

Umetna inteligenca pri nacionalnem preverjanju znanja iz slovenskega jezika



Šifra: ume12

Razred: 9. razred

Področje: slovenščina

Mestna občina Celje, Mladi za Celje

Celje, marec 2024

Povzetek

Raziskovalna naloga se ukvarja z vlogo umetne inteligence v izobraževanju, zlasti z uporabo modela ChatGPT pri reševanju nacionalnega preverjanja znanja (NPZ) iz slovenščine v 9. razredu. V tujini se je ChatGPT zelo izkazal na standardiziranih testih, v Sloveniji pa tega še nis(m)o preverili. Pri raziskavi je bil uporabljen anketni vprašalnik za devetošolce in izveden eksperiment, kjer je ChatGPT reševal NPZ. Ugotovitve kažejo, da devetošolci poznajo umetno inteligenco, prav tako model ChatGPT. Pri eksperimentu je ChatGPT dosegel nadpovprečne rezultate, vendar so se rezultati med leti razlikovali. Čeprav je večina devetošolcev pozitivno ocenila uporabo umetne inteligence, so precej zadržani glede njenega morebitnega nadomestila za učitelje.

Ključne besede: slovenščina, nacionalno preverjanje znanja, umetna inteligenca, ChatGPT

Abstract

The research deals with the role of artificial intelligence in education, especially with using the ChatGPT model in solving the national exam (NPZ) in Slovenia in the 9th grade. ChatGPT has proven itself in standardized tests abroad, but this still needs to be verified in Slovenia. Our research used a survey questionnaire for ninth graders and an experiment where ChatGPT solved NPZ. We found that ninth graders know artificial intelligence, and most also know ChatGPT and evaluate it positively. ChatGPT achieved above-average results in the experiment, but the results varied between years. Although most ninth graders were optimistic about using artificial intelligence, they are still hesitant about its potential replacement for teachers.

Keywords: Slovenian, national exam, artificial intelligence, ChatGPT

*Ker se opredeljujemo za inteligentne, četudi morda z neutemeljenimi razlogi,
skušamo inteligenco obravnavati kot neizogibno posledico evolucije ...*

Stephen Hawking

Kazalo vsebine

1	Uvod.....	7
1.1	Opredelitev raziskovalnega problema	7
1.2	Hipoteze	8
1.3	Metode dela	9
2	Teoretični del.....	10
2.1	Veliki jezikovni modeli	10
2.2	ChatGPT.....	12
2.2.1	Značilnosti	13
2.2.2	Navodila za pisanje dobrih besedilnih pozivov.....	14
2.2.3	Različice ChatGPT-ja	15
2.3	Nacionalno preverjanje znanja (NPZ)	16
2.4	Sorodne raziskave.....	17
3	Praktični del.....	20
3.1	Uvod.....	20
3.2	Predstavitev in utemeljitev rezultatov	21
3.2.1	Anketa	21
3.2.2	Analiza rezultatov NPZ-jev ChatGPT-ja.....	31
3.3	Izzivi pri raziskovalnem delu	44
4	Zaključek	45
5	Viri in literatura	48

Kazalo slik

Slika 1: Evolucija velikih jezikovnih modelov [6]	10
Slika 2: Uspešnost modela ChatGPT pri različnih problemih [20]	19
Slika 3: Zasnova raziskave	20
Slika 4: Razporejenost anketirancev glede na spol	21
Slika 5: Samoocena znanja slovenskega jezika.....	22
Slika 6: Samoocena znanja slovenskega jezika glede na to, ali je slovenščina materni jezik ..	22
Slika 7: Samoocena znanja slovenskega jezika glede na spol	22
Slika 8: Samoocena poznavanja umetne inteligence	23
Slika 9: Samoocena poznavanja umetne inteligence glede na spol	23
Slika 10: Poznavanje primerov modelov umetne inteligence	24
Slika 11: Jezik za postavljanje vprašanja modelu ChatGPT pri reševanju problema	25
Slika 12: Pričakovan skupen rezultat ChatGPT-ja pri reševanju NPZ-ja	26
Slika 13: Pričakovan rezultat ChatGPT-ja pri pravopisu	26
Slika 14: Pričakovan rezultat ChatGPT-ja pri oblikovanju besedila	27
Slika 15: Pričakovan rezultat ChatGPT-ja pri slogovni pravilnosti	27
Slika 16: Pričakovan rezultat ChatGPT-ja pri slovnici	27
Slika 17: Pričakovan rezultat ChatGPT-ja pri književnosti	28
Slika 18: Pričakovan rezultat ChatGPT-ja pri jezikovnih nalogah	28
Slika 19: Pričakovan rezultat ChatGPT-ja pri bralnem razumevanju.....	28
Slika 20: Ali bi bil svet z umetno inteligenco boljši?.....	29
Slika 21: Ali bi bil svet z umetno inteligenco boljši - glede na spol	29
Slika 22: Ali bi lahko umetna inteligenca v prihodnosti nadomestila učitelja?	29
Slika 23: Ali bi lahko umetna inteligenca v prihodnosti nadomestila učitelja - glede na spol.	30
Slika 24: Mnenje o uporabi umetne inteligence pri učenju	30
Slika 25: Oblak besed anketirancev pri opisu uporabe umetne inteligence pri učenju	31
Slika 26: Porazdelitev točk pri NPZ 2023 in ChatGPT 3.5 [26].....	34
Slika 27: Porazdelitev točk pri NPZ 2022 in ChatGPT 3.5 [27].....	36
Slika 28: Porazdelitev točk pri NPZ 2021 in ChatGPT 3.5 [28].....	38
Slika 29: Porazdelitev točk pri NPZ 2021 in ChatGPT 4 [28].....	40
Slika 30: Uspešnost ChatGPT 3.5 glede na območje nalog in leta	43
Slika 31: Primerjava ChatGPT 3.5 in 4 pri NPZ 2021.....	44

Kazalo tabel

Tabela 1: Primerjava različnih verzij ChatGPT [10]	16
Tabela 2: Sestavni deli NPZ-ja iz slovenščine	17
Tabela 3: Povzetek napak ChatGPT 3.5 pri NPZ 2023 [25]	35
Tabela 4: Povzetek napak ChatGPT 3.5 pri NPZ 2022 [24]	37
Tabela 5: Povzetek napak ChatGPT 3.5 pri NPZ 2021 [23]	39
Tabela 6: Povzetek napak ChatGPT 4 pri NPZ 2021 [23]	41
Tabela 7: Primerjava dosežka ChatGPT-ja pri NPZ	42

1 Uvod

Ste se kdaj spraševali, kako lahko računalniki pametno pomagajo pri reševanju problemov? To je možno zaradi umetne inteligence, ki predstavlja čarobno moč, s katero lahko računalniki razumejo stvari, se učijo in celo naredijo pametne odločitve. Bolj natančno pa so jo strokovnjaki opredelili takole: *“Umetna inteligenca je zmožnost stroja, da izkazuje človeške lastnosti, kot so mišljenje, učenje, načrtovanje in kreativnost.”* [1]

V zadnjih letih se je področje umetne inteligence močno razvilo, prav tako je postalo dosegljivo širši javnosti. Pojavlja se povsod – na socialnih omrežjih, pri prepoznavi obraza za odklepanje telefona, v zdravstvu, šolstvu in vsakdanjem življenju nasploh. Posledično se z njim srečujemo - ali pa se še bodo – tudi mladostniki.

V računalniških igrah npr. umetna inteligenca pomaga ustvariti pametne nasprotnike, ki se učijo in postajajo boljši. Tako je izziv v igri vedno zanimiv. Primere uporabe najdemo pri razvoju pametnih prometnih sistemov, ki lahko predvidijo gneče in optimizirajo promet ter celo pomagajo pri razvoju samovozečih avtomobilov.

Hkrati se že pojavljajo tudi številni pomisleki in strah pred širšo uporabo. Z avtomatizacijo delovnih mest se nekateri bojijo, da bodo izgubili službo, ker jih bodo nadomestili "pametni" roboti. Problematično je tudi zbiranje velikih količin podatkov, ki so potrebni za delovanje modelov umetne inteligence, zlasti če se ti podatki ne uporabljajo pravilno. Umetna inteligenca nima čustev, kar pomeni, da se odločitve sprejemajo zgolj logično in to lahko v kakšni situaciji povzroči etične dileme. Nekateri tudi trdijo, da lahko postanemo preveč odvisni od tehnologije in smo tako še bolj ranljivi v primeru nedelovanja ali napak tehnologije. [2, 3]

1.1 Opredelitev raziskovalnega problema

V raziskovalni nalogi sva skušali ugotoviti, kako uspešen je ChatGPT pri reševanju NPZ-ja iz slovenskega jezika ob koncu tretjega trilettja. Pri tem sva bili pozorni na to, katere tipe nalog rešuje uspešno in kakšen je njegov končni dosežek v primerjavi s povprečnim dosežkom devetošolcev pri posameznih nacionalnih preverjanjih znanja iz slovenščine.

Med devetošolci celjskih osnovnih šol sva izvedli anketo, kjer naju je zanimalo njihovo poznavanje umetne inteligence, modelov umetne inteligence, njihovo mnenje o tem, kako uspešen bi bil ChatGPT pri reševanju NPZ-ja iz slovenskega jezika, in njihovo mnenje glede uspešnosti reševanja ChatGPT-ja različnih tipov nalog (npr. pravopis, književnost, oblikovanje besedila, slovnica, bralno razumevanje ipd.).

Takšne raziskave v slovenskem prostoru še nisva zasledili. Raziskovalni problem se nama je tudi zato zdel zanimiv in aktualen.

1.2 Hipoteze

Pri raziskovalni nalogi sva zastavili hipoteze, ki so nama omogočale preučevanje različnih vidikov uporabe ChatGPT-ja pri reševanju NPZ-ja, in sicer:

- **Hipoteza 1:** Devetošolci poznajo umetno inteligenco in najbolj razširjen model ChatGPT.
- **Hipoteza 2:** Devetošolci menijo, da bo ChatGPT uspešno reševal naloge pri NPZ-ju iz slovenščine.
- **Hipoteza 3:** Učenci so bolj naklonjeni uporabi umetne inteligence kot učenke.
- **Hipoteza 4:** Devetošolci menijo, da bi umetna inteligenca v prihodnosti lahko nadomestila učitelja.
- **Hipoteza 5:** ChatGPT uspešno rešuje NPZ iz slovenščine in je pri tem boljši kot povprečen devetošolec.
- **Hipoteza 6:** ChatGPT ima pri NPZ-ju največ težav s težjimi nalogami, ki jih večinoma ne rešijo niti najboljši učenci.

1.3 Metode dela

Pri najinem delu sva uporabili več raziskovalnih metod, in sicer:

1. Pregled literature

S pomočjo podatkov v člankih in na spletnih straneh, kar je navedeno med viri, sva pripravili vsebino teoretičnega dela v poglavju 2. Med iskanjem literature sva našli veliko koristnega gradiva, s pomočjo katerega sva podrobneje raziskali področje. Rezultate pregleda literature sva neposredno uporabili za pripravo anketnega vprašalnika.

2. Metoda spraševanja – anketa

Anketni vprašalnik [4] sva razdelili na štiri sklope in ga posredovali devetošolcem celjskih osnovnih šol. Večina vprašanj je bilo zaprtega tipa z izbiro na lestvici od 1 do 5, z izjemo enega delno zaprtega vprašanja in enega vprašanja odprtega tipa. Odziv anketirancev (128 sodelujočih) nama je omogočal podrobno analizo odgovorov, ki je predstavljena v poglavju 3.2.1.

3. Eksperiment

Pri eksperimentu sva želeli preveriti, kako uspešen je ChatGPT pri reševanju zadnjih treh NPZ-jev (iz leta 2023, 2022 in 2021) iz slovenščine v 9. razredu. Vprašanja s praznih pol sva preko pozivov vnesli v ChatGPT in zabeležili odgovore, ki sva jih pregledali in ocenili s pomočjo enotnih in posebej pripravljenih merilih.

4. Analiza podatkov

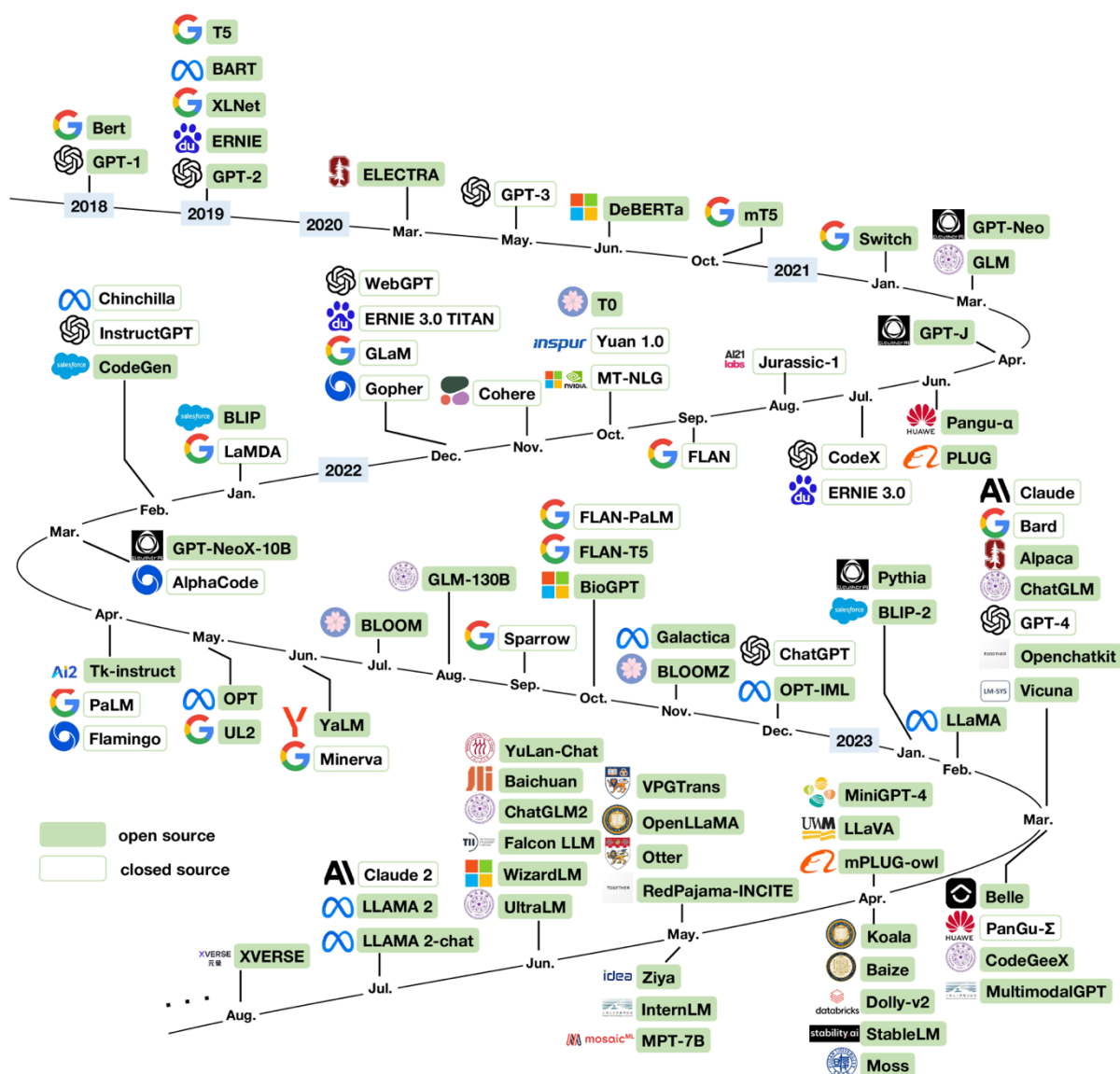
Analiza nama je omogočala razumevanje povezav ter ovrednotenje hipotez. Analizirali sva podatke vprašalnikov in reševanja NPZ-jev s strani ChatGPT-ja. Pri tem sva uporabili osnovne statistične značilnosti, kot je povprečje in porazdelitev odgovorov med različnimi možnostmi. Pri potrjevanju hipotez sva upoštevali tudi povezave med več spremenljivkami (npr. vpliv spola in predznanja). Primerjali sva uspešnost reševanja NPZ-ja, ki jo je dosegel ChatGPT, z dosežki devetošolcev pri reševanju enake testne pole. Rezultate analize sva prikazali v obliki grafov in tabel.

V nadaljevanju v poglavju 2 predstavljava teoretične osnove raziskovalne naloge, in sicer velike jezikovne modele z najbolj priljubljenim modelom ChatGPT. Sledi predstavitev nacionalnega preverjanja znanja, ki sva ga kasneje uporabili za preverjanje uspešnosti ChatGPT-ja. V poglavju 3 je predstavljen praktični del z anketo in reševanjem NPZ-ja s strani ChatGPT-ja, zadnjemu poglavju 4 pa sledijo zaključne ugotovitve in možnosti nadaljnjega dela.

2 Teoretični del

2.1 Veliki jezikovni modeli

Veliki jezikovni modeli (angl. *Large Language Models (LLM)*) so podpodročje umetne inteligence, ki uporablja velike količine besedilnih podatkov za simulacijo človeških pogovornih odzivov. Jezikovni modeli temeljijo na modelih globokega učenja, ki uporabljajo večplastne nevronske mreže za obdelavo, analizo in napovedovanje s kompleksnimi podatki za ustvarjanje odgovorov v naravnem jeziku. [5]



Slika 1: Evolucija velikih jezikovnih modelov [6]

Velike jezikovne modele lahko uporabljamo za analizo različnih vrst vhodnih podatkov, kot so: besedilo (Chat-GPT¹, Google Gemini²), slike (Dalle-E³, Leonardo⁴), video (Pictory⁵, Runway⁶), zvok (Speechify⁷, Murf⁸), podatki za trženje (Claude⁹, AdCreative¹⁰), oblikovanje (Canva¹¹, Adobe Firefly¹²), socialna omrežja (Lately¹³, Brandwatch¹⁴) in pisanje besedil (Grammarly¹⁵, WriteSonic¹⁶). Zelo hiter razvoj posameznih modelov prikazuje Slika 1, kjer lahko opazimo, da se obstoječi modeli več časa posodablajo, prav tako pa se pojavljajo novi modeli.

