

# UPORABNOST VELIKEGA JEZIKOVNEGA MODELA CHATGPT PRI OBRAVNAVI KLINIČNIH PRIMEROV PACIENTOV

## USEFULNESS OF THE LARGE LANGUAGE MODEL CHATGPT IN THE EVALUATION OF CLINICAL CASES

AVTORJI / AUTHORS:

asist. dr. Ana Kodrič, mag. farm.<sup>1</sup>

doc. dr. Lea Knez, mag. farm., spec.<sup>1, 2</sup>

doc. dr. Nejc Horvat, mag. farm.<sup>1</sup>

<sup>1</sup> Univerza v Ljubljani, Fakulteta za farmacijo,  
Aškerčeva cesta 7, 1000 Ljubljana

<sup>2</sup> Univerzitetna klinika za pljučne bolezni in alergijo Golnik,  
Golnik 36, 4204 Golnik

NASLOV ZA DOPISOVANJE / CORRESPONDENCE:

E-mail: ana.kodric@ffa.uni-lj.si



### POVZETEK

Na področju zdravstva, vključno s farmacijo, se pojavljajo možnosti uporabe velikih jezikovnih modelov, zato smo želeli oceniti uporabnost ChatGPT v klinični farmaciji. Pet kliničnih farmacevtk je ocenjevalo uspešnost ChatGPT (verzija GPT 3.5, marec 2023) pri obravnavi kliničnih primerov. ChatGPT se je izkazal kot uporaben pri prepoznavanju osnovnih težav, kot so interakcije med zdravili in neželeni učinki, vendar so bili odgovori včasih preveč splošni ali napačni. Pri splošnih vprašanjih je bila pravilnost odgovorov 29 %, pri specifičnih vprašanjih pa le 19 %, pri čemer je pogosto izostal kontekstualni vidik, kar je zmanjšalo uporabnost odgovorov. Klinične farmacevtke so odgovore ocenile kot jasne in jezikovno ustrezne, vendar so opozorile na pomanjkanje celovitosti in argumentacije. Mnenja o uporabi umetne inteligence v klinični praksi so bila deljena, saj so nekatere klinične farmacevtke prepoznale njen potencial pri podpori odločanju, druge pa so izpostavile omejeno pravilnost in pomanjkanje strokovne presoje. Čeprav ChatGPT lahko podpira klinične odločitve, trenutno še ne more nadomestiti strokovne presoje farmacevta.

### KLJUČNE BESEDE:

ChatGPT, klinična farmacija, pregled zdravljenja z zdravili, umetna inteligenca, veliki jezikovni modeli

### ABSTRACT

Opportunities for using large language models are emerging in healthcare, including pharmacy. To evaluate ChatGPT's (version GPT 3.5, march 2023) usefulness in clinical pharmacy, five clinical pharmacists assessed its performance on clinical cases. ChatGPT demonstrated utility in identifying basic issues, such as drug interactions and adverse effects, but its responses were often overly general or incorrect. Answer accuracy was 29% for general questions and only 19% for specific ones, with frequent lack of context reducing their practical value. While pharmacists found the responses linguistically clear, they criticised the lack of comprehensiveness and reasoning. Opinions on using artificial intelligence in clinical practice were mixed; some recognised its potential for supporting decision-making, while others emphasised its limited accuracy and



lack of professional judgement. Although ChatGPT can assist in clinical decisions, it currently cannot replace the professional judgement of a pharmacist.

#### KEY WORDS:

artificial intelligence, ChatGPT, clinical pharmacy, large language models, medication review

# 1 UVOD

Umetna inteligenca (UI) je po Slovarju slovenskega knjižnega jezika opredeljena kot »sposobnost stroja, računalnika, da rešuje umske probleme« (1). Sistemi UI vključujejo dele človeške inteligence, kot so sklepanje, reševanje problemov, načrtovanje, učenje, delovanje, odzivanje, razumevanje in ustvarjanje jezika (2). UI je obsežno in dinamično področje, ki vključuje različne discipline, med katerimi sta strojno in globoko učenje. UI temelji na raznovrstnih tehnikah in metodah, katerih cilj je razvoj inteligentnih sistemov, ki lahko izvajajo naloge, za katere je običajno potrebna človeška inteligenca, kot so prepoznavanje slik, govor, sprejemanje odločitev in obdelava naravnega jezika (3). Kot del širšega področja UI so veliki jezikovni modeli (angl. *large language models*, LLM) ena od najnaprednejših tehnologij, ki temeljijo na globokem učenju in analizi ogromnih količin besedilnih podatkov. Med najbolj prepoznanimi velikimi jezikovnimi modeli je ChatGPT, ki temelji na tehnikah globokega učenja in deluje kot napredno orodje za komunikacijo, pridobivanje informacij in reševanje različnih nalog (4). Iz tega razloga predstavlja privlačno orodje za izboljšanje stroškovne učinkovitosti in uspešnosti na različnih področjih, vključno z zdravstvom. ChatGPT je lahko uporaben pri podpori zdravnikom pri diagnosticiranju in zdravljenju bolezni, zagotavljanju visokokakovostnega medicinskega izobraževanja ter posredovanju medicinskega znanja širši javnosti. Tudi na področju farmacije se pojavljajo pomembne možnosti njegove uporabe, kjer se nekateri jezikovni modeli, kot je ChatGPT, uporabljajo za proučevanje terapevtskih možnosti, mehanizmov delovanja in interakcij med zdravili (5–7). Raziskave so pokazale tudi njegovo uporabnost v klinični praksi v smislu prepoznavanja težav, povezanih z zdravili, preprečevanja prekomernega predpisovanja, neustrezne izbire zdravil, interakcij med zdravili, napak pri odmerjanju ali izboljšanja sodelo-

vanja pri zdravljenju z zdravili (8–10). ChatGPT je s svojimi zmogljivostmi izjemno uporabno orodje za podporo raziskovalcem, zdravnikom, farmacevtom in drugim strokovnjakom, vendar pa je treba upoštevati tudi morebitna tveganja, povezana z varovanjem zasebnosti, varnostjo podatkov, širjenjem lažnih ali napačnih informacij ter etičnimi vprašanji (7).

V tej raziskavi smo želeli oceniti, kako uporaben je jezikovni model ChatGPT na področju klinične farmacije; natančneje, kako učinkovito model ChatGPT (verzija GPT 3.5) prepoznava in rešuje težave, povezane z zdravili, pri obravnavi kliničnih primerov.

