

CHATGPT: OTKRIJTE MAGIJU VELIKIH JEZIČKIH MODELA

CHATGPT: DISCOVER THE MAGIC BEHIND LARGE LANGUAGE MODELS

NIKOLA JELISAVAC¹

Pregledni rad
DOI: 10.5937/VI24069J

Rezime: Ovaj rad predstavlja pristupačan uvod u Velike Jezičke Modele (VJM) za inženjerske stručnjake izvan domena softverskog inženjerstva. Povlačeći paralele između ljudske kognicije i veštačkih neuronskih mreža, objašnjavamo fundamentalne koncepte koji stoje iza VJM-a, njihov proces treniranja i operativne karakteristike. Cilj rada je da demistifikuje ove kompleksne sisteme kroz pristupačne analogije i praktične primere, sa posebnim osvrtom na arhitekturu modela, proces treniranja i primenu u inženjerskoj praksi.

Glavne reči: Veštačka inteligencija (VI), Veliki jezički modeli (VJM), agenti

Abstract: This paper presents an accessible introduction to Large Language Models (LLMs) for engineering professionals outside the field of software engineering. By drawing parallels between human cognition and artificial neural networks, we explain the fundamental concepts behind LLMs, their training process, and operational characteristics. The aim of this paper is to demystify these complex systems through relatable analogies and practical examples, with special emphasis on model architecture, training process, and applications in engineering practice.

Key Words: Artificial Intelligence (AI), Large Language Models (LLM), Agents

1. Uvod

Pojava Velikih jezičkih modela (engl. Large Language Models – LLM), čiji je primer sistem ChatGPT (engl. Chat Generative Pre-Trained Model), predstavlja značajan napredak u veštačkoj inteligenciji u protekloj deceniji. Da bismo razumeli ove sofisticirane sisteme, ključno je ispitati njihovu arhitekturu, metodologiju treniranja i operativne principe na način koji će biti razumljiv široj inženjerskoj publici i van softverske industrije - koristeći analogije koje su bliske

¹ Nikola Jelisavac, Microsoft, Seattle, USA, nikola.jelisavac@gmail.com, ORCID 0009-0007-8937-172X

ljudskom umu. Povezujući koncepte sa svakodnevnim ljudskim iskustvima, nadam se da ćemo osvetliti osnovnu tehnologiju na pristupačan način, što će omogućiti praktično razumevanje ovih modela.

2. Arhitektura neuronske mreže: biološka paralela

U srži svakog velikog jezičkog modela nalazi se neuronska mreža koja funkcioniše slično kao neuroni u ljudskom mozgu. Zamislite milijarde međusobno povezanih čvorova koji prenose signale jedan drugom. Ovi čvorovi se prilagođavaju na osnovu ulaza i utiču na „odluke“ jedni drugih—baš kao što neuroni u našem mozgu reaguju kada su stimulisani. Kada je veliki jezički model inicijalno kreiran, on podseća na mozak novorođenčeta: mreža nasumičnih, besmislenih veza. U toj fazi, VJM ne zna ništa i proizvodi izlaze koji nemaju nikakvu koherenciju.

3. Trening faza 1: unos znanja

Prva kritična faza u razvoju VJM-a uključuje izlaganje neuronske mreže ogromnim količinama tekstualnih informacija. Ovo uključuje: tehničke članke, objavljene knjige, akademske radove, javno dostupan sadržaj. Metodologija treniranja koristi jedinstveni pristup fokusiran na predviđanje sledeće reči. Sistem uči tako što mu se pokaže deo teksta iz izvornog materijala i VJM treba da predvidi sledeću reč u sekvenci, na osnovu prethodnog konteksta. Na primer:

- Ulaz: „Sunce sija, nebo je ____“
- Očekivani izlaz: „plavo“

Proces obuke uključuje pravljenje predviđanja i njihovo ispravljanje—slično načinu na koji ljudi uče kroz pokušaje i greške. U početku, model će praviti greške, predviđajući pogrešne reči ili besmislene rečenice. Kada napravi ove greške, koristi se proces nazvan povratna propagacija (engl. Backpropagation). Povratna propagacija je proces koji omogućava modelu da prilagodi jačine veza između svojih čvorova na osnovu grešaka u predikciji, slično kao kada osoba ispravlja svoje zaključke kroz povratne informacije iz okruženja. Važno je razumeti da model ne memoriše reči tačno onako kako se pojavljuju, već stvara veze između koncepata. Kao što ljudi možda neće zapamtiti svaki tačan izraz iz knjige, ali mogu da se sete ključnih koncepata i da ih izraze svojim rečima, tako i VJM-i formiraju apstraktne predstave podataka iz tekstualnog materijala.

4. Igra „Sledeća reč“ i izgradnja koherencije

Osnovna funkcija VJM-a je da predviđa sledeću reč u nizu - što može zvučati obmanjujuće jednostavno ali zaista funkcioniše u praksi, gde kroz nastavljanje sadržaja predviđajući reč po reč, VJM mogu da formiraju kompleksne misli.

Da bismo ovo ilustrovali, setite se popularne igre reči koju ste možda igrali kao deca. U ovoj igri deca sede u krugu i svako izgovara po jednu reč, kako bi zajedno izgradili priču. Svako dete dodaje reč na osnovu onoga što je prethodno rečeno. Na primer:

- Prvo dete: „Danas“
- Drugo dete: „Danas je“
- Treće dete: „Danas je lep“
- Četvrto dete: „Danas je lep dan“

Ova jednostavna igra demonstrira kako sekvencijalno predviđanje reči može stvoriti koherentan i smislen sadržaj. Da bismo dodatno ilustrovali važnost znanja i konteksta, razmotrimo proširenje ove igre na grupu profesora fizike kojima je postavljeno pitanje o Ajnštajnovoj čuvenoj jednačini. Profesori, oslanjajući se na svoje stručno znanje, mogu uspešno konstruisati jednačinu reč po reč:

- Prvi profesor: „E“
 - Drugi profesor: „E jednako“
 - Treći profesor: „E jednako m“
 - Četvrti profesor: „E jednako m c“
 - Peti profesor: „E jednako m c kvadrat“
- Alternativno, mogu početi drugačije:
- Prvi profesor: „Ajnštajnova“
 - Drugi profesor: „Ajnštajnova čuvena“
 - Treći profesor: „Ajnštajnova čuvena formula“
 - Četvrti profesor: „Ajnštajnova čuvena formula glasi“
 - Peti profesor: „Ajnštajnova čuvena formula glasi E“ i tako dalje...

Ovaj primer ilustruje nekoliko ključnih tačaka:

- Sposobnost generisanja tačnih odgovora zavisi od posedovanja neophodnog znanja (deca koja bi igrala istu igru ne bi mogla da proizvedu odgovor o Ajnštajnovoj jednačini)
- Ista informacija se može izraziti počevši od različitih tačaka, pri čemu naredni učesnici prilagođavaju svoje doprinose u skladu sa tim
- Jednom kada se izabere određeni put, odgovori moraju slediti taj kontekst sekvencijalno

Ovo preslikava način na koji VJM-i funkcionišu - oni generišu tekst na osnovu svog treniranog znanja, mogu pristupiti temama iz različitih početnih tačaka, ali su ograničeni na kretanje napred sa izabranim putem, bez mogućnosti da se vrate unazad i revidiraju svoj pristup jednom kada je započet. Kvalitet ovih predviđanja u potpunosti zavisi od baze znanja stečene tokom treniranja.