Z velikimi jezikovnimi modeli se srečujemo tudi v:

- **zdravstvu:** takšni modeli lahko razumejo beljakovine, molekule in DNK. Ta položaj omogoča velikim jezikovnim modelom, da pomagajo pri razvoju cepiv, iskanju zdravil za bolezni in izboljšanju zdravil za preventivno nego,
- **pravu:** lahko pomagajo odvetnikom in pravnemu osebju,
- **bančništvu:** jezikovni modeli lahko podpirajo podjetja, ki poslujejo s kreditnimi karticami pri odkrivanju goljufij in
- **šolstvu:** takšni modeli lahko pomagajo pri reševanju matematičnih vprašanj in odgovarjanju na vprašanja.

Zgodovina jezikovnih modelov se je začela v petdesetih letih prejšnjega stoletja, ko je bilo prvič ustanovljeno področje umetne inteligence. Eden najzgodnejših primerov jezikovnega modela je bil ELIZA, program, ki ga je razvil Joseph Weizenbaum leta 1966. [7] V naslednjih nekaj desetletjih se je področje umetne inteligence še naprej razvijalo in napredovalo, pri čemer so raziskovalci razvijali jezikovne modele, ki bi lahko obravnavali bolj zapletene naloge. V devetdesetih letih prejšnjega stoletja je pojav globokega učenja utrl pot še naprednejšim jezikovnim modelom. Ena ključnih novosti pri razvoju velikih jezikovnih modelov je bila uporaba nenadzorovanega učenja, ki omogoča modelom učenje iz velikih količin nestrukturiranih besedilnih podatkov. To je raziskovalcem omogočilo učenje modelov na

¹ <https://chat.openai.com/>

² <https://gemini.google.com/>

³ <https://openai.com/dall-e>

⁴ <https://app.leonardo.ai/>

⁵ <https://pictory.ai/>

⁶ <https://runwayml.com/>

⁷ <https://speechify.com/voice-cloning/>

⁸ <https://murf.ai/>

⁹ <https://claude.ai/>

¹⁰ <https://www.adcreative.ai/>

¹¹ <https://www.canva.com/magic-design/>

¹² <https://www.adobe.com/products/firefly.html>

¹³ <https://www.lately.ai/>

¹⁴ <https://www.brandwatch.com/p/latest-ai-features/>

¹⁵ <https://www.grammarly.com/>

¹⁶ <https://writesonic.com/>

ogromnih količinah podatkov, kar je privedlo do razvoja velikih jezikovnih modelov, ki bi lahko ustvarili besedilo, podobno tistemu, ki ga pripravi človek. [8]

Veliki jezikovni modeli uporabljajo transformatorske modele in se usposabljaajo z uporabo velikih naborov podatkov. To jim omogoča prepoznavanje, prevajanje, predvidevanje ali ustvarjanje besedila ali druge vsebine. Transformatorski model je nevronska mreža, ki se uči konteksta in pomena s sledenjem razmerij v zaporednih podatkih. Transformator je sestavljen iz več transformatorskih blokov oz. plasti. Transformator ima npr. plasti za posredovanje in normalizacijske plasti, ki skupaj delujejo pri dešifriranju vhoda za napovedovanje tokov izhoda pri sklepanju. Transformatorje je Google prvič predstavil leta 2017. [9]

Pozitivne lastnosti velikih jezikovnih modelov so, da se nenehno izboljšujejo, več, kot se učijo, boljši so. Imajo pa tudi nekaj negativnih lastnosti. Pogosto dajo vtis, da razumejo pomen in se lahko nanj natančno odzovejo, lahko razkrijejo zasebne podatke ljudi, sodelujejo pri lažnem predstavljanju in ustvarijo neželjeno pošto, če niso pravilno upravljeni. Uporabniki z zlonamernimi nameni lahko reprogramirajo umetno inteligenco in prispevajo k širjenju lažnih informacij. Iz opisov fotografij lahko izluščijo osebne podatke, kar lahko ogrozi zasebnost.

Z vedno večjo priljubljenostjo velikih jezikovnih modelov se je sprožila tudi razprava o njihovi prihodnosti. Rast in izboljšanje znanja naravnega jezika velikih jezikovnih modelov povzročata zaskrbljenost glede posledic napredka na trgu dela. Veliki jezikovni modeli lahko povečajo produktivnost in učinkovitost, vendar se pojavljajo tudi etična vprašanja glede njihove uporabe v človeški družbi.

2.2 ChatGPT

ChatGPT je računalniški program, ki s pomočjo umetne inteligence ustvarja človeku podobne pogovore, odgovarja na vprašanja in pomaga pri različnih opravilih. Je kot pameten prijatelj, ki razume in uporablja jezik skoraj kot človek. [10]

Gre torej za napreden jezikovni model umetne inteligence. Uporablja sistem obdelave naravnega jezika (*angl. Natural Language Processing (NLP)*), ki razvijalcem omogoča ustvarjanje klepetalnih robotov, ki razumejo uporabnikove vhodne podatke in se nanje odzivajo naravno in pogovorno.

Novembra 2022 je bil ChatGPT izdan kot brezplačen raziskovalni prototip. Poganja ga model strojnega učenja, imenovan GPT-3, ki ga je razvilo podjetje OpenAI in velja za enega najnaprednejših modelov NLP doslej. OpenAI velja za eno najpomembnejših podjetij s področja umetne inteligence, ki sta ga ustanovila Elon Musk in Sam Altman, med prvimi vlagatelji pa je tudi Microsoft. [11] ChatGPT je bil naučen na ogromni količini internetnih besedil in posledično razglašen za alternativo Googlu, saj lahko zagotavlja odgovore na zapletena vprašanja, ustvarja kodo, v prihodnosti pa bo sposoben opraviti še veliko več.

Google je sicer na to odgovoril z lastnim klepetalnim robotom Google Gemini (takrat Bard), ki pa ni tako uspešen.

2.2.1 Značilnosti

ChatGPT se odlikuje po tem, da na podlagi svoje zmožnosti napovedovanja ustvarja skoraj človeško besedilo. Zanj velja naslednje [12]:

- **razumevanje konteksta:** ChatGPT je sposoben razumeti kontekst pogovora, kar mu omogoča ustvarjanje ustreznih odgovorov na pozive, ohranjanje poteka pogovora in sklicevanje na predhodno omenjene informacije.
- **jezikovno modeliranje:** usposobljen je za napovedovanje naslednje besede v stavku. Zasnovan je tako, da razume različne sestavine jezika, vključno s slovnico, skladnjo, reki in celo kulturnimi referencami.
- **generativna sposobnost:** za razliko od sistemov umetne inteligence, ki temeljijo na pravilih, je ChatGPT generativen, kar pomeni, da lahko ustvari odgovore na pozive in ne izbere vnaprej pripravljenega odgovora. Tako lahko sodeluje v odprtih pogovorih in obravnava najrazličnejše teme.
- **sposobnost prenosa učenja:** pri tem se znanje, pridobljeno v enem kontekstu, uporabi v novem kontekstu. Če je bil na primer naučen iz različnih podatkov iz knjig, spletnih mest in drugih besedil, lahko ChatGPT to široko znanje uporabi za odgovarjanje na široko paleto vprašanj ali pozivov.
- **razumevanje več jezikov:** ChatGPT je dokazal, da je sposoben razumeti in ustvariti besedilo v več jezikih, zato je uporabno orodje za prevajanje ali večjezično komunikacijo.
- **samopopravljanje:** ChatGPT se lahko v več ponovitvah sam popravi in izboljša svoje odzive. Med procesom usposabljanja uporablja povratne zanke za nenehno izboljševanje svojih zmožnosti ustvarjanja in razumevanja jezika.
- **razširljivost:** arhitektura ChatGPT ima visoko zmožnost prilagajanja, kar pomeni, da lahko opravlja vedno bolj zapletene in raznolike naloge, če je na voljo več računske moči ali podatkov.
- **prilagodljivo učenje:** ChatGPT lahko na podlagi pozivov in interakcij prilagodi svoje odzive, da se uskladijo z zahtevanim tonom, bodisi formalnim, neformalnim, humornim ali resnim.

Čeprav je ChatGPT zmogljiv, ima tudi omejitve. Ne vključuje osebnih izkušenj ali čustev. Ne dostopa do osebnih podatkov iz pogovorov in jih ne shranjuje, razen če so izrecno navedeni za trenutni pogovor, in včasih lahko zapiše napačne ali nesmiselne informacije. OpenAI je uvedel tudi varnostne ukrepe, ki zagotavljajo, da zavrača škodljive ali neprimerne napotke.

2.2.2 Navodila za pisanje dobrih besedilnih pozivov

Umetna inteligenca ni oseba, ki bi razmišljala sama od sebe, ampak uporablja veliko informacij, ki so jih drugi ljudje že kdaj objavili in shranili. Ko iščemo odgovor, je pomembno, da povemo, kaj nas zanima. To velja tudi v primeru, ko se pogovarjamo z računalniškimi programi, kjer je treba uporabiti dobra vprašanja in navodila. Namesto da se jezimo na npr. ChatGPT, ker ni ponudil pravega odgovora, mu raje bolje povejmo, kaj sploh želimo [13].

ChatGPT deluje po načinu štirih korakov ne glede na tip vprašanja ali naloge, ki mu jo zastavimo:

- **poziv:** uporabnik posreduje poziv ali vnos. To je lahko vprašanje, stavek, ki ga je treba dopolniti ali na primer tema, o kateri je treba pisati, itd.
- **analiza konteksta:** model analizira poziv skupaj z vso ustrezno zgodovino dialoga in tako ustvari tako imenovani kontekst. Pomembno je poudariti, da nima spomina na pretekle interakcije zunaj trenutnega pogovora. Pri reševanju NPZ-ja je kontekst I. ali II. del NPZ-ja z umetnostnim oz. neumetnostnim besedilom.
- **ustvarjanje odzivov:** ChatGPT na podlagi konteksta ustvari odziv. To počne besedo za besedo, kjer vsako naslednjo besedo predvidi na podlagi besed, ki jih je že ustvaril za odgovor.
- **izhod:** to postane del konteksta za naslednji krog vnosa podatkov s strani uporabnika.

Ekipa iz podjetja OpenAI je nedavno objavila nekaj priporočil [14] o tem, kako bolje uporabljati njihova orodja. Med nasveti so izpostavili, da računalniški programi ne morejo brati vaših misli. Če so odgovori predolgi, prosite za krajše odgovore. Če so odgovori preveč preprosti, prosite za bolj zapletene. Če vam ni všeč oblika, povejte, kako bi radi, da izgleda. Bolj jasno kot poveste, kaj hočete, večja je možnost, da boste dobili tisto, kar želite.

Podrobneje so nasveti naslednji:

1. Napišite jasna navodila s podrobnostmi

Če želite, da vam model vrne pravilen odgovor, morajo vaša vprašanja vsebovati vse pomembne podrobnosti. Ne smete dovoliti, da bi model zamenjal nejasnosti z lastnim ugibanjem.

- Slab primer: "Kdo je predsednik?"
- Dober primer: "Kdo je bil leta 2021 predsednik Slovenije in kako pogosto so volitve?"

2. Po korakih povejte, kaj naj model naredi

- Primer: Sledi naslednjim korakom, da odgovoriš na uporabnikovo vprašanje.
- 1. korak: Uporabnik bo napisal besedilo v narekovajih. Zapiši obnovo besedila v enem stavku z oznako "Povzetek".
- 2. korak: Prevedi obnovo iz 1. koraka v nemščino in ji dodaj oznako "Prevod".

3. Dodajte referenčno besedilo

Včasih se zgodi, da si model začne nekaj izmišljevati, če nima dovolj informacij, kar imenujemo haluciniranje. Težavno je, če si izmišljuje vire podatkov. Modelu lahko zato naročimo, naj uporabi citate iz knjige ali povezavo do spletne strani, ko sestavlja odgovor.

- Primer: Podal ti bom dokument, ki bo označen z dvojnimi narekovaji, in vprašanje. Tvoja naloga je odgovoriti na vprašanje le z uporabo tega dokumenta in navesti del dokumenta, ki si ga uporabil pri odgovoru. Če v dokumentu ne najdeš potrebnih podatkov, enostavno zapiši "Ni dovolj podatkov". Vedno ko vrneš odgovor, ga moraš potrditi z navedbo.

4. Razdelite težjo nalogo na več lažjih

Ko imate opravka z nalogami, kjer morate slediti različnim primerom in uporabiti več navodil, je koristno, da najprej te naloge razdelite na manjše dele. Glede na to, kaj je treba narediti, določite, katera dodatna navodila so potrebna. Kot pravijo razvijalci OpenAI, to zmanjša možnost napak. [3] Pri komunikaciji s ChatGPT upoštevajte, da je količina besedila, ki ga lahko vnesete, omejena, zato daljše dokumente razdelite na manjše dele.

5. Dajte modelu dovolj časa za razmišljanje

Včasih lahko dobimo boljši odgovor, če modelu damo priložnost, da razmisli o vprašanju, preden se prehitro odloči.

- Primer: Recimo, da imamo enostavno matematično nalogo: Miha je nabral kilogram kivija in ga začel prodajati, Mojca pa kilogram hrušk ...
- Vprašanje: Ali je Mojca zaslužila več?
- ChatGPT se lahko hitro odloči za odgovor, ki ni pravilen. Bolje je, da mu naročimo naslednje: "Najprej izračunaj, koliko je vsak zaslužil. Nato primerjaj rezultate zaslužkov in ugotovi, kdo je zaslužil več. Ne odloči se o pravilnosti odgovora, dokler ga ne analiziraš."

2.2.3 Različice ChatGPT-ja

ChatGPT je del družine modelov GPT, ki so jih razvili v OpenAI. Prvi model, imenovan GPT-1, je bil predstavljen leta 2018 in je bil takrat eden največjih modelov umetne inteligence, z več kot 100 milijoni parametrov. Naslednji model, GPT-2, je bil izdan leta 2019 in je bil še večji, z 1,5 milijardami parametrov. GPT-2 je bil znan po svoji sposobnosti generiranja prepričljivega in naravnega besedila ter je sprožil veliko pozornost v medijih in javnosti zaradi morebitne

zlorabe umetne inteligence. Zato so se odločili, da ne bodo izdali največje različice modela, GPT-3, za splošno uporabo. Kljub temu so razvijalci iz OpenAI zagotovili dostop do modela za raziskovalce in podjetja, ki se ukvarjajo z umetno inteligenco, in tako ustvarili možnost za nadaljnji napredek v razvoju naravne besedilne interakcije. ChatGPT je model, ki je bil naučen na podlagi GPT-3 arhitekture in ima manj parametrov kot GPT-3, vendar je še vedno sposoben ustvariti naravno besedilno interakcijo s človekom preko besedilnih sporočil. *Verzija GPT-3.5* je bila naučena s podatki do septembra 2021. To pomeni, da vključuje novejša trende in informacije. [12]

GPT-3.5 Turbo je najnovejša različica, ki predstavlja pomemben preskok v obdelavi in analizi naravnega jezika, ki gradi na temeljih, ki jih je postavil njen predhodnik GPT-3.5 [15].

Različica modela GPT (GPT-4) je najnovejša verzija v jezikovnih modelih umetne inteligence in je presegla vse prejšnje jezikovne modele. Obdela lahko več podatkov, je natančnejša, težje jo je pretentati, sprejema pozive v obliki slik in besedil, poleg tega ima bolj dodelan jezikovni slog in bolj globinsko obvlada več jezikov (glej primerjavo v Tabela 1).

Preizkus	ChatGPT-3.5	ChatGPT-4
Velikost	1,37 milijarde parametrov	175 milijard parametrov
Podatki za usposabljanje	500 GB	1,5 TB
Zmožljivosti	Ustvarjajte besedila, prevajajte jezike, pišite različne vrste ustvarjalnih vsebin in informativno odgovarjanje na vaša vprašanja	Najbolj napreden model, odličen za naloge, ki zahtevajo ustvarjalnost in napredno razmišljanje.
Razpoložljivost	Brezplačno	Chat GPT Plus (20 dolarjev/mesec)

Tabela 1: Primerjava različnih verzij ChatGPT [10]

2.3 Nacionalno preverjanje znanja (NPZ)

NPZ je način preverjanja uspešnosti doseganja standardov znanja po določenem obdobju, pri katerem vsi učenci v državi na isti dan rešujejo enake preizkuse znanja pod enakimi pogoji. Naloge pripravijo učitelji skupaj z drugimi strokovnjaki za posamezne predmete in strokovnjaki za sestavljanje preizkusov znanja. Preverja se le tisto znanje, ki ga vsebujejo učni načrti. To vključuje branje praviloma neznanega neumetnostnega besedila, preverjanje razumevanja prebranega neumetnostnega besedila, preverjanje jezikovnega znanja in zmožnosti rabe jezika, preverjanje zmožnosti tvorbe krajšega pisnega neumetnostnega besedila, branje praviloma neznanega umetnostnega besedila, preverjanje razumevanja in interpretacije prebranega umetnostnega besedila, preverjanje književnega znanja in preverjanje zmožnosti tvorbe krajšega pisnega besedila po branju umetnostnega besedila. Naloge se vrednotijo po enotnih, posebej pripravljenih merilih. V obstoječem šolskem sistemu Slovenije je NPZ pomemben element pri ugotavljanju in zagotavljanju kakovosti šolskega dela. Po sedanji zakonodaji se NPZ opravlja ob koncu 6. in 9. razreda osnovne šole, so pa na tem področju za

prihodnja leta že pripravljene spremembe. V šolskem letu 2017/18 se je začelo poskusno nacionalno preverjanje znanja tudi ob koncu 3. razreda osnovne šole. [16]

Ob koncu 6. razreda osnovne šole preverjanje znanja obsega naslednje obvezne predmete:

- slovenščino ali italijanščino oziroma madžarščino na narodno mešanih območjih,
- matematiko,
- prvi tuji jezik.

