



This work is licenced under a Creative Commons Attribution 4.0 International License
Ovaj rad dostupan je za upotrebu pod licencom Creative Commons Imenovanje 4.0. međunarodna

	UDK: 811.112.2'243	
PRETHODNO	004.8:373.3.016	Primljen: 25. 11. 2024.
PRIOPĆENJE	811.112.2(075)	Prihvaćen: 22. 1. 2025.
	DOL: https://doi.org/10.22210/strjez/53-2/4	

Primjena ChatGPT-ja u izradi zadataka razumijevanja tekstova za čitanje i slušanje

Lukas Paun

lukas.paun@gmail.com

OŠ Julija Kempfa, Požega

Modeli umjetne inteligencije kao ChatGPT sve su prisutniji u obrazovnom sustavu. Istraživanja su pokazala da takvi modeli s jedne strane pružaju mnoge mogućnosti kao što su izrada priprema za nastavu, izvedbenih planova za kolegije i sl., a s druge strane krše prava privatnosti, diskriminiraju, pojednostavljaju varanje i sl. U ovom će se radu istražiti može li se koristiti ChatGPT-jem za izradu zadataka za provjeru čitanja i slušanja s razumijevanjem u nastavi Njemačkoga jezika. Istraživanje je provedeno na uzorku učenika osnovnoškolske dobi. Jedni su zadatci proizvedeni ChatGPT-jem, a druge su zadatke priredile autorice udžbenika i sastavni su dio udžbenika. Istraživanjem se željelo utvrditi hoće li učenici imati bolje rezultate na zadacima koji su proizvedeni s pomoću jezičnoga modela ili na zadacima koje su sastavile autorice udžbeničkih tekstova. Rezultati su pokazali da su učenici imali više točnih odgovora na zadacima uz tekst za čitanje koje je proizveo ChatGPT te da su imali više točnih odgovora na zadatku uz tekst za slušanje koji su izradile autorice udžbenika. Rezultati upućuju na to da zadatci proizvedeni ChatGPT-jem mogu biti alternativa zadacima koji se nalaze u udžbenicima, posebice kada se dodatno želi uvježbati razumijevanje teksta na stranom jeziku.

Ključne riječi: *ChatGPT, njemački kao strani jezik, razumijevanje pročitanoga teksta, razumijevanje slušanoga teksta, udžbenik, veliki jezični modeli*

1. UVOD

S pojavom globalne pandemije SARS-CoV-2 i zatvaranjem obrazovnih ustanova u Republici Hrvatskoj te uvođenjem poučavanja na daljinu obrazovni je sustav pokazao pojačan interes za uporabu digitalnih alata u nastavnom procesu. Novija su istraživanja potvrdila da su dionici obrazovnog sustava naklonjeni uporabi informacijsko-komunikacijskih tehnologija u poučavanju te da će se određenim aplikacijama koristiti i nakon završetka mjera koje su ograničavale održavanje nastave u obrazovnim ustanovama (Ivić i Blagojević, 2022; Krnjus, 2022; Dadić Fruk, 2023). Jedna je od tih aplikacija jezični model ChatGPT koji se pojavio u svibnju 2022. Ovaj model pruža nove mogućnosti korištenja umjetne inteligencije (Haque i sur., 2022; Gašević i sur., 2023), ne samo u obrazovanju već i mnogim drugim granama gospodarstva.

ChatGPT je veliki jezični model (engl. *large language model*) koji je razvila tvrtka OpenAI. Dostupan je putem interneta i svatko se njime može koristiti nakon prethodne prijave. ChatGPT se temelji na takozvanoj arhitekturi generativnoga predtreniranog transformera - GPT (engl. *generative pretrained transformer*). Ovi modeli primjenjuju posebne tehnike učenja kako bi proizveli tekstove tako što previđaju sljedeću riječ na temelju svih prethodnih riječi. Trenirani su na iznimno velikom broju skupa podataka sačinjenoga od tekstova te mogu izvršavati određene jezične zadatke kao što su proizvodnja prirodnoga jezika, klasifikacija tekstova, strojno prevođenje i dr. (Ray, 2023; Roumeliotis i Tselikas, 2023).

Prvi je model generativnoga predtreniranoga transformera GPT-1 razvijen 2015. On je predstavljao prvi jezični model koji se temeljio na arhitekturi transformera. Predtreniran je na velikom korpusu sazdanom od neobjavljenih romana različitih žanrova zvanom BooksCorpus (Radford i sur., 2018). Predtrenoingom je model usvojio jezične obrasce i veze između riječi. Potom je dodatno prilagođen (engl. *fine-tuned*) na zadacima poput strojnoga prevođenja, analize sentimenta ili klasifikacije tekstova. Ovaj je relativno malen model imao samo 117 milijuna parametara, tj. ključnih postavki koje određuju kako će model reagirati na upute (npr. slučajnost odabira sljedeće riječi). Već je sljedeći model GPT-2 iz 2019. godine imao 1,5 milijardi parametara i predstavljao je jedan od najvećih velikih jezičnih modela do tada. Za razliku od GPT-1 treniran je na internetskim tekstovima s Reddita označenima s najmanje tri preporuke (Radford i sur., 2019). Ovaj je model mogao proizvesti duže i koherentnije tekstove koji su se teško razlikovali od onih koje bi sročio čovjek. Sljedeći je model GPT-3 predstavljen

2020. godine. Sadrži 175 milijardi parametara, što ga čini sto puta većim od prethodnika. Za razliku od prethodna dva modela ovaj je model treniran kako na tekstovima iz knjiga, tako i na tekstovima s interneta. Ovaj je model, međutim, pokazao dva nedostatka: (i) nije treniran na podacima koji su uključivali softverski kôd i nije mogao rješavati matematičke zadatke te (ii) ponekada nije mogao slijediti upute i proizveo je tekstove nepoželjnoga i neprikladnoga sadržaja. Model je, naime, treniran na tekstovima koji su sadržavali dezinformacije, zlouporabu, mržnju i sl., što se ogledalo i u odgovorima koje je davao korisnicima. Taj se problem riješio uvođenjem modela GPT-3.5. Drugim riječima, model GPT-3 mogao je proizvesti tekstove prirodnoga jezika, a GPT-3.5 uz navedeno, mogao je proizvesti i softverski kôd. Uz GPT-3.5 prvi je put predstavljeno sučelje ChatGPT, čime se htjelo omogućiti pristup širokom krugu korisnika, a ne samo ovlaštenim informatičkim stručnjacima. Posljednji model GPT-4 iz 2023. godine uz tekst može obrađivati i slike (Kalyan, 2023; Šušnjak, 2022; Ray, 2023; Roumeliotis i Tselikas, 2023).

Kako ističe Zhai (2022), ChatGPT može odgovarati na naknadna pitanja, priznati greške, ispraviti netočne pretpostavke i odbijati neprimjerene upute. Međutim, za razliku od uobičajenih internetskih tražilica kao što su Google i Bing, ChatGPT ne pretražuje internet. Štoviše, ChatGPT raspolaze informacijama samo do 2021. godine, a plaćena verzija ChatGPT Plus do travnja 2024¹.

