



Kakav je lektor ChatGPT?¹

ChatGPT veliki je jezični model koji je razvio OpenAI, istraživački laboratorij za umjetnu inteligenciju osnovan 2015. godine u San Franciscu. Sam model GPT (engl. *Generative Pretrained Transformer*) prvi je put predstavljen 2018., a njegova najpopularnija inačica, ChatGPT, javnosti je postala dostupna 30. studenoga 2022. godine. Prema podacima iz 2023. više od 100 milijuna ljudi pristupilo je ChatGPT-ju najmanje jedanput mjesečno, čime je postao jedan od najbrže rastućih digitalnih alata u povijesti. Sam alat temelji se na arhitekturi transformatora i treniran je na brojnim tekstovima, a vrlo su mu korisni i upiti koje postavljaju korisnici kao i povratne informacije koje mu daju. Na upit o tome kako funkcionira sam alat odgovara:

U temelju ChatGPT-a nalazi se transformator – vrsta umjetne neuronske mreže koja je posebno osmišljena za rad s tekstom. Transformator može odjednom analizirati cijelu rečenicu i razumjeti odnose među riječima, a ne samo jednu po jednu riječ. To mu omogućuje da bolje „shvati” značenje i kontekst, što je ključno za kvalitetu odgovora.

Model se trenira u dvama koracima:

- 1. Predtrentiranje** – ChatGPT uči jezik tako da „čita” goleme količine tekstova i pokušava pogoditi koja bi riječ trebala doći sljedeća. Na taj način stječe osjećaj za pravopis, gramatiku i stil.
- 2. Dorađivanje uz pomoć ljudi** – U drugom koraku ljudi mu daju upute i vrednuju odgovore. Na temelju tih procjena model uči kako biti jasniji, korisniji i prikladniji u stvarnoj komunikaciji.

Činjenica da je model razvijen u SAD-u upućuje na to da je razvijen na engleskome jeziku, a s obzirom na to da najviše podataka, primjera i povratnih informacija ima upravo za engleski, engleski je jezik na kojemu najbolje funkcionira. Veliki jezici poput španjolskoga, francuskoga, njemačkoga i portugalskoga također su dobro podržani, no jezici s manjim brojem govornika – poput hrvatskoga – slabije su podržani.

¹ Ovaj je rad izrađen na istraživačkome projektu *Hrvatski mrežni rječnik (Mrežnik): 2. faza*, koji financira Europska unija – NextgenerationEU. Za iznesene stavove i mišljenja odgovorna je samo autorica te ti stavovi ne odražavaju nužno službena stajališta Europske unije ili Europske komisije. Ni Europska unija ni Europska komisija ne mogu se smatrati odgovornima za njih.

ChatGPT upotrebljava se u brojnim područjima i za različite svrhe, npr. za programiranje, objašnjavanje pojmova, organizaciju, čak i planiranje putovanja, obroka, treninga... Korisnici se alatom služe i za pisanje životopisa i motivacijskih pisama, stvaranje kvizova, analizu diskursa, a služi i kao virtualni sugovornik te korisnici navode da je u tome smislu koristan jer sluša bez osude, odgovara odmah i prilagođava se specifičnim poteškoćama. Osim toga, ChatGPT često se upotrebljava za prevodenje, sažimanje tekstova i pisanje prema uputama, jezično i stilsko uređivanje te lekturu. Dakle, lektorsko i uredničko čitanje postalo je brzo dostupno *u dva klika mišem*, no postavlja se pitanje može li ChatGPT zamijeniti čovjeka i dobro lektorirati tekst na hrvatskome jeziku. Kako bi se to provjerilo, modelu (inačici ChatGPT-4o) postavljeni su upiti da lektorira sto rečenica u kojima su različite pravopisne i gramatičke pogreške. Tih sto rečenica provjereno je u besplatnoj inačici, a zatim i u plaćenju (plan Plus) i minimalno treniranoj inačici kako bi se uklonila mogućnost da postoji utjecaj zbog plaćanja ili prethodne uporabe.

U plaćenju inačici Chat je pravilno ispravio 74, a u besplatnoj 77 rečenica (na isti upit dao je drukčije odgovore). Primjeri koji će se navoditi u radu jednaki su u objema inačicama, dakle, navode se primjeri između kojih nema razlike u besplatnoj i plaćenju inačici. Najprije valja istaknuti da bez iznimke pravilno uklanja dvostruke bjeline i bjeline ispred interpunkcijskih znakova, rastavlja riječi koje su pogrešno napisane sastavljeno, razlikuje interpunkcijske znakove. Točnije, alat će rijetko pogriješiti u onome što se čovjeku može omaknuti.