Ob koncu 9. razreda osnovne šole preverjanje znanja obsega naslednje obvezne predmete:

- slovenščino ali italijanščino oziroma madžarščino na narodno mešanih območjih,
- matematiko,
- tretji predmet, ki ga vsako leto določi minister, pristojen za šolstvo, in se po šolah razlikuje.

Temeljni cilj nacionalnega preverjanja znanja je pridobiti dodatno informacijo o znanju učencev, ki je namenjena učencem, njihovim staršem, učiteljem, šolam in sistemu na nacionalni ravni.

Pri slovenskem jeziku je NPZ sestavljen iz dveh delov, ki sta ovrednotena v razmerju 40 % in 60 %. I. del vsebuje vprašanja o umetnostnem besedilu, v II. delu pa vprašanja o neumetnostnem besedilu (glej Tabela 2). Za reševanje obeh delov imajo učenci na razpolago 60 minut.

Preizkus	Opis	Delež v skupnem številu točk	Predvideni čas
I. del	Eno ali več neznanih umetnostnih besedil ter naloge za preverjanje doseganja ciljev in standardov znanja iz UN.	40 %	60 minut (za oba dela skupaj).
II. del	Eno ali več neznanih neumetnostnih besedil ter naloge za preverjanje doseganja ciljev in standardov znanja iz UN.	60 %	

Tabela 2: Sestavni deli NPZ-ja iz slovenščine

2.4 Sorodne raziskave

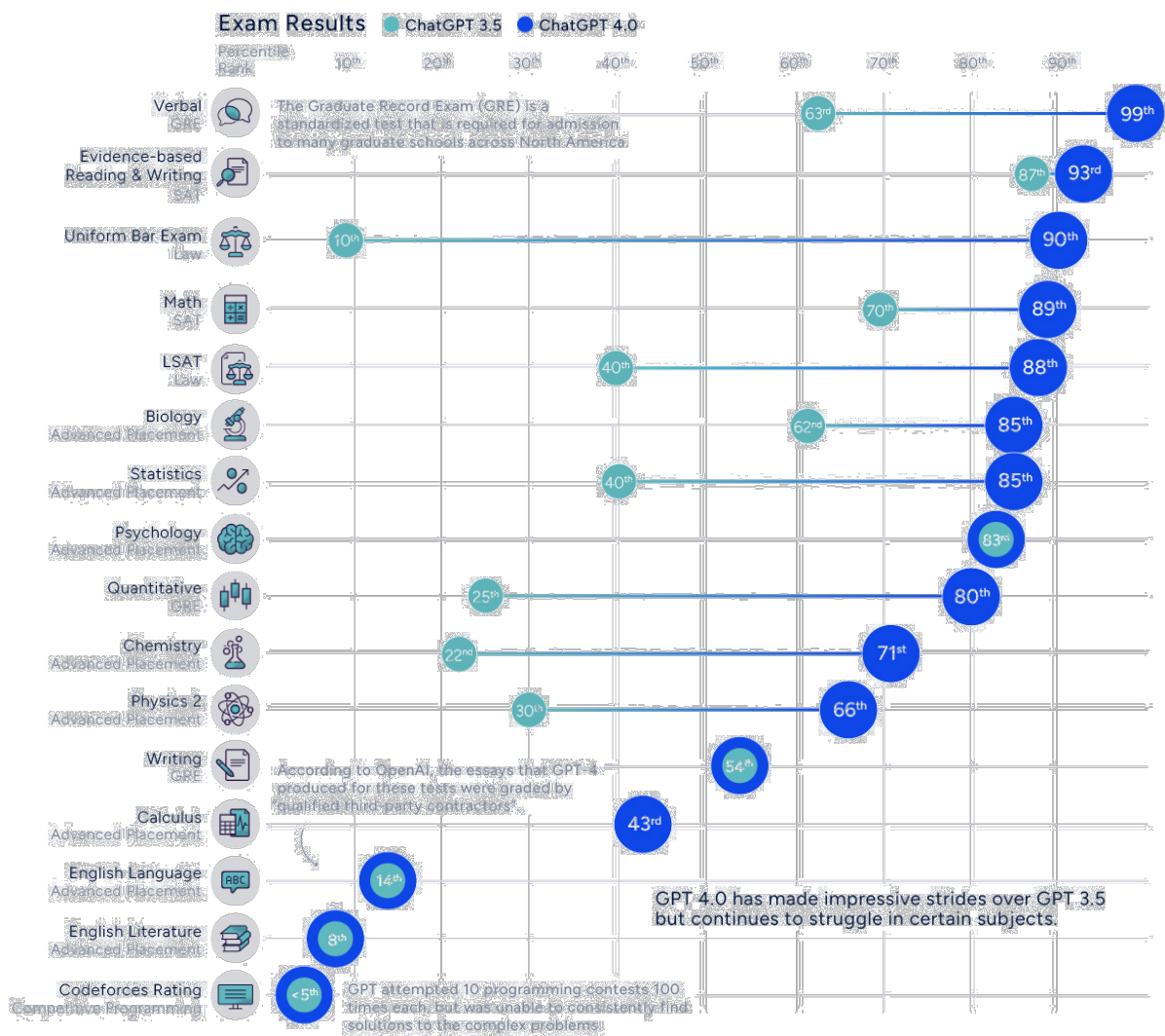
Kolikor nama je znano, raziskave o uporabi ChatGPT-ja pri reševanju nacionalnega preverjanja znanja iz slovenščine v Sloveniji še ni bilo. Obstajajo pa sorodne raziskave, ki sva jih zasledili v literaturi.

Leta 2022 so na Nizozemskem izvedli raziskavo [17], kjer je bil uporabljen model ChatGPT 3.5 za reševanje državnih izpitov za srednje šole na temo angleškega bralnega razumevanja. ChatGPT 3.5 je dosegel povprečno oceno 7,30 na nizozemski lestvici od 1 do 10, medtem ko

je bila povprečna ocena dijakov 6,99. Vendar je bil občasno potreben ponovni poziv ChatGPT-ju, da so prišli do jasnega odgovora. Brez ponovnih pozivov je bila skupna ocena ChatGPT-jevih odgovorov 6,50. Po izidu nove različice ChatGPT-ja, GPT 4, so mu predložili iste izpite, kjer je dosegel oceno 8,30, brez kakršnekoli potrebe po ponovitvi pozivov.

V Turčiji so raziskovali akademske dosežke ChatGPT-ja na državnem izpitu iz matematike [18]. V raziskavi sta bili uporabljeni različici GPT 3.5 in GPT 4. Kot rezultat analize podatkov je bilo ugotovljeno, da je ChatGPT 4 bolj uspešen pri izpitu v primerjavi z različico ChatGPT 3.5. Bolje je razumel zastavljena vprašanja, kot tudi navodila in vključil več podrobnosti v odgovor vprašanja, hkrati pa sta obe različici naredili pogoste in raznolike napake. V študiji so ugotovili, da je ChatGPT včasih deloval zelo dobro, včasih le dobro in včasih neuspešno. Raziskovalci predlagajo uporabo različic ChatGPT pri poučevanju matematike predvsem za pridobitev osnovnih informacij.

Pri podjetju OpenAI, ki skrbi za razvoj ChatGPT, so izdali tehnično poročilo [19], kjer so predstavili napredek njihovega modela skozi čas in se osredotočili predvsem na modela ChatGPT 3.5 in ChatGPT 4. Ugotovili so, da je postal zelo učinkovit in da ga tudi učenci uporabljajo pri domačih nalogah. Posledično so številne šole v ZDA blokirale dostop do modela ChatGPT iz njihovega omrežja, tako da ga učenci ne morejo uporabljati v šoli. Slika 2 prikazuje rezultat modelov ChatGPT 3.5 in 4 na različnih standardiziranih testih v ZDA. Opazimo lahko, da se ChatGPT zelo izkaže pri sprejemnih testih različnih področij, od prava, matematike, biologije, statistike, psihologije itd.



Slika 2: Uspešnost modela ChatGPT pri različnih problemih [20]

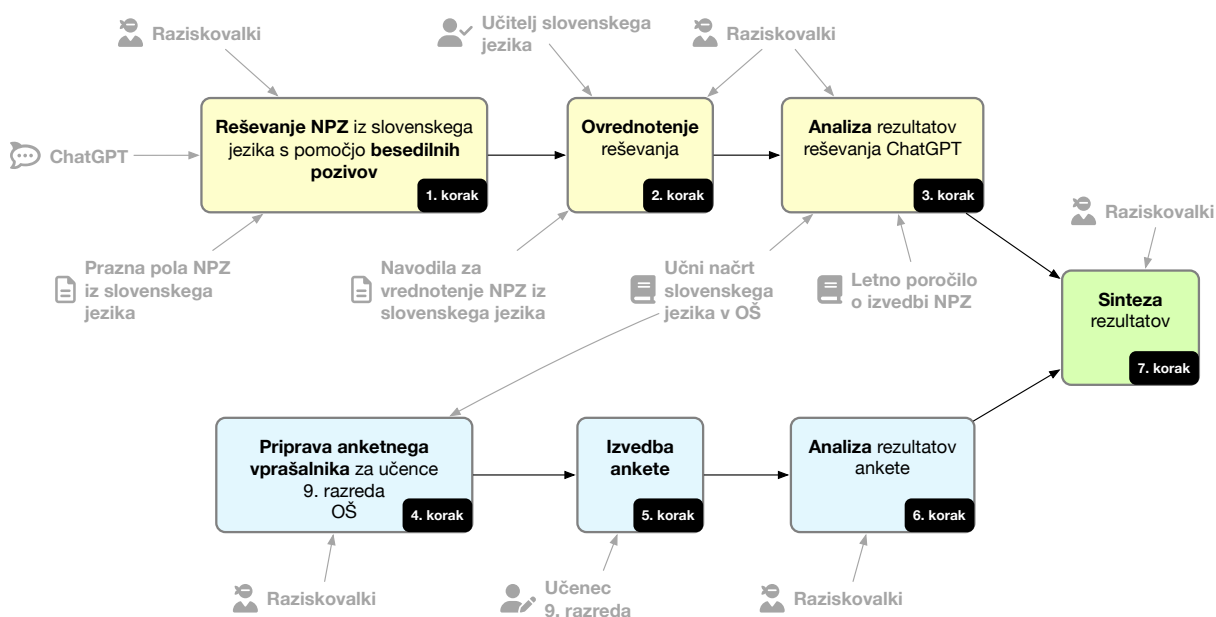
Na podlagi pregleda literature sva ugotovili, da so velike jezikovne modele in predvsem ChatGPT že uspešno uporabili pri različnih standardiziranih testih, ki pa so predvsem v angleškem jeziku.

V slovenskem jeziku takšnih velikih jezikovnih modelov nisva zasledili.

3 Praktični del

3.1 Uvod

V okviru raziskovalne naloge sva praktični del razdelili na dva dela, in sicer na anketni vprašalnik in na reševanje NPZ-ja za 9. razred s ChatGPT-jem. Slika 3 prikazuje vse korake, ki sva jih izvedli. V nadaljevanju rezultate analize podrobno opiševa in jih grafično predstaviva z grafi in s tabelami.



Slika 3: Zasnova raziskave

Aktivnosti, povezane z anketnim vprašalnikom, so na Slika 3 prikazane z modro barvo, in sicer sva najprej pripravili vprašalnik in ga posredovali učencem 9. razreda osnovne šole. Sledila je analiza rezultatov, ki je predstavljena v poglavju 3.2.1.

Poleg izvedbe ankete sva model ChatGPT uporabili za reševanje NPZ-jev iz slovenskega jezika, kar je na Slika 3 prikazano z rumeno barvo. Za ChatGPT sva najprej pripravili ustrezne pozive z vprašanji iz NPZ-ja ter mu podali besedilo, na katerega se vprašanje nanaša. Sledilo je ovrednotenje reševanja, za kar sva uporabili enotna merila (rešitve) za vrednotenje NPZ. Nato sva rezultate analizirali, kar predstavlja v poglavju 3.2.2.

3.2 Predstavitev in utemeljitev rezultatov

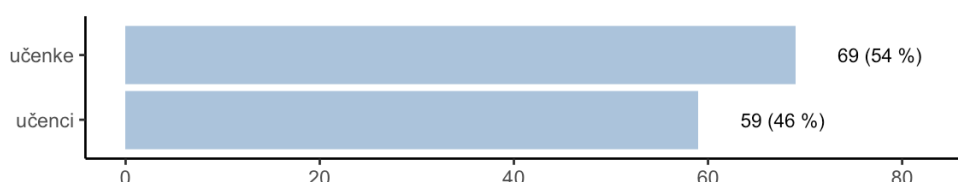
3.2.1 Anketa

Spletni anketni vprašalnik, ki je bil na voljo na [4], so reševali devetošolci celjskih osnovnih šol. Razdeljen je bil razdeljen na štiri vsebinske sklope: (1) podatki anketiranja, (2) predznanje, (3) model umetne inteligence ChatGPT in NPZ ter (4) vpliv umetne inteligence. Skupno je anketo izpolnilo 128 udeležencev.

3.2.1.1 Podatki anketirancev

V prvem sklopu so anketiranci navedli osnovne podatke; spol, starost in osnovno šolo, ki jo obiskujejo. Anketiranci so bili stari med 14 in 16 let ter so obiskovali 9. razred ene izmed štirih celjskih osnovnih šol, in sicer II. OŠ Celje, III. OŠ Celje, OŠ Lava in OŠ Frana Roša.

Slika 4 prikazuje razporejenost anketirancev glede na spol, kjer je v raziskavi sodelovalo nekoliko več učenk (69 oz. 54 %) kot učencev (59 oz. 46 %).



Slika 4: Razporejenost anketirancev glede na spol

3.2.1.2 Predznanje

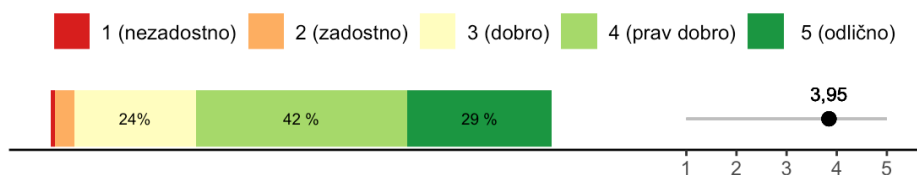
V tem sklopu so anketiranci odgovarjali na tri vprašanja zaprtega tipa in enega delno zaprtega, delno odprtega, in sicer:

1. Ali je slovenski jezik vaš **materni jezik**?
2. Kako bi **ocenili** svoje **znanje slovenskega jezika**?
3. Kako dobro poznate **umetno inteligenco**?
4. Označite, katere **primere modelov umetne inteligence** poznate.

Pri prvem vprašanju so imeli anketiranci dve možni izbiri (da/ne), pri drugem in tretjem vprašanju pet možnih izbir (ocenjevalna lestvica v razponu 1-5, pri čemer je 1 pomenilo nezadostno, 5 pa odlično), pri četrtem pa sedem možnih izbir in možnost lastnega vnosa.

Med sodelujočimi je bilo **80 %** učenk in učencev, za katere je slovenščina njihov **materni jezik**.

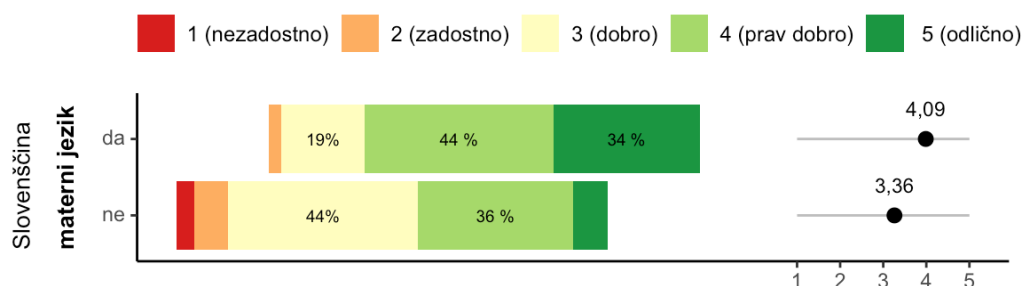
Slika 5 prikazuje, da so sodelujoči **lastno znanje slovenskega jezika** pri drugem vprašanju ocenili z visoko **povprečno oceno 3,95**.



Slika 5: Samoocena znanja slovenskega jezika

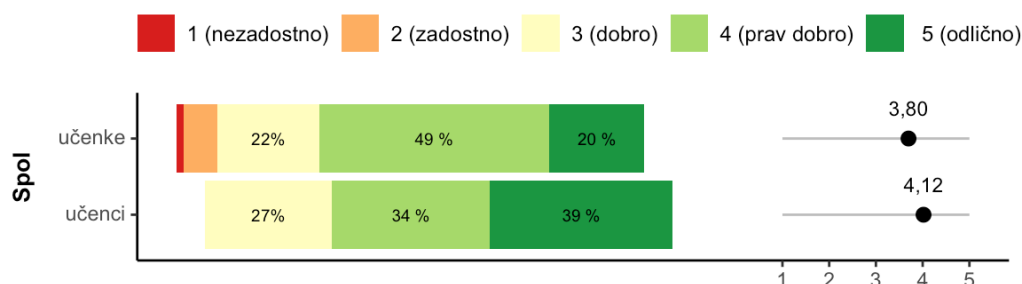
Če odgovore na to drugo vprašanje prikažemo v odvisnosti od prvega vprašanja, lahko opazimo povezavo med ocenjenim znanjem slovenskega jezika in slovenščino kot maternim jezikom.