5. Trening faza 2: usklađivanje modela za interakciju sa ljudima

Prva faza obuke uči VJM ogromnoj količini znanja, ali to samo po sebi nije dovoljno da bi model bio koristan za efektivnu interakciju sa ljudima. U ovom stadijumu VJM-i mogu generisati odgovore koji su činjenično tačni, ali neusa- glašeni sa očekivanjima tipičnih razgovora. Na primer, ako se postavi pitanje „Koji je glavni grad Francuske?“, model bi mogao pratiti obrazac i odgovoriti sa „Koji je glavni grad Srbije?“, umesto odgovora koji bi većina ljudi očekivala na postavljeno pitanje – „Pariz“. Slično, model bi mogao nastaviti odgovor sa pes- mom, dugačkim tekstom gde bi kratak odgovor bio dovoljan, odgovoriti drugim jezikom i ostalim opcijama koje bi formirale tačne rečenice ali ne bi bile očekivani primereni odgovor.

Da bi ovi modeli mogli smisleno da komuniciraju sa ljudima, istraživači uvode fazu dodatnog treninga da bi naučili model da usaglasi odgovore sa ljudskim očekivanjima. Tokom ove faze, model se obučava na hiljadama primera ručno napisanih razgovora između čoveka i zamišljenog korisnog asistenta. Ovi razgovori su ručno kreirani kako bi precizno izrazili način na koji bi koristan asistent trebalo da komunicira - prikazujući odgovarajući ton, dužinu odgovora i sadržaj za različite tipove upita. Nakon ove faze, model uči da pruža korisne, sažete i relevantne odgovore, kao i da ljubazno odbije neprikladne zahteve.

6. Dvojna priroda ljudskog razmišljanja i VJM-a

Iz psihologije, ljudska kognicija se može grubo kategorizovati na dva načina razmišljanja: „Sistem 1“ i „Sistem 2“. Sistem 1 je brzo, intuitivno razmišljanje - ono koje koristimo da odmah odgovorimo na osnovu iskustva ili „osećaja iz stomaka“. Sistem 2 je sporije, namerno razmišljanje koje koristimo za složeno rasuđivanje. VJM-i, onakvi kakvi postoje danas, rade slično ljudskom „Sistem 1“ načinu razmišljanja. Oni generišu odgovore odmah na osnovu prethodnog učenja, bez dubokog promišljanja ili planiranja. Iako je ovo dovoljno za mnoge zadatke, dolazi sa određenim ograničenjima. Ako VJM počne da ide pogrešnim putem u svom odgovoru, ne može se vratiti unazad i ispraviti se.

7. Ključne karakteristike i ograničenja

Na osnovu dosadašnjih objašnjenja kako funkcionišu VJM-i rezimiraćemo par njihovih ključnih karakteristike i ograničenja, a zatim ćemo razmotriti nekoliko strategija za prevazilaženje datih ograničenja.

1. Statična baza znanja

- Slično kako student nakon završenog fakulteta prestaje da se fokusira na učenje i umesto toga se fokusira na primenu stečenog znanja u poslu, VJM

završavaju svoje učenje nakon faze treninga i ne stižu novo znanje u interakcijama sa ljudima.

- Da bi stekli novo znanje, model mora da se ponovo trenira na dodatnom tekstu.

2. Dominacija „Sistema 1“

- Odgovori postojećih VJM-a su trenutni i intuitivni, slično kako ljudi razmišljaju po „Sistemu 1“ većim delom svog dana.
- Iako je ovo praktično u većini situacija, ovaj način razmišljanja donosi i ograničenja u vidu mogućnost revizije ili preispitivanja početnih izlaza potrebno za rešavanje jako kompleksnih problema.

3. Standardizovani obrasci interakcije

- Odgovori su kalibrisani prema prosečnim stilovima ljudske interakcije, sa konzistentnim nivou učtivosti i formalnosti, i predodređenim rukovanjem osetljivim temama.
- Ovim pristupom je interakcija sa VJM olakšana jer sa vrlo malo informacija mogu da odgovore na dobar standardizovani način. Međutim, nekad bi VJM mogao da pruži bolje odgovor koji bi bio više prilagođen datom problemu i znanju korisnika.