# 2 METODE

ChatGPT smo uporabili pri prepoznavanju težav, povezanih z zdravili, in iskanju predlogov za njihovo rešitev v okviru petih kliničnih primerov ter prosili klinične farmacevtke (KF) za oceno pravilnosti, razumljivosti, celovitosti, jezikovne ustreznosti ter argumentacije teh odgovorov.

## 2.1 UPORABNOST CHATGPT PRI OBRAVNAVI KLINIČNIH PRIMEROV

Klinične primere smo oblikovali v slovenskem jeziku tako, da pokrivajo različna farmakoterapevtska področja, pri čemer vključujejo variabilnost v značilnostih bolnikov ter v glavnih in spremljajočih diagnozah. Soavtorica članka (LK), ki deluje kot klinična farmacevtka in kot visokošolska učiteljica, je oblikovala klinične primere na osnovi značilnosti bolnikov, ki jih srečuje v praksi v bolnišnici in v ambulantah farmacevta svetovalca. Pri oblikovanju primerov niso bili uporabljeni podatki, ki bi omogočali identifikacijo posameznih bolnikov. Gre za hibridne primere, saj združujejo elemente realnih kliničnih situacij in didaktične modifikacije. Primeri niso neposredno povzeti iz realnih primerov, temveč predstavljajo kombinacijo elementov več pacientov, nato pa so bili dodatno prilagojeni za izobraževalne namene. Oblikovani primeri so bili validirani in več let uporabljeni v pedagoškem procesu. Zahtevnost primerov je bila prilagojena znanju in kompetencam študentov zadnjega letnika enovitega magistrskega študijskega programa Farmacija. Primere smo uredili v enotno strukturo, ki je vključevala opis bolnika, spremljajoče bolezni, zgodovino zdravljenja z zdravili ter rezultate opravljenih preiskav. Zdravila

Preglednica 1: Povzetki opisov kliničnih primerov.

Table 1: Summaries of clinical cases.

Klinični primer	Povzetek opisa kliničnega primera
1	<p>Gospod Leon (75 let, 185 cm, 90 kg) je bil sprejet zaradi dihalne odpovedi ob akutni okužbi dihal. Ima produktiven kašelj, bolečino v prsih, mrzlico, vročino nad 38 °C in saturacijo s kisikom (sO<sub>2</sub>) 88 %. Rentgenska slika je pokazala obsežne intersticijske spremembe, potrjena je pljučnica (sum na bakterijsko). Zdravljen je z amoksicilinom/klavulansko kislino 1,2 g i.v. na 8 ur ter simptomatsko terapijo s kisikom in antipiretiki. Pred boleznijo je bil fizično aktiven, samostojen. Ima arterijsko hipertenzijo in benigno hiperplazijo prostate. V redni terapiji prejema perindopril/indapamid tablete 4 mg/1,25 mg, 1 tableta 1-krat/dan, doksazosin SR tablete 4 mg, 1 tableta 1-krat/dan in ibuprofen tablete 400 mg, 1 tableta po potrebi. Alergij na zdravila nima. Laboratorijske preiskave: serumski kreatinin (S-Kr) 170 µmol/L in kalij v serumu 5,6 mmol/L.</p>
2	<p>Gospa Leopolda (67 let, 170 cm, 66 kg) je bila sprejeta zaradi dihalne odpovedi ob akutni okužbi dihal. Ima vročino nad 38 °C, mrzlico, glavobol, bolečine v mišicah in prsih, nosni izcedek ter izrazito utrujenost. Izmerjena sO<sub>2</sub> je 88 %. Potrjena je gripa tipa A s pljučnico, na rentgenski sliki so vidne obsežne intersticijske spremembe brez zgostitev. Prejema kisik, paracetamol in oseltamivir 75 mg, 1 tableta 2-krat/dan. Ima arterijsko hipertenzijo in sladkorno bolezen tipa 2 brez zapletov. V redni terapiji prejema perindopril tablete 8 mg, 1 tableta 1-krat/dan, doksazosin SR tablete 4 mg, 1 tablete 1-krat/dan in metformin tablete 850 mg, 1 tableta 3-krat/dan. Uporabo drugih zdravil ali pripravkov v samozdravljenju zanika. Laboratorijske preiskave: S-Kr 110 µmol/L in glukoza v serumu na tešče 11 mmol/L.</p>
3	<p>Gospa Borka (55 let, 165 cm, 60 kg) je v zadnjih dveh letih shujšala 15 kg. Zaradi hude bolečine v hrbtenici redno jemlje ibuprofen (Diverin<sup>®</sup>) 400 mg, paracetamol ji ne pomaga, drugih analgetikov še ni poskusila. Počuti se utrujeno in brezvoljno, vendar zavrača diagnozo depresije ter si želi manj zdravil. Čaka na operacijo hrbtenice in ima ishemično srčno bolezen, arterijsko hipertenzijo ter hipotirozo. V redni terapiji ima pantoprazol (Controloc<sup>®</sup>) tablete 40 mg, 1 tableta 1-krat/dan; levotiroksinat (Euthyrox<sup>®</sup>) tablete 100 µg (pon, sre, pet, ned), in 125 µg (tor, čet, sob), občasno pozabi vzeti zdravilo; perindopril (Prenessa<sup>®</sup>) tablete 8 mg, 1 tableta 1-krat/dan; acetilsalicilno kislino (Aspirin<sup>®</sup>) GR tablete 100 mg, 1 tableta 1-krat/dan in rosuvastatin (Crestor<sup>®</sup>) tablete 20 mg, 1 tableta 1-krat/dan. Krvni tlak: 120/80 mm Hg, samo-meritve doma: 110–130/75–80 mm Hg; občasno vrtočlavice ob spremembah položaja.</p>
4	<p>Gospa Cvetka (47 let, 170 cm, 75 kg) ima težave z menopavzo (vročinski oblivi, nespečnost, motnje koncentracije) in je potrta ob nedavni diagnozi raka dojke. Rak dojke je bil operiran, zdravljena je bila z obsevanjem in prejema dopolnilno hormonsko zdravljenje, za katerega slabo razume namen zdravljenja. Zaradi depresije prejema antidepresiv, ki ji pomaga, zaradi nespečnosti pa uspavalo, po katerem navaja predolgo spanje in moteno zbranost. Po operaciji raka dojke je doživela globoko vensko trombozo. Jemlje tamoksifen (Nolvadex<sup>®</sup>) tablete 20 mg, 1 tableta 1-krat/dan; cipramil (Cipramil<sup>®</sup>) tablete 20 mg, 1 tableta 1-krat/dan; kvetiapin (Kvelux<sup>®</sup>) tablete 25 mg, 1 tableta 4-krat/dan in dalteparin (Fragmin<sup>®</sup>) 15.000 i.e. subkutana injekcija 1-krat/dan. Alergij na zdravila nima. Testi ledvične funkcije: S-Kr 70 mmol/L; sečnina v serumu 5,0 mmol/L.</p>
5	<p>Gospod Dragan (67 let, 180 cm, 65 kg) ima napredovalega raka pljuč z zasevki v kosteh in jetrih. V zadnjih dveh mesecih je izgubil 8 kg in ima hude bolečine pod desno lopatico zaradi zasevkov (ocena 8/10). Bolečina je bila na začetku dobro nadzorovana s kombinacijo oksikodon/nalokson (Targinact<sup>®</sup>) SR tablete 20/10 mg, 1 tableta 2-krat/dan, sedaj jih slabo obvladuje s fentanil (Durogesic<sup>®</sup>) transdermalnimi obliži 25 µg, zato 3–5-krat dnevno jemlje morfin (Sevredol<sup>®</sup>) tablete 10 mg, ki mu začasno pomaga. Diklofenak (Naklofen<sup>®</sup>) mu malo lajša bolečine, pregabalin (Lyrica<sup>®</sup>) ne. Alprazolama (Xanax<sup>®</sup>) ne jemlje zaradi strahu pred odvisnostjo, občasno uporablja konopljinno smolo za spanje. V terapiji ima še kalcijev karbonat tablete 1 g, 1 tableta 1-krat/dan; holecalfiferol (Plivit D3<sup>®</sup>) 35 kapljic/teden in denosumab (Prolia<sup>®</sup>) 60 mg subkutana injekcija 1-krat/mesec. Vse ostale težave so ob sedanjih bolečinah manj pomembne.</p>