Slika 6 prikazuje, da je samoocena slovenskega jezika višja pri udeležencih, kjer je slovenščina zanje materni jezik, in sicer **4,09**, v primerjavi s **3,36** pri udeležencih, kjer slovenščina zanje ni materni jezik. Razlika v ocenjenem znanju slovenskega jezika med udeleženci s slovenščino kot maternim jezikom in ostalimi je tako znašala malo več kot pol ocene oz. **0,73**.



Slika 6: Samoocena znanja slovenskega jezika glede na to, ali je slovenščina materni jezik

Opazimo lahko razlike v ocenjenem znanju slovenskega jezika med učenci in učenkami. Slika 7 prikazuje višjo samooceno znanja slovenskega jezika pri učencih, in sicer **4,12**, v primerjavi s **3,80** pri učenkah. Razlika v ocenjenem znanju slovenskega jezika med učenci in učenkami tako znaša **0,32**.

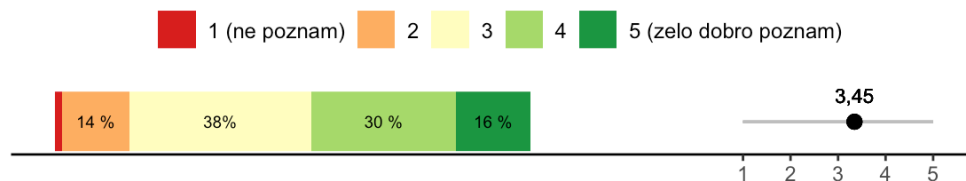


Slika 7: Samoocena znanja slovenskega jezika glede na spol

Rezultati NPZ-ja iz slovenskega jezika ne potrjujejo [21] pridobljenih anketnih podatkov. Velja namreč, da je povprečna ocena učenk na NPZ-ju višja od povprečne ocene učencev.

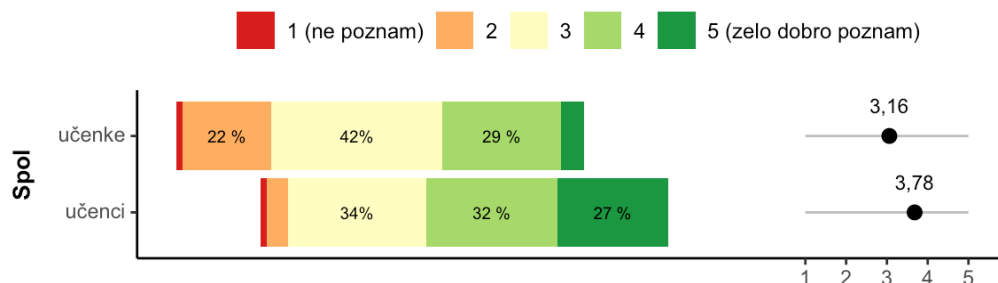
Pri tretjem vprašanju so imeli učenci ob vprašanju kot pomoč pri reševanju opredeljen pojem umetne inteligence, in sicer *“Umetna inteligenca je zmožnost stroja, da izkazuje človeške lastnosti, kot so mišljenje, učenje, načrtovanje in kreativnost.”* [1] Dodatno so bili podani tudi primeri umetne inteligence, in sicer: *avtomatsko prevajanje, pametno ogrevanje stanovanja, samovozeči avtomobil, digitalni asistenti na telefonu.*

Slika 8 prikazuje, da sodelujoči pojem umetne inteligence poznajo, in sicer je povprečna ocena znašala **3,45**.



Slika 8: Samoocena poznavanja umetne inteligence

Ugotovimo lahko, da obstajajo razlike v poznavanju umetne inteligence med učenkami in učenci. Slika 9 izpostavlja, da so učenci bolj samozavestni pri poznavanju umetne inteligence, saj je njihova povprečna ocena **3,78** v primerjavi s **3,16** pri učenkah. Razlika v stopnji poznavanja umetne inteligence med učenci in učenkami je znašala **0,62**.



Slika 9: Samoocena poznavanja umetne inteligence glede na spol

S tem sva **potrdili prvi del hipoteze 1**, da devetošolci poznajo umetno inteligenco in pri tem še dodatno izpostavili razlike med poznavanjem, kjer učenci umetno inteligenco poznajo bolje kot učenke.

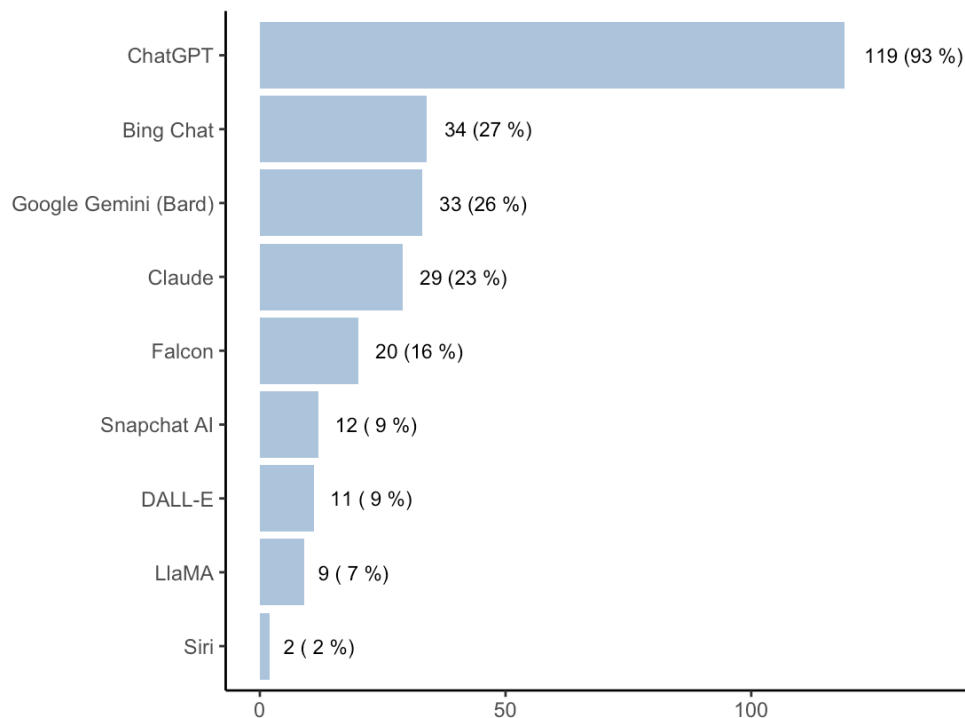
Hipoteza 1: Devetošolci poznajo umetno inteligenco in najbolj razširjen model ChatGPT.

Pri četrtem vprašanju so anketiranci izmed vnaprej določenih modelov umetne inteligence izbrali tiste, za katere so že slišali ali jih že uporabili. Prav tako so imeli možnost dodati nov model. Pri analizi rezultatov sva upoštevali vse odgovore, ki sta jih izbrala vsaj 2 sodelujoča.

Slika 10 z rezultati najbolj poznanih modelov umetne inteligence med učenci prikazuje, da velika večina sodelujočih (**119 oz. 93 %**) pozna model **ChatGPT**.

S tem sva **potrdili drugi del hipoteze 1**, da učenci poznajo model ChatGPT.

Hipoteza 1: Devetošolci poznajo umetno inteligenco in najbolj razširjen model ChatGPT.



Slika 10: Poznavanje primerov modelov umetne inteligence

Glede preostalih modelov umetne inteligence je približno 30 udeležencev odgovorilo, da poznajo **Bing Chat (27 %)**, **Google Gemini (Bard) (26 %)** in **Claude (23 %)**. Sledijo še ostali modeli, in sicer **Falcon**, **Snapchat AI**, **DALL-E**, **LLaMA** in **Siri**. Modela **Snapchat AI** in **Siri** so sodelujoči izpostavili sami in nista bila del prvotnega nabora, med katerimi so lahko izbirali.

3.2.1.3 Model umetne inteligence ChatGPT in NPZ

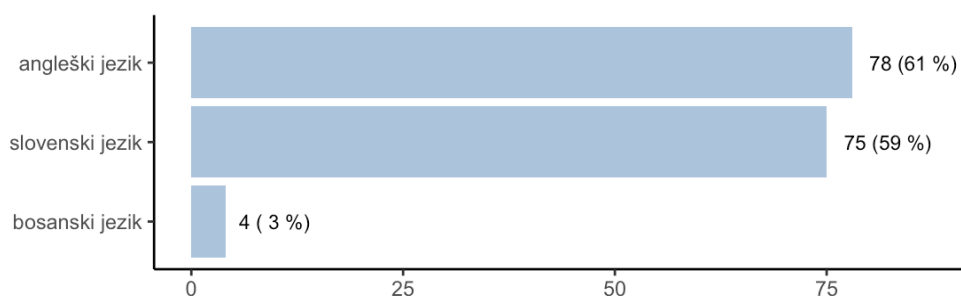
V tretjem sklopu vprašalnika so anketiranci odgovarjali na vprašanja o modelu ChatGPT pri uporabi za reševanje NPZ-ja iz slovenskega jezika. Oba pojma sta bila anketirancem na kratko predstavljena, in sicer kot: "*ChatGPT je računalniški program, ki s pomočjo umetne inteligence ustvarja človeku podobne odgovore, odgovarja na vprašanja in pomaga pri različnih opravilih. Je pametni digitalni asistent, ki razume in uporablja jezik skoraj tako kot človek*" [10] in "*Nacionalno preverjanje znanja (NPZ) se izvaja ob koncu 6. in 9. razreda osnovne šole in med drugim obsega tudi obvezni predmet slovenskega jezika.*" [22]

Nato so anketiranci odgovarjali na devet vprašanj, in sicer:

1. V **katerem jeziku** bi **postavili vprašanje** modelu ChatGPT, če bi ga potrebovali pri reševanju kakšnega problema?
2. Kakšen **skupni rezultat** bi po vašem mnenju dosegel **ChatGPT** pri **reševanju** nacionalnega preverjanja znanja (**NPZ**) iz **slovenskega jezika**?
3. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **bralnega razumevanja umetnostnih in neumetnostnih besedil** (*npr. odgovori na vprašanja, povezana s priloženim besedilom*)?
4. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **književnosti** (*npr. analiza umetnostnega besedila, določitev teme besedila, pesniških sredstev (npr. primera, okrasni pridevek, rima, poosebitev)*)?
5. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **reševanja jezikovnih nalog** (*npr. določanje glavnega in odvisnega stavka v povedi, stopnjevanje pridevnikov, spreganje glagolov, poznavanje frazemov itd.*)?
6. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **oblikovanja zaključenega besedila** (*npr. pisanje vabila, kratkega življenjepisa ipd.*)?
7. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **slovnice** (*npr. pravilno sklanjanje besed, raba predlogov, stopnjevanje pridevnikov*)?
8. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **pravopisa** (*npr. pravilen zapis besed, velike začetnice, raba ločil (npr. vejica, pika ipd.)*)?
9. Kako uspešen bi po vašem mnenju bil ChatGPT z vidika **slogovne pravilnosti** (*npr. jasno in natančno izražanje, uporaba knjižnega jezika in pretežno nevtralnega besedišča*)?

Pri prvem vprašanju so imeli anketiranci navedeni dve možnosti, lahko pa so dodali še svojo. Pri ostalih vprašanjih so imeli pet možnih izbir (ocenjevalna lestvica 1-5, pri čemer je 1 pomenilo zelo slabo, 5 pa zelo dobro).

Slika 11 prikazuje jezika, ki sta anketirancem najbližja pri pozivih modelu ChatGPT, in sicer **angleški jezik (61 %)** in **slovenski jezik (59 %)**. Pri analizi rezultatov smo upoštevali vse odgovore, ki sta jih sta izbrala vsaj 2 sodelujoča.

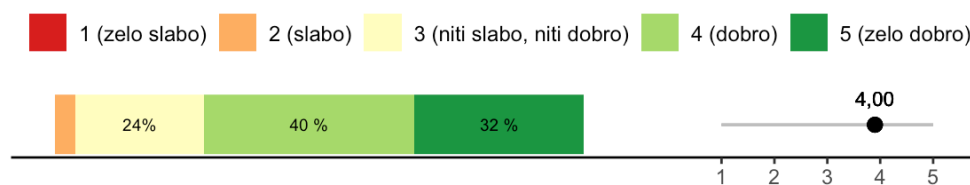


Slika 11: Jezik za postavljanje vprašanja modelu ChatGPT pri reševanju problema

Slika 12 prikazuje mnenje anketirancev o pričakovani učinkovitosti ChatGPT-ja pri reševanju NPZ-ja. Opazimo lahko, da nihče izmed sodelujočih ni na vprašanje odgovoril z 1 (zelo slabo) in da je povprečno **zaupanje v ChatGPT pri reševanju NPZ-ja visoko**, in sicer s povprečno oceno **4,00**.

Tako lahko potrdiva **hipotezo 2**, saj anketiranci z visoko povprečnim odgovorom 4,00 od 5 menijo, da bo ChatGPT uspešno reševal naloge pri NPZ-ju.

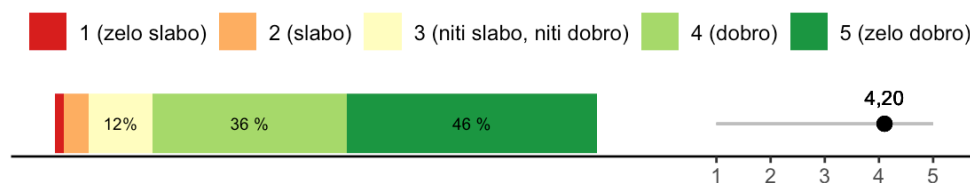
Hipoteza 2: Devetošolci menijo, da bo ChatGPT v splošnem uspešno reševal naloge pri NPZ-ju iz slovenščine.



Slika 12: Pričakovan skupen rezultat ChatGPT-ja pri reševanju NPZ-ja

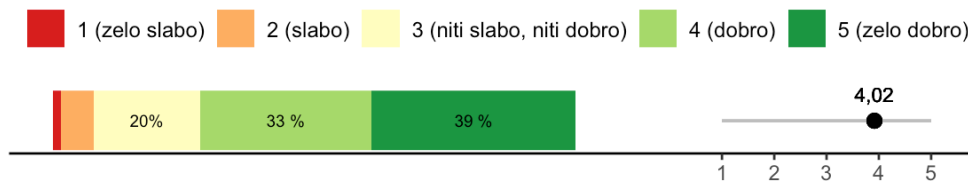
Podrobneje naju je zanimalo mnenje anketirancev o učinkovitosti ChatGPT-ja po posameznih vidikih NPZ-ja. V nadaljevanju so prikazani rezultati po padajočem vrstnem redu pomembnosti, kot so ga podali anketiranci, in sicer: **pravopis** (Slika 13), **oblikovanje besedila** (Slika 14), **slogovna pravilnost** (Slika 15), **slovnica** (Slika 16), **književnost** (Slika 17), **jezikovne naloge** (Slika 18) in **bralno razumevanje** (Slika 19).

Slika 13 prikazuje rezultate pričakovane uspešnosti ChatGPT-ja z vidika **pravopisa**, ki je s povprečno oceno **4,20** najvišje ocenjena. Anketiranci menijo, da se bo ChatGPT najbolje odrezal pri pravilnem sklanjanju besed, rabi predlogov, stopnjevanju pridevnikov ipd.



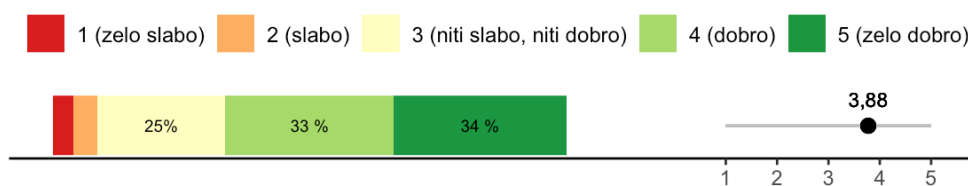
Slika 13: Pričakovan rezultat ChatGPT-ja pri pravopisu

Slika 14 prikazuje vidik **oblikovanja besedila**, ki je s povprečno oceno **4,02** ravno tako visoko ovrednoten. Anketiranci menijo, da bo ChatGPT uspešen pri pisanju vabil, kratkega življenjepisa ipd.



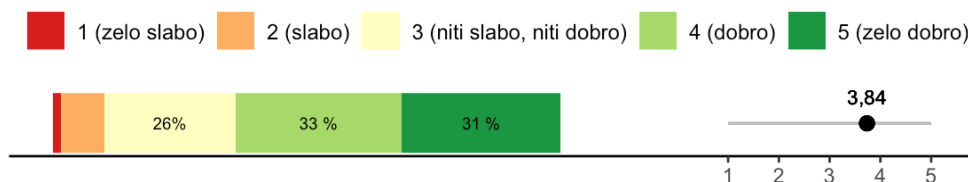
Slika 14: Pričakovan rezultat ChatGPT-ja pri oblikovanju besedila

S povprečno oceno **3,88** na Slika 15 sledi **slogovna pravilnost**, ki so ga anketiranci postavili na tretje mesto. Pri tem so upoštevali jasnost in natančnost izražanja, uporabo knjižnega jezika in pretežno nevtralnega besedišča.