7. Napredni saveti za korišćenje VJM-a

Na osnovu dosadašnjeg učenja o VJM-a, možemo primeniti nekoliko trikova kako bismo izvukli maksimum iz ovih modela.

Prva stvar koju treba imati na umu jeste da ovi modeli operišu u „Sistemu 1“ razmišljanja, što znači da će se osloniti na prvu misao koja im padne na pamet. Kao i kod ljudi, prva ideja nije uvek najbolja, stoga umesto da tražite samo jedno rešenje, možete tražiti pet ili deset predloga. VJM-ima je lako da generišu više rešenja, a možete ih i usmeriti upitom da želite da budu različita i originalna, i da VJM sagleda problem iz različitih perspektiva. Ova strategija posebno je korisna u kreativnim zadacima i rešavanju kompleksnih problema, gde različite perspektive mogu dovesti do inovativnijih rešenja.

Sledeća tehnika se odnosi na ulogu modela. Iako su VJM-i trenirani da odgovaraju na pitanja na „prosečan“ način, možete im dodeliti specifičnu ulogu. Na primer, možete ih zamoliti da odgovore kao naučnik iz oblasti elektroinženjeringa ili kao vrhunski geolog. Ovim postupkom fokusirate model na znanje koje poseduje iz te specifične oblasti, što može rezultirati preciznijim i relevantnijim odgovorima.

Treća tehnika je personalizacija odgovora prema nivou razumevanja korisnika. Iako su modeli trenirani da pružaju „prosečno najbolje“ odgovore, mogu se fino prilagoditi specifičnim potrebama korisnika. Ako ste novi u nekoj oblasti i želite osnovno objašnjenje, možete reći modelu da objasni pojmove kao da priča sa srednjoškolcem. S druge strane, ako ste stručnjak, možete tražiti od modela da koristi stručne termine i pruži duboko tehničko objašnjenje. Takođe, možete odrediti stil odgovora da li treba da bude napisan kao tehnički rad, email ili tekst za objavu na društvenim mrežama.

Još jedna važna tehnika je davanje primera modelu. Iako VJM-ovi ne mogu trajno učiti iz interakcija, mogu efikasno adaptirati svoje odgovore prema datim primerima unutar jedne sesije. Na primer, ako želite da napišete email koji zvuči slično vašim prethodnim emailovima, možete modelu pružiti nekoliko primera i reći mu da se ugleda na te stilove. Model će prilagoditi svoj odgovor kako bi odgovarao uzorcima koje ste mu dali.

Peti savet je podstaći model na više iteracija odgovora. Kao što često ljudi rade u praksi, prvo pravimo grubi nacrt teksta, zatim ga pročitamo i poboljšavamo. Slično, možete zamoliti model da nakon prvog odgovora, iskritikuje ono što je napisao, a zatim da napravi novu verziju baziranu na kritikama i sugestijama prethodne verzije. Na ovaj način, dobijate značajno unapređen odgovor.

Zanimljiva tehnika koja podstiče model na dublje razmišljanje za teže probleme jeste sledeći primer. Slično kao kod ljudi, strategija da prvo razbijemo problem na osnovne komponente na višem nivou pre nego što razmotrimo detalje dovodi do boljeg rešenja, ovo možemo primeniti i na VJM.

Jednostavnim zahtevom posle pitanja da želimo da VJM reši problem „korak po korak“, gde se od modela eksplicitno traži da razloži problem na manje celine i metodički pristupi rešenju, dokazano dovodi do značajnog unapređenja u rezultatima.