Do sada se točnost modela iz skupine GPT istraživala na sljedećim zadacima: rješavanje stručnih ispita u medicini (Kasai i sur., 2023), pravosudnoga ispita u SAD-u (Bommarito i sur., 2022), ispita iz pojedinih kolegija na prijediplomskom studiju iz programiranja (Đambić, 2023; Radovan 2023), za strojno prevođenje (Klimova i sur., 2024) i mnoge druge zadatke (za pregled vidi Kalyan, 2023). U nekim se istraživanjima poput onoga koje su proveli Yeo i sur. (2023) pokazalo da ChatGPT ima slabije rezultate od GPT-4, a u drugima da jezik na kojem se postavljaju upute značajno utječe na to kakav će odgovor dati model (Ando i sur., 2024, Robinson i sur., 2023). Ta je pak činjenica utjecala i na to kakav će odnos pojedini ispitanici imati prema takvim modelima. Haque i sur. (2022) su, primjerice, u svojem istraživanju sentimenta prema ChatGPT-ju analizirali objave na internet-skoj društvenoj mreži X (negdašnji Twitter). Ustanovili su da rani korisnici ovoga modela imaju uglavnom pozitivan stav prema tom modelu, ali da su

¹ Ova informacija vrijedi u trenutku sastavljanja ovoga teksta i može se promijeniti.

svjesni i etičkih problema. Pokazalo se da korisnici imaju najveće dvojbe kod primjene ChatGPT-ja u obrazovnom sustavu. Ističu moguće poteškoće s kojima će se susresti učitelji, nastavnici i profesori, a tiču se ocjenjivanja i vrednovanja pismenih uradaka učenika i studenata, prepisivanja, akademske čestitosti i integriteta, ali i toga da će se previše osloniti na model te će tako sve manje moći kritički razmišljati (Klimova i sur., 2024; Šušnjak, 2022; Brown i sur., 2020; Tlili i sur., 2023; Kasneci i sur., 2023).

Kako je pristup ChatGPT-ju omogućen tek prije dvije godine, ne postoji mnogo radova koji se bave njegovom primjenom u nastavi. Malobrojni radovi na koje smo naišli uglavnom obrađuju pozitivne i negativne implikacije primjene modela u nastavi (npr. Kasneci i sur., 2023; Baidoo-Anu i Ansah, 2023). Stoga ćemo u ovom radu istražiti primjenu modela ChatGPT u učenju njemačkoga jezika na osnovnoškolskoj razini.

U sljedećem dijelu rada osvrnut ćemo se na dosadašnja istraživanja koja su se koristila ChatGPT-jem u obrazovanju. Također ćemo kratko opisati koji su rizici i izazovi uporabe ovoga modela. Potom će se prikazati metodologija istraživanja i njegovi rezultati. Rad ćemo zaključiti raspravom te ujedno ukazati na moguća ograničenja istraživanja.

2. CHATGPT U OBRAZOVANJU

Mnogi radovi koji se bave ChatGPT-jem u nastavi su teorijske prirode. Drugim riječima, iznose se mogućnosti i načini uporabe modela, ali bez empirijskoga istraživanja. Primjerice, Kasneci i sur. (2023) utvrdili su da ChatGPT pruža različite mogućnosti učenicima i učiteljima. Istaknuli su da ovaj računalni model kod osnovnoškolskih učenika može razviti jezične vještine čitanja i pisanja tako što im može predložiti sintaktičke strukture ili ispravljati gramatičke greške. Za srednjoškolske učenike model bi mogao oblikovati kvizove i praktične problemske zadatke koji će im olakšati shvaćanje i predočiti svrsishodnost gradiva koje trenutačno uče. Studenti bi se ChatGPT-jem mogli koristiti kako bi razvili vještine za rješavanje problema, a model bi im istodobno ukazao na pojedine teme čije neriješene ili dvojbene aspekte treba istražiti. ChatGPT-jem se također mogu koristiti učitelji i nastavnici tako što bi mogli učenicima dati povratnu informaciju vezano uz njihove pismene uratke, nadalje za izradu priprema za nastavu, za poučavanje jezičnih zakonitosti, pripremu za istraživački rad i pisanje te profesionalni razvoj učenika. Na taj bi se način smanjilo radno opterećenje učitelja i više bi se vremena moglo posvetiti učenicima i njihovim potrebama (Tlili i sur., 2023; Berg i Plessis, 2023; Farrokhnia i sur., 2024).

Malobrojni istraživači koji su do sada proučavali uporabu ChatGPT-ja u nastavi uglavnom su istraživali vlastito obrazovno okružje. Kako su autori većinom sveučilišni nastavnici, istraživali su, prema tome, moguću primjenu modela u visokoškolskom obrazovanju. Primjerice, Kostikova i sur. (2024) upotrijebili su ChatGPT kako bi s pomoću njega izradili izvedbeni plan za sveučilišni kolegij engleskoga kao jezika struke. Nadalje, Berg i Plessis (2023) proučavali su uporabu ChatGPT-ja prilikom izrade priprema za nastavni sat. Postavili su modelu zadatak izrade pripreme za sat engleskoga kao stranoga jezika u šestom razredu osnovne škole. Tema sata bila je uporaba prijedloga mjesta i vremena. Zatražili su od modela da izradi prezentaciju i pripadajuće radne listiće. Zaključili su da je priprema jasna i dobro može poslužiti učitelju početniku koji se još nije snašao u tom zanimanju. Međutim, iskusniji bi učitelji pripremu mogli doraditi ovisno o potrebama svojih učenika. Berg i Plessis (2023) također navode kako je uloga učitelja ključna jer su odgovorni za odabir nastavnoga materijala koji je prikladan za određeno kulturno okružje.

Dok su Berg i Plessis (2023) samo teorijski procjenjivali kakvoću pripreme, Karaman i Göksu (2024) u nastavi su s učenicima trećega razreda osnovne škole primijenili pripreme koje je proizveo model umjetne inteligencije. S pomoću modela ChatGPT-4 izradili su nastavne pripreme za gradivo iz matematike. Zaključili su da su pripreme koje je proizveo model bile jednako učinkovite kao i one koje je osmislio učitelj što učitelju može skratiti vrijeme pisanja priprema za nastavu.

U svojem je preglednom radu o ulozi ChatGPT-ja u nastavi engleskoga kao stranoga jezika Balci (2024) zaključio da su se istraživanja koja je analizirao uglavnom usredotočila na jezičnu vještinu pisanja. Ističe da ChatGPT ima pozitivne učinke na koheziju i koherentnost pisanih uradaka učenika, pruža im podršku u učenju gramatike, povećava vokabular na stranom jeziku, daje neposrednu povratnu informaciju vezanu uz pisani uradak prosljeđen jezičnom modelu i sl. Istodobno ističe i negativne strane koje ovaj model ima, a to su pretjerano oslanjanje na njega, što može imati za posljedicu manjak kreativnosti, ali i to da se više ne uzimaju u obzir uobičajeni izvori znanja. Athanassopoulos i sur. (2023) istražili su kako se ChatGPT može primijeniti za poboljšanje pisanja u višejezičnom i višekulturnom okružju. Naime, domaćim učenicima iz Grčke te izbjeglicama i migrantima pokušali su s pomoću modela ispraviti najčešće gramatičke i semantičke greške u zadatku pisanja elektroničke poruke na njemačkom jeziku upućenu prijatelju iz Njemačke. Pokazalo se da intervencijom modela učenici nisu uspjeli samo povećati broj jedinstvenih riječi u elektroničkoj poruci već i broj ukupnih riječi u poruci.

