvatskog jezika i osiguranje njegove dosljedne upotrebe u svim sferama komunikacije.

Slika 3. Podatci koje je ChatGPT dao za normativne priručnike hrvatskoga jezika (24. 7. 2023.)

Pogrešno navođenje podataka svojstveno je ChatGPT-ju i naziva se *halucinacija*.

Kad isti zadatak rješava leksikograf i ChatGPT

Dana 21. travnja 2023. godine u povodu Noći knjige 2023., manifestacije na kojoj je ove godine u žarištu interesa bila umjetna inteligencija, leksikografi iz Leksikografskoga zavoda Miroslava Krleža postavili su ChatGPT-ju nekoliko pitanja o životu i djelu

Miroslava Krleže i Antuna Gustava Matoša. Na ta pitanja odgovore je dao i leksikograf ne znajući što je ChatGPT odgovorio. Odgovori ChatGPT-ja i leksikografa za usporedbu su navedeni na slici, jedan do drugoga, te su slike objavljene na Facebookovim stranicama Leksikografskoga zavoda. Uglavnom su tekstovi koje je sastavio ChatGPT bili dulji od tekstova leksikografa te sadržavali dosta nepotrebnih i u nekim slučajevima netočnih informacija ili nisu dali precizan odgovor na postavljeno pitanje. Primjerice, na pitanje *Koliko je novela objavio Antun Gustav Matoš?* ChatGPT je odgovorio s *nekoliko* te nabrojio djela koja nisu novele te ih nije napisao Matoš (npr. roman *U registraturi* Ante Kovačića, ep *Judita* Marka Marulića), a leksikograf je odgovorio na pitanje sa *šezdesetak* te spomenuo da je u trima zbirka objavio oko polovice. Unatoč tomu što nije nabrojio naslove poznatih novela, ipak je dao točan odgovor za razliku od ChatGPT-ja, koji je u svojem pokušaju da da potpuni odgovor dao pogrešne informacije. ChatGPT katkad može pogriješiti i u nekim informacijama, npr. za Krležine *Balade Petrice Kerempuha* odgovorio je da su tiskane prvi put u Zagrebu 1936., a točno je da su tiskane u Ljubljani. Treba napomenuti da su odgovori koje je ChatGPT dao najčešće razumljivi tako da ljudi koji nemaju dovoljno znanja o nekoj temi mogu smatrati da su točni. Ipak, katkad ima i nejasnih ili nespretno sročениh rečenica (npr. *Krleža je podrijetlom iz građanske obitelji s nekoliko sjevernokrvačkih nasljednih odvjetaka*).

Zaključak

Umjetna inteligencija nema svijest ni sposobnost razvoja izvan okvira unutar kojega je programirana te skupa podataka na temelju kojih je trenirana. ChatGPT i druge umjetne inteligencije u budućnosti će se poboljšavati na temelju naprednije arhitekture i tehnike učenja te većega skupa podataka. Osim toga, postupno će se poboljšavati sučelje i način interakcije između umjetne inteligencije i korisnika.

ChatGPT trenutačno može sastaviti tekst na hrvatskome jeziku, ali se on ne bi trebao koristiti u profesionalne svrhe bez iscrpne ljudske provjere i lekture. Na prethodno spomenutim primjerima možemo vidjeti da ChatGPT daje mnogo nepotpunih i pogrešnih podataka (halucinacija). Predviđa se da će se u budućnosti program razvijati te će moći dati preciznije i točnije odgovore, ali sve što stvori umjetna inteligencija i dalje treba pregledavati i ispravljati. Trenutačno ChatGPT može pomoći u pronalaženju podataka koji se mogu upotrijebiti za daljnju ljudsku obradu. Primjerice, može se zatražiti od njega da pronade popis određenih riječi koje se mogu koristiti za izradu oblaka riječi, može pomoći u pronalaženju izvora (ali u nekim slučajevima može navesti pogrešne izvore ili druge podatke) ili kao alternativni prevoditelj tekstova na hrvatski jezik (u nekim slučajevima radi bolje od Googleova prevoditelja). Unatoč tim nedostacima ChatGPT mnoge zadatke može napraviti mnogo brže od čovjeka te time ubrzati izvršavanje nekoga zadatka. Međutim, u svim su slučajevima ljudski pregled i znanje i dalje potrebni da bi se podatci koje generira ChatGPT mogli iskoristiti.