

**(DE)KONSTRUKCIJA INFORMACIJSKOG
PRETRAŽIVANJA: PRIMJENA ALATA GENERATIVNE
UMJETNE INTELIGENCIJE PRI OBLIKOVANJU
PRETRAŽIVAČKIH NIZOVA ZA BAZU PODATAKA
MEDLINE/PUBMED**

**(DE)CONSTRUCTION OF INFORMATION RETRIEVAL: THE
USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS
IN DESIGNING SEARCH STRINGS FOR THE MEDLINE/
PUBMED DATABASE**

Ivana Majer  Središnja medicinska knjižnica,
Medicinski fakultet Sveučilišta u Zagrebu
ivana.majer@mef.hr

Lea Škorić  Središnja medicinska knjižnica,
Medicinski fakultet Sveučilišta u Zagrebu
lea.skoric@mef.hr

UDK / UDC: 004.8:[61:004.65]:[004.78:025.4.036]

Izvorni znanstveni rad / Original scientific paper

<https://doi.org/10.30754/vbh.68.3.1585>

Primljeno / Received 18. 8. 2025.

Prihvaćeno / Accepted 3. 11. 2025.



Sažetak

Cilj: Cilj je istraživanja ispitati potencijal i učinkovitost alata generativne umjetne inteligencije (ChatGPT) pri tematskom pretraživanju, odnosno oblikovanju pretraživačkih nizova za bazu podataka MEDLINE/PubMed.

Metodologija: Istraživanje je provedeno sa studentima studija medicine u dvije sesije, između kojih ih se educiralo o pravilima kreiranja *promptova* i pretraživačkih nizova. Studenti su u obje sesije kreirali *promptove* za alat ChatGPT namijenjene generiranju pretraživačkog niza za bazu MEDLINE/PubMed. Kvaliteta *promptova* i generiranih nizova potom je ocjenjivana pomoću posebno osmišljenih evaluacijskih in-

strumenata (zlatni standardi i provjereni popisi). U analizi je kombiniran kvantitativan (bodovanje kvalitete *promptova* i nizova) i kvalitativan pristup (uočavanje pogrešaka i obrazaca).

Rezultati: Uočena je pozitivna korelacija između kvalitete *promptova* (kako ukupne kvalitete tako i korištenja pojedinih specifičnih elemenata) i ukupne kvalitete nizova. Druga sesija pokazala je značajan porast ukupne kvalitete *promptova* i pretraživačkih nizova, čime se sugerira važnost provedene edukacije. Nadalje, identificirani su pojedini elementi *prompta* koji pozitivno koreliraju s višom kvalitetom generiranih nizova.

Ograničenja: Korišten je samo jedan alat generativne umjetne inteligencije (ChatGPT-3.5) i promatrana je samo jedna baza podataka (MEDLINE/PubMed).

Praktična primjena: Razvijeni evaluacijski instrumentarij može poslužiti kao početna točka za razvoj sustavne metodologije evaluacije rezultata alata generativne umjetne inteligencije u informacijskom pretraživanju. Uočeni obrasci korisničkog ponašanja mogu biti osnova za osmišljavanje naprednih programa edukacije za samostalno pretraživanje baza podataka.

Društveni značaj: Pokazan je potencijal za uključivanje tema o upotrebi alata generativne umjetne inteligencije i oblikovanja *promptova* u programe informacijske pismenosti, pri čemu su upravo knjižničari prepoznati kao ključni edukatori.

Originalnost/vrijednost: Prema saznanjima autorica, ovo je prva studija koja kombinira istraživanje potencijala alata generativne umjetne inteligencije pri tematskom pretraživanju i utjecaj edukacije korisnika na njihovu učinkovitost. Razvijeni su posebni evaluacijski instrumenti za sustavnu evaluaciju *promptova* i generiranih pretraživačkih nizova namijenjenih medicinskom području čime se otvara rasprava o standardizaciji metodologije vrednovanja rezultata generiranih putem alata generativne umjetne inteligencije u akademskoj i stručnoj zajednici u području informacijskih znanosti.

Ključne riječi: baza podataka MEDLINE/PubMed; ChatGPT; generativna umjetna inteligencija; informacijsko pretraživanje; oblikovanje pretraživačkog niza

Abstract

Purpose: This paper examines the potential and effectiveness of generative artificial intelligence tools (ChatGPT) in designing search strings for the MEDLINE/PubMed database.

Methodology: The study was conducted in two sessions with medical students, with training on prompt and search string design provided between sessions. In both sessions, students created prompts for ChatGPT to generate a search string for MEDLINE/PubMed. The quality of the prompts and the generated strings was evaluated using specially developed instruments (gold standards and checklists). The analysis combined a quantitative approach (scoring the quality of prompts and strings) and a qualitative approach (identifying errors and recurring patterns).

Findings: A positive correlation was found between prompt quality (both overall and in terms of specific elements) and the overall quality of search strings. The second

session demonstrated significant improvement in both prompts and strings, indicating the value of the training. Certain prompt elements were identified as positively correlating with higher-quality generated strings.

Limitations: The study employed one generative AI tool (ChatGPT-3.5) and examined a single database (MEDLINE/PubMed).

Practical implications: The developed evaluation toolkit can serve as a foundation for developing systematic methodologies to assess generative AI outputs in information retrieval. Identified user behaviour patterns provide a basis for designing training programs for independent database searching.

Social implications: The findings highlight the potential for integrating topics on generative AI use and prompt design into information literacy programs, positioning librarians as key educators.

Originality/value: To the best of authors' knowledge, this is the first study to combine research on the potential of generative AI tools for thematic searching and the impact of user education on their effectiveness. It introduces tailored evaluation instruments for the systematic assessment of prompts and generated search strings in the medical domain, opening a discussion on standardization of methodologies for evaluating generative AI outputs within the academic and professional information science community.

Keywords: ChatGPT; generative artificial intelligence; information retrieval; MEDLINE/PubMed database; search string design

1. Uvod

U današnjem ubrzanom znanstvenom i stručnom okruženju svjedoci smo golemom porastu količine dostupnih informacija i objavljene literature. Posebno je taj trend vidljiv u biomedicini, propulzivnom području koje se odlikuje čestim objavljivanjem i brzim akumuliranjem znanstvenih spoznaja, ali i njihovim jednako brzim zastarijevanjem, uz iznimnu društvenu značajnost na međunarodnoj razini. Riječ je o znanjima za koja su jednako zainteresirani i stručnjaci i šira, često nestručna javnost.

Različiti informacijski sustavi i baze podataka ključne su točke u procesu pronalaska relevantnih i pouzdanih medicinskih informacija. Učinkovito informacijsko pretraživanje zahtijeva poznavanje načela izrade pretraživačkih nizova, što uključuje upotrebu pojmova iz kontroliranih rječnika, njihovo kombiniranje sa slobodno oblikovanim ključnim riječima te primjenu logičkih i sintaktičkih pravila pojedine baze.

Razvoj alata generativne umjetne inteligencije (GUI) otvara nove mogućnosti podrške korisnicima pri informacijskom pretraživanju, osobito pri izradi pretraživačkih nizova pomoću posebno osmišljenih *promptova*, tj. uputa prema kojima

alat GUI-a generira traženi odgovor. Iako je GUI vrlo aktualna tema u stručnoj i široj javnosti, broj radova koji ispituju i evaluiraju potencijal alata GUI-a u izradi pretraživačkih nizova za baze podataka – osobito iz perspektive krajnjeg korisnika, često neiskusnoga u izradi i *promptova* i pretraživačkih nizova – još uvijek je ograničen.

1.1. Informacijsko pretraživanje u području biomedicine i zdravstva

Iako je eksponencijalan rast broja objavljenih radova dokaz dinamične prirode medicinskih znanosti, on stvara i svojevrсни paradoks: unatoč sve većoj količini dostupnih informacija i literature, korisnicima postaje sve teže naći odgovor na informacijski upit. Primjerice, sustav PubMed (NLM, 2025) specijaliziran za pretraživanje medicinske literature, sadrži više od 38 milijuna referencija. Godine 2023. u njega je uključeno više od 1,5 milijuna referencija, a napravljeno je više od 3,66 milijardi pretraživanja (NLM, 2024).

Pojmovi PubMed i MEDLINE često se koriste kao sinonimi jer se referiraju na međusobno povezane izvore i alate koji pokrivaju literaturu iz biomedicinskog i javnozdravstvenog područja, a koje razvija i održava Nacionalna medicinska knjižnica SAD-a (*U. S. National Library of Medicine* – NLM). PubMed je besplatno dostupan sustav, tj. sučelje za pretraživanje medicinskih informacija iz više izvora, među kojima je najvažniji bibliografska baza podataka MEDLINE koja indeksira oko 5200 probranih časopisa i sadrži oko 31 milijun referencija (NLM, 2023). Pojednostavljeno, MEDLINE je izvor informacija, a PubMed je alat kojim se te informacije pretražuju. U literaturi se često koriste i kombinacije naziva poput MEDLINE/PubMed i PubMed/MEDLINE, kojima se ističe način pristupanja MEDLINE-u putem PubMeda, no zbog terminološke jednostavnosti i u skladu s ustaljenom praksom imenovanja u akademskoj zajednici, u ovom se radu naziv PubMed koristi kao sveobuhvatan pojam koji označava i sučelje i bazu podataka.