Opomba: V primerih, ko smo pri vnosu v ChatGPT uporabili lastniška imena zdravil, so v preglednici 1 le-ta zapisana v oklepaju po nelastniških imenih.



smo v nekaterih primerih zapisali z generičnimi imeni, v drugih pa z lastniškimi imeni, da bi preverili sposobnost ChatGPT pri prepoznavanju lastniških imen. Povzetke opisov kliničnih primerov prikazuje preglednica 1. Besedila opisov primerov pred izvedbo raziskave niso bila vnesena oz. pregledana s strani ChatGPT ali drugih velikih jezikovnih modelov.

Za uporabnost v raziskavi smo preizkusili tri jezikovne modele UI (ChatGPT GPT 3.5, Bing Chat, sedaj Microsoft Copilot, ki je uporabljal Microsoftovo prilagojeno različico modela GPT-4, in Perplexity AI, ki je uporabljal interne modele in modele OpenAI), ki so bili na voljo v času izvedbe raziskave (marec 2023), in ugotovili, da je bil za namen naše raziskave zaradi zmožnosti vnosa velike količine podatkov ter povezovanja pozivov uporaben zgolj prvi. Pri vseh kliničnih primerih smo ChatGPT najprej postavili dve splošni vprašanji: »Kakšne težave, povezane z zdravili, lahko prepoznamo?« in »Kakšne intervencije bi predlagal?«, potem pa še šest do deset specifičnih vprašanj, ki so se med primeri razlikovala (preglednica 2). Ta so se nanašala na poznavanje specifičnih zdravil, najustrežnejšega režima odmerjanja ter njihovih neželenih učinkov, interakcij z drugimi zdravili ter na predloge za izbiro ustreznejše terapije. Besedila vprašanj smo dokončno oblikovali pred izvedbo raziskave. Preizkus različnih pozivov je potekal v pripravljani fazi z namenom preverjanja razumljivosti in jasnosti vprašanj.

Na podlagi odgovorov ChatGPT smo razvili ankete v 1KA (<https://www.1ka.si>). Pred posredovanjem KF smo jih pilotno testirali z vidika jasnosti, razumljivosti, splošnega vtisa, zmožnosti ocenjevanja, težav pri odgovarjanju in časovne ustreznosti ter skladno z ugotovitvami ustrezno nadgradili. Uspešnost ChatGPT pri prepoznavanju in reševanju težav, povezanih z zdravili, je vrednotilo pet KF. Vključevanje KF v raziskavo je potekalo v juniju in juliju 2023. Kriteriji za povabilo k sodelovanju v raziskavi so bili 1) specializacija iz klinične farmacije, 2) aktivno izvajanje kognitivnih storitev klinične farmacije, 3) aktivnosti v strokovnih izobraževalnih procesih. Vse klinične farmacevtke, ki smo jih povabili k sodelovanju, so se povabilu odzvale. Dve izmed njih delujeta na terciarnem nivoju, ena na sekundarnem in dve na primarnem nivoju. Tri farmacevtke – med njimi obe, ki delujeta na primarni ravni – so izvajale ali izvajajo tudi farmakoterapijske preglede v okviru ambulantne farmacevta svetovalca. Z njimi smo se najprej osebno dogovorili za sodelovanje v raziskavi, potem pa smo jim poslali e-pošto s povezavo do ankete. Za sodelovanje v raziskavi KF niso prejele finančne kompenzacije. Reševanje anket je potekalo od 3. 7. 2023 do 7. 8. 2023. Vsaka KF je neodvisno ovred-

notila uspešnost ChatGPT pri obravnavi dveh od petih kliničnih primerov in vsak klinični primer sta ovrednotili dve KF. KF so ocenjevale pravilnost, razumljivost, celovitost, jezikovno skladnost in argumentiranost vsakega odgovora ChatGPT na petstopenjski lestvici: (1) se ne strinjam, (2) delno se ne strinjam, (3) sem neodločen, (4) delno se strinjam, (5) se strinjam. Pri ocenjevanju pravilnosti so KF presojele, ali so informacije, ki jih je podal ChatGPT, točne in skladne z dejstvi. Pri splošnih in specifičnih vprašanjih so vsako trditev ChatGPT ocenjevale posebej. ChatGPT je pri posameznem vprašanju namreč podal več zaznanih težav ali predlagal več intervencij, pri čemer so se te trditve razlikovale po pravilnosti. Celovitost odgovora so ocenjevale glede na to, ali ChatGPT na postavljeno vprašanje odgovori celovito (ali le delno) ter ali je odgovor oblikovan kot zao-krožena in smiselna celota. Svoje ocene so lahko dodatno pojasnile v obliki prostega besedila.