8. Nova faza u razvoju VJM-a: „Duboko razmišljanje“

Do sada smo govorili o „Sistemu 1“ razmišljanja koji VJM-i koriste, ali istraživači rade na sledećem nivou pristupa, razvijajući modele koji mogu uključiti i više faze razmišljanja koja će biti slična ljudskom „Sistemu 2“ razmišljanja. OpenAI, kompanija koja je tvorac ChatGPT-a, je nedavno lansirala ranu probnu verziju modela po imenu „O-1“ koji ima novu fazu obuke koja uči model tehnikama automatskog rezonovanja, kao i fazu „razmišljanja“ pre davanja konačnog odgovora na korisnikovo pitanje. U ovoj fazi, model može unapred planirati, analizirati i revidirati svoj odgovor pre nego što ga pruži korisniku. Ovaj proces pomaže da se poveća tačnost i kvalitet rešenja za složenije probleme.

9. Integracija sa alatima i praktična primena

Značajan napredak u razvoju VJM-a predstavlja njihova sposobnost interakcije sa eksternim alatima.

Da bi razmotrili ovaj slušaj i zašto bi ovo bilo korisno, ilustrovaćemo na primeru matematičkih operacija. Zamislimo situaciju kada je potrebno izračunati kvadrat broja. Kada je u pitanju jednostavan zadatak poput „3²“, većina ljudi može trenutno odgovoriti „9“ koristeći memorisano znanje. Međutim, kada se suočimo sa zadatkom poput „3.12345²“, situacija je značajno drugačija. Iako razumemo koncept kvadriranja, većina ljudi će posegnuti za kalkulatorom. Slično tome, VJM-ovi poseduju razumevanje matematičkih koncepata, ali mogu biti precizniji kada im se omogući pristup računarskim alatima za izvršavanje složenih kalkulacija.

Međutim do sada smo pričali samo o mogućnostima da VJM može tekstom da odgovori na dati upit, onda se postavlja pitanje kako je moguće da damo VJM moć da pokreću alate? Da bismo razumeli kako bi ovo mogli da ostvarimo, osvrnimo se opet na jedan primer iz ljudske interakcije. Posebno je interesantna analogija sa situacijom pružanja tehničke podrške preko telefona. Zamislite scenario gde pokušavate pomoći nekome da reši problem na računaru putem telefonskog razgovora.

Ne možete direktno videti ili kontrolisati njihov računar, ali možete davati precizna uputstva tekstom i primati povratne informacije o rezultatima svake akcije. Integracija VJM-a sa eksternim alatima funkcioniše na sličan način. U uputstvu za VJM pruža se opis dostupnih alata. U slučaju da VJM želi da koristi alat za određeni problem, u odgovoru će dati precizne instrukcije kako koristiti alat. Sistem koji koristi VJM potom izvršava te instrukcije i vraća rezultat kao odgovor VJM-u, koji dalje interpretira dobijene rezultate. Ovaj pristup značajno proširuje VJM-ove praktične mogućnosti.

Ova integracija sa alatima već je implementirana u vodećim VJM platformama. Na primer, ChatGPT kroz svoju funkciju „Code Interpreter“ (interpreter računarskog koda) ili „Advanced Data Analysis“ (napredna analiza podataka) omogućava modelu da direktno izvršava programski kod za rešavanje kompleksnih zadataka. Korisnik može zatražiti ne samo izračunavanje složenih matematičkih izraza, već i kreiranje vizuelizacija podataka. Na primer, model može napisati Python kod (Python je popularni programski jezik) koji će izračunati vrednost složenog izraza i prikazati rezultate kroz graf, što demonstrira sposobnost kombinovanja matematičke preciznosti sa vizuelnom prezentacijom podataka. Napomena: ova funkcionalnost je dostupna samo u plaćenj verziji ChatGPT-a.

Slika 1 i 2 prikazuju kako ChatGPT generiše kod za crtanje grafa kvadratne funkcije i izračunavanje vrednosti za 3.14. Korisnik može kliknuti na oznaku „Analyzed“ ili „[>_]“ da bi pregledao kod koji je ChatGPT generisao i koji je izvršen u pozadini, nakon čega je rezultat vraćen ChatGPT-u radi završetka odgovora.