U kontekstu informacijske preopterećenosti ključnu važnost imaju vještine odabira najprikladnijih načina pronalaženja informacija, znanje oblikovanja strategije pretraživanja, kritičkog vrednovanja i uspješna upotreba informacija – što su odlike informacijski pismenog pojedinca (Petrak, 2019: 164) – i to ne samo za stručnu i znanstvenu medicinsku zajednicu nego i za studente koji se tijekom obrazovanja također oslanjaju na kvalitetnu literaturu, osobito tijekom izrade diplomskih i doktorskih radova. „Informacijsko pretraživanje“ (engl. *information retrieval*) (Leščić, 2010: 71) odnosi se na aktivnosti povezane s organizacijom, obradom i pristupom informacijama različitih oblika i formata. Riječ je o procesu pretraživanja i dohvaćanja relevantnih informacija iz različitih skupova podataka (baza podataka, knjižničnih kataloga itd.) u skladu s korisničkim informacijskim potrebama (Chowdhury, 2010: 1–2). Razlikujemo računalno usmjeren pristup in-

formacijskom pretraživanju koji se bavi izgradnjom učinkovitih računalnih sustava s fokusom na obradu upita, algoritme rangiranja i prikaz rezultata pretraživanja, a korisnički usmjeren pristup odnosi se na razumijevanje korisničkoga informacijskog ponašanja (Chowdhury, 2010: 11). Tu je stoga potrebno spomenuti i blizak pojam „informacijskog traženja“ (engl. *information searching*) (Leščić, 2010: 71) koji obuhvaća ponašanje u interakciji s informacijskim sustavom, i to od prepoznavanja informacijske potrebe i oblikovanja pitanja, do formiranja i prilagodbe upita te vrednovanja i odabira rezultata (Wilson, 2000: [49]). U ovom je radu fokus međutim stavljen na pojam informacijskog pretraživanja, i to na njegov računalni aspekt, s obzirom na to da se istraživanje bavi evaluacijom kvalitete pretraživačkih nizova i učinkovitosti alata GUI-a u njihovu oblikovanju.¹

Informacijsko pretraživanje oslanja se na strategiju pretraživanja, čiju okosnicu čine pretraživački nizovi. Strategiju pretraživanja (engl. *search strategy*) promatramo kao cjelokupni put traženja informacija, od postavljanja istraživačkog pitanja, definiranja teme, odabira ključnih riječi i/ili predmetnica iz kontroliranih rječnika, izvora i baza podataka koje će se pretraživati, postavljanja granica, pa do bilježenja dobivenih rezultata itd. (Petrak, 2019: 167; Poropat, 2023: 50–51). S druge strane pretraživački niz (engl. *search string*) predstavlja formuliranje te strategije u konkretan upit koji se unosi u bazu podatka i/ili u tražilicu. Za potrebe, npr., tematskog pretraživanja, pretraživački se nizovi oblikuju od skupova termina iz kontroliranih rječnika i/ili slobodno oblikovanih ključnih riječi, gdje se uključuju sinonimi i varijante pisanja pojmova, uparenih u logičke cjeline pomoću Booleovih operatora, a pritom se mogu koristiti i odabrana metapodatkovna polja (npr. pretraživanje naslova i/ili sažetka) i dodatne mogućnosti (npr. pretraživanje fraza, kraćenje/korjenovanje ključnih riječi) usklađene sa specifičnostima svake pretraživane baze podataka (Lefebvre i sur., 2025; Petrak, 2019: 167–171).

Ključan element pretraživačkih nizova u području biomedicine i zdravstva čini upotreba kontroliranog rječnika „Medical Subject Headings“ (MeSH), predmetnog sustava namijenjenog indeksiranju i pretraživanju medicinske literature, koji razvija spomenuta knjižnica NLM (NLM, 2025a). Pod sadržajnim indeksiranjem (engl. *indexing*) podrazumijevamo iscrpno specifično tematsko opisivanje sadržaja članaka iz časopisa indeksiranih u bazi MEDLINE (Škorić, 2013: 107–108). Preciznost i konzistentnost označivanja sadržaja, zasnovana na NLM-ovim pravilima, osigurava iznimno učinkovito pretraživanje i pronalaženje relevantne literature. Tezaurus MeSH sastoji se od tri vrste zapisa, a to su predmetnice (predmetne

¹ Istraživanje je provedeno sa studentima studija medicine kao ispitanicima čija je uloga bila demonstrirati kako krajnji korisnici, bez stručnog znanja o pretraživanju i kreiranju pretraživačkih nizova te izradi *promptova*, oblikuju upite namijenjene upotrebi u alatu generativne umjetne inteligencije. Fokus istraživanja nije dakle bio na ponašanju korisnika, nego na tehničkim ishodima, tj. kvaliteti *promptova* i pretraživačkih nizova generiranih pomoću alata GUI-a.

odrednice, deskriptori; engl. *headings, descriptors*), potpredmetnice (pododrednice, kvalifikatori; engl. *subheadings, qualifiers*) koje se kombiniraju s predmetnicama da se podrobnije označi o kojem se aspektu raspravlja te dopunski pojmovi *Supplementary Concept Records* (NLM, 2025b; Škorić, 2011: 97–98). Izdanje MeSH-a za 2025. godinu broji 30 956 predmetnica, 76 potpredmetnica i 323 939 dopunskih SCRs-pojmova (NLM, 2025c, 2025d). Svaka MeSH-predmetnica donosi i popis sinonima (tzv. *entry terms*), što omogućuje da se pri pretraživanju obuhvate i različite varijante izraza unutar istoga tematskog koncepta, tj. jedinstvene semantičke cjeline koja obuhvaća neku temu (NLM, 2025e, 2025f). Također, hijerarhijska struktura tezaurusa osigurava i proširivanja i/ili sužavanja pretraživanja, ovisno o tome izabire li se tematski relevantan, i/ili nadređeni ili podređeni koncept.

Knjižničari medicinskih knjižnica imaju ključnu ulogu u izradi kompleksnih pretraživačkih nizova, osobito pri izradi sustavnih pregleda gdje se zahtijeva iznimno visoka preciznost, osjetljivost i sveobuhvatnost pretraživanja literature u više izvora i baza podataka (Spencer i Eldredge, 2018). Takvi pregledi predstavljaju standard pri sintetiziranju znanstvenih spoznaja o nekoj kliničkoj temi, a pritom se informacijski upiti strukturiraju u formi kliničkog pitanja i zatim prevode u strategiju pretraživanja. Taj se proces obavlja kroz različite modele, među kojima je najpoznatiji PICO, koji definira četiri elementa: pacijent/populacija/problem, intervencija, usporedna intervencija i ishod (engl. *Patient/Population/Problem, Intervention, Comparison i Outcome*) (Gosak i sur., 2025; Lefebvre i sur., 2025; Poropat, 2023: 50). Razdvajanje upita u jasne tematske koncepte olakšava izradu pretraživačkog niza i posljedično povećava vjerojatnost pronalaženja relevantne literature (Schardt i sur., 2007).

Izrada takvih preciznih pretraživačkih nizova iziskuje tehnička znanja (razumijevanje načela funkcioniranja baze podataka i pretraživačke sintakse u pozadini), a od knjižničara se očekuje i da u određenoj mjeri razumije kliničku problematiku kako bi mogao konstruirati upit koji vjerno odražava stvarne korisničke potrebe. Kao podršku u tom procesu, spomenut ćemo neke od uputa za izradu i izvještavanje o strategijama pretraživanja te njihovu evaluaciju u sustavnim pregledima. Sveobuhvatan vodič za izradu sustavnih pregleda *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins i sur., 2024) sadrži i iscrpno poglavlje o pretraživanju sa savjetima o informacijskim izvorima, postavljanju strategije pretraživanja, potrebi njezine evaluacije itd. (Lefebvre i sur., 2025). Posebno se naglašava i važnost dokumentiranja te detaljnog i transparentnog izvještavanja o korištenim strategijama pretraživanja, što je također načelo i smjernica PRISMA – *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA Executive, 2025).

Budući da je kvaliteta pretraživačkih nizova ključna za valjanost sustavnih pregleda, u stručnoj je zajednici razvijena praksa njihove provjere i validacije primjenom standardiziranih smjernica za procjenu kvalitete i preciznosti strategija pretraživanja (Lefebvre i sur., 2025; Neilson, 2021). Među njima ističu se smjernice *Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration* (PRESS E&E) (McGowan i sur., 2016) koje pružaju strukturirani okvir za evaluaciju pretraživačkih nizova, čime se osigurava njihova metodološka ispravnost prije primjene i objave u sustavnim pregledima.