## 2.2 ODNOS KLINIČNIH FARMACEVTOV DO UPORABE UMETNE INTELIGENCE

Odnos do uporabe umetne inteligence smo vrednotili na istem vzorcu KF, ki so vrednotile uspešnost modela ChatGPT pri prepoznavanju in obravnavi težav, povezanih z zdravili. Anketi, ki je bila namenjena oceni uporabnosti ChatGPT, smo dodali sklop petih vprašanj, ki so se nanašala na odnos KF do uporabe umetne inteligence pri njihovem delu, glede pravilnosti UI pri predlaganju kliničnih intervencij, vloge in vpliva UI na njihovo delo ter potrebe po usposabljanju KF za delo z UI. Raziskali smo tudi, ali obstaja povezava med odnosom KF do UI in njihovimi ocenami pravilnosti odgovorov modela ChatGPT.

## 2.3 ANALIZA

Za vsako posamezno vprašanje smo izračunali povprečje dveh ocen KF. Za ugotavljanje skladnosti ocen dveh KF, ki sta vrednotili isti primer, smo uporabili Cohenov koeficient kapa, ki se pogosto uporablja kot merilo strinjanja med izvajalci. Višji koeficient pomeni večjo stopnjo strinjanja. Cohenov koeficient kapa smo izračunali čez ocene vseh treh kategorij vprašanj skupaj (splošno vprašanje o prepoznanih težavah, splošno vprašanje o predlaganih intervencijah, specifična vprašanja). S pomočjo testa Mann-Whitney smo primerjali ocene pravilnosti odgovorov ChatGPT med KF, ki podpirajo uporabo UI v praksi, in tistimi, ki ji niso naklonjeni.

Preglednica 2: Specifična vprašanja pri posameznem kliničnem primeru.

Table 2: Specific questions for each clinical case.

Klinični primer	Specifična vprašanja
1	<ol style="list-style-type: none"> <li>1. Ali je izbira antibiotičnega zdravljenja (zdravilna učinkovina, režim odmerjanja, način aplikacije) pravilna?</li> <li>2. Ali je potrebna prilagoditev zdravljenja z zdravili glede na ledvično funkcijo?</li> <li>3. Kakšen je kreatininski očistek gospoda Leona?</li> <li>4. Ali bi bila zaradi takšnega kreatininskega očistka potrebna prilagoditev odmerkov zdravil, ki jih gospod Leon jemlje?</li> <li>5. Kaj pa odmerek antibiotika?</li> <li>6. Ali so vrednosti kalija znotraj normalnih vrednosti?</li> <li>7. Ali je potrebna prilagoditev zdravljenja z zdravili glede na hiperkalemijo?</li> <li>8. Ali obstajajo interakcije med zdravili, ki jih gospod Leon jemlje?</li> </ol>
2	<ol style="list-style-type: none"> <li>1. Kako deluje metformin in zakaj bi naj povzročal tveganje za hipoglikemijo?</li> <li>2. Kaj lahko sklepamo iz vrednosti glukoze gospe Leopolde?</li> <li>3. Ali je izbira protivirusnega zdravljenja (zdravilna učinkovina, režim odmerjanja, način aplikacije) pravilna?</li> <li>4. Ali obstajajo interakcije med zdravili, ki jih gospa jemlje?</li> <li>5. Ali je potrebna prilagoditev zdravljenja z zdravili glede na ledvično funkcijo?</li> <li>6. Ali je potrebna prilagoditev zdravljenja z zdravili glede na krvni tlak, izmerjen na oddelku?</li> </ol>
3	<ol style="list-style-type: none"> <li>1. Kako deluje Euthyrox® in kdaj tekom dneva je priporočljivo, da gospa Borka vzame zdravilo Euthyrox®?</li> <li>2. Kdaj tekom dneva je priporočljivo, da gospa Borka vzame preostala zdravila?</li> <li>3. Zakaj naj bi običajno vzela Aspirin® GR100 mg zvečer?</li> <li>4. Ali naj gospa Borka vzame Aspirin® GR 100 mg in Crestor® 20 mg pred, med ali po jedi?</li> <li>5. Zakaj je priporočljivo, da gospa Borka vzame 100 mg Aspirina® in Crestor® 20 mg po obroku?</li> <li>6. Kateri je najbolj primeren analgetik za gospo Borko ob degenerativnih spremembah hrbtenice?</li> <li>7. Katere analgetike še poznamo, ki bi jih lahko zdravnik predpisal gospe Borki?</li> <li>8. Kateri bi bil najbolj primeren antidepresiv za gospo Borko?</li> </ol>
4	<ol style="list-style-type: none"> <li>1. Ali so predpisana ustrezna zdravila in režimi odmerjanja?</li> <li>2. Gospa Cvetka ima predpisana tudi Nolvadex® in Fragmin®. Ali sta predpisana ustrezna zdravila in režima odmerjanja?</li> <li>3. Ali obstajajo interakcije med zdravili, ki jih gospa jemlje?</li> <li>4. Ali so pri gospe Cvetki interakcije med predlaganim hormonskim nadomestnim zdravljenjem (estrogeni in progestogeni) in ostalimi zdravili?</li> <li>5. Kdaj in zakaj se odločamo za hormonsko nadomestno zdravljenje?</li> <li>6. Kateri so primeri zdravil, ki se uporabljajo za hormonsko nadomestno zdravljenje in kako delujejo?</li> <li>7. Ali bi bile prisotne interakcije med zdravili za hormonsko nadomestno zdravljenje (estrogeni in progestogeni) in zdravili, ki jih ima gospa Cvetka predpisane?</li> <li>8. Katera je zdravilna učinkovina zdravila Kvelux®? Kakšne neželene učinke lahko povzroča? Kakšne so interakcije med kvetiapiinom in drugimi zdravili?</li> <li>9. Ali lahko pride do interakcije med citalopramom in kvetiapiinom?</li> <li>10. Kateri drug antidepresiv bi lahko gospa Cvetka uporabljala namesto Cipramila®, da bi se izognili interakciji QTc s kvetiapiinom?</li> </ol>
5	<ol style="list-style-type: none"> <li>1. Katere so zdravilne učinkovine zdravil, ki jih jemlje gospod Dragan in kako delujejo?</li> <li>2. Kaj pa Prolia®?</li> <li>3. Kdaj tekom dneva je priporočljivo, da gospod Dragan vzame zdravila?</li> <li>4. Kaj bi lahko storil gospod Dragan, da bi imel bolj urejeno bolečino?</li> <li>5. Ali obstajajo interakcije med zdravili, ki jih jemlje gospod Dragan?</li> <li>6. Ali bi lahko prišlo do interakcije med zdravili, če bi gospodu Draganu dodatno uvedli terapijo s pembrolizumabom?</li> <li>7. Ali bi lahko prišlo še do kakšne druge interakcije med pembrolizumabom in drugimi zdravili, ki jih jemlje gospod Dragan?</li> <li>8. Kakšni so možni neželeni učinki protibolečinskih zdravil, ki jih gospod Dragan uporablja?</li> <li>9. Kakšni so možni neželeni učinki Sevredola®, Durogesica® in Lyrice®?</li> <li>10. Kaj bi lahko priporočali gospodu Draganu glede nespečnosti?</li> </ol>