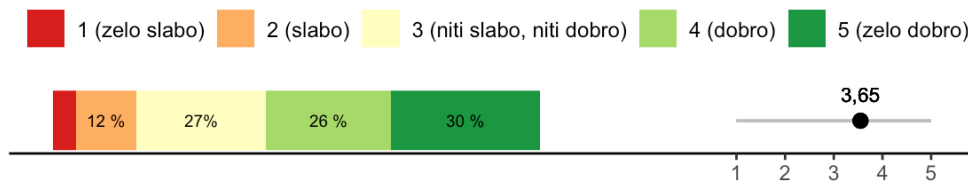
Slika 15: Pričakovan rezultat ChatGPT-ja pri slogovni pravilnosti

Slika 16 prikazuje vidik **slovnice**, ki je dosegel zgolj nekoliko slabšo povprečno oceno **3,84** v primerjavi s predhodnikom. Anketirani so se osredotočili na pravilno sklanjanje besed, rabo predlogov, stopnjevanje pridevnikov ipd.



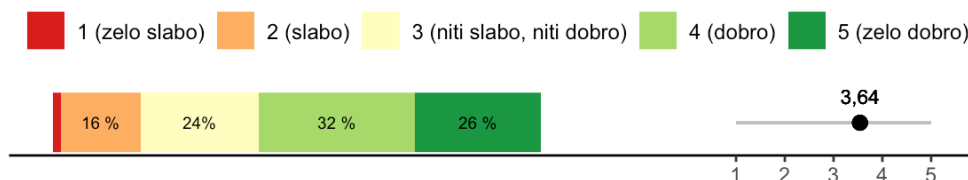
Slika 16: Pričakovan rezultat ChatGPT-ja pri slovnici

Sledi **književnost** na Slika 17, ki je glede na mnenje anketirancev med slabše ocenjenimi. Dosegla je povprečno oceno **3,65**. Pri tem vprašanju sva anketirance prosili, da naj ocenijo, kako uspešen bi po njihovem mnenju bil ChatGPT pri analizi neumetnostnega besedila, določitvi teme besedila, pesniških sredstev (npr. primere, okrasnega pridevka, rime, poosebitve) ipd.

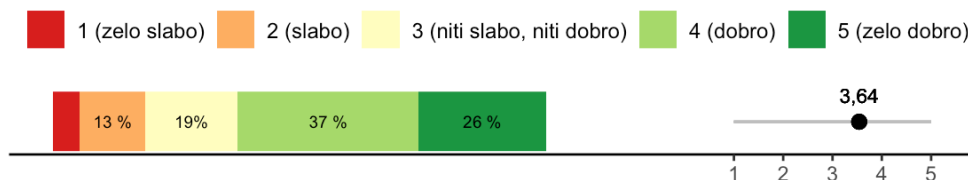


Slika 17: Pričakovan rezultat ChatGPT-ja pri književnosti

S podobno povprečno oceno **3,64** sledita področji **jezikovne naloge** (Slika 18) in **bralno razumevanje** (Slika 19). Pri jezikovnih nalogah so se anketiranci morali osredotočiti na določanje glavnega in odvisnega stavka v povedi, stopnjevanje pridevnikov, spreganje glagolov, poznavanje frazemov ipd. Pri bralnem razumevanju umetnostnih in neumetnostnih besedil pa na odgovore na vprašanja, povezana s priloženim besedilom.



Slika 18: Pričakovan rezultat ChatGPT-ja pri jezikovnih nalogah



Slika 19: Pričakovan rezultat ChatGPT-ja pri bralnem razumevanju

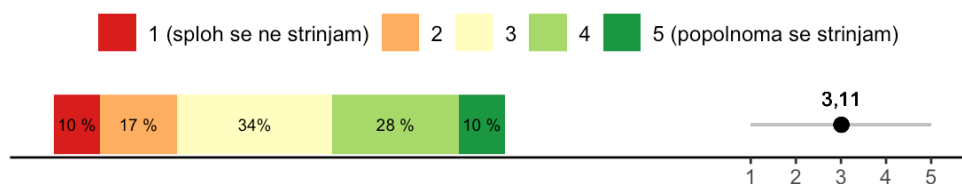
3.2.1.4 Vpliv umetne inteligence

V zadnjem sklopu vprašalnika so anketiranci odgovarjali na tri vprašanja o vplivu umetne inteligence, in sicer:

1. **Svet bi bil z umetno inteligenco boljši.**
2. **Umetna inteligenca bi v prihodnosti lahko nadomestila učitelja.**
3. Kakšno je vaše mnenje o **uporabi umetne inteligence pri učenju?**

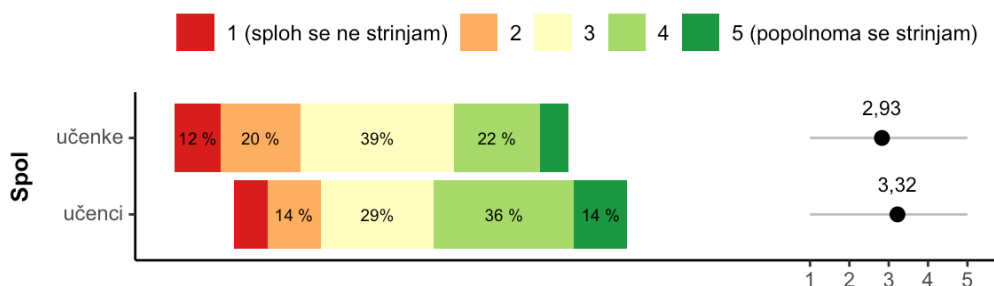
Pri prvih dveh vprašanjih je bilo možnih pet izbir (ocenjevalna lestvica 1-5; pri čemer 1 pomeni sploh se ne strinjam, 5 pa pomeni popolnoma se strinjam), pri zadnjem vprašanju pa je bil predviden kratek besedilni odgovor.

Slika 20 prikazuje neodločenost anketirancev pri prvem vprašanju, kjer naju je zanimalo njihovo mnenje, ali bi bil svet z umetno inteligenco boljši. Povprečna ocena 3,11 je zelo blizu odgovora, kjer se anketiranci niti ne strinjajo, niti strinjajo s trditvijo.



Slika 20: Ali bi bil svet z umetno inteligenco boljši?

Slika 21 prikazuje, da so **učenke** s povprečno vrednostjo **2,93 bližje nestrinjanju** s trditvijo in **učenci** s povprečno vrednostjo **3,32 bližje strinjanju** s trditvijo. Medsebojna razlika med omenjenima ocenama znaša **0,39** in nakazuje, da so učenci bolj naklonjeni ideji, da bi bil svet z umetno inteligenco boljši.

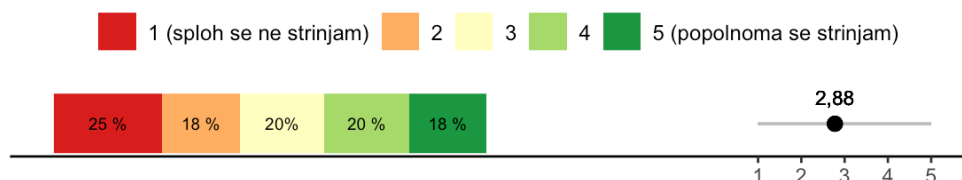


Slika 21: Ali bi bil svet z umetno inteligenco boljši - glede na spol

S tem lahko **potrdiva hipotezo 3**.

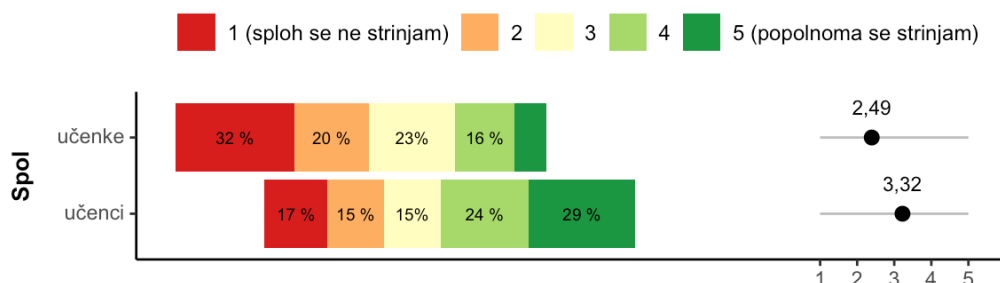
Hipoteza 3: Učenci so bolj naklonjeni uporabi umetne inteligence kot učenke.

Pri vprašanju, ali bi umetna inteligenca v prihodnosti lahko nadomestila učitelja, so anketiranci s povprečnim odgovorom **2,88** izrazili rahlo nestrinjanje s trditvijo (Slika 22). Opazimo lahko tudi zelo visok delež najbolj negativnih odgovorov (1 oz. izrazito nestrinjanje s trditvijo).



Slika 22: Ali bi lahko umetna inteligenca v prihodnosti nadomestila učitelja?

Pri podrobni analizi odgovorov glede na spol anketiranca se je pokazala velika razlika, kot to prikazuje Slika 23. Učenke se izrazito manj strinjajo s trditvijo (odgovor s povprečno vrednostjo **2,49**), medtem ko se učenci s trditvijo rahlo strinjajo (odgovor s povprečno vrednostjo **3,32**). Medsebojna razlika med učenkami in učenci tako znaša kar **0,83**.



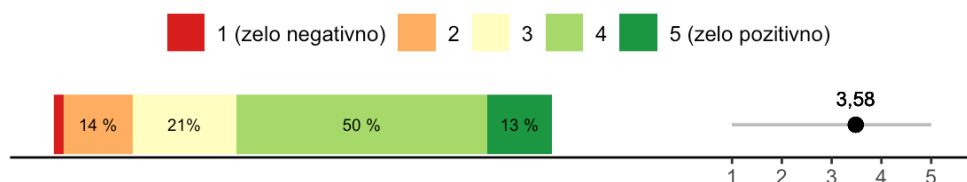
Slika 23: Ali bi lahko umetna inteligenca v prihodnosti nadomestila učitelja - glede na spol

Hipoteza 4 ne moremo potrditi. Povprečni odgovor devetošolcev je 2,88, kar je rahlo nestrinjanje, kjer so učenci bolj naklonjeni umetni inteligenci v vlogi učitelja (povprečna ocena 3,32), medtem ko pri učenkah velja ravno nasprotno (povprečna ocena 2,49).

Hipoteza 4: Devetošolci menijo, da bi umetna inteligenca v prihodnosti lahko nadomestila učitelja.

Pri tretjem vprašanju, ki se je nanašalo na mnenje o uporabi umetne inteligence pri učenju, so imeli anketiranci prosto možnost odgovora, kjer le-ta ni bil obvezen.

Z veljavnim odgovorom je na vprašanje odgovoril **101 anketiranec**, kjer sva odgovore ročno pregledali in jih vsebinsko razdelili v **pet kategorij mnenja o uporabi umetne inteligence pri učenju**, in sicer: 1 (zelo negativno), 2 (negativno), 3 (nevtravno), 4 (pozitivno) in 5 (zelo pozitivno). Slika 24 prikazuje, da je bila povprečna vrednost 3,58, kar predstavlja rahlo pozitiven odnos do uporabe umetne inteligence pri učenju.



Slika 24: Mnenje o uporabi umetne inteligence pri učenju

Komentarje o uporabi umetne inteligence pri učenju sva vsebinsko združili v sorodne sklope, kjer so bili najbolj izraziti naslednji:

- **pomaga, skrajša čas pri razumevanju, razloži** (12,50 % vseh anketirancev),

- **previdno z uporabo, kot pomočnik in potem sam preveri odgovor** (6,25 % vseh anketirancev),
- **domača naloga, esej** (3,91 % vseh anketirancev),
- **predmeti (FIZ, KEM, MAT, BIO, ANG)** (3,91 % vseh anketirancev),
- **strah, kaj lahko naredi** (3,12 % vseh anketirancev),
- **goljufanje, regulacija** (0,78 % vseh anketirancev).

Na Slika 25 so v obliki oblaka besed prikazane najpogostejše uporabljene besede pri komentarju o uporabi umetne inteligence pri učenju. V skladu s prejšnjimi vprašanji lahko opazimo rahlo pozitiven odnos, kjer so najbolj pogoste besede: pomaga, dobro, lahko, zelo, včasih, v redu itd.



Slika 25: Oblak besed anketirancev pri opisu uporabe umetne inteligence pri učenju

3.2.2 Analiza rezultatov NPZ-jev ChatGPT-ja

V drugem sklopu praktičnega dela sva raziskovali, kako uspešno ChatGPT rešuje NPZ. Za reševanje sva uporabili ChatGPT 3.5, zadnje tri NPZ-je iz prejšnjih let, in sicer 2023, 2022 in 2021. Naloge iz posameznega NPZ-ja sva pretvorili v pozive ChatGPT in od modela pridobili odgovore. Odgovore sva ovrednotili s pomočjo navodil za vrednotenje [23, 24, 25], pri čemer sva pri nekaterih nalogah, kjer sva bili pri vrednotenju v dilemi, za pomoč zaprosili učitelja slovenščine. Pridobljene rezultate sva primerjali s povprečnim rezultatom devetošolcev, ki so pisali NPZ.

Med najinim raziskovanjem nama je šola ponudila možnost uporabe plačljive verzije ChatGPT-ja – ChatGPT 4. ChatGPT 4.0 sva uporabili za reševanje NPZ-ja, ki ga je ChatGPT 3.5 najslabše rešil, torej NPZ iz leta 2021. Nato sva primerjali rezultate, ki sta jih dosegla ChatGPT 3.5 in ChatGPT 4.0 pri reševanju NPZ-ja iz leta 2021, in skušali ugotoviti, ali je ChatGPT 4 uspešnejši pri reševanju tega NPZ-ja.

3.2.2.1 Težavnost nalog

Naloge pri NPZ-ju so označene z barvnimi območji, ki pomenijo uspešnost reševanja, kjer je nalogo pravilno rešilo vsaj 65 % učencev z dosežki v danem območju in velja za vse NPZ-je. To pomeni, da so učenci z višjimi dosežki uspešno rešili tudi naloge iz nižjih območij. V vsakem območju nalog je po 10 % vseh učencev, ki so pisali NPZ. [21]

V **zelenem območju** je najslabših 25 % učencev, katerih dosežki so višji od spodnjih 20 % in nižji od 70 % preostalih dosežkov.

V **rumenem območju** se nahaja meja med polovicami (50 %) dosežkov. Njihovi dosežki so višji od spodnjih 45 % in tudi nižji od 45 % preostalih dosežkov.

V **rdečem območju** je najboljših 25 % učencev, kar pomeni, da so njihovi dosežki višji od 70 % spodnjih in nižji od 20 % preostalih dosežkov.

Modro območje označuje najboljših 10 % učencev, katerih dosežki so višji od 90 % dosežkov.

Nad modrim območjem so naloge nad modrim območjem, ki jih v 65 % primerov ne rešijo niti učenci z najvišjimi skupnimi dosežki.

Za vsako nacionalno preverjanje znanja je določen tudi **indeks težavnosti (IT)**, ki ga vrednotimo z izračunom deleža pravilnih odgovorov na posamezno nalogo.

- Če je indeks težavnosti naloge večji od 90 % (**IT > 0,9**), pomeni, da je naloga zelo lahka.
- Če ima naloga indeks težavnosti večji od 70 % (**IT > 0,7**), pomeni, da je naloga lahka.
- Če ima naloga indeks težavnosti večji od 30 % (**IT > 0,3**), je naloga srednje težka.
- Če ima naloga indeks težavnosti večji od 10 % (**IT > 0,1**), je naloga težka.

3.2.2.2 Reševanje NPZ-ja s ChatGPT-jem

Slika 3 prikazuje vse korake, ki sva jih izvedli pri reševanju NPZ-ja s ChatGPT-jem. Za vsak del NPZ-ja (I. in II. del) sva ChatGPT-ju posredovali besedilo (umetnostno ali neumetnostno), na katerega so se navezovala vprašanja iz NPZ-ja. Nato sva oblikovali pozive za ChatGPT, kjer sva upoštevali priporočila iz poglavja 2.2.2.

Primer pogovora s ChatGPT-jem:

- **ChatGPT poziv:**
 - Določi vrsto pripovedovalca v odlomku. Obkroži odgovor in ga utemelji.
 - A Prvoosebni pripovedovalec.
 - B Vsevedni pripovedovalec.
- **ChatGPT odgovor:**
 - A Prvoosebni pripovedovalec.

Sledil je pregled ChatGPT-jevega reševanja NPZ-jev in vrednotenje s pomočjo enotnih meril. [25]. Tako sva lahko določili, kako uspešen je bil ChatGPT pri reševanju.

Pri analizi sva se osredotočili na posamezne vidike NPZ-ja, in sicer: bralno razumevanje umetnostnega in neumetnostnega besedila, književnost, jezikovne naloge, tvorjenje zaključnega besedila ipd. Posebno pozornost sva namenili nalogam, pri katerih ChatGPT ni bil povsem uspešen.