S druge strane korisnici, poput studenata koji traže literaturu za diplomski rad ili liječnika u svakodnevnoj praksi, baze podataka uglavnom pretražuju samostalno, što često doživljavaju kao apstraktan i tehnički zahtjevan pothvat, osobito ako nisu upoznati sa specifičnostima rada u bazama i metodologijom izrade pretraživačkih nizova. Za razliku od informacijskih i knjižničarskih stručnjaka, oni se oslanjaju na jednostavnije oblikovane upite, izrečene prirodnim, ponekad i laičkim jezikom, te svoj informacijskih upit posljedično „prevode“ u jednostavnije strategije pretraživanja (Roos, 2015). Zbog toga je potrebno razvijati alate i metode koji će im olakšati snalaženje u složenom informacijskom okruženju, posebice u formuliranju učinkovitih pretraživačkih nizova za pronalaženje relevantnih i provjerenih izvora.

Upravo u potpori korisnicima u izražavanju informacijskih potreba i prevodeanju njihova spontanoga prirodnog jezika u strojno razumljive forme otvara se prostor za upotrebu alata umjetne inteligencije (dalje: alata UI-a). Riječ je o području informacijskog pretraživanja za koje je još 1990-ih K. Spärck Jones istaknula potencijal umjetne inteligencije u njegovu unaprjeđenju (Jones, 1991: 559, 561). Recentna istraživanja o primjeni alata GUI-a u kreiranju pretraživačkih nizova uglavnom su se usmjerila na složene nizove za potrebe sustavnih pregleda (Alshami i sur., 2023; Bourgeois i Ellingson, 2025; Parisi i Sutton, 2024). Rezultati pokazuju da alati GUI-a mogu pomoći u inicijalnoj izradi pretraživačkog niza, no ističe se nužnost ljudske provjere zbog pojave mogućih nepouzdanosti, poput pogrešno generiranih predmetnica iz kontroliranih rječnika, preopćenitih ključnih riječi ili sintaktičkih pogrešaka u strukturiranju niza itd. Međutim, u dostupnoj literaturi i dalje nedostaju istraživanja koja razmatraju korisničku perspektivu – primjerice, studenta ili liječnika bez stručnog iskustva u naprednom pretraživanju baza podataka – čije su informacijske potrebe i razina znanja o pretraživanjima i kreiranju pretraživačkih nizova bitno drukčije od onih informacijskih i knjižničarskih stručnjaka. Upravo će se na taj kontekst osloniti ciljevi i istraživačka pitanja ovog istraživanja.

2. Generativna umjetna inteligencija i izrada promptova

Generativna umjetna inteligencija (engl. *generative artificial intelligence* – GUI) odnosi se na vrstu umjetne inteligencije u kojoj se, na temelju uzoraka naučenih iz velikih količina podataka upotrijebljenih u procesu treniranja, mogu stvarati (generirati) novi sadržaji, npr. tekst, slika, zvuk itd. (Meskó, 2023). U području obrade teksta, pozadinu tehnologija GUI-a čine veliki jezični modeli (engl. *large language models*) koji prepoznaju obrasce prirodnog jezika i stvaraju koherentne i kontekstualno primjerene sadržaje (Giray, 2023; Meskó, 2023). Zbog te sposobnosti rekreiranja ljudske komunikacije, alati u obliku *chatbotova* (ChatGPT, Gemini, Claude itd.), posljednjih godina privlače veliku pozornost.

Preliminarni pregled literature pokazao je da ChatGPT koji je razvila tvrtka OpenAI (OpenAI, 2025) ima vodeću poziciju među jezičnim alatima korištenima u području biomedicine i zdravstva. U kolovozu 2023. godine u PubMedu je indeksirano više od 1000 radova u kojima se spominje ChatGPT (Temsah i sur., 2023), dok je u srpnju 2025. godine broj porastao na oko 7000 radova. Taj porast potvrđuje snažan interes znanstvenika i stalno rastući utjecaj alata GUI-a na promjene paradigmi u medicinskom istraživanju, zdravstvenom sustavu i skrbi za pacijenta, ali i u načinima obrazovanja medicinskih i zdravstvenih profesionalaca. Također, budući da je ChatGPT prvi alat GUI-a temeljen na velikim jezičnim modelima koji je bio predstavljen javnosti krajem 2022. godine (OpenAI, 2022), a usto je besplatno dostupan te za njegovo korištenje nije bilo potrebno tehničko predznanje, upravo je njime pokrenut trend šire primjene tehnologije GUI-a među različitim korisničkim skupinama.

Osnovna komponenta komunikacije s alatima GUI-a jest *prompt*, tj. tekstualna uputa koju korisnik unosi u alat kako bi ga usmjerio prema generiranju željenog odgovora (Giray, 2023; Heston i Khun, 2023; Zamfirescu-Pereira i sur., 2023). Kvaliteta *prompta* izravno utječe na kvalitetu izlaznog odgovora: slabo definiran *prompt* često rezultira površnim, nepreciznim i izmišljenim odgovorima, dok promišljeni *prompt* može potaknuti alat na preciznije i relevantnije rezultate. Neki od elemenata koji se mogu u njega ugraditi jesu jasno opisivanje zadatka (opisati što alat treba napraviti, npr. sažeti tekst, generirati popis natuknica, prevesti tekst...), specificiranje konteksta (navesti informacije o svrsi zadatka, korisniku odgovora, specifičnosti teme...) i definiranje izlaznog rezultata (definirati izgled krajnjeg rezultata, tzv. *output*, npr. popis, tablica, duljina generiranog teksta...) (Gao i sur., 2024: 1921–1922; Giray, 2023; Heston i Khun, 2023; Meskó, 2023). Uz osnovna načela valja spomenuti i napredne tehnike kreiranja *promptova* koje dodatno povećavaju kvalitetu izlaza. To su zadavanje uloge i razine stručnosti alatu, postavljanje zamišljenog scenarija koji je potrebno razriješiti, može se odrediti stil i ton odgovora, itd. (Maaz i sur., 2025; Meskó, 2023). Različiti principi komuniciranja s alatima GUI-a podrazumijevaju npr. *zero-shot*, *few-shot*, *chain-of-thought* i druge

načine *promptanja*. Kod *zero-shot promptanja* alatu GUI-a daju se samo osnovne upute, bez navođenja primjera željenog izlaza/rezultata, što polazi od pretpostavke da model može izvesti zadani zadatak samo na temelju svojega implicitnog znanja (Gao i sur., 2024: 1922; Gotlib-Małkowska i sur., 2025: 11; Heston i Khun, 2023). S druge je strane *few-shot* kreiranje *promptova* u koje se ugrađuju konkretni primjeri očekivanog rezultata kako bi se alat GUI-a usmjerio k željenom odgovoru (Gao i sur., 2024: 1922; Heston i Khun, 2023). *Chain-of-thought promptanjem* alat se navodi da postavljeni zadatak rješava postupno, kroz više koraka, što poboljšava sposobnost njegova zaključivanja (Gotlib-Małkowska i sur., 2025: 11).

Unatoč velikom potencijalu, pozornost se mora skrenuti i na izazove rada s alatima GUI-a i iznimnu potrebu kritičkog vrednovanja generiranih rezultata. Budući da je riječ o statističkim modelima koji izlazni sadržaj kreiraju na temelju vjerojatnosti, moguće su tzv. halucinacije, tj. generiranje izmišljenih i netočnih odgovora (Giray, 2023; OpenAI, 2022). Osim toga, alati GUI-a funkcioniraju po principu „crne kutije“ (engl. *black box*), što znači da korisnik nema informaciju o tome na kojim su podacima oni trenirani niti koje su zakonitosti njihova rada (Gotlib-Małkowska i sur., 2025: 15; Wang i sur., 2023: [9]). Zbog te se netransparentnosti postavljaju pitanja etičke prirode kao što su odgovornost za potencijalno širenje dezinformacija, pristranost u odgovaranju, pitanje privatnosti, neetična upotreba alata u školstvu i znanosti, itd. (Miltenoff, 2024).

Osim što postoje različiti načini vrednovanja performansi alata GUI-a koje se odnose na generirane rezultate, vrednovanje je moguće i na razini samog *prompta*, gdje se procjenjuje kako je on strukturiran i može li kao takav potaknuti željeni odgovor. Međutim, zasad nema univerzalnog standarda za procjenu njihove kvalitete te se često pribjegava razvoju *ad hoc* pristupa prilagođenih specifičnom istraživačkom ili radnom zadatku (Bakst, 2025). Upravo stoga razvoj vještine izrade kvalitetnih *promptova* – uz obavezno vrednovanje dobivenih generiranih odgovora – postaje dijelom informacijske pismenosti, osobito u profesionalnom okruženju u kojem preciznost upita postavljenoga alatu GUI-a izravno utječe na kvalitetu i korisnost generiranih informacija.