# 3 REZULTATI Z RAZPRAVO

V tem poglavju predstavljamo rezultate raziskave, ki zajemajo oceno pravilnosti, razumljivosti in uporabnosti odgovorov ChatGPT pri obravnavi kliničnih primerov ter odnos kliničnih farmacevtov do uporabe UI v svoji praksi.

## 3.1 UPORABNOST CHATGPT PRI OBRAVNAVI KLINIČNIH PRIMEROV

### 3.1.1 Pravilnost odgovorov

Klinične farmacevtke so v povprečju nekoliko bolje ocenile pravilnost odgovorov ChatGPT na splošna vprašanja v primerjavi s specifičnimi vprašanji. Po oceni KF je bilo na splošna vprašanja napačnih (izbira odgovora KF glede pravilnosti »Se ne strinjam« ali »Se deloma ne strinjam«) 30 %

odgovorov, na specifična vprašanja pa 37 % odgovorov. Povprečne ocene pravilnosti odgovorov ChatGPT na splošna in specifična vprašanja prikazuje preglednica 3. Pri vsakem primeru je podana povprečna ocena vseh trditev, ki jih je pri posameznem vprašanju podal ChatGPT. Največje ujemanje med ocenami KF smo izračunali pri kliničnem primeru št. 1, saj je bila vrednost Cohenovega koeficienta kapa 0,70, kar pomeni precejšnje strinjanje v skladu z usmeritvami Landisa in Kocha (11). Pri preostalih kliničnih primerih je bilo opaziti zmerno strinjanje med ocenjevalkama, saj so vrednosti Cohenovega koeficienta kapa znašale med 0,41 in 0,51 (preglednica 4) (12).

ChatGPT je pokazal razmeroma dobro sposobnost kontekstnega razumevanja, kar pomeni, da je bil zmožen uporabiti informacije iz prejšnjih vprašanj in odgovorov, čeprav pri tem ni bil popolnoma konsistenten. Pri odgovorih na zdravstvena vprašanja je pogosto podajal splošne nasvete za ohranjanje zdravja in dosledno poudarjal pomembnost posveta z zdravnikom ali farmacevtom, kar je ključno, če bi ChatGPT uporabljal pacient. V splošnem je uspešno razumel in izvajal naloge, vendar pa so bili odgovori, čeprav

*Preglednica 3: Povprečne ocene pravilnosti odgovorov na splošna in specifična vprašanja.*

*Table 3: Average ratings of answer correctness to general and specific questions.*

Št. kliničnega primera	Povprečna ocena (razpon ocen)		
	Splošno vprašanje o prepoznanih težavah	Splošno vprašanje o predlaganih intervencijah	Specifična vprašanja
1	3,9 (1–5)	4,3 (4–5)	3,1 (1–5)
2	2,8 (1–4)	3,9 (2–5)	3,6 (1–5)
3	3,1 (1–5)	2,3 (1–5)	2,1 (1–5)
4	2,6 (1–5)	3,2 (1–5)	3,0 (1–5)
5	3,9 (2–5)	4,1 (2–5)	3,3 (1–5)

*Opomba: Število ocenjenih trditev, podanih s strani ChatGPT pri posamezni kategoriji, je bilo od 4 do 31.*

*Preglednica 4: Ujemanje med ocenami KF, izraženo s Cohenovim koeficientom kapa.*

*Table 4: Agreement between clinical pharmacists' ratings, measured using Cohen's kappa coefficient.*

Klinični primer	Ocenjevalki	Cohenov koeficient kapa
1	KF1 in KF4	0,695
2	KF1 in KF2	0,446
3	KF2 in KF3	0,489
4	KF3 in KF5	0,505
5	KF4 in KF5	0,414

ne napačni, večkrat pomanjkljivi, splošni in nenatančni ter po mnenju KF zato pogosto neuporabni. V določenih primerih ni uspel povezati lastnih sklepov v različnih odgovorih ali pa je imel težave pri prepoznavanju »celotne slike« pacienta oz. upoštevanju le-te pri podajanju specifičnih nasvetov.