Istraživači i korisnici ChatGPT-ja već su rano uočili problem nedostatka akademske čestitosti i poštenja u pisanju pisanih radova. Tlili i sur. (2023) smatraju, primjerice, da modeli poput ChatGPT-ja mogu poticati učenike na plagiranje radova i varanje na ispitima, osobito kod onih pisanih uradaka koje trebaju samostalno pripremiti kod kuće.

Kasneci i sur. (2023) uočili su da ChatGPT može biti pristran zbog toga što je model uglavnom treniran na tekstovima engleskoga govornog područja. To će dovesti govornike drugih jezika u neravnopravan položaj u odnosu na govornike engleskoga jezika. Istraživanjima je također potvrđeno da upute ChatGPT-ju na ne-engleskom jeziku polučuju slabije rezultate od onih koje su dane na engleskom jeziku (Yeo i sur., 2023; Ando i sur., 2024). Robinson i sur. (2023) ustanovili su da su klasični modeli strojnoga prevođenja za jezike s malom zastupljenošću na internetu (engl. *low-resource languages*) bolji od GPT-ja, dok su kod jezika s velikom zastupljenošću na internetu (engl. *high-resource languages*) modeli GPT-ja i strojnoga prevođenja podjednako dobri.

U nastavku rada obrazložit će se cilj koji smo imali istražujući ovo područje. Također će se navesti dva istraživačka pitanja na koja smo ovim radom pokušali odgovoriti. Nakon toga će se opisivati metodologija rada.

3. CILJ I ISTRAŽIVAČKA PITANJA

Svrha je ovoga rada opisati kako se modelom ChatGPT može koristiti u nastavi ne bi li se proizveli zadatci za provjeru razumijevanja tekstova i time rasteretilo učitelje i nastavnike u pripremi zadataka i drugih dodatnih nastavnih materijala.

U ovom smo se radu vodili dvama istraživačkim pitanjima:

1. Hoće li učenici nakon čitanja udžbeničkoga teksta biti uspješniji na popratnim zadacima čitanja s razumijevanjem u udžbeniku ili na zadacima koje je na osnovi teksta proizveo ChatGPT?
2. Hoće li učenici nakon slušanja zvučnoga zapisa koji prati udžbenik biti uspješniji na popratnim zadacima slušanja s razumijevanjem u udžbeniku ili na zadacima koje je na osnovi slušnoga teksta proizveo ChatGPT?

Kako bismo odgovorili na ova dva istraživačka pitanja, proveli smo istraživanje koje ćemo u nastavku detaljnije opisati.

4. METODOLOGIJA

4.1. Ispitanici

U istraživanju je sudjelovalo 24 ispitanika, odnosno učenika (14 ženskih i 10 muških) osmoga razreda jedne osnovne škole na istoku Republike Hrvatske. Dobivena je privola za sudjelovanje učenika u istraživanju. Učenici su zadatke rješavali anonimno i njihovo se sudjelovanje nije vrednovalo. Većina je učenika u dobi od 14 godina. Učenici su u poznavanju njemačkog jezika uglavnom na razini A1 prema ZEROJ-u (Vijeće Europe, 2001). Učenici su učili njemački jezik od prvoga razreda, a trenutačno pohađaju osmi razred. U prvim četirima razredima učenici su imali tjedno dva sata nastave njemačkog jezika, a od petoga do osmoga razreda tri sata tjedno. Prvi im je jezik hrvatski, a treći im je jezik engleski. Učenici su počeli učiti njemački uglavnom u dobi od šest godina kada su i pošli u prvi razred osnovne škole. Učenici su također bar jedan dio svojega školovanja učili i engleski jezik kao J3 u okviru izborne nastave. U upitniku su istaknuli da uče strani jezik u školi i kod kuće. Učenici nisu pohađali dodatnu nastavu stranoga jezika u školi stranih jezika kao izvanškolsku aktivnost. To znači da su učenici formalno izloženi njemačkom jeziku samo u školi.

Učenici su upitnikom LEAP-Q (Marian i sur., 2007) samoprocjenjivali svoje znanje iz triju jezičnih vještina Likertovom ljestvicom od pet stupnjeva. Tom se ljestvicom mjeri stav ispitanika prema nekom objektu stava. U ovom je istraživanju odabrana ljestvica s pet stupnjeva jer zrcali sustav ocjenjivanja u Hrvatskoj na koji su se učenici već naviknuli. Najmanja vrijednost na ljestvici predstavila je vrlo slabo poznavanje određene jezične vještine, a najveća vrijednost iznimno dobro poznavanje te jezične vještine. Samoprocjenu znanja može se vidjeti u Tablici 1.

Tablica 1.
Samoprocjena znanja njemačkoga J2 u jezičnim vještinama slušanja, govorenja i čitanja. U zagradi su standardne devijacije.

Table 1.
Listening, speaking and reading self-assessed proficiency in German L2. Standard deviations are in brackets.

Jezične vještine	Vrijednosti
slušanje	3,083 (1,176)
govorenje	2,583 (1,139)
čitanje	2,833 (1,465)

Iz Tablice 1 može se vidjeti kako su se učenici procijenili najboljima u vještini slušanja, dok su u govorenju procijenili da su najslabiji. Prema Marian

i sur. (2007) upravo procjena u jezičnoj vještini slušanja najbolje ukazuje na njihovu kompetenciju na stranom jeziku. Nadalje, učenici su s pomoću Likertove ljestvice od pet stupnjeva procjenjivali što je najviše utjecalo na njihovo znanje njemačkoga te u kojim su situacijama najviše izloženi njemačkom jeziku (v. Tablicu 2).

Tablica 2.

Samoprocjena čimbenika koji su utjecali na učeničko znanje njemačkoga J2 te gdje su učenici najviše izloženi njemačkom J2.

Table 2.

Self-assessment of the factors that influenced the students knowledge of German L2 and where students are most exposed to German L2.

	Interakcija s prijateljima i obitelji	Slušanje glazbe i radija	Gledanje televizije, filmova i serija	Čitanje	Samostalno učenje
Utjecaj njemačkoga	1,500 (,722)	1,292 (,690)	1,917 (1,176)	2,125 (1,116)	2,542 (1,444)
Izloženost njemačkom	1,708 (,907)	1,708 (1,160)	2,042 (1,301)	2,042 (1,042)	2,500 (1,383)

Iz Tablice 2 može se vidjeti da je na učeničko znanje njemačkoga jezika najviše utjecalo samostalno učenje. Pokazalo se, nadalje, da su učenici samostalnim učenjem najviše izloženi njemačkom jeziku. Najmanje je na poznavanje njemačkoga jezika utjecalo slušanje glazbe i radija, a i učenici su procijenili da su u toj kategoriji kao i u kategoriji interakcije s prijateljima i članovima obitelji najmanje izloženi njemačkom jeziku.