3. Istraživanje

3.1. Ciljevi istraživanja

Temeljni cilj ovog istraživanja bio je ispitati potencijal i učinkovitost alata GUI-a – konkretno alata ChatGPT – u procesu oblikovanja strategija pretraživanja i pretraživačkih nizova za bazu podatka PubMed, na primjeru studenata studija medicine kao korisnika alata.

Specifični ciljevi su:

1. procijeniti učinkovitost alata GUI-a (ChatGPT) u procesu izrade pretraživačkih nizova za bazu podataka PubMed, na temelju analize razine kvalitete generiranih nizova u odnosu na zlatni standard i provjereni popis

2. istražiti povezanost između kvalitete *prompta* i kvalitete generiranoga pretraživačkog niza, s naglaskom na prepoznavanju elemenata *prompta* koji imaju najveći utjecaj na točnost, potpunost i sintaktičku ispravnost niza

3. evaluirati kvalitetu studentskih *promptova* i pretraživačkih nizova, prije i nakon edukacije studenata, s ciljem utvrđivanja učinka ciljane edukacije o principima izrade kvalitetnih i učinkovitih *promptova* za alate GUI-a i oblikovanja kvalitetnih pretraživačkih nizova za bazu podataka PubMed.

3.2. Metodologija

3.2.1. Ustroj istraživanja

Istraživanje je koncipirano tako da uključuje dvije točke mjerenja: provedeno je u dvije odvojene sesije istraživanja, u koje je bila pozvana ista skupina studenata, a održane su prije i nakon edukacije o principima izrade pretraživačkih nizova i kreiranja *promptova*. Time je stvoren referentni okvir za ocjenjivanje napretka studenata u dijelu koji se tiče kreiranja *promptova* te povećanja kvalitete pretraživačkih nizova koje generiraju alati GUI-a.

U svakoj sesiji studenti su kreirali *promptove* za alat GUI-a na temelju istoga kliničkog pitanja, nakon čega je alat GUI-a generirao pretraživačke nizove. Autorice su provele evaluaciju svih izrađenih *promptova* i pretraživačkih nizova zasniivajući ocjenjivanje na prethodno izrađenim zlatnim standardima i provjerenim popisima posebno osmišljenima za potrebe ovog istraživanja.

3.2.2. Istraživački instrumenti i materijali

3.2.2.1. Informacijski upit

Za potrebe istraživanja osmišljen je informacijski upit temeljen na stvarnim studentskim informacijskim upitima vezanima za teme diplomskih radova na studiju medicine, tj. primjerima iz prakse informacijske službe Središnje medicinske knjižnice Medicinskog fakulteta Sveučilišta u Zagrebu. Odabran je upit iz interne medicine s obzirom na to da je riječ o području s najzastupljenijim brojem tema studentskih ocjenskih radova, što je vidljivo i u Digitalnom repozitoriju Medicinskog fakulteta Sveučilišta u Zagrebu „Dr Med“ (2025).

Budući da se studenti pri pisanju diplomskih radova, a jednog dana i kao liječnici u svom profesionalnom radu, susreću upravo s klinički formuliranim upitima, u istraživanju je odlučeno da će se informacijski upit predstaviti u obliku kliničkog

pitanja jer će tako pružiti temelj za realistično testiranje učinkovitosti alata GUI-a u kreiranju relevantnih pretraživačkih nizova.

Kliničko pitanje koje je u ovom istraživanju bilo temelj za sve istraživačke zadatke – od izrade zlatnih standarda *prompta* i pretraživačkog niza autorica rada pa do studentskih zadataka vezanih za rad s ChatGPT-em tijekom dvije sesije – glasi: „Smanjuje li primjena statina rizik od moždanog udara kod odraslih osoba s hiperlipidemijom?“ (vidi i Prilog 1. *Kliničko pitanje*).

3.2.2.2. *Alat generativne umjetne inteligencije ChatGPT*

Alat generativne umjetne inteligencije korišten tijekom obje sesije bio je ChatGPT, i to besplatna verzija, model GPT-3.5. Većina se studenata prijavila u sustav, dok ih je nekolicina upotrijebila verziju alata dostupnu bez prijave.

Široka dostupnost, potaknuta mogućnošću besplatne upotrebe, njegovo rašireno korištenje i već uspostavljen interes u znanstvenoj zajednici bili su razlozima odluke da se u ovom istraživanju upravo ChatGPT koristi kao reprezentativan primjer alata GUI-a.

3.2.2.3. *Zlatni standardi prompta i pretraživačkog niza*

U kontekstu informacijskih znanosti, zlatni standardi odnose se na referentne modele za usporedbu rezultata stvarnih informacijskih upita s unaprijed definiranim optimalnim rješenjima. Oba zlatna standarda – i *prompta* i pretraživačkog niza – autorice su izradile prije početka analize i ocjenjivanja, u skladu s najboljim principima oblikovanja kvalitetnih upita za alate GUI-a te praksama u području informacijskog pretraživanja i kreiranja kvalitetnih pretraživačkih nizova.

Zlatni standard prompta

Zlatni standard *prompta* predstavlja optimalno formuliran *prompt* za alat GUI-a s ciljem izrade pretraživačkog niza za pretraživanje baze PubMed (vidi Prilog 2. *Zlatni standard prompta* za alat GUI-a). Oblikovan je na temelju principa kreiranja kvalitetnih upita za alate GUI-a, kao što su jasna uputa, definiranje očekivanog rezultata, zadavanje uloge alatu i opisivanje konteksta, što sve služi usmjeravanju alata k željenom odgovoru. Budući da je riječ o *promptu* koji ima cilj generiranja pretraživačkog niza, u njega su uključene i upute koje se tiču specifičnih zahtjeva vezanih za rad s bazom PubMed.

Zlatni standard služio je kao referentni okvir za analizu i ocjenjivanje studentskih *promptova* u obje sesije. Svaki je *prompt* evaluiran uspoređivanjem s optimalnim zlatnim *promptom*, a zadovoljavanje kriterija i odstupanja bilježena su i bodovno pomoću posebnoga provjerenog popisa.

Zlatni standard pretraživačkog niza

Zlatni standard pretraživačkog niza optimalno je oblikovan pretraživački niz namijenjen pretraživanju baze PubMed, formuliran sukladno zadanom kliničkom pitanju (vidi Prilog 3. *Zlatni standard pretraživačkog niza za bazu PubMed*). Oblikovan je kombiniranjem sadržajnih i sintaktičkih zahtjeva, kao što su upotreba tezaurusa MeSH i ključnih riječi, Booleovih operatora te specijalnih znakova za pretraživanje fraza i kraćenje ključnih riječi u bazi PubMed. Navedeni niz, utemeljen na standardiziranim principima i iskustvima autorica u kreiranju složenih pretraživačkih nizova za različite baze podatka, služio je kao referentni okvir za evaluaciju pretraživačkih nizova generiranih pomoću alat GUI-a tijekom obje sesije. Svaki je niz uspoređen sa zlatnim nizom te bodovan pomoću posebnoga provjerenog popisa.

3.2.2.4. Provjereni popisi za ocjenjivanje kvalitete prompta i pretraživačkog niza

Provjereni popis (engl. *checklist*) alat je koji se koristi za sustavno vrednovanje neke pojave, što je omogućeno standardizacijom kriterija i elemenata koji se evaluiraju. Za ovo su istraživanje razvijena dva istraživačka instrumenta: provjereni popis za ocjenjivanje *promptova* i provjereni popis za ocjenjivanje pretraživačkih nizova. Razvijani su komplementarno uz zlatne standarde za *prompt* i pretraživački niz, pri čemu zlatni standardi definiraju optimalne osobitosti, tj. vršna su referentna vrijednost pri ocjenjivanju kvalitete, a provjereni popisi služe kao instrument za mjerenje, evaluaciju poštivanja ili odstupanja od zadanih osobitosti. Bodovno vrednovanje osiguralo je preciznu kvantitativnu analizu i mogućnost usporedbe različitih *promptova* i pretraživačkih nizova.

Provjereni popisi sastoje se od popisa kriterija, tj. elemenata koji se ocjenjuju, te opisa pojedinačnih kriterija kao svojevrsnog vodiča za dosljedno postupanje pri ocjenjivanju elemenata *promptova* i pretraživačkih nizova. Osim smjernica za bodovanje, doneseni su i primjeri dobre i loše prakse.

Provjereni popisi i sustavi bodovanja kreirani su prije početka prikupljanja podatka i analize studentskih zadataka kako bi se osigurala objektivnost evaluatora pri ocjenjivanju.