Koristen je bil pri prepoznavanju osnovnih težav, povezanih z zdravili, in pri splošnem usmerjanju glede intervencij za njihovo reševanje. Izjemno se je izkazal pri opisu zdravlilnih učinkovin z vidika namena uporabe in načina delovanja ter svetovanju glede izbire terapije, npr. ustreznega antibiotika ali podpornega zdravljenja, in poudarjanja pomena rednega jemanja predpisane terapije ter spremljanja kliničnih parametrov, npr. krvnega tlaka in glukoze v krvi. Zelo uspešno je navajal potencialne neželene učinke zdravil, čeprav ob tem ni podajal njihove pojavnosti in s pomanjkljivimi nasveti za njihovo spremljanje ter obvladovanje, če te informacije niso bile izrecno zahtevane. V določenih primerih ni uspel prepoznati, kateremu izmed pacientovih zdravil lahko pripišemo določen neželeni učinek, ali pa je ob navajanju neželenih učinkov izpostavil takšnega, ki za navedeno zdravilo ni značilen, npr. pojav hipoglikemije zaradi jemanja metformina. Večinoma zelo uspešen je bil pri prepoznavanju interakcij med zdravili, čeprav je večkrat manjkala presoja klinične pomembnosti, v določenih primerih pa ni bilo jasno, ali je pri poročanju interakcij upošteval le kombinacijo zdravil ali v presojo vključil tudi laboratorijske vrednosti in rezultate drugih preiskav, za katere je imel na voljo podatke, npr. EKG. Nekoliko slabše se je ChatGPT odrezal pri določanju ocene glomerulne filtracije in podajanju nasvetov glede odmerjanja zdravil, npr. amoksicilin/klavulanska kislina, dokozosin in oseltamivir v skladu z ledvično funkcijo. V nekaterih primerih je podal netočne informacije o odmerjanju, npr. kakšen je ustrezen odmerek pregabalina pri zdravljenju nevropatske bolečine, ali času jemanja zdravil, npr. nasvet naj pacient jemlje acetilsalicilno kislino in rosuvastatin zvečer. Težave je imel tudi s prepoznavanjem zdravlilnih učinkovin v generičnih zdravilih, npr. zdravilu Kvelux® je najprej pripisal učinkovino zolpidem, kasneje pa kvercetin, kar je negativno vplivalo na pravilnost njegovih kasnejših odgovorov. Literatura sicer navaja, da je ChatGPT občasno nagnjen k podajanju netočnih informacij ali halucinacijam, zato je ključnega pomena, da pridobljene informacije obravnavamo kritično in jih preverimo tudi z drugimi viri (13).

Kot primer navajamo odgovore ChatGPT na splošno vprašanje glede prepoznavanja težav, povezanih z zdravili, ter ocene KF na podane odgovore pri kliničnem primeru št. 3 (preglednica 5). Kot je povzela KF3: »ChatGPT je izpostavil veliko število težav, tudi z napačnimi poudarki (npr. morebitni

zadržki ob jemanju ibuprofena zaradi možne odvisnosti). Ne poda razlage, kaj je vzrok ("mehanizem") posamezne težave.« Kot težavo je izpostavil tudi previsok odmerek ibuprofena, čeprav njegov maksimalen odmerek (2400 mg) ni presežen. Ob oceni je KF2 dodala: »Ob rednem jemanju višjih odmerkov zdravila ibuprofen ter sočasnem prejemanju acetilsalicilne kisline se bojimo predvsem tveganja za krvavitve, zlasti iz prebavil, in zmanjšanega kardioprotektivnega učinka acetilsalicilne kisline.« ChatGPT je težave, ki jih gospa navaja (brezvoljnost, utrujenost), pripisal neželenim učinkom rosuvastatina, ki so sicer navedeni med možnimi neželenimi učinki zdravila s to učinkovino, vendar po mnenju KF2 pri gospe najbrž niso povezani z njim. Predvidevamo namreč lahko, da gospa prejema rosuvastatin že več let, glede na to, da ji je bila žilna opornica vstavljena pred dvema letoma. Uspešneje je ChatGPT prepoznal pomen izpuščanja odmerkov Euthyroxa® in zvišane ravni tirotropina (TSH) v krvi, kar kaže na neustrezno raven ščitničnih hormonov in potencialen vpliva na počutje pacientke. Ob oceni je KF2 dodala: »Pomembno je, če se pogosto zgodi, da pozabi vzeti levotiroksin. Če se to zgodi enkrat mesečno, pa ni tako problematično.« Kot težavo je ChatGPT pravilno prepoznal tudi, da ima gospa le redke meritve krvnega tlaka pri zdravniku. Občasne vrtoglavice bi lahko kazale na neustrezno nadzorovan krvni tlak, zato bi bila smiselna meritve krvnega tlaka pri zdravniku ali farmaceutu. Ob tem je KF2 dodala: »Glede na dostopne podatke je krvni tlak sicer v mejah normale in si ga gospa meri tudi doma.«

### 3.1.2 Razumljivost, celovitost, jezikovna ustreznost in argumentiranost odgovorov

V večini so se KF strinjale, da so odgovori ChatGPT jasni, razločni in nedvoumni. S trditvijo, da je odgovor razumljiv, so se strinjale pri 31 % in delno strinjale pri 41 % primerov odgovorov. Pretežno pozitivno so ocenile tudi jezikovno ustreznost odgovorov, saj je bilo po njihovem mnenju 68 % odgovorov jezikovno dobro podanih. Po drugi strani so slabše ocenile celovitost in argumentiranost odgovorov ChatGPT, saj je bila po njihovem mnenju le približno četrтина odgovorov celovitih (23 %) in dobro argumentiranih (26 %). Pri testiranju uporabnosti ChatGPT smo uporabili slovenski jezik, zato je možno, da bi bili rezultati drugačni, če bi primere in vprašanja zastavili v angleščini, saj je njegova uporabnost oz. natančnost odgovorov lahko odvisna od jezika, v katerem se uporablja (14). K uporabnosti ChatGPT pomembno prispeva tudi način podajanja vprašanj. Raziskava Roosana in sod. je pokazala, da je točnost njegovih odgo-



Preglednica 5: Odgovori ChatGPT na splošno vprašanje ter ocene dveh kliničnih farmacevtk (KF) o pravilnosti podanih odgovorov.  
Table 5: ChatGPT's responses to general questions and clinical pharmacists' ratings of the correctness of the provided answers.

Odgovor ChatGPT na splošno vprašanje »Kakšne težave, povezane z zdravili, lahko prepoznamo?«	Oceni* KF in (povprečna ocena)
Prevelik odmerek Diverina®: Gospa Borka jemlje precej visok odmerek Diverina®, kar lahko povzroči neželene učinke, kot so slabost, bruhanje, driska, omotica in glavoboli. Če se zdravilo jemlje dlje časa, lahko povzroči tudi poškodbe ledvic.	1, 2 (1,5)
Možna odvisnost od Diverina®: Gospa Borka jemlje Diverin® vsak dan že zadnji mesec, kar povečuje možnost, da postane odvisna od zdravila.	1, 1 (1)
Interakcije med zdravili: Gospa Borka jemlje več različnih zdravil, kar povečuje možnost, da lahko zdravila medsebojno vplivajo. Na primer, ibuprofen lahko poveča tveganje za krvavitve, če se jemlje skupaj z Aspirinom®. Prav tako lahko Controloc GR® poveča raven nekaterih drugih zdravil v krvi.	4, 4 (4)
Neustrezno jemanje Euthyroxa®: Gospa Borka občasno pozabi vzeti Euthyrox®, kar lahko vpliva na njeno zdravljenje hipotiroze.	4, 5 (4,5)
Neustrezna raven ščitničnih hormonov: Gospa Borka ima zvišano raven TSH, kar kaže na neustrezno raven ščitničnih hormonov v krvi, kar lahko vpliva na njeno počutje in razpoloženje.	5, 5 (5)
Potencialni neželeni učinki statina Crestor®: Statini, kot je Crestor®, lahko povzročijo neželene učinke, kot so mišična bolečina, šibkost in utrujenost.	1, 3 (2)
Neustrezna kontrola krvnega tlaka: Gospa Borka ima le redke meritve krvnega tlaka pri zdravniku ali farmacevtu, zato ni jasno, ali je njeno zdravljenje hipertenzije ustrezno. Občasne vrtoglavice lahko kažejo na neustrezno nadzorovan krvni tlak.	2, 5 (3,5)