Claude.ai je platforma sa vodećim VJM modelom koja demonstrira jedinstveni pristup integraciji alata putem sistema „artefakata“. Ovaj sistem omogućava kreiranje interaktivnih web komponenti direktno u razgovoru, što je posebno korisno za izradu specijalizovanih mini aplikacija za određene domene. Na primer, model može kreirati:

- Interaktivne 3D reprezentacije molekula za hemičare,
- Kalkulatore specifičnih formula za geologe,
- Vizuelizacije podataka za analitičare.

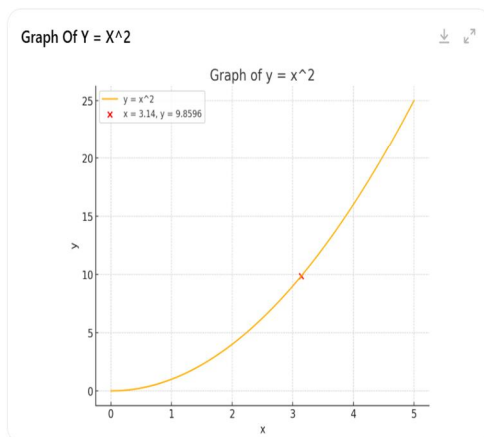
Ove komponente mogu se deliti sa drugim korisnicima putem jedinstvenih URL-ova, čime se omogućava saradnja i širenje znanja unutar stručnih zajednica.

ChatGPT 4o

Share

iskoristi tvoj Code Interpreter tool da izracunas x na kvadrat za vrednost 3.14 i nacrtaj graf za tu funkciju

Analyzed



Rezultat za x^2 kada je $x = 3.14$ iznosi 9.8596. Na grafikonu je prikazana funkcija $y = x^2$, a tačka za $x = 3.14$ je označena crvenom bojom. [↩]

Slika 1. ChatGPT koristi Code Interpreter alat (izvor: ChatGPT)



Analyzed ^

```
python

import numpy as np
import matplotlib.pyplot as plt

# Define the value
x_value = 3.14
# Calculate x squared
x_squared = x_value ** 2

# Create an array for x values
x = np.linspace(0, 5, 100)
# Calculate y = x^2 for the range of x values
y = x ** 2

# Plotting
plt.figure(figsize=(8, 6))
plt.plot(x, y, label='y = x^2')
plt.scatter(x_value, x_squared, color='red', label=f'x = {x_value}, y = {x_squared}')
plt.xlabel('x')
plt.ylabel('y')
plt.title('Graph of y = x^2')
plt.legend()
plt.grid(True)
plt.show()

x_squared
```

Always show details  Copy codeResult
9.8596

Slika 2. Programski kod koji je ChatGPT napisao za predhodni problem (izvor: ChatGPT)

10. Agentni sistemi u praksi

Korišćenje alata predstavlja dodatnu mogućnost primene veštačke inteligencije. Umesto pružanja jednostavnih odgovora korisniku, kombinacija alata omogućava kreiranje sistema u kojem veštačka inteligencija dobija instrukcije za obavljanje kompleksnih zadataka iz određenih oblasti, čime postaje „agent“ zadužen za izvršenje tih uloga. Takođe, do sada su navedeni primeri podrazumevali da korisnik uvek inicira interakciju sa VJM-om. Međutim, moguće je uspostaviti automatsku integraciju sa VJM-om koja će se aktivirati prilikom

dostupnosti novih podataka. Najjednostavnije je ilustrovati ove koncepte na primerima.

Jednostavan primera agenata se može ilustrovati kroz konkretan primer automatske obrade sastanaka u Microsoft Teams-u. Nakon snimljenog sastanka, sistem automatski šalje transkript VJM-u, koji zatim generiše sažetak ključnih tačaka, identifikuje akcione stavke i označava odgovorne osobe. Ovo je efikasan primer kako VJM može delovati autonomno, preuzimajući ulogu virtuelnog sekretara koji obrađuje informacije i proizvodi korisne rezultate bez direktne ljudske intervencije.