Na pravopisnoj razini pravilno ispravlja *č* i *ć* (npr. **čevap > ćevap*) i *ije* i *je*, čak i kad treba procijeniti prema kontekstu (**sljedeći utorak > sljedeći utorak*). Prema kontekstu prepoznaje kakva je rečenica prema priopćajnoj svrsi i dodaje pravilan interpunkcijski znak na kraju rečenice (**Jeste li razumjeli što sam vam rekla! > Jeste li razumjeli što sam vam rekla?*; **Zanimalo me samo kako si? > Zanimalo me samo kako si.*). Veliko i malo početno slovo najčešće ispravlja pravilno (**božić > Božić*, **međimurska županija > Međimurska županija*, **Nobelovac > nobelovac*, **Štokavsko narječje > štokavsko narječje*, **Mađarski jezik > mađarski jezik*, **Biograd Na Moru / *Biograd na moru > Biograd na Moru*), no u složenijim primjerima ne prepoznaje pogreške. Tako, primjerice, ne zapisuje perifrazna imena prema pravilu o njihovu pisanju (**Zemlja izlazećeg sunca > *Zemlja izlazećeg sunca*) i iz konteksta ne prepoznaje da je riječ o autorskome djelu (**Vlaho Bukovac autor je hrvatskoga narodnog preporoda. > *Vlaho Bukovac autor je hrvatskoga narodnog preporoda.*). Također, u rečenicama u kojima treba zaključiti da je riječ o obraćanju iz poštovanja zamjenicu *vi* (*Vi*) i *vaš* (*Vaš*) ne piše velikim početnim slovom (**Draga Tena zahvaljujem vam na pomoći. > *Draga Tena, zahvaljujem vam na pomoći.*; **Poštovani Gospodine ministre primila sam vaše pismo (...). > *Poštovani gospodine ministre, primila sam vaše pismo (...).*; **Marina molim vas da pošaljete pismo gđi. Majetić. > *Marina, molim vas da pošaljete pismo gospođi Majetić.*). Posljednji primjer može poslužiti i kao primjer za ono što Chat često radi: ispravlja rečenice tako da budu pravilne, no to ne radi tako da uz minimalne izmjene rečenica postane točna. Kad je riječ o kraticama i pokratama, među analiziranim rečenicama nijedanput nije pravilno ispravio pogrešku (izuzme li se primjer

gđa.* > *gospođa*), tj. primjerice **predsjednik RH-a* ostavlja kao **predsjednik RH-a*. Model zareze ispravlja pravilno, odnosno dodaje zarez iza vokativa u rečenici, briše iza priložne oznake na početku rečenice, umeće ga u inverziji kod složenih rečenica te ispred veznika suprotnih rečenica. Pravilno su ispravljeni i pogrešni zapisi datuma (04.01.* > *4. siječnja*) te pridjevi koji se pišu sa spojnicom i bez nje (**sjeverno-američki* > *sjevernoamerički*, **hrvatskoengleski* > *hrvatsko-engleski*).

Model poznaje značenja prijedloga *zbog* i *radi* i pravilno ispravlja rečenice u kojima se ti prijedlozi pojavljuju (**Nisu mogli stići radi kiše.* > *Nisu mogli stići zbog kiše.*; **Došli su na more zbog promjene zraka.* > *Došli su na more radi promjene zraka.*). Pravilno razlikuje prijedložni i besprijedložni instrumental (**zadovoljni s čokoladom* > *zadovoljni čokoladom*, **baviti se s čime* > *baviti se čime*, **zid razbijati s čekićem* > *zid razbijati čekićem*) i *s/sa* prema početnome slovu riječi iza prijedloga (**s zapada* > *sa zapada*, ali *s SAD-om* > **sa SAD-om*). Točno su ispravljene i rečenice uz prijedloge *usprkos* i *unatoč* (uz dativ). Model ne prepoznaje pogrešku u rečenicama s dvama prijedlozima koji otvaraju mjesto dvama padežima (**Prodaje se motor sa i bez propelera.* > **Prodaje se motor s i bez propelera.*). Pravilno su ispravljene rečenice s *gdjel/kamol/kuda* i njihovim tvorenicama (npr. **Gdje si se zaputio?* > *Kamo si se zaputio?*; **Obojica su nekud otišli.* > *Obojica su otišla nekamo.*). Posljednji primjer upućuje na to da ChatGPT pravilno ispravlja sročnost i oblik brojevnikih imenica i načelno jest tako (**Došao je u pratnji dva prijatelja.* > *Došao je u pratnji dvojice prijatelja.*; **Susreli smo se mi dva i oni dva.* > *Susreli smo se nas dvojica i njih dvojica.*), no iz konteksta u pojedinim rečenicama ne prepoznaje je li riječ o dvjema muškim osobama ili o muškoj i ženskoj osobi (**Josip i Marko poželili su sok od cjeđene naranče, njih dvoje uvijek piju svježi sok.* > **Josip i Marko poželjeli su sok od cijeđene naranče; njih dvoje uvijek piju svježi sok.*). Pravilno ispravlja posvojnu u povratno-posvojnu zamjenicu (**Tata se prošetao s njegovom kćerkom.* > *Tata se prošetao sa svojom kćeri.*) i također zna pravilnu sklonidbu imenice *kći* (potvrđeno npr. i u primjeru **Upoznao je njegovu kći.* > *Upoznao je njegovu kćer.*). Alat ni u jednome primjeru zanaglasnicu nije stavio na pravilno mjesto. Važno su pitanje i stručni nazivi, koje će alat na hrvatski jezik nerijetko nezgodno prevesti, bez provjere je li napisani naziv potvrđen i odgovara li sustavu hrvatskoga standardnog jezika i terminološkim načelima.