4.2. Materijal i instrumenti

Za potrebe ovoga istraživanja odabran je po jedan tekst za čitanje i jedan za slušanje iz udžbenika za sedmi razred osnovne škole koje je odobrilo Ministarstvo znanosti, obrazovanja i mladih Republike Hrvatske (Troha i Valjak Ilić, 2021). Odabrani su tekstovi namijenjeni dodatnom učenju i produbljivanju znanja iz četiriju jezičnih vještina. Ti su tekstovi odabrani zato što bi po razini zahtjevnosti trebali biti razumljivi učenicima i zato što se ne nalaze u trenutačnim udžbenicima kojima se koriste. Tema teksta za čitanje bio je neandertalac Ötzi koji je 1991. godine pronađen u južnom Tirolu, a tema teksta za slušanje bila su iskustva dvoje učenika glede druženja s prijateljima. Svaki je od dvaju tekstova imao prateće zadatke u udžbeniku i pripadajućoj radnoj bilježnici (Troha i Valjak Ilić, 2021 i 2023).

Odabrana su dva zadatka koja prate tekst za čitanje. U jednom su zadatku učenici trebali pročitati tvrdnje za provjeru razumijevanja i odrediti jesu li točne ili ne, tj. odrediti odgovaraju li sadržaju teksta ili ne. U drugom su

zadatku trebali povezati početak i kraj rečenice tako da budu smisleni s obzirom na sadržaj pročitanoga teksta. Isto su tako odabrana dva zadatka koja prate tekst za slušanje. U prvom su zadatku učenici trebali odrediti jesu li dane tvrdnje točne ili ne, a u drugom su zadatku trebali odrediti na koju se osobu odnosi određena tvrdnja, tj. trebali su odrediti je li tu tvrdnju izrekla djevojčica ili dječak koji su govorili o svojim iskustvima.

Tekstovi su za čitanje i slušanje zatim digitalizirani kako bi ih se moglo prenijeti u model ChatGPT. Budući da je riječ o tekstu na njemačkom jeziku, upute su također bile na njemačkom jeziku kako ne bi došlo do mijenjanja jezika. Povelj smo se primjerom Rüdiana i Pinkwarta (2023) koji su učinili isto. Svaki je zadatak oblikovan s pomoću modela ChatGPT-ja imao jednak broj rečenica kao i izvorni zadatak iz udžbenika. Modelu su potom dane upute: Treba se postaviti u položaj da je učitelj njemačkoga jezika koji predaje učenicima sa slabo razvijenom vještinom čitanja tekstova napisanih na njemačkom jeziku. Za prvi je zadatak model trebao oblikovati osam tvrdnji, a učenici su trebali odrediti jesu li točne ili ne, a za drugi zadatak model je trebao razdvojiti šest rečenica koje su učenici potom ponovno povezivali. Slično kao u istraživanju Rüdiana i Pinkwarta (2023), pri čemu je model oblikovao rečenice suprotne uputama, tako je i ovdje model u prvom zadatku oblikovao rečenice s glagolskim oblicima koji se ne uče u osnovnoj školi (pasiv, preterit), iako je bilo navedeno da je trebao oblikovati rečenice na razini A1 prema ZEROJ-u. Modelu se trebalo stoga izričito napomenuti koja je rečenica neprikladna i što u njoj treba promijeniti. U drugom zadatku, u kojem su učenici trebali povezati početak i kraj rečenice, model nije imao poteškoća s oblikovanjem rečenica.

Zatim se prešlo na oblikovanje zadataka koji prate tekst za slušanje. Stoga je započeta nova sesija s modelom ne bi li se na taj način izbjegao mogući negativan utjecaj prethodnoga razgovora s modelom vezan uz oblikovanje zadataka uz tekst za čitanje, a istodobno se htjelo spriječiti da se sadržaj dvaju tekstova pomiješa. Prilikom oblikovanja zadataka koji su pratili tekst za slušanje, postupilo se jednako kao i kod zadataka uz tekst za čitanje. Za prvi smo zadatak zamolili model da proizvede devet tvrdnji za provjeru razumijevanja vezanih uz tekst za slušanje, za koje su učenici trebali odrediti jesu li točne ili ne. Međutim, ChatGPT nije označio točne odgovore pa ga se moralo zasebnim upitom podsjetiti na to. U drugom je zadatku model trebao oblikovati tvrdnje za provjeru razumijevanja koje su vezane za pojedinu osobu, a učenici su trebali označiti je li ju izrekla djevojčica ili dječak. Međutim, model ni nakon više od deset pokušaja nije mogao proizvesti zadovoljavajući zadatak. Naime, rečenice su često bile dvoznačne i mogle

su se pripisati obama govornicima. Budući da model nije bio u stanju proizvesti rečenice na zadovoljavajući način, u daljnjem razmatranju i analizi drugi zadatak uz tekst za slušanje nije se više uzimao u obzir.

Kako bi procijenili svoje jezično iskustvo i kompetenciju, učenici su ispunili anketni upitnik LEAP-Q (Marian i sur., 2007).

4.3. Postupak

Istraživanje je provedeno tijekom dva sata njemačkoga jezika u školskoj godini 2023./2024. Morali smo uključiti dovoljan vremenski razmak kako bi se što više izbjegao negativan utjecaj poznavanja teksta na rezultate koje će učenici postići rješavanjem ponuđenih zadataka. Prvi je sat tako održan početkom listopada 2023. godine. Učenici su tada rješavali zadatke uz tekst za čitanje i slušanje koje su pripremile autorice udžbenika, a krajem svibnja 2024. rješavali su zadatke koje je proizveo ChatGPT. U istraživanju čitanja s razumijevanjem koje su proveli Tibken i sur. (2022) na uzroku ispitanika otprilike jednake dobi kao i ispitanici u ovom istraživanju između dvaju trenutaka provođenja ispitivanja prošlo je osam mjeseci. Stoga smo se odlučili za jednak vremenski razmak između dvaju trenutaka provjere razumijevanja teksta za čitanje i slušanje.

Tijekom oba sata postupak je bio isti. Učenici su prvo trebali pročitati tekst za čitanje vezan uz pronalazak neandertalca u Alpama, nakon čega su dobili listiće sa zadacima. Kada su bili gotovi s jednim zadatkom, dobili su listić s drugim zadatkom vezanim uz tekst za čitanje. Pošto su učenici završili s drugim zadatkom vezanim uz tekst za čitanje, učitelj je skupio listiće sa zadacima i podijelio nove listiće sa zadacima koji prate tekst za slušanje. Prije nego što su poslušali tekst za slušanje, učenici su pročitali tvrdnje za provjeru razumijevanja slušanjem na listiću sa zadacima. Učenici su dva puta poslušali tekst za slušanje. Nakon toga su rješavali zadatak. Budući da jezični model nikako nije mogao proizvesti zadovoljavajući drugi zadatak uz tekst za slušanje, tijekom sata u svibnju riješili su samo prvi zadatak dok su tijekom sata u listopadu riješili oba zadatka.