Provjereni popis za ocjenjivanje kvalitete prompta

U svrhu kvantitativnog ocjenjivanja kvalitete *promptova* razvijen je provjereni popis namijenjen standardiziranom bodovanju elemenata *prompta* koji utječu na generiranje kvalitetnih odgovora u alatima GUI-a (vidi Prilog 4. *Provjereni popis za ocjenjivanje kvalitete prompta za alat GUI-a namijenjenog izradi pretraživačkog niza za bazu PubMed*). Budući da ne postoji standardizirani okvir za procjenu

promptova u kontekstu generiranja pretraživačkih nizova za baze podatka, osmišljen je popis prilagođen specifičnom kontekstu ovog istraživanja.

Uz kategorije temeljene na općim principima kreiranja kvalitetnih uputa za alate GUI-a, uvršteni su i elementi važni za generiranje pretraživačkog niza za bazu PubMed, s usmjerenjem na korisnike bez prethodne edukacije o pretraživanju baza podatka i kreiranju pretraživačkih nizova.

Provjereni popis sastoji se od 11 kategorija koje uključuju zadavanje uloge alatu GUI-a, opis konteksta upita, navođenje naziva PubMed i MeSH, spominjanje fraza „kliničko pitanje“ i „pretraživački niz“, korištenje ključnih riječi i upotrebu Booleovih operatora, pretraživanje fraza, kraćenje riječi te pretraživanje po naslovu i sažetku. Jedan bod dodijeljen je ako je neki kriterij iz popisa zadovoljen, a nula ako nije zadovoljen. Uz navedeno, dodana je i kategorija za evidentiranje jezičnih pogrešaka u *promptovima* iako nije uzimana u obzir pri bodovanju.

Provjereni popis za ocjenjivanje kvalitete pretraživačkog niza

Za objektivnu procjenu kvalitete pretraživačkih nizova kreiranih pomoću alata GUI-a razvijen je provjereni popis za ocjenjivanje nizova namijenjenih za pretraživanje baze PubMed (vidi Prilog 5. *Provjereni popis za ocjenjivanje kvalitete pretraživačkog niza za bazu PubMed*). Utemeljen je u stručnim smjernicama (npr. McGowan i sur., 2016: 37–40, tablice 8 i 9) i na praktičnom iskustvu autorica u pretraživanju baza podataka.

Prilagođen je za evaluaciju nizova koje kreiraju korisnici baza koji nisu upoznati s njihovim naprednim odlikama i pravilima oblikovanja složenih pretraživačkih nizova. Kriteriji u provjerenom popisu stoga su usmjereni na funkcionalnu ispravnost i sadržajnu primjerenost kreiranoga pretraživačkog niza.

Provjereni popis podijeljen je u šest glavnih kategorija, s dodatnim potkategorijama: 1) „Preoblikovanje kliničkog pitanja u tematske koncepte“ (npr. zastupljenost i ispravnost koncepata reprezentiranih u kliničkom pitanju), 2) „Predmetnice iz tezaurusa Medical Subject Headings (MeSH)“ (npr. ispravnost i relevantnost MeSH-predmetnica), 3) „Ključne riječi“ (npr. zastupljenost ključnih riječi i dodatnih sinonima), 4) „Sintaksa pretraživačkog niza“ (npr. upotreba Booleovih operatora, korištenje zagrada za grupiranje pojmova, navodnika za pretraživanje fraza, kraćenje riječi zvezdicom itd.), 5) „Granice pretraživanja i filtri“ (prema dobnim odlikama, vremenu objave rada itd.) i 6) „Dodatni pokazatelji kvalitete“ (npr. prepoznavanje i strukturiranje upita prema modelu PICO).

Svaka kategorija sastoji se od više kriterija uz koje je dodana i rubrika s njihovim opisom i smjernicama za postupanje pri ocjenjivanju. Većina kriterija ocjenjuje se binarno, tj. dodjeljivanjem bodova 1 ili 0, dok je u kriterijima koji se tiču sadržajnog aspekta predviđena mogućnost dodjeljivanja i negativnih bodova

u slučajevima kad korištenje nekog elementa rezultira neodgovarajuće postavljanim pretraživačkim nizom (npr. preširoke/preuske ključne riječi, izmišljene MeSH-predmetnice itd.).

3.2.3. Sudionici istraživanja

Sudionici istraživanja bili su studenti 5. godine integriranoga prijediplomskog i diplomskog studija medicine na Medicinskom fakultetu Sveučilišta u Zagrebu koji polaze obavezni kolegij „Uvod u znanstveni rad“. U izvođenju nastave kolegija sudjeluju i knjižničari Središnje medicinske knjižnice koji studente poučavaju o medicinskim informacijama, znanstvenim publikacijama, informacijskim izvorima, radu u bazama podataka itd. (Škorić i sur., 2012).

Istraživanje je provedeno tijekom vježbi i seminara koje vode knjižničari Središnje medicinske knjižnice. Sudionici su prije početka obaviješteni o svrsi i načinu provođenja istraživanja. Sudjelovanje je bilo dobrovoljno, uz mogućnost odustajanja od sudjelovanja u istraživanju bez posljedica. Istraživanje je provedeno poštujući etička načela, podaci su se prikupljali i analizirali anonimno te su poduzete mjere za zaštitu privatnosti i osiguranje povjerljivosti prikupljenih podataka.

U istraživanju su sudjelovali polaznici 4. turnusa kolegija „Uvod u znanstveni rad“ tijekom ak. god. 2024./25., od 23. lipnja do 11. srpnja 2025. godine. Istraživanje se sastojalo od dvije sesije koje su se provodile u razmaku od dva tjedna, a poziv za sudjelovanje u svakoj sesiji upućen je istoj skupini studenata. Zbog organizacijskih okolnosti vezanih za izvođenje nastave i različitog stupnja studentskog odaziva, sesije nisu homogene što se tiče broja ispitanika. Istraživanju je pristupilo 28 studenata tijekom prve i 53 studenta tijekom druge sesije. Na temelju unaprijed određenih kriterija isključivanja, u analizu je uključeno 20 studenata iz prve i 46 studenata iz druge sesije (prva sesija: $N = 20$, isključeno 8; druga sesija: $N = 46$, isključeno 7). Razlozi isključivanja bili su nepravilno i/ili nepotpuno prepisano kliničko pitanje, nepostojanje izričite upute u *promptu* kojom se zahtijeva izrada pretraživačkog niza ili strategije pretraživanja, dostavljanje necjelovitog sadržaja *prompta* i nepostojanje generiranoga pretraživačkog niza u odgovoru.

3.3. Tijek istraživanja

3.3.1. Prva sesija

U prvoj sesiji studentima je predstavljen kontekst zadatka, a zatim se od njih zatražilo da – na temelju unaprijed zadanoga kliničkog pitanja – samostalno svojim riječima osmisle *prompt* kojim će ChatGPT-u zadati uputu da generira pretraživački niz za bazu PubMed. Prva je sesija održana na samom početku nastave spomenutog kolegija te studenti nisu pohađali prethodnu edukaciju o pretraživanju baza podataka i oblikovanju pretraživačkih nizova, kao ni o kreiranju *promptova*.

Studenti su rezultate dostavljali elektroničkom poštom, i to dijeljenjem poveznice na razgovor (engl. *thread*) ili kopiranjem sadržaja razgovora s ChatGPT-em, a prikupljali su se podaci i o njihovu *promptu* i o cjelovitom odgovoru dobivenom od ChatGPT-a.

3.3.2. Druga sesija

Drugo sesiji istraživanja prethodila je faza edukacije u kojoj se studente tijekom nastave poučavalo o izvorima medicinskih informacija i radu s bazom PubMed, što je uključivalo osnove rada i jednostavno pretraživanje, složeno tematsko pretraživanje i kreiranje pretraživačkih nizova, poučavanje o razlici između predmetnica iz tezaurusa MeSH i slobodno oblikovanih ključnih riječi, Booleovoj logici, modelu PICO u kontekstu medicine utemeljene na dokazima itd. Predstavljena su im osnovna pravila kreiranja učinkovitih *promptova* za alate GUI-a, kao i primjeri dobrih praksi preuzeti iz recentne literature.

U drugoj sesiji sudjelovali su isti studenti koji su imali isti zadatak i upotrebljavali su isti alat GUI-a, ali ovaj put formulirajući razrađeniji *prompt* temeljen na saznanjima s nastave o oblikovanju pretraživačkih nizova i kreiranju *promptova*. Rezultate su na isti način ponovno dostavljali autoricama, nakon čega su one pristupile njihovoj analizi.

3.4. Evaluacija i analiza podataka

Autorice ovog rada sve su *promptove* i pretraživačke nizove neovisno ocijenile prema unaprijed izrađenim provjerenim popisima, uspoređujući *promptove* i nizove prema također unaprijed izrađenim zlatnim standardima i *prompta* i pretraživačkog niza. U slučaju neslaganja u rezultatima evaluacije, pristupilo su dodatnim provjerama u cilju osiguravanja konsenzusa oko ocjenjivanja.