\*možne ocene: 1–5

vorov pri kompleksnih primerih pacientov podobna točnosti pri enostavnih kliničnih primerih, vendar so potrebna številčnejša in bolj specifična vprašanja (15).

### 3.2 ODNOS KLINIČNIH FARMACEVTOV DO UPORABE UMETNE INTELIGENCE

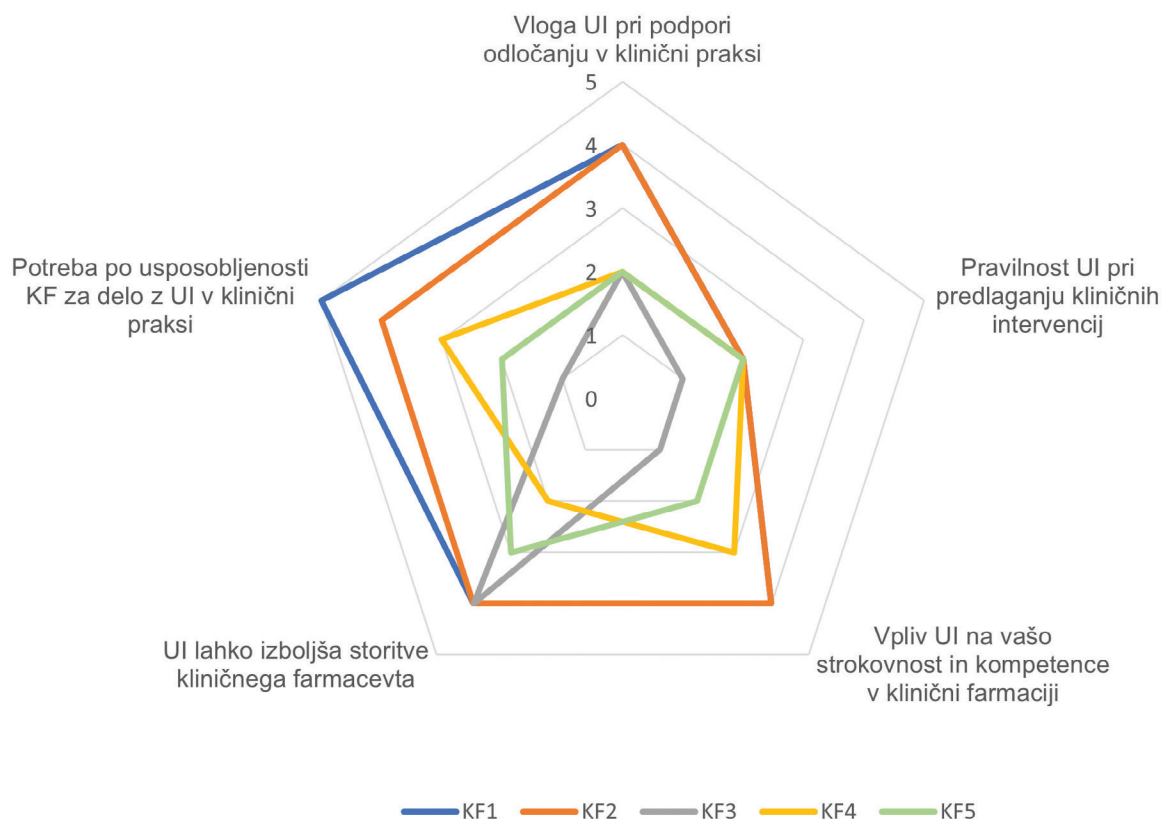
Klinične farmacevtke, vključene v raziskavo, niso imele enotnega mnenja glede vloge UI pri podpori odločanja v klinični praksi. Dve sta izrazili naklonjenost uporabi UI, tri pa so izrazile, da njeni uporabi niso naklonjene. Njihovi pogledi glede vpliva na strokovnost in kompetence KF ter zmožnost UI, da izboljša storitve KF so bili zelo heterogeni, sta pa bolj pozitivno mnenje glede potencialnega pozitivnega vpliva UI na delo KF izražali klinični farmacevtki, ki sta uporabi UI v klinični praksi naklonjeni. Po drugi strani so vse delile mnenje o omejeni pravilnosti UI pri predlaganju

kliničnih intervencij (slika 1). Povprečna ocena pravilnosti odgovorov ChatGPT s strani KF, ki podpirajo uporabo UI (na vprašanje glede vloge UI izbrani odgovor »Podpiram uporabo UI«), je bila 3,29 in tistih, ki je ne podpirajo (izbrani odgovor »Ne podpiram uporabe UI«), 3,26; med skupinama ni bilo statistično značilne razlike ( $p = 0,824$ ).