Nakon što su učenici završili s rješavanjem zadataka u svibnju, učenici su ispunjavali anketni upitnik o svojem jezičnom iskustvu i kompetenciji LEAP-Q (Marian i sur., 2007).

5. REZULTATI

U Tablici 3 prikazani su rezultati deskriptivne statistike za točnost odgovora s obzirom na podrijetlo zadataka (udžbenik odnosno ChatGPT) i je-

zičnu vještinu (čitanje odnosno slušanje). Prikazane su aritmetičke sredine i standardne devijacije. Svi su podatci obrađeni u programskom okružju za statističke izračune *R* i to verzijom 4.0.2 (R Core Team, 2020).

Tablica 3.
Mjere deskriptivne statistike za točnost odgovora s obzirom na podrijetlo zadataka i jezičnu vještinu (N = 24)

Table 3.
Descriptive statistics measures on accuracy of responses regarding task origin and language skill (N = 24)

Podrijetlo zadataka	Jezična vještina	Aritmetička sredina (M)	Standardna devijacija (SD)
Udžbenik (udžbenički zadatci)	Čitanje s razumijevanjem	7,042	2,493
	Slušanje s razumijevanjem	5,792	1,103
ChatGPT (jezičnim modelom proizvedeni zadatci)	Čitanje s razumijevanjem	9,583	2,466
	Slušanje s razumijevanjem	5,125	1,393

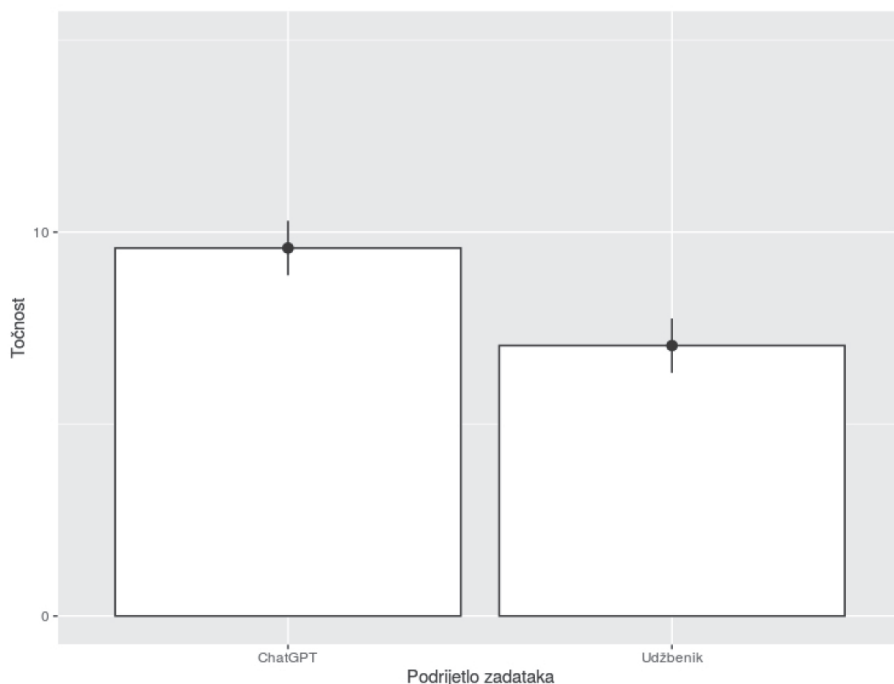
Kako bi se provjerilo prvo istraživačko pitanje hoće li učenici imati više točnih odgovora u zadacima uz tekst za čitanje koji su proizvedeni s pomoću modela ChatGPT-ja ili u zadacima koje su formulirale autorice udžbenika, napravljen je Wilcoxonov test ekvivalentnih parova. Shapiro-Wilkovim *W* testom se, naime, pokazalo da podatci nisu normalno distribuirani ($W = ,885, p = ,010$). U testu koji je provjeravao razliku između točno riješenih odgovora u zadacima uz tekst za čitanje s razumijevanjem učenici su prosječno i medijalno imali statistički značajno više točnih odgovora u zadacima koje je proizveo ChatGPT ($M = 9,583, SD = 2,466, Mdn = 9$) nego u zadacima koje su osmislile autorice udžbenika ($M = 7,042, SD = 2,493, Mdn = 6$) $V = 27, p < ,01, r = -,631$. Slika 1 prikazuje dijagram koji pokazuje razliku točnosti riješenih odgovora u zadacima uz tekst za čitanje s razumijevanjem prema podrijetlu zadatka (udžbenik i ChatGPT). Učenici su imali prosječno 2,5 točna odgovora više u zadacima proizvedenima s pomoću jezičnoga modela nego u zadacima koje su izradile autorice osnovnoškolskoga udžbenika.

Slika 1.

Točnost u zadacima uz tekst za čitanje s razumijevanjem s obzirom na podrijetlo zadataka (ChatGPT – eksperimentalnim jezičnim modelom proizvedeni zadatci i udžbenik – kontrolni udžbenički zadatci)

Figure 1.

Accuracy on reading comprehension tasks regarding their source (ChatGPT – experimental tasks generated by the language model and textbook – textbook tasks)



Kako bi se provjerilo drugo istraživačko pitanje hoće li učenici imati više točnih odgovora u zadatku uz tekst za slušanje koji je proizveden s pomoću modela ChatGPT-ja ili u zadatku koji su formulirale autorice udžbenika, proveden je Wilcoxonov test ekvivalentnih parova. Unatoč tome što je ova puta Shapiro-Wilkov W test pokazao da podatci jesu normalno distribuirani ($W = ,951, p = ,292$), pribjegli smo navedenom testu zbog maloga broja ispitanika. U testu koji je provjeravao razliku između broja točno riješenih odgovora u zadatku uz tekst za slušanje s razumijevanjem učenici su prosječno i medijalno imali statistički značajno više točnih odgovora u zadatku koji su oblikovale autorice udžbenika ($M = 5,792, SD = 1,103, Mdn = 6$) nego u zadatku koji je proizveo jezični model ($M = 5,125, SD = 1,393, Mdn = 5$) $V = 120,5, p = ,033, r = -,435$. Slika 2 prikazuje dijagram razlike točnosti riješenih odgovora u zadatku uz tekst za slušanje s razumijevanjem prema podrijetlu zadatka (udžbenik i ChatGPT). Učenici su imali prosječno ,667 više točnih odgovora u zadatku koji su izradile autorice udžbenika nego u