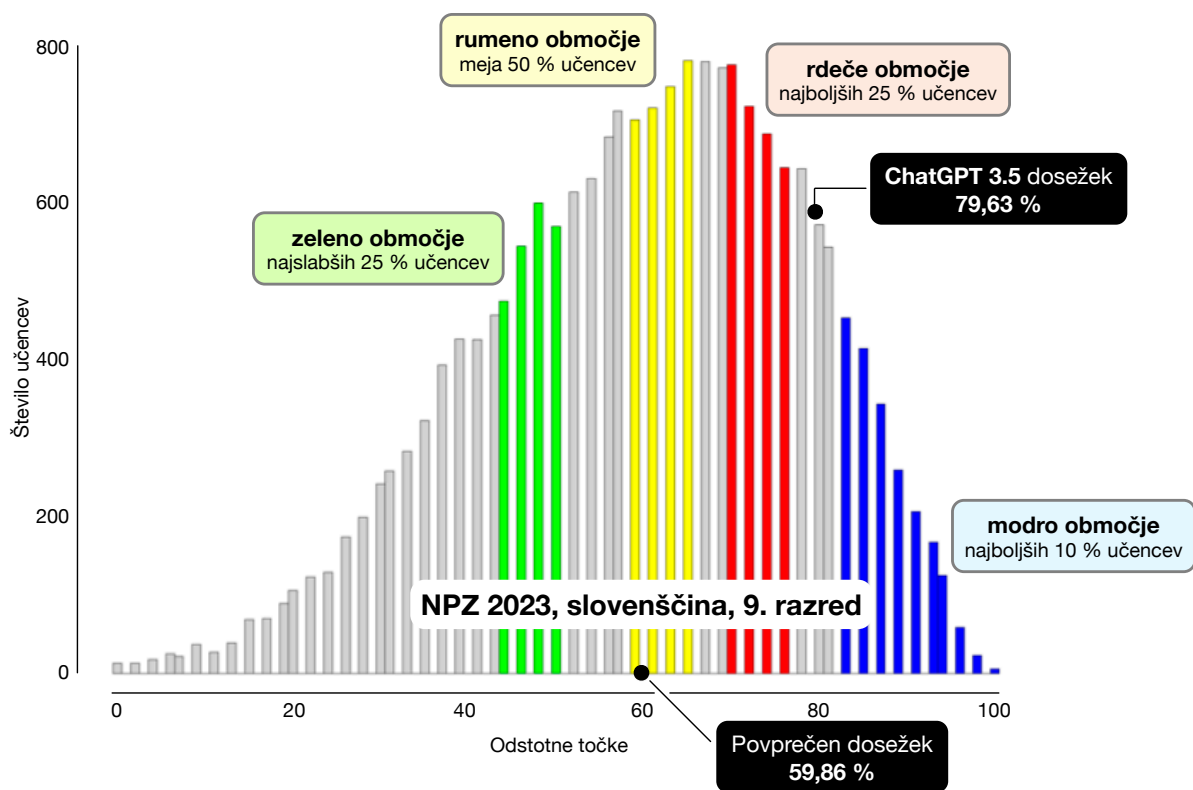
V nadaljevanju so predstavljeni rezultati ChatGPT-jevega reševanja NPZ-jev v zadnjih treh letih od 2021 do 2023.

3.2.2.3 NPZ 2023, ChatGPT 3.5

V preizkusu leta 2023 je bilo **27 nalog** in vseh **možnih točk 54**. Učenke in učenci so v **povprečju** dosegli **32,32 točke (59,86 %)**, kar je za približno 10 % višje kot prejšnja leta. **Indeks težavnosti (IT)** je znašal **0,60**, kar predstavlja nekoliko lažje naloge kot v preteklih letih. [26]

Izhodiščno umetnostno besedilo (I. del) je bil odlomek iz mladinskega romana Aksinje Kermauner **Berenikini kodri**. Izhodiščno neumetnostno besedilo (II. del) je bilo prirejeno besedilo Alenke Veler z naslovom **Nepozabne Slovenke**, objavljeno v reviji PIL (številka 4, leto 2020). Dodan mu je bil še odlomek besedila s spletne strani založbe Mladinska knjiga. [26]

Od **54 možnih točk** je ChatGPT 3.5 dosegel **43 točk (79,63 %)**, kar je za **približno 20 % več** od povprečnega dosežka učenk in učencev, ki znaša 59,86 %. Slika 26 prikazuje, da ChatGPT 3.5 pri tem NPZ-ju po uspešnosti spada med najboljših 10 % (rdeče območje) in 25 % učencev (modro območje).



Slika 26: Porazdelitev točk pri NPZ 2023 in ChatGPT 3.5 [26]

V Tabela 3 so povzete vse napake, ki jih je ChatGPT 3.5 naredil, glede na pravilne rešitve NPZ-ja [25].

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	9.B	-	znanje, uporaba	modro	-1
I. del	10.A	Pozna, razume in uporablja literarnovedne izraze.	znanje	rumeno	-1
II. del	2	Določi sporočevalčev namen.	razumevanje	zeleno	-1
II. del	4	Povzame bistvene podatke.	razumevanje	zeleno	-2
II. del	5	Povzame bistvene podatke.	razumevanje	zeleno	-1
II. del	11	K danim besedam doda njihove knjižne podpomenke in nadpomenke.	znanje	rumeno	-1
II. del	12.B	Določi stavčne člene.	znanje	modro	-1
II. del	13.A	Dane besede iz besedila uvrsti v besedne vrste skladno z opredelitvijo v učnem načrtu.	znanje, uporaba	rumeno	-1
II. del	14.C	V zvezi dveh enostavnih povedi prepozna vzorčno-posledično razmerje, ga izrazi z veznikom in združi zvezo povedi v skladenjsko pravilno dvostavčno poved.	uporaba	modro	-1

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
II. del	14.C	-	znanje, uporaba	modro	-1

Tabela 3: Povzetek napak ChatGPT 3.5 pri NPZ 2023 [25]

V **I. delu** (umetnostno besedilo) je ChatGPT 3.5 **izgubil** zgolj **2 točki** od 22 točk, in sicer je pri 2 vprašanjih odgovoril delno pravilno, medtem ko je pri preostalih 9 vprašanjih odgovoril pravilno. Učenci so naloge I. dela reševali uspešneje kot sicer. [26] Podobno sva ugotovili tudi pri ChatGPT-ju, ki je bil pri tem delu zelo uspešen.

V **II. delu** (neumetnostno besedilo) je **izgubil 9 točk** od 32 točk, in sicer je pri 4 vprašanjih odgovoril napačno in pri 4 vprašanjih delno pravilno, medtem ko je pri preostalih 8 vprašanjih odgovoril pravilno.

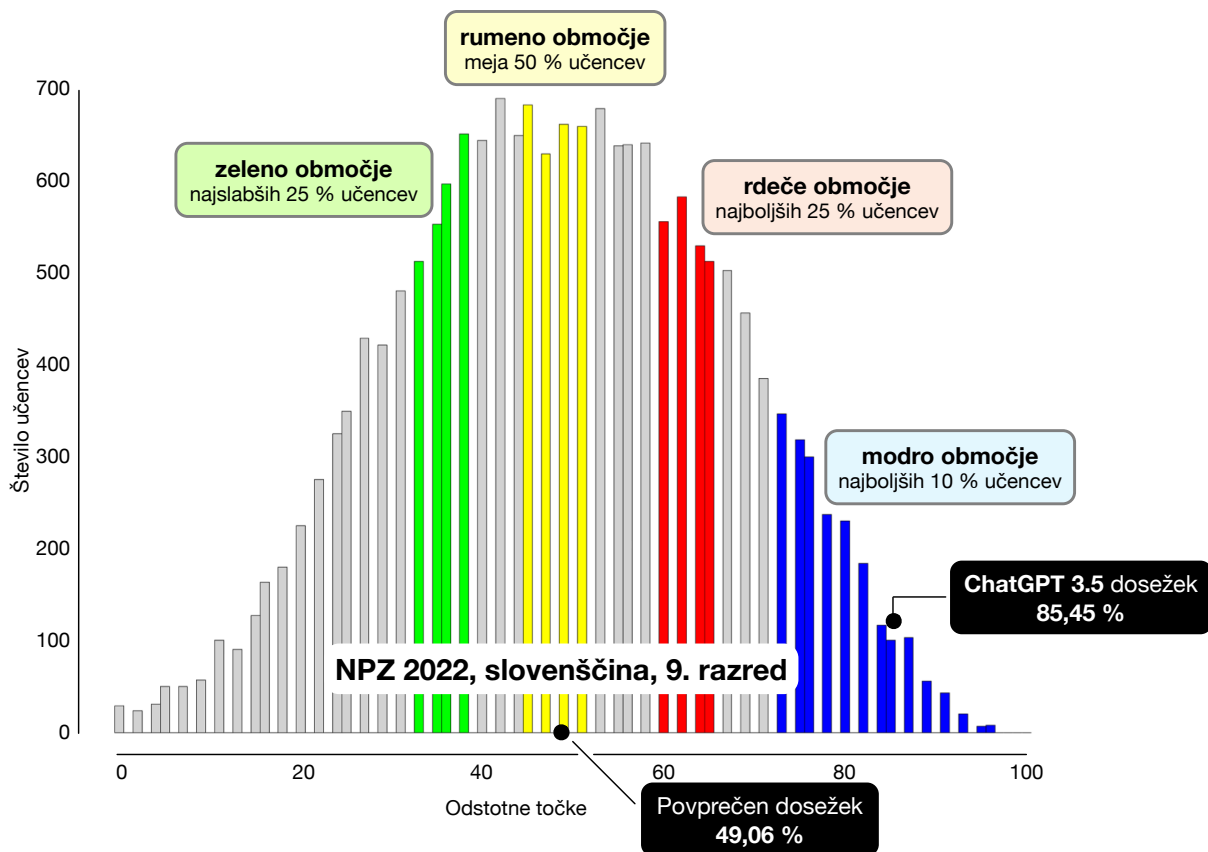
Med napakami lahko zasledimo po 3 naloge iz zelenega (najlažje naloge), rumenega (srednje zahtevne naloge) in modrega območja (zelo zahtevne naloge).

3.2.2.4 NPZ 2022, ChatGPT 3.5

V preizkusu leta 2022 je bilo **32 nalog** in vseh **možnih točk 55**. Učenke in učenci so v povprečju dosegli **26,98 točk (49,06 %)**, kar je primerljivo z dosežki prejšnjih let. **Indeks težavnosti (IT)** je znašal **0,49**, kar predstavlja nekoliko težje naloge kot leta 2023, a primerljive s preteklimi leti. [27]

Izhodiščno umetnostno besedilo (I. del) je bil odlomek iz mladinskega romana Cvetke Bevc **Škampi v glavi**. Izhodiščno neumetnostno besedilo (II. del) je bil prirejen intervju Jureta Krefta z režiserjem Mitjo Okornom **Leto življenja** (PIL, februar 2021, str. 30–31) [27].

Od **55 možnih točk** je ChatGPT 3.5 dosegel **47 točk (85,45 %)**, kar je **približno 35 % več** od povprečnega dosežka učenk in učencev, ki znaša 49,06 %. Slika 27 prikazuje, da ChatGPT 3.5 pri tem NPZ-ju po uspešnosti spada med najboljših 10 % učencev (modro območje).



Slika 27: Porazdelitev točk pri NPZ 2022 in ChatGPT 3.5 [27]

V Tabela 4 so povzete vse napake, ki jih je ChatGPT 3.5 naredil, glede na pravilne rešitve NPZ-ja [24].

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	2	Ubesedi svojo predstavo dogajalnega prostora in časa.	razumevanje	rumeno	-1
I. del	7	Izrazi svoje doživljanje, razumevanje in vrednotenje posameznih prvin besedila.	razumevanje	rumeno	-1
I. del	12.A	V sprejetem UB loči glavno književno osebo in stransko, opiše izbrano osebo, predstavi njen družbeni položaj.	razumevanje, vrednotenje, sinteza	rdeče	-1
II. del	4	Povzame podteme.	razumevanje	zeleno	-1
II. del	12.A	Danim besedam iz besedila določi oblikovne lastnosti.	znanje, uporaba	nad modrim	-1
II. del	15.A	Vpraša po danih delih povedi ter tvori pomensko in oblikovno ustrezne vprašalne povedi.	znanje, uporaba	zeleno	-1
II. del	16	Iz danih skladenjskih podstav tvori tvorjenke.	uporaba	nad modrim	-1

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
II. del	18	Pravilno piše vrstne pridevnike, izpeljane iz zemljepisnih lastnih imen in eno- in večbesedna zemljepisna lastna imena.	znanje, uporaba	nad modrim	-1

Tabela 4: Povzetek napak ChatGPT 3.5 pri NPZ 2022 [24]

V **I. delu** (umetnostno besedilo) je ChatGPT **izgubil 1 točko** od 22 točk, in sicer je pri 1 vprašanju odgovoril delno pravilno, medtem ko je pri preostalih 12 vprašanjih odgovoril pravilno.

V **II. delu** (neumetnostno besedilo) je **izgubil 5 točk** od 33 točk, in sicer je pri 3 vprašanjih odgovoril napačno in pri 2 vprašanjih delno pravilno, medtem ko je pri preostalih 14 vprašanjih odgovoril pravilno.

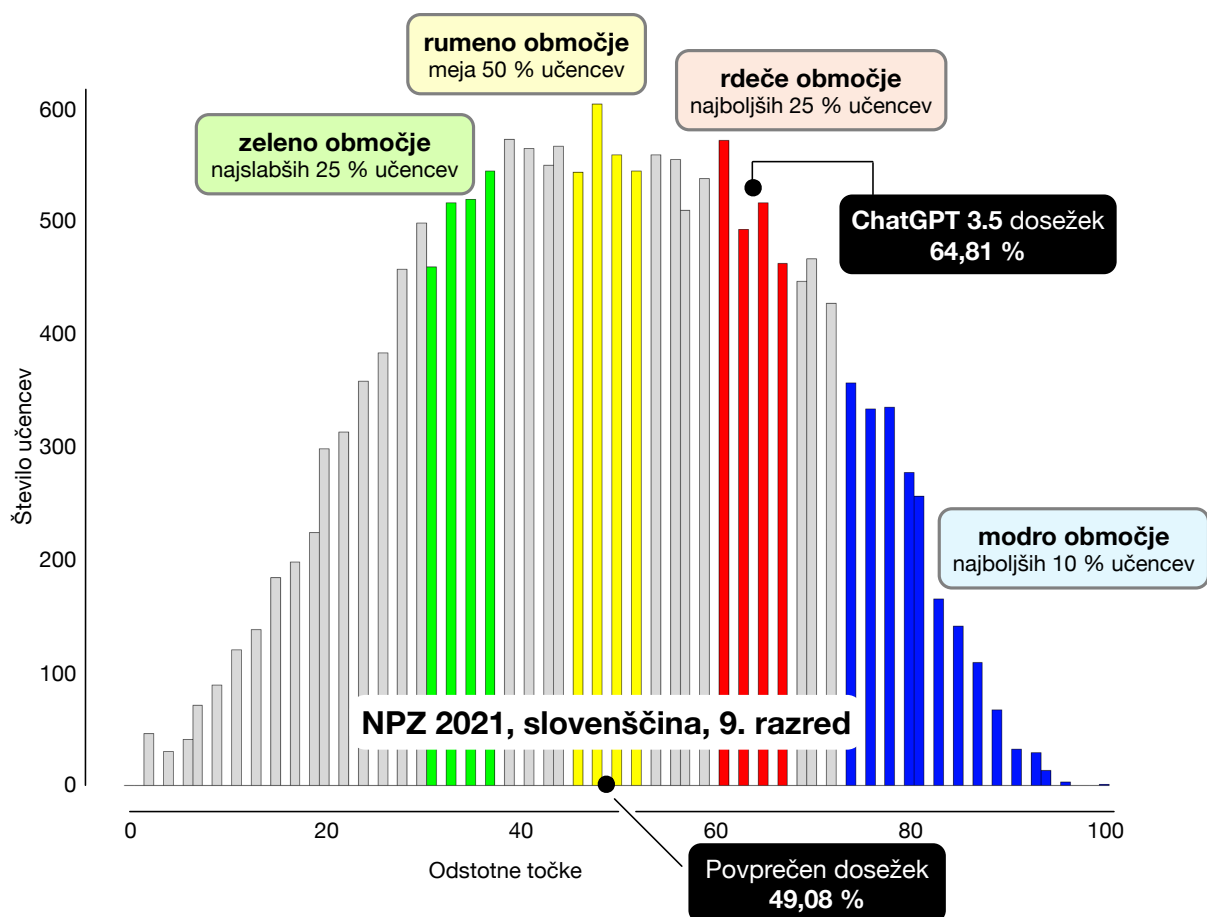
Med napakami lahko zasledimo 2 nalogi iz zelenega območja (najlažje naloge), 2 nalogi iz rumenega območja (srednje zahtevne naloge), 1 nalogo iz rdečega območja (zahtevne naloge) in 3 naloge iz območja nad modrim, ki predstavlja najbolj zahtevne naloge, ki jih večinoma ne rešijo niti najboljši učenci.

3.2.2.5 NPZ 2021, ChatGPT 3.5

V preizkusu leta 2021 je bilo **27 nalog** in vseh **možnih točk 54**. Učenke in učenci so v povprečju dosegli **26,50 točk (49,08 %)**, kar je primerljivo z dosežki prejšnjih let. **Indeks težavnosti (IT)** je znašal **0,49**, kar predstavlja primerljivo težavnost s preteklimi leti. [23]

Izhodiščno umetnostno besedilo (I. del) je bil odlomek iz mladinskega romana Smiljana Rozmana **Ta glavna Urša**. Izhodiščno neumetnostno besedilo (II. del) je bilo prirejeno besedilo Vesele šole z naslovom **Ludwig van Beethoven – dramatični "tonski" pesnik**, ki ga je napisal Peter Grum (PIL, februar 2020, str. 41–49).

Od **54 možnih točk** je ChatGPT 3.5 dosegel **35 točk (64,81 %)**, kar je **približno 15 % več** od povprečnega dosežka učenk in učencev, ki znaša 49,08 %. Slika 28 prikazuje, da ChatGPT 3.5 pri tem NPZ-ju po uspešnosti spada med najboljših 25 % učencev (rdečega območja).



Slika 28: Porazdelitev točk pri NPZ 2021 in ChatGPT 3.5 [28]

V Tabela 7 so povzete vse napake, ki jih je ChatGPT 3.5 naredil, glede na pravilne rešitve NPZ-ja [23].