Dobiveni rezultati opisani su deskriptivnom statistikom. Uz frekvenciju, izračunata je prosječna vrijednost i pripadajuća standardna devijacija za ukupan broj bodova *promptova* i ukupan broj bodova pretraživačkih nizova. Normalna distribucija ukupnog broja bodova *promptova* i pretraživačkih nizova, kao i broja bodova svakog elementa pretraživačkog niza provjerena je Shapiro-Wilkovim testom, dok je homogenost varijanci procijenjena Leveneovim testom. Nadalje, navedeni podaci analizirani su kao nezavisni zbog različitog stupnja odaziva studenata u dvije sesije istraživanja, odnosno za testiranje značajnosti korišten je Mann-Whitneyev U test. Analiza rang korelacije prema Spearmanu provedena je kako bi se utvrdilo postoji li povezanost između pojedinih elemenata *prompta* i kvalitete pretraživačkog niza. Za obradu i analizu podataka korišten je program za statističku analizu JASP, verzija 0.19.3 (JASP Team, 2024).

Osim kvantitativne analize, pristupilo se i kvalitativnoj analizi koja je pridonijela prepoznavanju obrazaca pogrešaka, boljem razumijevanju korisničkih izazova pri upotrebi alata GUI-a te identifikaciji elemenata čijim bi se unaprjeđenjem mogla posredno povećati kvaliteta generiranih pretraživačkih nizova.

4. Rezultati i rasprava

Ukupno je analizirano 20 *promptova* i pretraživačkih nizova iz prve te 46 *promptova* i pretraživačkih nizova iz druge sesije istraživanja. Prikaz rezultata za obje vrste analiziranih podataka i za obje sesije, njihova interpretacija i rasprava u nastavku poglavlja organizirani su u nekoliko cjelina: kvaliteta *promptova* i pretraživačkih nizova prema ostvarenom ukupnom broju bodova, analiza elemenata *promptova* i nizova prema elementima iz provjerenih popisa i njihova interpretacija te rasprava o povezanosti između elemenata *promptova* i pretraživačkog niza.

4.1. Evaluacija kvalitete *promptova*

Pri evaluaciji *promptova* korišten je zlatni standard *prompta* (vidi Prilog 2.) koji služi kao pokazatelj maksimalne kvalitete i referentna točka za procjenu studentskih *promptova*, a komplementaran je provjerenom popisu za *promptove* (vidi Prilog 4.). Zbroj bodova svih kriterija unutar provjerenog popisa predstavljao je ukupnu ocjenu pojedinog *prompta*, pri čemu je najveći mogući broj bodova po *promptu* bio 11 bodova.

U nastavku slijedi prikaz ostvarenih bodova za svaku sesiju istraživanja te analiza elemenata *promptova* prema kriterijima provjerenog popisa kako bi se procijenile razlike u kvaliteti *promptova* između dviju sesija i njihov mogući utjecaj na kvalitetu pretraživačkih nizova te ispitalo postoji li učinak edukacije između dviju sesija.

4.1.1. Ukupno ostvaren broj bodova prema provjerenom popisu

U prvoj sesiji *promptovi* su ocijenjeni u rasponu od 1 do 4 boda, pri čemu ih je čak 75 % (15/20) ostvarilo samo 2 boda. U drugoj sesiji distribucija bodova znatno je šira: *promptovi* su ocijenjeni u rasponu od 2 do 9 bodova, a gotovo polovica (45,6 %; 21/46) biva ocijenjena sa 6 i/ili više bodova. Detaljan prikaz distribucije bodova u obje sesije prikazan je u tablici 1.

Tablica 1. Distribucija bodova kvalitete *promptova* koje su studenti zadali alatu ChatGPT u svrhu izrade pretraživačkog niza za bazu PubMed (Sesija 1 i Sesija 2)

Ostvaren broj bodova po <i>promptu</i> (najviše 11 bodova – zlatni standard)	f (Sesija 1)	f (Sesija 2)	% (Sesija 1)	% (Sesija 2)
11	0	0	0,0 %	0,0 %
10	0	0	0,0 %	0,0 %
9	0	5	0,0 %	10,9 %
8	0	7	0,0 %	15,2 %
7	0	3	0,0 %	6,5 %
6	0	6	0,0 %	13,0 %
5	0	6	0,0 %	13,0 %
4	2	4	10,0 %	8,7 %
3	2	7	10,0 %	15,2 %
2	15	8	75,0 %	17,4 %
1	1	0	5,0 %	0,0 %
0	0	0	0,0 %	0,0 %
Ukupno <i>promptova</i>	20	46	100,0 %	100,0 %
Ukupno bodova po sesiji	45	241		
Prosjeak bodova po <i>promptu</i>	2,25 ^a	5,24 ^b		

Napomena: a, b – različiti eksponenti u istom retku upućuju na statistički značajne razlike (Mann-Whitneyev U test, $p < 0,001$)

Statistička usporedba pokazuje značajan porast ukupne kvalitete *promptova* u drugoj sesiji. Širi raspon i viši prosjek ostvarenih bodova po pojedinom *promptu* u drugoj sesiji (5,24 bodova u drugoj naspram 2,25 bodova u prvoj sesiji; Mann-Whitney U = 117; $p < 0,001$; rang biserijalna korelacija = -0,75) sugeriraju značajan napredak u strukturiranju kvalitetnih *promptova* nakon provedene edukacije.

4.1.2. Analiza elemenata *prompta* i interpretacija rezultata

U nastavku detaljno analiziramo pojedine elemente *promptova* i njihovu zastupljenost prema kriterijima iz provjerenog popisa da bi se bolje razumjelo koji su aspekti utjecali na ukupno ostvarene bodove i njihovu kvalitetu.

Ključna komponenta interakcije s alatima GUI-a – očekivani izlaz (*output*) – definiran je u svim uključenim *promptovima* obiju sesija. Tražili su se „pretraživački niz“, „strategija pretraživanja“ i „search string“, ali i „pomoć u pretraživanju baze PubMed“.

Detaljna usporedba rezultata dviju sesija pokazuje da su *promptovi* iz druge sesije bili strukturiraniji i sadržavali veći broj elemenata iz provjerenog popisa. Fraza „kliničko pitanje“ u prvoj je sesiji navedena u četiri *prompta* (20 %; 4/20), a u drugoj u 18 *promptova* (39,1 %; 18/46). U drugoj se sesiji posebno ističe izričito navođenje upotrebe MeSH-predmetnica (67,4 %; 31/46) i ključnih riječi (41,3 %; 19/46) što nije bilo prisutno u *promptovima* prve sesije. Osim toga, za razliku od prve sesije, dio *promptova* druge sesije spominje i upute povezane s naprednim mogućnostima pretraživanja u bazi PubMed, u skladu s njezinim sintaktičkim pravilima, a to su upotreba Booleovih operatora (41,3 %; 19/46), kraćenje pretraživanih riječi zvjezdicom (37 %; 17/46), pretraživanje fraza pomoću navodnika (32,6 %; 15/46) i mogućnost pretraživanja u određenom PubMedovu polju „Title/Abstract“ (21,7 %; 10/46).

Jezično i terminološki neprecizne pojmove primijećene u prvoj sesiji – „search stream“, „search streak“ i „hiperloipidemija“ – alat ChatGPT uspješno je svladao, tj. točno je interpretirao upit i shvatio što se očekuje u generiranom odgovoru. To potvrđuje visoku toleranciju alata GUI-a, temeljenih na velikim jezičnim modelima, na korisničke pogreške i varijacije u izražavanju zbog implicitnog znanja i naučenih obrazaca, što omogućuje rekonstruiranje korisničkih namjera i ispravljanje nepreciznih unosa. Time se može tumačiti i činjenica da je alat generirao pretraživački niz sintaktički adekvatan za bazu PubMed, iako u jednom *promptu* druge sesije njezin naziv nije bio eksplicitno spomenut.

Zadavanje uloge alatu GUI-a u prvoj se sesiji javlja samo u dva od 20 *promptova* (10 %; 2/20), dok je u drugoj ipak primjetan pomak prema češćem dodjeljivanju uloge ChatGPT-u (26,1 %; 12/46) i opisu konteksta (15,2 %; 7/46). Recentna istraživanja o različitim načinima naprednog kreiranja *promptova* pri informacijskom pretraživanju i generiranju pretraživačkih nizova pokazuju da jasno zadavanje uloge i pojašnjenje konteksta povećavaju preciznost odgovora alata GUI-a (Gao i sur., 2024: 1921; Heston i Khun, 2023; Meskó, 2023; Robinson, Bontekoe i Muellenbach, 2025: 186–187), što u ovom istraživanju nije bilo dovoljno iskorišteno.