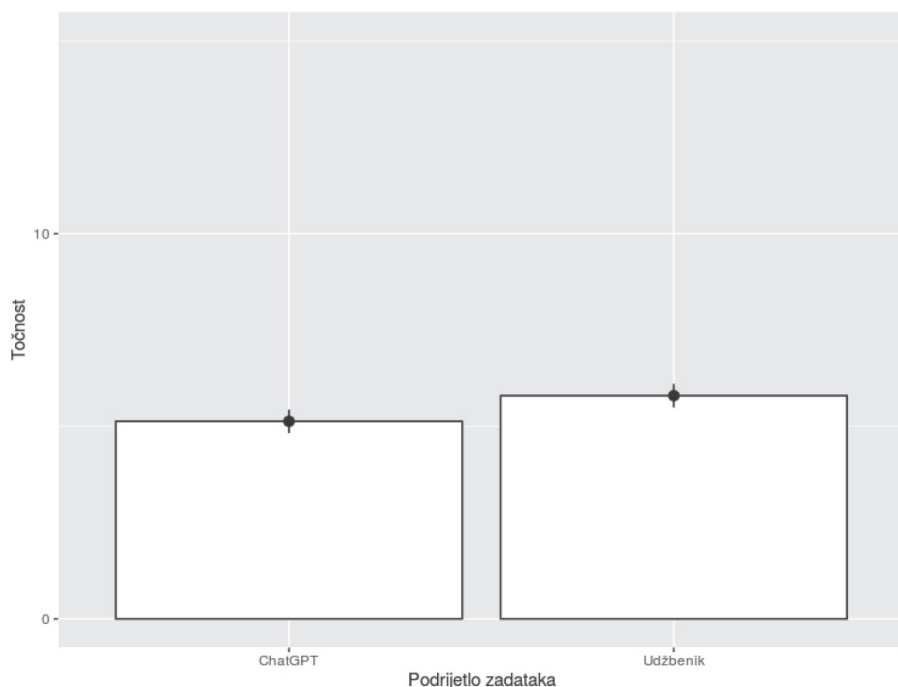
zadatku koji je proizveo s pomoću modela. Unatoč tomu što je razlika mala, ona je ipak statistički značajna, kako pokazuju rezultati Wilcoxonova testa ekvivalentnih parova.

Slika 2.

Točnost u zadacima uz tekst za slušanje s razumijevanjem s obzirom na podrijetlo zadataka (ChatGPT – eksperimentalnim jezičnim modelom proizvedeni zadatci i udžbenik – kontrolni udžbenički zadatci)

Figure 2.

Accuracy on listening comprehension tasks regarding their source (ChatGPT – experimental tasks generated by the language model and textbook – textbook tasks)



6. RASPRAVA

U ovom se radu željelo utvrditi koliko učinkovito učenici koji uče njemački jezik kao J2 rješavaju zadatke predložene u udžbeniku i one koje je proizveo jezični model ChatGPT. Drugačije rečeno, ovaj se rad vodio dvama istraživačkim pitanjima: prvo, hoće li učenici imati više točnih odgovora u zadacima uz tekst za čitanje koje su oblikovale autorice udžbenika ili koje je proizveo model umjetne inteligencije te drugo, u kojim će zadacima vezanima uz tekst za slušanje učenici imati više točnih odgovora? Uzimajući u obzir opisane rezultate dvaju statističkih testova, ustanovljeno je da su učenici polučili bolje rezultate u zadacima uz tekst za čitanje ako ih je proizveo

ChatGPT. Nasuprot tomu, pokazalo se da su učenici u zadatku koji prati tekst za slušanje imali statistički značajno bolje rezultate u zadatku koji su oblikovale autorice udžbenika u odnosu na onaj zadatak koji je proizveo jezični model. Rezultati ovih dvaju statističkih testova pokazuju disparitet u rezultatima zadataka uz tekst za čitanje i tekst za slušanje, a sugeriraju da model proizvodi zadatke uz tekst za čitanje u kojima su učenici bili uspješniji, a nasuprot tome model nije uspio proizvesti zadatke uz tekst za razumijevanje slušanja koji su učenici rješavali uspješnije. No budući da je poznato da na rezultate može utjecati način postavljanja uputa modelu, što bi svakako trebalo dalje istražiti, moguće je da je drugi zadatak uz tekst za slušanje mogao biti drugačije oblikovan i poboljšati rezultate. Ovaj rezultat svakako upućuje na to da model ima potencijal za primjenu u nastavi. Stoga možemo zaključiti da zadatci koje je proizveo ChatGPT mogu poslužiti kao alternativa zadacima koji prate tekst iz udžbenika. Budući da Kurikulum nastavnoga predmeta Njemački jezik za osnovne škole i gimnazije (MZO, 2019) zahtijeva da učenici mogu djelovati u stvarnim komunikacijskim situacijama, udžbenici ponekada ne pružaju dovoljan broj leksičkih sadržaja koji su primjereni dobi i razvoju učenika koje učitelj poučava. Stoga je potrebno da ih učitelj prilagođava s obzirom na dužinu, složenost i kompetenciju učenika ne bi li učenici tako mogli uvježbati globalno, selektivno i detaljno razumijevanje koje je ispitano zadacima u ovom istraživanju. Kako su učitelji k tomu sve češće zauzeti brojnim drugim obvezama pred koje ga postavlja suvremeni odgojno-obrazovni sustav (npr. sve više administrativnih obveza, zahtjevi roditelja, nadređenih i sl.), to može dovesti do stresnih situacija. Rüdian i Pinkwart (2023) dijele takvo mišljenje te ističu da se korištenjem modela umjetne inteligencije može smanjiti radno opterećenje učitelja jer ne moraju toliko vremena i kreativnosti utrošiti na osmišljavanje zadataka, priprema i sl. Međutim, prekomjerno oslanjanje na takve modele može izazvati suprotan učinak te će učitelji zapostaviti svoju kreativnost, a uz to može negativno utjecati na učiteljevu interakciju s učenicima (Kasneć i sur., 2023; Farrokhnia i sur., 2023).

Tlili i sur. (2023) utvrdili su da model ponekada na upite daje netočne ili ograničene odgovore. Na taj se način otežava učiteljev rad umjesto da ga se olakša jer učitelj tada dodatno mora provjeravati i prepravljati sadržaj tekstova, pitanja i dr. Tlili i sur. (2023) također navode da model katkada daje kontradiktorne izjave u odnosu na odgovore koje je prije dao. Slično se dogodilo prilikom proizvodnje drugoga zadatka za slušanje u ovom radu. Nakon višestrukoga postavljanja pitanja i upućivanja na popravak odgovora daljnjim uputama, model nije bio u stanju oblikovati rečenice koje

jednoznačno upućuju na osobu koja je izrekla tvrdnju navedenu u zadatku. Takvo ponašanje jezičnoga modela potvrđuju Rüdian i Pinkwart (2023) koji su zabilježili da model umjetne inteligencije nije mogao oblikovati rečenice s određenim riječima uz unaprijed definirano ograničenje razine poznavanja jezika makar je isprva tvrdio da postoji semantička povezanost među njima.