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	1.B	Pozna, razume in uporablja literarnovedne izraze, besedila uvrsti v časovni okvir njihovega nastanka in utemlji svojo izbiro.	znanje, razumevanje	rdeče	-1
I. del	3	Slovnična in pravopisna pravilnost ter slogovna ustreznost.	znanje, uporaba	modro	-1
I. del	5	Imenuje avtorja, prepozna in poimenuje prvoosebne pripovedovalca.	znanje, razumevanje	rdeče	-1
I. del	6	Razloži dane frazeme iz besedila.	razumevanje	rumeno	-1
I. del	7.A	V pesmi najde pesniško sredstvo (primero, inverzijo) in ga poimenuje; najde in poimenuje ponavljanje, prepozna rimo.	znanje, uporaba	modro	-1
I. del	7.B	Predstavi vlogo pesniških sredstev (primera, ponavljanje, rime, inverzija).	znanje, razumevanje	nad modrim	-1

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	8	Povzame sporočilo književnega besedila.	razumevanje	nad modrim	-1
I. del	11	Pozna umetnostna besedila obvezno obravnavanih avtorjev.	znanje	rumeno	-1
II. del	1	Določi okoliščine nastanka sprejetega besedila.	znanje	modro	-1
II. del	2.A	Strukturirano povzame bistvene podatke in logična razmerja med njimi.	razumevanje	nad modrim	-1
II. del	2.B	Slovnična in pravopisna pravilnost ter slogovna ustreznost.	znanje, uporaba	nad modrim	-1
II. del	4	Strukturirano povzame bistvene podatke.	razumevanje	rdeče	-1
II. del	8	Slovnična in pravopisna pravilnost ter slogovna ustreznost.	znanje, uporaba	rdeče	-1
II. del	9	Danim besedam, besednim zvezam določi slogovno vrednost.	razumevanje	nad modrim	-1
II. del	10.A	K danim besedam doda njihove knjižne sopomenke.	znanje	rdeče	-1
II. del	12.A	Dopolni enostavno poved s pomensko in oblikovno ustreznimi deli.	uporaba	rdeče	-1
II. del	13.A	Razume, uporablja in ponazori jezikoslovne izraze, navedene v učnem načrtu.	znanje	nad modrim	-1
II. del	14.A	Razume jezikoslovne izraze.	znanje	zeleno	-1
II. del	14.B	Vpraša po danih delih povedi ter tvori pomensko in oblikovno ustrezne vprašalne povedi.	razumevanje	modro	-1

Tabela 5: Povzetek napak ChatGPT 3.5 pri NPZ 2021 [23]

V **I. delu** (umetnostno besedilo) je ChatGPT 3.5 **izgubil** 8 točk od 22 točk, in sicer pri 4 vprašanjih je odgovoril napačno in pri 3 vprašanjih delno pravilno, medtem ko je pri preostalih 4 vprašanjih odgovoril pravilno.

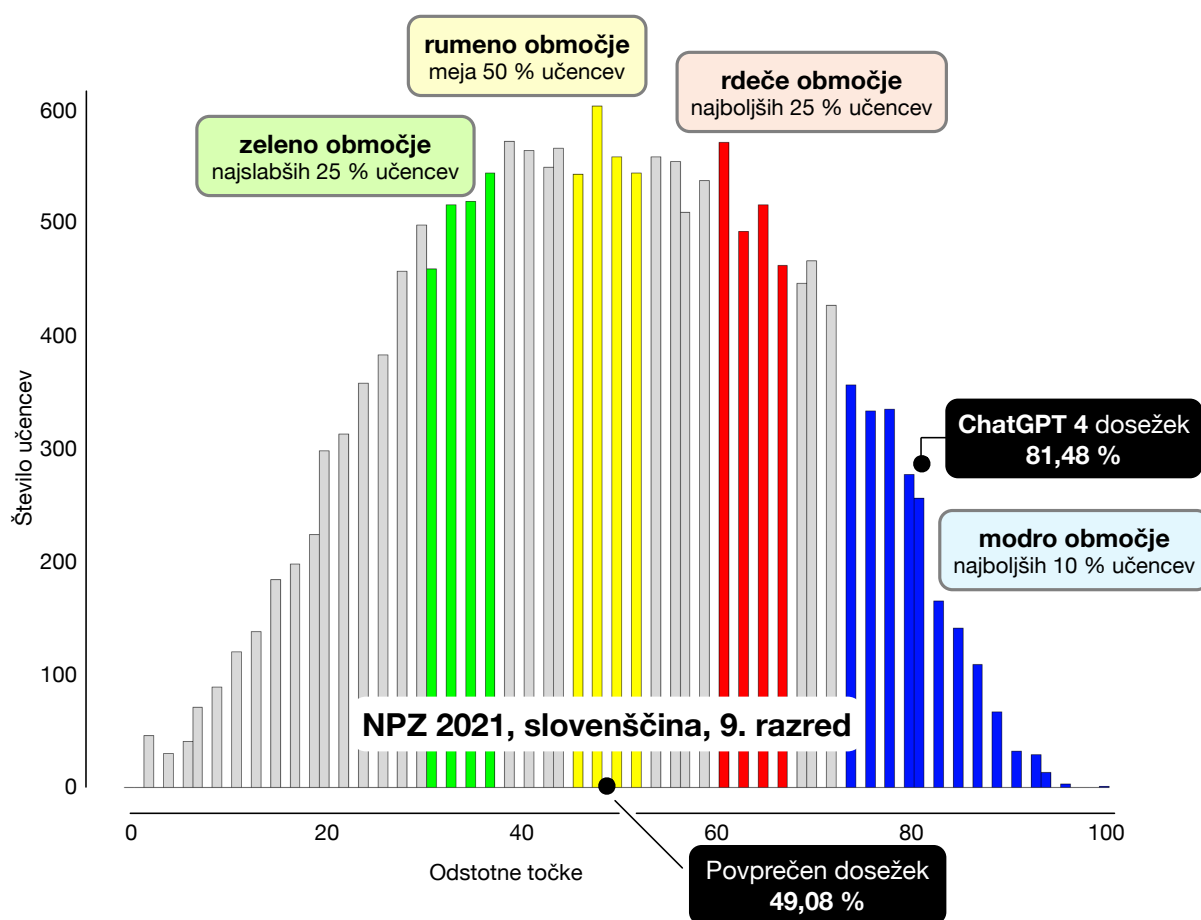
V **II. delu** (neumetnostno besedilo) je izgubil 11 točk od 32 točk, in sicer je pri 4 vprašanjih odgovoril napačno in pri 5 vprašanjih delno pravilno, medtem ko je pri preostalih 7 vprašanjih odgovoril pravilno.

Med napakami lahko zasledimo 1 nalogo iz zelenega območja (najlažje naloge), 2 nalogi iz rumenega območja (srednje zahtevne naloge), 5 nalog iz rdečega območja (zahtevne naloge), 4 naloge iz modrega območja (zelo zahtevne naloge) in 5 nalog iz območja nad modrim, ki predstavljajo najbolj zahtevne naloge, ki jih večinoma ne rešijo niti najboljši učenci.

3.2.2.6 NPZ 2021, ChatGPT 4

ChatGPT 3.5 je NPZ 2021 rešil najslabše (glej podrobnosti v prejšnjem poglavju 3.2.2.5), zato sva se odločili, da poskusiva reševanje še z najnovejšim modelom ChatGPT 4. Omenjena različica je sicer plačljiva, zato sva jo preverili zgolj pri enem NPZ-ju.

Od **54 možnih točk** je ChatGPT 4 dosegel **44 točk (81,48 %)**, kar je **približno 30 % več** od povprečnega dosežka učenk in učencev, ki znaša 49,08 %. Prav tako je omenjen dosežek modela ChatGPT 4 za **približno 15 % boljši** od modela ChatGPT 3.5. Slika 29 prikazuje, da ChatGPT pri tem NPZ-ju po uspešnosti spada med najboljših 10 % učencev (modro območje).



Slika 29: Porazdelitev točk pri NPZ 2021 in ChatGPT 4 [28]

V Tabela 6 so povzete vse napake, ki jih je ChatGPT 4 naredil, glede na pravilne rešitve NPZ-ja [23].

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	7.A	V pesmi najde pesniško sredstvo (primero, inverzijo) in ga poimenuje; najde in poimenuje ponavljanje, prepozna rimo.	znanje, uporaba	modro	-1

Del	Naloga	Standard znanja	Taksonomska stopnja	Območje	Izgubljene točke
I. del	7.B	Predstavi vlogo pesniških sredstev (primera, ponavljanje, rime, inverzija).	znanje, razumevanje	nad modrim	-1
I. del	11	Pozna umetnostna besedila obvezno obravnavanih avtorjev.	znanje	rumeno	-1
II. del	3.A	Določi okoliščine nastanka sprejetega besedila in poveže, iz katerih prvin besedila jih je prepoznal.	razumevanje	modro	-1
II. del	8	Slovnična in pravopisna pravilnost ter slogovna ustreznost.	znanje, uporaba	rdeče	-1
II. del	9	Danim besedam, besednim zvezam določi slogovno vrednost.	razumevanje	nad modrim	-1
II. del	11.B	V besedilu slogovno zaznamovane besede, besedne zveze zamenja z nezaznamovanimi.	razumevanje	nad modrim	-1
II. del	13.A	Razume, uporablja in ponazori jezikoslovne izraze, navedene v učnem načrtu.	znanje	nad modrim	-1
II. del	14.A	Razume jezikoslovne izraze.	znanje	zeleno	-1
II. del	14.B	Vpraša po danih delih povedi ter tvori pomensko in oblikovno ustrezne vprašalne povedi.	razumevanje	modro	-1

Tabela 6: Povzetek napak ChatGPT 4 pri NPZ 2021 [23]

V **I. delu** (umetnostno besedilo) je ChatGPT 4 **izgubil** 3 točki od 22 točk, in sicer je pri 2 vprašanjih odgovoril delno pravilno, medtem ko je pri preostalih 9 vprašanjih odgovoril pravilno.

V **II. delu** (neumetnostno besedilo) je izgubil 7 točk od 32 točk, in sicer je pri 1 vprašanju odgovoril napačno in pri 5 vprašanjih delno pravilno, medtem ko je pri preostalih 10 vprašanjih odgovoril pravilno.

Med napakami lahko zasledimo 1 nalogo iz zelenega območja (najlažje naloge), 2 nalogi iz rumenega območja (srednje zahtevne naloge), 1 nalogo iz rdečega območja (zahtevne naloge), 3 naloge iz modrega območja (zelo zahtevne naloge) in 4 naloge iz območja nad modrim, ki predstavljajo najbolj zahtevne naloge, ki jih večinoma ne rešijo niti najboljši učenci.

3.2.2.7 Povzetek ChatGPT-jevih reševanj NPZ-jev

Tabela 7 prikazuje uspešnost ChatGPT-ja pri reševanju NPZ-ja, hkrati pa je dodan povprečen odstotek uspešnosti učenk in učencev, tj. kako uspešno so preverjanje reševali devetošolci.

Leto	ChatGPT	Dosežek ChatGPT	Povprečje devetošolcev
2023	3,5	79,63 %	59,86 %
2022	3,5	85,45 %	49,06 %
2021	3,5	64,81 %	49,10 %
2021	4	81,48 %	49,10 %

Tabela 7: Primerjava dosežka ChatGPT-ja pri NPZ

Ugotovimo lahko, da je bila pri reševanju pri vseh obravnavanih NPZ-jih umetna inteligenca v obliki modela ChatGPT uspešnejša kot povprečen devetošolec, s čimer **potrjujeva hipotezo 5**.

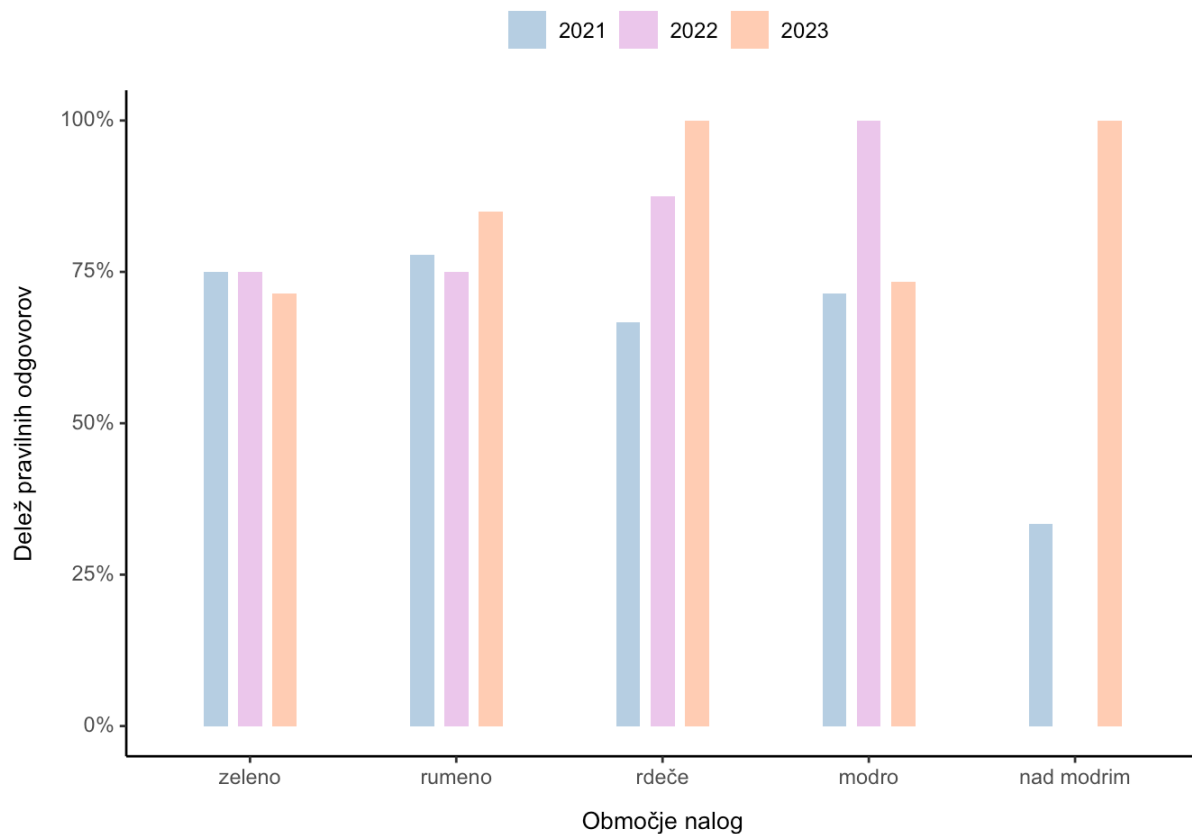
Hipoteza 5: ChatGPT uspešno rešuje NPZ iz slovenščine in je pri tem boljši kot povprečen devetošolec.

Zanimala naju je uspešnost reševanja ChatGPT-ja glede na težavnost nalog. Kot sva že predstavili v poglavju 3.2.2.1, je vsaka naloga pri NPZ-ju uvrščena v območje težavnosti [28, 27, 26], in sicer:

- **zeleno** območje (najlažje naloge),
- **rumeno** območje (srednje zahtevne naloge),
- **rdeče** območje (zahtevne naloge),
- **modro** območje (zelo zahtevne naloge),
- **nad modrim** območjem (najbolj zahtevne naloge).

Slika 30 prikazuje uspešnost ChatGPT-ja 3.5 pri vsaki izmed omenjenih skupin. Opazimo lahko, da je model umetne inteligence pri vseh treh NPZ-jih (od 2021 do 2023) enako uspešen pri nalogah v zelenem in rumenem območju, kjer v povprečju dosega okrog 75 %. V rdečem in modrem območju so razlike med leti sicer nekoliko večje, vendar je uspešnost ChatGPT-ja še vedno visoka, ponovno okrog 75 %.

Nekoliko drugače se ChatGPT odreže pri najbolj zahtevnih nalogah, ki spadajo v območje nad modrim, in sicer se pri zadnjem NPZ-ju iz leta 2023 pri takšni nalogi ni zmotil. Leta 2022 tako zahtevnih nalog na NPZ-ju ni bilo, medtem ko je ChatGPT NPZ iz leta 2021 podpovprečno uspešen, saj je pravilno rešil zgolj okrog 33 % nalog.

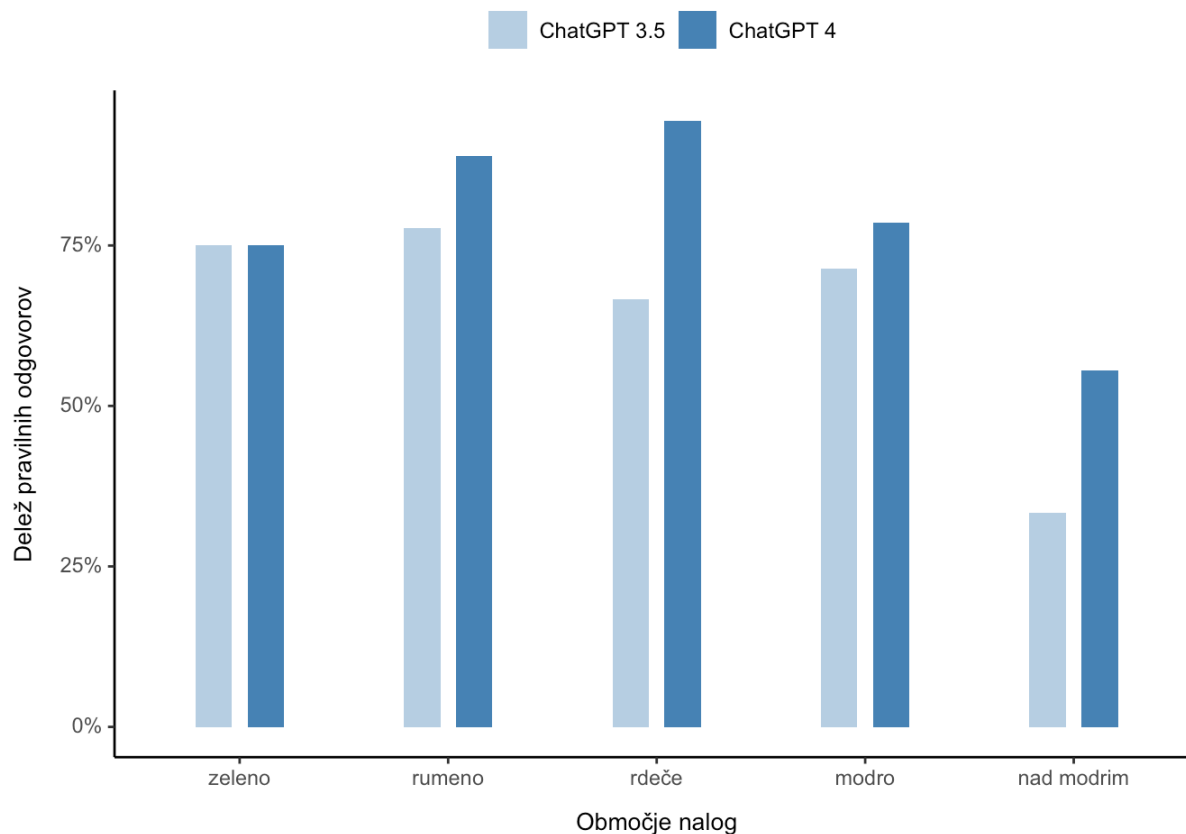


Slika 30: Uspešnost ChatGPT 3.5 glede na območje nalog in leta

S tem lahko **potrdiva hipotezo 6**, saj so bili rezultati pri težjih nalogah bolj raznoliki med NPZ-ji. Najtežje naloge v območju nad modrim je namreč ChatGPT pri NPZ-ju iz leta 2021 rešil zelo slabo. Prav tako je povprečen delež pravih odgovorov pri težjih nalogah nižji kot pri lažjih.