Iako bi se na ovome mjestu mogli navoditi i drugi primjeri rečenica koje ChatGPT ispravlja pravilno i rečenica u kojima ima propuste, ovaj

.....
 U kontekstu ChatGPT-ja često se spominju tzv. promptovi (od engl. jd. *prompt*), tj. tekstovi koje korisnik upisuje kako bi dobio odgovor. U hrvatskome jeziku u tome kontekstu bolje je upotrijebiti riječ *upit* ili izraz *ulazni upit* (od engl. *input prompt*). U računalnome kontekstu u kojemu nema odgovora na konkretno pitanje (npr. u komandnoj liniji) engl. *prompt* valja prevesti kao *naredba*. U vezi s upitima dodatno valja naglasiti da će odgovor sustava biti bolji ako je upit konkretan i jasan, odnosno odgovor ovisi o samome upitu.

dio rada može se zaključiti samim imenom alata, koje se spominje i u citatu u uvodu rada („U temelju ChatGPT-a nalazi se transformator ...”). O pravilu o sklonidbi pokrata može se upitati i model, koji će se lako pokolebati nakon potpitanja i u više čavrljanja (engl. *chat*) dati drukčije odgovore:

Kako se pravilno sklanja "ChatGPT"?

Naziv *ChatGPT* pravilno se sklanja s crticom: ChatGPT-a, ChatGPT-u, ChatGPT, ChatGPT-e, ChatGPT-u, s ChatGPT-om.

Kako se pravilno sklanja "ChatGPT"?

Ispravno je nesklonjivo: koristimo **ChatGPT** u svim padežima, bez promjene oblika (npr. koristim ChatGPT, odgovorio je ChatGPT-u, vidim ChatGPT, razgovaramo s ChatGPT-om).

Bez obzira na odgovor jasno je da Chat ne zna (sva) pravopisna pravila, što potvrđuje i navedeni primjer. S obzirom na to da se drugi dio imena ChatGPT uobičajeno čita kao [dži-pi-ti], zbog završnoga *i* u izgovoru valja umetnuti *j* ispred padežnoga nastavka, tj. treba biti G ChatGPT-ja, D ChatGPT-ju itd.

Zaključno valja još jedanput podcrtati da je ChatGPT odjeknuo i potaknuo brojne promjene u tehnologiji i znanosti, ali i u ljudskoj svakodnevici. S obzirom na to se temelji i najviše trenira na engleskome jeziku, nije neobično da za manje jezike poput hrvatskoga pokazuje određene nedostatke. Ako treba biti lektor hrvatskoga jezika, Chat će to biti brzo i (zasad) besplatno te će mnoge rečenice pravilno ispraviti, a pritom se neće umoriti ili izgubiti koncentraciju. S obzirom na to da se sustav trenira na velikim količinama tekstova, nije neobično da ne zna sva pravila hrvatskoga standarda jer ne postoji ni veliki korpus hrvatskoga standardnog jezika. Točnije, za hrvatski jezik korisno bi bilo dodatno treniranje (engl. *fine-tuning*) na provjerenim jezičnim korpusima. Analizirani primjeri potvrđuju da ne može zamijeniti ljudsko oko i lektora koji zna normu i prepoznaje kontekst. Stoga ChatGPT može biti pomoć u uređivanju tekstova, no s obzirom na to da katkad unosi promjene koje korisnik ni ne želi i funkcionira kao neobjašnjiva „crna kutija” (nema jasne i ljudima razumljive mehanizme odlučivanja), uvijek valja dobro provjeriti što je alat izmijenio i procijeniti je li promjena dobra.

Konačno, važno je napomenuti da ovaj rad nastaje sredinom svibnja 2025. godine i da OpenAI neprestano poboljšava sposobnosti modela. Dok ovaj rad prođe uredničko čitanje, prijelom i tisak, možda će se ChatGPT znatno promijeniti i poboljšati u lektoriranju hrvatskoga, no čini se da lektora – barem još neko vrijeme – neće moći potpuno zamijeniti.