Promptovi u prvoj sesiji bili su kreirani minimalistički, upotrebom prirodnog jezika, izbjegavajući stručnu terminologiju i metajezik vezan za područje informacijskog pretraživanja, upotrebljavajući općenite upute davane alatu i bez opisa konteksta i uloge. U nekim slučajevima *promptovi* su kreirani nasumično, oslanjajući se na intuiciju u radu s digitalnim alatima i na metodu pokušaj–pogreška. U literaturi se potvrđuje da je zapravo riječ o prilično čestim izazovima s kojima se neiskusni korisnici susreću pri radu s alatima GUI-a, ali koje je moguće prevladati edukacijom (Zamfirescu-Pereira i sur., 2023).

Značajno bolja kvaliteta *promptova* druge sesije, vidljiva iz prosječnog broja bodova ostvarenog po *promptu* obiju sesija (vidi tablica 1), upućuje i na to da čak

i kratka edukacija o principima izrade kvalitetnih iscrpnijih i ciljanih *promptova* može imati konkretan utjecaj na način na koji su studenti formulirali upite za alat GUI-a. Dobiveni rezultati upućuju informacijske stručnjake na potrebu poučavanja studenata radu s novim alatima kako neformalno, putem radionica ili individualnog pristupa, tako i unutar formalne nastave. Time se otvara pitanje uključivanja *promptanja* kao novog elementa informacijske pismenosti, odnosno – specifičnije – pismenosti u području umjetne inteligencije (engl. *AI literacy*; Miltenoff, 2024: 300).

I recentna literatura ističe važnost uključivanja GUI-a u medicinsku nastavu, naglašavajući veliko značenje optimizacije *promptova* kako bi se alati koristili sigurno i etički ispravno (Heston i Khun, 2023). Uvođenje novih sadržaja s jedne bi strane studentima poslužilo kao priprema za rad u kliničkom okruženju u kojem su alati umjetne inteligencije sve prisutniji (Meskó, 2023), a s druge bi im pomoglo već u nastavnom procesu, jer se upravo edukacijom i zajedničkim radom kako nastavnog i znanstvenog osoblja neke visokoškolske ustanove pa tako i studenata – a ne zabranama – mogu poticati kritičko razmišljanje te kreirati smjernice i pravila za ispravno korištenje tih alata u znanstvenim i obrazovnim krugovima (Abd-Alrazaq i sur., 2023; Chaudhuri i Terrones, 2024: 22–23; Heston i Khun, 2023). Upravo se tu otvara prostor u kojem medicinski knjižničari mogu imati ključnu ulogu, a to je pružanje podrške u korištenju alata GUI-a utemeljene na etičkim načelima i elementima informacijske pismenosti, što su postulati kojima oni svakodnevno i suvereno barataju.

4.2. Evaluacija kvalitete generiranih pretraživačkih nizova

Kvaliteta generiranih pretraživačkih nizova ocijenjena je također pomoću provjerenog popisa (vidi Prilog 5.), a glavna referentna točka pri ocjenjivanju bio je zlatni standard pretraživačkog niza (vidi Prilog 3.). Zbroj bodova svih kriterija, tj. kategorija i potkategorija, unutar provjerenog popisa činio je ukupnu ocjenu pretraživačkog niza, uz najviše 25 bodova po nizu.

Slijedi predstavljanje rezultata za svaku sesiju, s prikazom ostvarenih bodova i analizom elemenata nizova, kako bi se omogućila usporedba i prepoznavanje mogućih obrazaca, s posebnim naglaskom na vezu kvalitete *promptova* i kvalitete pretraživačkih nizova, prepoznavanje elemenata *prompta* koji utječu na kvalitetu niza te procjenu učinka edukacije između sesija.

4.2.1. Ukupno ostvaren broj bodova prema provjerenom popisu

U prvoj sesiji pretraživački nizovi ocjenjivani su od 8 do 22 boda, s time da ih se 75 % (15/20) kreće od 17 do 22 boda. Najviše pretraživačkih nizova u prvoj sesiji (25 %, 5/20) ocijenjeno je sa 17 bodova. U drugoj sesiji distribucija bodova znatno je pomaknuta prema višim vrijednostima na bodovnoj ljestvici: pretraži-

vački nizovi ocijenjeni su od 13 do 24 boda, te ih gotovo tri četvrtine (73,8 %; 34/46) biva ocijenjeno od 21 do 24 boda. Najviše pretraživačkih nizova u drugoj sesiji (30,4 %, 14/46) ocijenjeno je s 22 boda. Detaljan prikaz distribucije bodova obiju sesija donosi se u tablici 2.

Tablica 2. Distribucija bodova kvalitete pretraživačkih nizova za bazu PubMed koje je generirao alat GUI-a ChatGPT (Sesija 1 i Sesija 2)

Ostvaren broj bodova po pretraživačkom nizu (najviše 25 bodova – zlatni standard)	f (Sesija 1)	f (Sesija 2)	% (Sesija 1)	% (Sesija 2)
25	0	0	0,0 %	0,0 %
24	0	6	0,0 %	13,0 %
23	0	6	0,0 %	13,0 %
22	3	14	15,0 %	30,4 %
21	2	8	10,0 %	17,4 %
20	2	3	10,0 %	6,5 %
19	2	1	10,0 %	2,2 %
18	1	3	5,0 %	6,5 %
17	5	0	25,0 %	0,0 %
16	0	0	0,0 %	0,0 %
15	0	2	0,0 %	4,3 %
14	2	1	10,0 %	2,2 %
13	0	2	0,0 %	4,3 %
12	2	0	10,0 %	0,0 %
11	0	0	0,0 %	0,0 %
10	0	0	0,0 %	0,0 %
9	0	0	0,0 %	0,0 %
8	1	0	5,0 %	0,0 %
Ukupno pretraživačkih nizova	20	46	100,0 %	100,0 %
Ukupno bodova po sesiji	349	961		
Prosjek bodova po nizu	17,45 ^a	20,89 ^b		

Napomena: a, b – različiti eksponenti u istom retku upućuju na statistički značajne razlike (Mann-Whitneyev U test, $p < 0,001$)

Kratka deskriptivna statistička usporedba pokazuje porast ukupne kvalitete pretraživačkih nizova u drugoj sesiji. Viši prosjek ostvarenih bodova po pojedinom pretraživačkom nizu u drugoj sesiji (20,89 bodova u drugoj naspram 17,45 bodova u prvoj sesiji; Mann-Whitney U = 187,5; $p < 0,001$; rang biserijalna ko-

relacija = -0,59) pokazuje značajan napredak u općoj kvaliteti nizova generiranih u alatu GUI-a. To dodatno naglašava snažan učinak edukacije na kvalitetu kako kreiranja *prompta* tako posljedično i na kvalitetu generiranog niza.

4.2.2. Analiza elemenata pretraživačkog niza i interpretacija rezultata

Rezultati analize elemenata pretraživačkih nizova i njihove zastupljenost u obje sesije prema kriterijima iz provjerenog popisa pokazuju jasan napredak u kvaliteti nizova druge sesije: nizovi su bili elaboriraniji i sadržavali veći broj elemenata iz provjerenog popisa.

Koncepti. Četiri ključna koncepta iz zadanoga kliničkog pitanja – hiperlipidemija, statini, moždani udar i odrasle osobe – ispravno su prepoznata u velikom dijelu pretraživačkih nizova u obje sesije. U prvoj sesiji prepoznati su u 90 % (18/20), a u drugoj u 93,5 % (43/46) nizova, dok je u preostalim slučajevima nedostajao po jedan koncept. Neodgovarajuće postavljene preuski i/ili preširoki koncepti u prvoj se sesiji javljaju u 40 % (8/20), a u drugoj u samo 10,9 % (5/46) nizova, što upućuje na povećanu preciznost u „prevođenju“ informacijskog upita u strategiju pretraživanja i posljedično preveniranje sužavanja ili širenja rezultata pretraživanja baze izvan optimalnog opsega.

MeSH-predmetnice. Tri MeSH-predmetnice promatrane u ovoj kategoriji – *Hyperlipidemias*, *Hydroxymethylglutaryl-CoA Reductase Inhibitors* i *Stroke* – naslanjaju se na tri već definirana koncepta: hiperlipidemija, statini i moždani udar. Iako su bile upotrebljavane i u pretraživačkim nizovima proizašlima iz prve sesije, upotreba MeSH-predmetnica u nizovima ipak je češća u drugoj sesiji: 95,7 % (44/46) nizova druge naprama 85 % (17/20) nizova prve sesije.

Izrazi i oznake PubMedova polja specijaliziranog za pretraživanje po MeSH-predmetnicama („[MeSH Terms]“ ili „[Mesh]“) potpuno su točno upotrijebljeni u 12 slučajeva iz prve (60 %; 12/20) i čak 40 slučajeva druge (87 %; 40/46) sesije.