Kao u istraživanju Rüdiana i Pinkwarta (2023) i u našem se istraživanju pokazalo da jezični model ima poteškoće s oblikovanjem tvrdnji za provjeru razumijevanja koje su na početničkoj razini poznavanja jezika. Tako se, primjerice, ChatGPT prilikom oblikovanja tvrdnji za provjeru razumijevanja koristio preteritom i pasivom. Nakon upute da jezični model promijeni te rečenice, oblikovao je rečenice koje su, premda gramatički valjane, besmislene i netočne (npr. *Ötzi hat die Alpen gefunden. - Ötzi je pronašao Alpe.*). Stoga se modelu trebalo jasno ukazati na to koje rečenice treba promijeniti i u kojem ih smislu treba izmijeniti.

Tijekom proizvodnje tvrdnji za provjeru razumijevanja za pojedine zadatke zamijećeno je da je ChatGPT zaboravio navesti pripadajuće odgovore. Stoga ga se u sljedećem upitu moralo upozoriti na to. Tlili i sur. (2023) tu su pojavu povezali s time da model nije u cijelosti pouzdan što može dovesti do manipulacije i do toga da se nekim korisnicima modela daju bolji i pregledniji odgovori od drugih.

7. OGRANIČENJA

Ograničenja ovoga rada proizlaze iz toga što se u istraživanju koristilo prilično malim uzorkom ispitanika ($N = 24$). Nadalje, učenici su bili iz iste države i istoga razreda te čitali i slušali isti tekst. Stoga u sljedećim istraživanjima valja uzeti u obzir tu činjenicu. Isto tako treba proširiti istraživanja na druge jezike, druge države i druga sociokulturna okružja. Također bi valjalo ispitati kako učenici rješavaju zadatke vezane uz tekstove koji su različite tematike. Daljnje je ograničenje upotrijebljena verzija jezičnoga modela. Upotrijebljena je besplatna verzija 3.5, za razliku od verzije 4 koja se naplaćuje. Iako su Yeo i sur. (2023) pokazali da su rezultati bolji s verzijom 4, odlučili smo se za besplatnu verziju 3.5 jer smatramo da je ta verzija dostupnija učiteljima i drugim dionicima obrazovnoga sustava. Također, razmatrali smo samo jedan jezični model u ovom istraživanju. Odlučili smo se za ChatGPT jer je medijski najzastupljeniji i među učiteljima najpoznatiji. Međutim, u sljedećim bi se istraživanjima trebalo uključiti i druge modele poput Googleova *Geminia*, Microsoftova *Copilota* ili neki drugi. Isto tako

valja uzeti u obzir i način postavljanja uputa modelu jer i najmanje varijacije u postavljanju uputa mogu promijeniti ishod (tzv. *prompt engineering*). Stoga je bitno da se jezičnom modelu pruža jasan kontekst za što su namijenjeni odgovori koje će proizvesti. Isto se tako modelu može dati nekoliko primjera očekivanih rezultata. Istraživanja su pokazala da tako jezični model može proizvesti tekstove i zadatke visoke kakvoće (Lee i sur., 2023; Octavio i sur., 2024; Maity i sur., 2025). Nadalje, među ograničenja treba navesti i činjenicu kako se jedan zadatak za provjeru razumijevanja slušanjem nije analizirao jer jezični model nije proizveo tvrdnje koje su se jednoznačno mogli pridjenuti jednoj osobi.

8. ZAKLJUČAK

Može se zaključiti da je ChatGPT vrlo koristan alat za nastavu uz navedena ograničenja. Njime se mogu vrlo brzo i jednostavno proizvesti pitanja i zadatci za provjeru razumijevanja teksta za čitanje i slušanje kao što su zadatci višestrukoga ili alternativnoga odgovora, zadatci povezivanja, zadatci kratkoga odgovora i sl. (Lee i sur., 2023; Maity i sur., 2025). Stoga može poslužiti kao alternativa zadatcima koji su unaprijed postavljeni u udžbenicima. Također se njime mogu oblikovati dodatna pitanja za one učenike koji žele znati više ili s kojima treba dodatno učvrstiti gradivo. Nadalje ga se može upotrijebiti kako bi se učenike pripremilo za natjecanje tako što će se usložnjavati tekstovi iz udžbenika i na temelju njih zadati teži zadatci.

Valja istaknuti da je učitelj i dalje neophodan kako bi provjerio jesu li odgovori modela primjereni i kako bi učinio potrebne preinake. Nadalje, treba dodatno obrazovati učitelje kako bi učinkovito mogli iskoristiti mogućnosti jezičnih modela postavljanjem jasnih uputa modelu. Isto tako, jezični su modeli zauzeli svoje mjesto u digitalnom svijetu i više neće nestati, stoga ih obrazovni sustav treba prihvatiti, a njegove dionike primjereno uputiti kako se njima koristiti. Također treba u svako doba imati na umu ograničenja i rizike koje ovi modeli imaju te ništa ne treba nekritički prihvatiti (Tlili i sur., 2023).

LITERATURA

- Ando, K., Sato, M., Wakatsuki, S., Nagai, R., Chino, K., Kai, H., Sasaki, T., Kato, R., Nguyen, T. P., Guo, N. i Sultan, P. (2024). A comparative study of English and Japanese ChatGPT responses to anaesthesia-related medical questions. *BJA Open*, 10. <https://doi.org/10.1016/j.bja.2024.100296>
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K. i Komis V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural

- classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818-824. <https://doi.org/10.25082/AMLER.2023.02.009>
- Baidoo-Anu, D. i Ansah, L. O. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Balci, Ö. (2024). The Role of ChatGPT in English as a Foreign Language (EFL) Learning and Teaching: A Systematic Review. *International Journal of Current Education Studies*, 3(1), 66-82. <https://doi.org/10.5281/zenodo.12544675>
- Berg, G. v. d. i Plessis, E. d. (2023). ChatGPT and Generative AI: Possibilities for Its Contribution to Lesson Planning, Critical Thinking and Openness in Teacher Education. *Education Science*. <https://doi.org/10.3390/educsci13100998>
- Bommarito II, M. i Katz, D. M. (2022). *GPT Takes the Bar Exam*. <https://arxiv.org/abs/2212.14402> [15.11.2024.]
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Brener, C., McCandlish, S., Radford, S., Sutskever, I. i Amodei, D. (2020). Language Models are Few-Shot Learners. *Computation and Language*. <https://arxiv.org/abs/2005.14165> [15.11.2024.]
- Dadić Fruk, L. (2023). Digitalni alati korišteni u online nastavi za vrijeme pandemije virusa COVID-19. *Zbornik radova Ekonomskog fakulteta Sveučilišta u Mostaru*, 29, 105-122. doi.org/10.46458/27121097.2023.29.105
- Đambić, G. (2023). Upotreba ChatGPT-a za rješavanje ispita iz uvoda u programiranje. *Politehnika i dizajn*, 11(4), 253-259. <https://doi.org/10.19279/TVZ.PD.2023-11-4-05>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O. i Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460-474. <https://doi.org/10.1080/14703297.2023.2195846>
- Gašević, D., Siemens, G. i Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Artificial Intelligence*, 4. <https://doi.org/10.1016/j.caeai.2023.100130>
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N. i Ahmad H. (2022). "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. <https://arxiv.org/abs/2212.05856> [15.11.2024.]
- Ivić, V. i Blagojević, L. (2022). Korištenje informacijsko-komunikacijskom tehnologijom učenika i studenata prije i tijekom pandemije COVID-19. *Napredak: časopis za interdisciplinarna istraživanja u odgoju i obrazovanju*, 163(1-2), 93-113. orcid.org/0000-0002-0722-4029
- Kalyan, K. S. (2024). A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4. *Natural Language Processing Journal*, 6. <https://doi.org/10.1016/j.nlp.2023.100048>
- Karaman, M. R. i Göksu, I. (2024). Are lesson plans created by ChatGPT more effective? An experimental study. *International Journal of Technology in Education*, 7(1), 107-127. <https://doi.org/10.46328/ijte.607>
- Kasai, J., Kasai, Y., Sakaguchi, K., Yamada, Y. i Radev, D. (2023). *Evaluating GPT-4 and ChatGPT on Japanese Medical Licensing Examinations*. <https://arxiv.org/abs/2303.18027> [15. 11. 2024.]
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel,