Još jedan praktičan primer je automatizacija email komunikacije, gde agent može samostalno:

- Proceniti prioritet dolazne poruke
- Odgovoriti na rutinska pitanja koristeći postojeću dokumentaciju
- Proslediti kritične poruke odgovarajućim menadžerima
- Generisati predloge odgovora za složenije upite

Velike firme su zainteresovane za korišćenje VJM-a za stvaranje agenata koji unapređuju rad i automatizuju procese. Microsoft je uveo opciju jednostavnog definisanja agenta u sklopu Microsoft Copilot Studio platforme.

11. Korišćenje multimodalnih modela

Jedan od poslednjih napredaka u oblasti VJM-a je integracija različitih modaliteta. U novijim verzijama, modeli su obogaćeni sposobnošću da razumeju slike, video i zvuk pored teksta.

Ovo znači da model može interpretirati šta se nalazi na slici ili razumevati emocionalni kontekst u audio zapisu, i na osnovu toga pružiti odgovor. Pored toga, modeli sada mogu direktno generisati audio izlaze koji zvuče prirodnije, sa bogatijom intonacijom i emocijom.

Ove sposobnosti otvaraju vrata za potpuno nove scenarije, kao što su prirodne interakcije sa korisnicima preko telefona, automatizovano zakazivanje sastanaka, ili čak inicijalna komunikacija sa klijentima pre nego što se uključi prava osoba.

Razumevanje slika takođe može doprineti naprednim mogućnostima kao što su automatsko prepoznavanje i analiza dokumenata, dijagnostika u medicini kroz analizu medicinskih snimaka, kao i prepoznavanje objekata i situacija na sigurnosnim kamerama radi poboljšanja sigurnosti. Iako je ovo bilo moguće u određenoj meri i ranije, sada postaje jednostavnije da jedan VJM pokrije sve ove primene. Na primeru prikazanom na slici 3, može se videti kako Claude.ai može precizno identifikovati geološki naziv minerala sa slike.



Slika 3. Claude.ai odgovor na upit za prepoznavanje kamena na slici (izvor: Claude.ai)

12. Zaključak

Razumevanje principa rada VJM-a, njihovih mogućnosti i ograničenja ključno je za njihovu efektivnu primenu u praksi. Kroz pravilnu primenu opisanih strategija i tehnika, možemo značajno unaprediti kvalitet i korisnost njihovih odgovora. Integracija sa alatima i proširene mogućnosti interpretacije različitih medija transformišu način na koji možemo koristiti ove sisteme.

Kontinuirani razvoj u oblastima naprednog rezonovanja, zajedno sa mogućnostima autonomnog delovanja kroz agentne sisteme, otvara put ka još sofisticiranijim primenama. Budućnost će doneti dalja unapređenja u oblasti autonomije, kontekstualnog razumevanja i prirodnosti interakcije, čineći ove sisteme još moćnijim alatima u rukama stručnjaka različitih profila.

13. Literatura

- [1] *Introducing OpenAI o1* Dostupno na: <https://openai.com/o1/> (Pristupljeno: 6.11.2024), 2024.
- [2] Takeshi K, Shixiang Shane G, Machel R, Yutaka M, Yusuke I, *Large Language Models are Zero-Shot Reasoners*, NIPS'22: Proceedings of the 36th

- International Conference on Neural Information Processing Systems (2022), arXiv:2205.11916, 2022
- [3] *New autonomous agents scale your team like never before* (2024). Dostupno na: <https://blogs.microsoft.com/blog/2024/10/21/new-autonomous-agents-scale-your-team-like-never-before/> (Pristupljeno: 6.11.2024)
 - [4] Kahneman D, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, New York, 2011
 - [5] *Function calling*. Dostupno na: <https://platform.openai.com/docs/guides/function-calling> (Pristupljeno: 6.11.2024)
 - [6] *ChatGPT can now see, hear, and speak* (2023). Dostupno na: <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/> (Pristupljeno: 6.11.2024)