### 3.3 OMEJITVE RAZISKAVE

Različica ChatGPT 3.5, ki smo jo uporabljali, je bila usposobljena na podatkih, zbranih do septembra 2021, kar pomeni, da je ChatGPT morda primanjkovalo razumevanje nekaterih najnovejših spoznanj. Različica, ki smo jo uporabljali, je bila aktualna ob izvajanju raziskave, trenutno pa je na voljo novejša različica, ki je bila večkrat posodobljena in bi morda generirala drugačne odgovore. V raziskavi smo testirali uspešnost ChatGPT pri reševanju kliničnih primerov v slovenščini, vendar pa je njegova učinkovitost dokazano





**Slika 1:** Odnos kliničnih farmacevtk (KF) do uporabe umetne inteligence (UI). Višje vrednosti pomenijo pozitivnejši odnos do UI.  
**Figure 1:** Clinical pharmacists' attitudes towards the use of artificial intelligence (AI). Higher values indicate a more positive attitude towards AI.

različna glede na jezik in rezultati torej niso posplošljivi na uporabo ChatGPT v angleščini. Uporabljena shema pozivov (dve splošni in šest do deset specifičnih vprašanj) je bila heterogena med kliničnimi primeri, kar lahko vpliva na primerljivost rezultatov. Čeprav so bila vprašanja dokončno določena pred začetkom zbiranja podatkov, ni mogoče povsem izključiti t. i. pristranosti »prompt-tuning«, pri kateri lahko preizkusne iteracije vplivajo na končne odzive modela. Dodatna pomanjkljivost raziskave je bila uporaba zgolj enega jezikovnega modela. Raziskava torej ni vključevala primerjave z alternativnimi UI-orodji ali z obstoječimi kliničnimi podporami, kar bi omogočilo bolj objektivno vrednotenje uporabnosti UI. Dodatni omejitvi sta tudi majhen vzorec kliničnih farmacevtk in njihova morebitna pristranskost. Slednja zaradi subjektivne interpretacije lahko vpliva na ocene pravilnosti in uporabnosti odgovorov. V raziskavi smo uporabili omejen nabor kliničnih primerov, ki v celoti ne zajame raznolikosti realne prakse. Vse naštetu lahko bistveno omejuje posplošljivost ugotovitev.

## 4 SKLEP

ChatGPT je lahko v pomoč kliničnim farmacevtom pri prepoznavanju osnovnih težav, povezanih z zdravili, kot so neželeni učinki in interakcije med zdravili, pri posameznem pacientu, vendar v mnogih primerih poda nepopolne ali včasih celo napačne odgovore. Izjemno uporaben je kot bližnjica pri iskanju informacij, ki so razlage mehanizmov delovanja zdravil ali najpogostejših neželenih učinkov, manj uporaben je pri iskanju konkretnih usmeritev za optimalno obravnavo pacienta. Klinične farmacevtke, sodelujoče v raziskavi, so sicer izrazile pozitiven odnos do umetne inteligence, vendar so poudarile nujnost strokovne presoje. Rezultati raziskave nakazujejo, da veliki jezikovni modeli lahko podpirajo klinične odločitve, vendar ne morejo zamenjati kompetentnega kliničnega farmacevta.

# 5 ZAHVALA

Avtorji se zahvaljujemo Žanu Vodopivcu za opravljeno delo v okviru magistrske naloge. Zahvaljujemo se tudi kliničnim farmacevtkam Barbari Koder, mag. farm., spec. klin. farm., Veroniki Pelicon, mag. farm., spec. klin. farm., Mateji Stopinšek, mag. farm., spec. klin. farm. in Tini Kajan, mag. farm., spec. klin. farm., ki so ocenile odgovore ChatGPT.

# 6 LITERATURA

1. Slovar slovenskega knjižnega jezika. [Internet] Inštitut za slovenski jezik Frana Ramovša: Slovar slovenskega knjižnega jezika [cited 2024 Dec 11] Available from: <https://fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika/3608157/umeten?View=1&Query=umetna+inteligenca&S=1&All=umetna+inteligenca&FilteredDictionaryIds=130>.
2. Jeste DV, Graham SA, Nguyen TT, Depp CA, Lee ELE, Kim HC. Beyond artificial intelligence: exploring artificial wisdom. *Int Psychogeriatr*. 2020;32(8):993-1001.
3. Vodanovic M, Subasic M, Milosevic D, Pavicin IS. Artificial intelligence in medicine and dentistry. *Acta Stomatol Croat*. 2023;57(1):70-84.
4. Slovensko izobraževalno omrežje. Umetna inteligenca v izobraževanju: Chat GPT [Internet] Ministrstvo za vzgojo in izobraževanje: Slovensko izobraževalno omrežje; 2023 [cited 2024 Dec 11]. Available from: <https://novice.sio.si/2023/05/05/ui-v-izobrazevanju-chat-gpt/>.
5. Huang XR, Estau D, Liu XN, Yu Y, Qin JG, Li ZJ. Evaluating the performance of ChatGPT in clinical pharmacy: A comparative study of ChatGPT and clinical pharmacists. *Brit J Clin Pharmacol*. 2024;90(1):232-8.
6. Meng XB, Yan XY, Zhang K, Liu D, Cui XJ, Yang YD, et al. The application of large language models in medicine: A scoping review. *Iscience*. 2024;27(5).
7. Zhang Y, Pei H, Zhen S, Li Q, Liang F. Chat generative pre-trained transformer (ChatGPT) usage in healthcare. *Gastroenterol Endosc*. 2023;1(3):139-43.
8. Levivien C, Cavagna P, Grah A, Buronfosse A, Courseau R, Bezie Y, et al. Assessment of a hybrid decision support system using machine learning with artificial intelligence to safely rule out prescriptions from medication review in daily practice. *Int J Clin Pharm*. 2022;44(2):459-65.
9. van Nuland M, Erdogan A, Avarsigmaar C, Contrucci R, Hilbrants S, Maanach L, et al. Performance of ChatGPT on factual knowledge questions regarding clinical pharmacy. *J Clin Pharmacol*. 2024;64(9):1095-100.
10. Damiani G, Altamura G, Zedda M, Nurchis MC, Aulino G, Heidar Alizadeh A, et al. Potentiality of algorithms and artificial intelligence adoption to improve medication management in primary care: a systematic review. *BMJ Open*. 2023;13(3):e065301.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
12. DATAtab Team. Online Statistics Calculator: Cohen's Kappa [Internet]. DATAtab e.U. Graz, Austria; 2024 [cited 2024 Dec 12] Available from: <https://datatab.net/tutorial/cohens-kappa>.
13. Alkaiissi H, McFarlane S I (February 19, 2023) Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* 15(2): e35179.
14. Ozturk N, Yakak I, Ag MB, Aksoy N. Is ChatGPT reliable and accurate in answering pharmacotherapy-related inquiries in both Turkish and English? *Curr Pharm Teach Learn*. 2024;16(7):102101.
15. Roosan D, Padua P, Khan R, Khan H, Verzosa C, Wu Y. Effectiveness of ChatGPT in clinical pharmacy and the role of artificial intelligence in medication therapy management. *J Am Pharm Assoc* (2003). 2024;64(2):422-8 e8.