- C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J. i Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103. <http://dx.doi.org/10.1016/j.lindif.2023.102274>
- Klimova, B., Pikhart, M. i Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2024.1269319>
- Kostikova, I., Holubnycha, L., Besarab, T., Moshynska, O., Moroz, T. i Shamaieva, I. (2024). ChatGPT for Professional English Course Development. *International Journal of Interactive Mobile Technologies*, 18(2), 68-81. <https://doi.org/10.3991/ijim.v18i02.46623>
- Krnjusi, A. (2022). *Utjecaj pandemije na uporabu digitalnih nastavnih materijala u razrednoj nastavi*. Diplomski rad. <https://urn.nsk.hr/urn:nbn:hr:189:928081>
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J. i Kim, H. (2023). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*, 29, 11483-11515. <https://doi.org/10.1007/s10639-023-12249-8>
- Marian, V., Blumenfeld, H. K. i Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940-967. [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Maity, S., Deroy, A. i Sarkar, S. (2025). Can large language models meet the challenge of generating school-level questions? *Computers and Education: Artificial Intelligence*, 8, 1-13. <https://doi.org/10.1016/j.caeai.2025.100370>
- Ministarstvo znanosti i obrazovanja (2019). *Kurikulum nastavnoga predmeta Njemački jezik za osnovne škole i gimnazije*. Ministarstvo znanosti i obrazovanja.
- Octavio, M. M., Gonz  lez Arg  ello, M. V. i Pujol  , J.-T. (2024). ChatGPT as an AI L2 teaching support: A case study of an EFL teacher. *Special Issue on Language Teaching and Learning with ChatGPT*, 6(1), 1-25. <https://doi.org/10.29140/tlt.v6n1.1142>
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Radford, A., Narasimhan, K., Salimans, T. i Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI Blog*. [15. 11. 2024.]
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. i Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI Blog*. [15. 11. 2024.]
- Radovan, A. (2023). Analiza utjecaja korištenja ChatGPT alata kod podu  avanja i savladavanja vještina programiranja na visokom u  ilištu u Republici Hrvatskoj. *FIP - Financije i pravo*, 11(2), 127-140.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Robinson, N., Ogayo, P., Mortensen, D. R. i Neubig, G. (2023). ChatGPT MT: Competitive for High- (but not Low-) Resource Languages. U: P. Koehn, B. Haddow, T. Kocmi i C. Monz (ur.), *Proceedings of the Eighth Conference on Machine Translation*. Association for Computational Linguistics, 392-418. <https://doi.org/10.18653/v1/2023.wmt-1.40>

- Roumeliotis, K. I. i Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, 15(6). <https://doi.org/10.3390/fi15060192>
- Rüdian, S. i Pinkwart, N. (2023). Auto-generated language learning online courses using generative AI models like ChatGPT. U: 21. *Fachtagung Bildungstechnologien (DELFI)*. Gesellschaft für Informatik e.V., 65-76. <https://doi.org/10.18420/delfi2023-14>
- Šušnjak, T. (2022). *ChatGPT: The End of Online Exam Integrity?* <https://arxiv.org/abs/2212.09292> [15. 11. 2024.]
- Tibken, C., Richter, T., Wannagat, W., Schmiedeler, S., Linden, N. v. d. i Schneider, W. (2022). Measuring Comprehension Monitoring with the Inconsistency Task in Adolescents: Stability, Associations with Reading Comprehension Skills, and Differences Between Grade Levels. *Discourse Processes*, 59(5–6), 439-461. <https://doi.org/10.1080/0163853X.2022.2073736>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R. i Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(15). <https://doi.org/10.1186/s40561-023-00237-x>
- Troha, J. i Valjak Ilić, I. (2021). *Gut gemacht! 7: Udžbenik njemačkog jezika u sedmom razredu osnovne škole*. Školska knjiga.
- Troha, J. i Valjak Ilić, I. (2023). *Gut gemacht! 7: Radna bilježnica za njemački jezik u sedmom razredu osnovne škole*. Školska knjiga.
- Vijeće Europe (2005). *Zajednički europski referentni okvir za jezike: učenje, poučavanje, vrednovanje*. Školska knjiga.
- Yeo, Y. H., Samaan, J. S., Ng, W. H., Xiaoyan, M., Ting, P.-S., Kwak, M.-S., Panduro, A., Lizaola-Mayo, A., Trivedi, H., Vipani, A., Ayoub, W., Yang, J. D., Liran, O., Spiegel, B. i Kuo, A. (2023). GPT-4 outperforms ChatGPT in answering non-English questions related to cirrhosis. *medRxiv*. <https://doi.org/10.1101/2023.05.04.23289482>
- Zhai, X. (2022). ChatGPT User Experience: Implications for Education. *SSRN*. <https://dx.doi.org/10.2139/ssrn.4312418>

Application of ChatGPT in creating reading and listening comprehension tasks

Lukas Paun

Artificial intelligence models such as ChatGPT are becoming increasingly present in the education system. Research has shown that such models offer many opportunities, such as the creation of teaching materials, lesson plans for various school types, etc., but also, they violate privacy, discriminate, facilitate cheating, etc. This paper investigates whether ChatGPT can be used to create tasks for assessing comprehension of reading and listening texts in classes of German as a foreign language. The study was conducted with elementary school students. One task was generated by ChatGPT, and the other tasks were prepared by the textbook authors. The texts are part of the textbook. The goal of this paper was to find out whether the students will achieve better results in the tasks generated by the ChatGPT or in the tasks created by the textbook authors. The results showed that the students gave more correct answers to the ChatGPT-generated tasks on the reading text and that they gave more correct answers in the task on the listening text created by textbook authors. This indicates that the tasks generated by ChatGPT can be an alternative to the tasks in the textbooks, especially if students should improve their understanding of particular texts.

Keywords: *ChatGPT, German as a foreign language, large language models, listening comprehension, reading comprehension, textbook*