Hipoteza 6: ChatGPT ima pri NPZ-ju največ težav s težjimi nalogami, ki jih večinoma ne rešijo niti najboljši učenci.

Slika 31 prikazuje reševanje NPZ-ja iz leta 2021, ki ga je model ChatGPT 3.5 rešil najslabše. Ta NPZ sva dali še v dodatno reševanje najnovejšemu plačljivemu modelu ChatGPT 4. Rezultati so pokazali, da je najnovejša verzija ChatGPT-ja enako ali bolj uspešna od predhodnika prav pri vseh različnih zahtevnostih nalogah. Pri najlažjih nalogah (zeleno območje) sta modela enako uspešna, pri zahtevnejših nalogah pa je (rumeno, rdeče, modro in območje nad modrim) najnovejši model ChatGPT 4 boljši od modela ChatGPT 3.5.



Slika 31: Primerjava ChatGPT 3.5 in 4 pri NPZ 2021

3.3 Izzivi pri raziskovalnem delu

V raziskovalni nalogi se nisva soočili z večjimi težavami, kljub temu pa sva naleteli na nekaj izzivov.

Eden je bil ta, da se na anketni vprašalnik niso odzvali učenci vseh šol, ki sva jih nameravali vključiti v raziskavo. Od devetih šol so odgovore na anketo podale štiri, kar je, kljub temu, znašalo 128 sodelujočih.

Veliko dela sva imeli tudi z vnosom prazne pole NPZ-ja v obliko pozivov ChatGPT-ja. Najprej sva namreč morali celotno vsebino pretvoriti v obliko besedila in jo vnesti. Pri določenih vprašanjih sva morali zastaviti vprašanje na način, kot to predlagajo v priporočilih, ki sva jih predstavili v poglavju 2.2.2.

Zahtevno je bilo tudi vrednotenje ChatGPT-jeve uspešnosti reševanja NPZ-ja, kjer sva si pomagali z navodili za vrednotenje [23, 24, 25], pri nekaterih nalogah pa sva ne glede na dane rešitve navodil za vrednotenje potrebovali pomoč učitelja slovenščine.

4 Zaključek

V raziskovalni nalogi sva se ukvarjali z uporabo umetne inteligence in njeno vlogo v izobraževanju. Osredotočili sva se na uspešnost reševanja modela umetne inteligence ChatGPT-ja iz nacionalnega preverjanja znanja iz slovenščine v 9. razredu. Po pregledu literature sva ugotovili, da v Sloveniji takšne raziskave še niso izvedli, v tujini pa je ChatGPT na standardiziranih testih dosegel zelo dobre rezultate. Zastavili sva 6 hipotez, ki sva jih preverjali z analizo vprašalnika in ChatGPT-jevega reševanja NPZ-jev.

Praktičen del raziskave sva izvedli v dveh delih. Pripravili sva anketni vprašalnik za devetošolce, kjer naju je zanimalo njihovo predznanje o slovenščini ter modelu umetne inteligence ChatGPT. Prav tako sva želeli pridobiti njihovo mnenje o uspešnosti reševanja ChatGPT-ja pri NPZ-ju ter v kakšni vlogi vidijo umetno inteligenco v prihodnosti. V drugem delu raziskave sva izvedli eksperiment - kako uspešno ChatGPT rešuje NPZ-je iz slovenščine v 9. razredu zadnjih treh let.

Spletni anketni vprašalnik je reševalo 128 devetošolcev. Z analizo vprašalnika sva potrdili hipotezo 1, s katero sva preverjali, ali devetošolci poznajo umetno inteligenco in najbolj razširjen model ChatGPT, kar je večina vseh (93 %) potrdila. Potrdili sva tudi hipotezo 2, kjer naju je zanimalo mnenje devetošolcev glede uspešnosti reševanja ChatGPT-ja nalog NPZ-ja iz slovenščine. S povprečno oceno 4 od 5 so ga anketiranci ocenili zelo pozitivno, medtem ko so področja, na katerih naj bi ChatGPT uspešno reševal NPZ, ovrednotili v naslednjem vrstnem redu od najboljšega do najslabšega: pravopis (4,20), oblikovanje besedila (4,02), slogovna pravilnost (3,88), slovnica (3,84), književnost (3,65), jezikovne naloge (3,64) in bralno razumevanje (3,64). Odgovori anketirancev na vprašanje, ali bi bil svet z umetno inteligenco boljši, so bili neodločeni (3,11), pri čemer so bile učenke bolj nenaklonjene (2,93) kot učenci (3,32), s čimer sva potrdili hipotezo 3 o razlikah pri naklonjenosti uporabi umetne inteligence med učenkami in učenci. Hipoteze 4, kjer sva anketirance spraševali, ali bi lahko umetna inteligenca v prihodnosti nadomestila učitelja, nisva mogli potrditi (2,88), saj so bile učenke pri odgovorih izrazito nenaklonjene temu, da bi umetna inteligenca lahko v prihodnosti nadomestila učitelja (2,49) v primerjavi z učenci (3,32), ki so se s tem zgolj rahlo strinjali.

Pri eksperimentu, kjer je ChatGPT reševal zadnje tri NPZ-je, sva ugotovili, da je bil ChatGPT pri prav vseh reševanjih zelo uspešen in je dosegel nadpovprečne rezultate (79,63 %, 85,45 % in

64,81 %) glede na povprečnega devetošolca. Po primerjavi s povprečnimi rezultati učenk in učencev lahko zatrdiva, da se je ChatGPT uvrstil med 10 % do 25 % najboljših devetošolcev. S tem lahko potrdiva hipotezo 5, s katero sva preverjali, ali ChatGPT uspešno rešuje NPZ iz slovenščine in je pri tem boljši kot povprečen devetošolec. Podrobneje sva pogledali tudi, kako se ChatGPT obnese pri nalogah različne težavnosti (zeleno, rumeno, rdeče, modro in območje nad modrim). Izkazalo se je, da je najuspešnejši pri lažjih nalogah (npr. zeleno in rumeno območje), medtem ko pri težjih nalogah njegovi dosežki nihajo in se tudi nekoliko znižajo. S tem sva potrdili hipotezo 6, s katero sva preverjali, ali ima ChatGPT pri NPZ-ju največ težav s težjimi nalogami, ki jih večinoma ne rešijo niti najboljši učenci. S tem ko sva preverili uspešnost reševanja nalog najnovejšega modela ChatGPT 4, lahko trdimo, da je uspešnejši kot brezplačni model ChatGPT 3.5 pri vseh tipih nalog, še zlasti pri težjih.

Čeprav je ChatGPT pri vseh nalogah, ki jih je reševal, dosegel nadpovprečne rezultate, ni vselej najbolj zanesljiv, saj so se rezultati reševanja pri različnih NPZ-jih kar precej razlikovali (npr. od 64,81 % pri NPZ 2021 do 85,45 % pri NPZ 2022). Odgovori na vsebinska vprašanja so pogosto točna, natančna in gostobesedna. Odgovori so z redkimi odstopanji jezikovno dobro zapisani, saj upoštevajo jezikovno ustreznost. Pri vsebinskih vprašanjih, ki zahtevajo bralno razumevanje izhodiščnega besedila, je ChatGPT večinoma uspešen. Nekoliko slabše odgovarja na vprašanja, ki zahtevajo poznavanje strokovnega (jezikoslovnega) besedišča in uporabo slovničnega znanja, npr. pojasnilo rabe začetnice v besedah hollywoodski in Hollywood, ustreznost vprašalnic po podčrtanih stavčnih členih, poznavanje in poimenovanje stavčnih členov. ChatGPT 3.5 prav tako ne prepozna nekaterih pojmov, kot na primer "sodobna književnost" in zato napačno odgovori.

Obravnavali sva zanimivo in aktualno tematiko. Meniva, da sva jo podrobno predstavili. Pri njenem raziskovanju se nama je porodilo še nekaj idej, je pa za njihovo izvedbo zmanjkalo časa; ChatGPT bi preverili še na večjem naboru NPZ-jev in ne zgolj iz zadnjih 3 let, lahko bi se še bolj poglobili v pisanje pozivov in poskušali z drugačnim načinom pisanja pozivov oz. vnosa vprašanj NPZ-ja v ChatGPT izboljšati število pravilnih odgovorov, zanimivo bi bilo preveriti, kako se ChatGPT obnese pri preizkusih znanja iz drugih predmetov, npr. matematike, angleščine, kemije ipd., kako uspešen bi bil ChatGPT pri reševanju NPZ-jev za 6. razred in pri maturitetnih preizkusih na različnih predmetnih področjih.

Nekdo izmed anketirancev je o umetni inteligenci zapisal *"Ni mi všeč! Občutek imam, da bodo kmalu roboti prevzeli svet."* Sami z njim ne deliva njegovega razmišljanja.

Prepričani sva, da bo umetna inteligenca našla svoje mesto ob delovanju človeka in mu bo predvsem v pomoč v vsakdanjem življenju, le naučiti se bo moral sobivati z njo tako, da jo bo znal (koristno) uporabljati. Človeka pa bo, tako menijo tudi anketiranci, težko (ali sploh) povsem nadomestiti.

5 Viri in literatura

- [1] Evropski parlament, „Kaj je umetna inteligenca in kako se uporablja v praksi?“, 26. 3. 2021. [Elektronski]. Dostopno: <https://www.europarl.europa.eu/topics/sl/article/20200827STO85804/kaj-je-umetna-inteligenca-in-kako-se-uporablja-v-praksi>. [Poskus dostopa 1. 3. 2024].
- [2] Evropski parlament, „Umetna inteligenca: Priložnosti in tveganja“, 22. 5. 2022. [Elektronski]. Dostopno: <https://www.europarl.europa.eu/topics/sl/article/20200918STO87404/umetna-inteligenca-priloznosti-in-tveganja>. [Poskus dostopa 3. 3. 2024].
- [3] STA, „Prednosti in slabosti umetne inteligence v umetnosti“, 15. 8. 2023. [Elektronski]. Dostopno: <https://ptujinfo.com/novica/globalno/prednosti-slabosti-umetne-intelligence-v-umetnosti/206368>. [Poskus dostopa 3. 3. 2023].
- [4] P. Lavbič in N. Rakuša, „Umetna inteligenca pri nacionalnem preverjanju znanja (NPZ) iz slovenskega jezika“, 2024. [Elektronski]. Dostopno: <https://forms.gle/1idUThkgeU2qzMow5>. [Poskus dostopa februar 2024].
- [5] G. Zakrajšek, „Evolucija umetne inteligence: od Turinga do velikih jezikovnih modelov“, 4. 7. 2023. [Elektronski]. Dostopno: <https://gamechanger.frodx.com/blog/evolucija-umetne-intelligence-od-turinga-do-velikih-jezikovnih-modelov/>. [Poskus dostopa 3. 3. 2024].
- [6] K. Gao, S. He, Z. He in J. Lin, „Examining User-Friendly and Open-Sourced Large GPT Models: A Survey on Language, Multimodal, and Scientific GPT Models“, arXiv, 2023.
- [7] Wikipedia, „Eliza“, 22. 2. 2024. [Elektronski]. Dostopno: <https://en.wikipedia.org/wiki/ELIZA>. [Poskus dostopa 3. 3. 2024].
- [8] Shaip, „Veliki jezikovni modeli (LLM): Celoten vodnik leta 2023“, 2024. [Elektronski]. Dostopno: <https://sl.shaip.com/blog/a-guide-large-language-model-llm/>. [Poskus dostopa 3. 3. 2024].
- [9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser in I. Polosukhin, „Attention is all you need“, arXiv, 2017.
- [10] Joker, „Kaj je Chat GPT in 10 najboljših primerov uporabe“, 13. 6. 2023. [Elektronski]. Available: <https://www.joker.si/kaj-je-chatgpt-in-najboljsi-primeri-uporabe/>. [Poskus dostopa 1. 3. 2024].
- [11] Wikipedia, „OpenAI“, <https://en.wikipedia.org/wiki/OpenAI>, 2024.

- [12] B. Abe, „Eno leto ChatGPT: kako je spremenil življenja in pet nasvetov, kako napisati dober besedilni poziv,“ *Moje finance*, 18. 1. 2024. [Elektronski]. Dostopno: <https://mojefinance.finance.si/razno/eno-leto-chatgpt-kako-je-spremenil-zivljenja-in-pet-nasvetov-kako-napisati-dober-besedilni-poziv/a/9020276>. [Poskus dostopa 27. 1. 2024].
- [13] OpenAI, „Prompt engineering,“ 2024. [Elektronski]. Dostopno: <https://platform.openai.com/docs/guides/prompt-engineering>. [Poskus dostopa 27. 1. 2024].
- [14] Wikipedia, „ChatGPT,“ 2024. [Elektronski]. Dostopno: <https://en.wikipedia.org/wiki/ChatGPT>. [Poskus dostopa 3. 3. 2024].
- [15] A. Jošovc, „Umetna inteligenca in vsakdanja uporaba,“ 2. 4. 2023. [Elektronski]. Dostopno: <https://uporabaui.splet.arnes.si/2023/04/02/chat-gpt/>. [Poskus dostopa 1. 3. 2024].
- [16] Državni izpitni center, „Splošne informacije o nacionalnem preverjanju znanja,“ 2022. [Elektronski]. Dostopno: <https://www.ric.si/nacionalno-preverjanje-znanja/splosne-informacije/>. [Poskus dostopa 3. 3. 2024].
- [17] J. de Winter, „Can ChatGPT pass high school exams on English Language Comprehension?,“ *International Journal of Artificial Intelligence in Education*, 2023.
- [18] N. K. Güler in Z. G. Dertli, „An artificial intelligence application in mathematics education: Evaluating ChatGPT's academic achievement in a mathematics exam,“ *Pedagogical Research*, Izv. 9, št. 2, 2024.
- [19] Državni izpitni center, „Poročila, analize, raziskave nacionalnega preverjanja znanja,“ 2023. [Elektronski]. Dostopno: <https://www.ric.si/nacionalno-preverjanje-znanja/porocila--analize--raziskave/>. [Poskus dostopa 1. marec 2024].
- [20] Državni izpitni center, „Splošne informacije o nacionalnem preverjanju znanja,“ 2024. [Elektronski]. Dostopno: <https://www.ric.si/nacionalno-preverjanje-znanja/splosne-informacije/>. [Poskus dostopa 1. 3. 2024].
- [21] Državni izpitni center, „Navodila za vrednotenje NPZ-ja pri slovenščini v 9. razredu,“ 4. 5. 2023. [Elektronski]. Dostopno: <https://www.ric.si/mma/N231-101-3-2.pdf/2023061822173800/?m=1687119458>. [Poskus dostopa 2. 3. 2024].
- [22] Državni izpitni center, „Letno poročilo o izvedbi v šolskem letu 2022/2023,“ 7. 12. 2023. [Elektronski]. Dostopno:

https://www.ric.si/mma/Letno_poro__ilo_NPZ_2023_za_objavo.pdf/2023120710032739/?m=1702372117. [Poskus dostopa 2. 3. 2024].

- [23] Državni izpitni center, „Letno poročilo o izvedbi v šolskem letu 2021/2022,“ 14. 11. 2022. [Elektronski]. Dostopno: https://www.ric.si/mma/Letno_poro__ilo_NPZ_2022_za_objavo.pdf/2023033014090546/?m=1680178145. [Poskus dostopa 3. 3. 2024].
- [24] Državni izpitni center, „Navodila za vrednotenje NPZ-ja pri slovenščini v 9. razredu,“ 14. 11. 2022. [Elektronski]. Dostopno: <https://www.ric.si/mma/N221-101-3-2.pdf/2022061322452094/?m=1655153120>. [Poskus dostopa 2. 3. 2024].
- [25] Državni izpitni center, „Navodila za vrednotenje NPZ-ja pri slovenščini v 9. razredu,“ 8. 11. 2021. [Elektronski]. Dostopno: <https://www.ric.si/mma/n211-101-3-2-pdf/2021061414180366/?m=1623673083>. [Poskus dostopa 3. 3. 2024].
- [26] Državni izpitni center, „Letno poročilo o izvedbi v šolskem letu 2020/2021,“ 8. 11. 2021. [Elektronski]. Dostopno: https://www.ric.si/mma/Letno_poro__ilo_NPZ_2021_Objava.pdf/20220221141506/?m=1645449307. [Poskus dostopa 3. 3. 2024].