Što se tiče prisutnosti točnih MeSH-predmetnica u nizovima, uz usporednu ispravnu upotrebu oznake za polje specijalizirano za pretraživanje po MeSH-u („[MeSH Terms]“ ili „[Mesh]“), u drugoj sesiji isto je tako primjetan porast u odnosu na prvu sesiju. U drugoj sesiji i MeSH-predmetnice i oznake polja ispravno su navedeni u 40 od 44 pretraživačkih nizova s MeSH-predmetnicama (90,9 %), dok je u prvoj sesiji sve točno navedeno u samo 12 od 17 nizova s MeSH-om (70,6 %).

U obje se sesije javljaju pretraživački nizovi u kojima su dodane i neodgovarajuće, preusko i/ili preširoko postavljene MeSH-predmetnice: 23,5 %, tj. 4 slučaja od 17 nizova s MeSH-om u prvoj sesiji, te 22,7 %, tj. 10 slučajeva od 44 niza s MeSH-om u drugoj sesiji. Također, u obje su sesije zabilježene halucinacije u ge-

neriranim odgovorima, no primjetan je trend smanjenja izmišljenih MeSH-predmetnica u drugoj sesiji: 6,8 %, tj. 3 slučaja od 44 nizova s MeSH-om u drugoj, naspram 29,4 %, tj. 5 slučajeva od 17 nizova u prvoj sesiji. Zanimljivo je da je u svih osam nizova halucinirana uvijek ista MeSH-predmetnica – *Hydroxymethylglutaryl-CoA Reductase Inhibitors* – i da je bila jednako neispravno navođena kao termin „statins“, uvriježen u komunikaciji i u literaturi.

Kako su se svi pretraživački nizovi analizirani u ovom istraživanju manualno unosili u bazu PubMed radi povjere njihove sintaktičke ispravnosti, zanimljivo je da je izmišljena/halucinirana MeSH-predmetnica imala vrlo ograničen utjecaj na funkcionalnost pretraživačkog niza. Razlog tomu vjerojatno leži u činjenici što je termin „statins“ službeno zabilježen kao sinonim (tzv. *entry term*) za MeSH-predmetnicu *Hydroxymethylglutaryl-CoA Reductase Inhibitors*. Kao takvoga, baza PubMed automatski ga u pozadini svojim pretraživačkim algoritmima i postavkama mapira, tj. „prevodi“ u odgovarajući MeSH-termin, stoga halucinirani termin nije uzrokovao sintaktičku pogrešku u generiranom pretraživačkom nizu.

Pojava izmišljenih MeSH-predmetnica u ovom istraživanju u skladu je s nalazima u literaturi koji upozoravaju na mogućnost haluciniranja alata GUI-a pri generiranju MeSH-termina za potrebe izrade pretraživačkih nizova (Bourgeois i Ellingson, 2025: 5–6; Gosak i sur., 2025; Gotlib-Małkowska i sur., 2025: 16; Reed i Barnett, 2025; Wang i sur., 2023: [8–9]). U njima se također stoga opetovano spominje važnost ljudske validacije takvih pretraživačkih nizova i osobito MeSH-termina kao njihove okosnice. Naše istraživanje tu međutim pokazuje i da viša kvaliteta *prompta*, osobito kad se u *promptu* eksplicitno sugerira upotreba tezaurusa MeSH, spomene fraza „kliničko pitanje“ i upotreba ključnih riječi, smanjuje učestalost halucinacija i posljedično povećava navođenje točnih MeSH-termina (više u poglavlju 4.3.).

Ključne riječi. Analiza upotrebe ključnih riječi u generiranom pretraživačkom nizu također pokazuje napredak u drugoj sesiji. U svim pretraživačkim nizovima obiju sesija korištena je barem po jedna ključna riječ kao pandan trima očekivanim MeSH-predmetnicama. Dodatno naveden bar po još jedan sinonim za svaku od tri spomenute ključne riječi u prvoj se sesiji javlja u 75 % (15/20), a u drugoj u 82,6 % (38/46) nizova. Također je primjetno i smanjenje navođenja neodgovarajućih preuskih i/ili preširokih ili potpuno neodgovarajućih ključnih riječi u drugoj sesiji: u 21,7 % (10/46) u drugoj nasuprot 40 % (8/20) nizova u prvoj sesiji. U nekim nizovima ChatGPT generirao je i do 25 ključnih riječi, povezanih termina i sinonima, što potvrđuje inherentnu sposobnost alata GUI-a temeljenih na velikim jezičnim modelima za leksičku obradu. Ta odlika, već opisana u literaturi, potvrđuje potencijal pri osmišljavanju učinkovitih strategija pretraživanja (Alshami i sur., 2023; Gitman, Maxwell i Gamble, 2025: 5).

Sintaksa pretraživačkog niza. Napredne funkcionalnosti povezane s kreiranjem pretraživačkih nizova i posebnosti rada u bazi PubMed također su češće u nizovima druge sesije. Kraćenje riječi zvjezdicom u prvoj je sesiji upotrijebljeno u tri slučaja (15 %; 3/20), a u drugoj u 25 nizova (54,3 %; 25/46). Pretraživanje fraza pomoću navodnika u prvoj sesiji koristi se u 10 nizova (50 %; 10/20), a u drugoj u 39 nizova (84,8 %; 39/46). Navođenje oznake polja „[Title/Abstract]“ ili „[tiab]“ u pretraživačkom nizu, čime se ključne riječi pretražuju u PubMedovu polju specijaliziranom za pretraživanje po naslovu i sažetku publikacije, u prvoj je sesiji navedeno u tri niza (15 %; 3/20), a u drugoj u njih 17 (37 %; 17/46). Booleovi operatori (AND i OR) i zagrade, čija je funkcija logičko grupiranje i razdvajanje koncepata, ispravno su korišteni i konzistentno primijenjeni u svim nizovima obiju sesija. Iako su generirani nizovi uključivali već spominjane značajke – npr. kraćenje riječi, pretraživanje fraza – i generirali zadovoljavajuće nizove, nisu dostigli razinu postavljenu zlatnim standardom te smatramo da bi iskusni stručnjak vrlo vjerojatno dodatno prilagodio sintaksu pretraživačkog niza da bi se postigla veća specifičnost pretraživanja. Slični stavovi spominju se i u literaturi (Alaniz, Vu i Pfaff, 2023; Gosak i sur., 2025; Gotlib-Małkowska i sur., 2025: 16), gdje se također navodi da takvi nizovi i dalje predstavljaju pomoć stručnjacima jer povećavaju učinkovitost procesa izrade strategije pretraživanja i smanjuju vrijeme potrebno za izradu inicijalnoga pretraživačkog niza.

Granice pretraživanja i filtri. U popratnom odgovoru, generiranom uz pretraživački niz, ChatGPT samostalno je sugerirao primjenu dodatnih filtera koji nisu izričito navedeni u postavljenom kliničkom pitanju, kao što su dob, spol, vrsta rada, jezik rada i vrijeme objave rada. Navedeno je učestalije u drugoj (93,5 %; 43/46) nego u prvoj (75 %; 15/20) sesiji.

Pozornost je posvećena i načinu na koji ChatGPT tretira koncept „odrasle osobe“ (engl. *adult*) zadan kliničkim pitanjem korištenim u istraživanju. U bazi PubMed dobne odlike uobičajeno se primjenjuju kao filtri dostupni u sučelju baze, i to nakon izvršenog pretraživanja, a ne kao dio pretraživačkog niza. U analizi je stoga upotreba toga pojma tretirana na tri načina: optimalna kombinacija MeSH-predmetnice *Adult* i ključne riječi „adult“, pretraživanih u polju „Title/Abstract“ i povezanih Booleovim operatorom OR, vrednovana je jedinim bodom; s nula bodova, tj. neutralno, ocijenjeni su slučajevi u kojima se navodi ili samo MeSH-predmetnica *Adult* ili samo ključna riječ „adult“ pretraživana u svim poljima baze PubMed (oznaka polja [All fields]), s obzirom na to da ne utječu na kvalitetu rezultata pretraživanja; po jedan bod oduziman je pri navođenju ključne riječi „adult“ pretraživane samo u polju „Title/Abstract“, što dovodi do iznimnog sužavanja rezultata pretraživanja. Neodgovarajuća, negativno bodovana upotreba pojma „adult“ u pretraživačkom nizu javlja se u podjednakom postotku u obje sesije: u prvoj u 15 % (3/20), a u drugoj u 17 % (8/46) nizova. Međutim, u drugoj sesiji porast preciznosti očituje se u upotrebi optimalne kombinacije MeSH-pred-

metnice i ključne riječi (povezanih operatorom OR) u devet nizova (19,6 %; 9/46), za razliku od prve sesije gdje se ta kombinacija ne pojavljuje.

Prepoznavanje PICO modela. U drugoj sesiji model PICO spontano je u ChatGPT-evu odgovoru spomenut u više od polovice promatranih nizova (54,3 %; 25/46), dok je u prvoj sesiji bio prisutan u manjem broju pretraživačkih nizova (30 %; 6/20). Povećano spominjanje PICO-a u drugoj sesiji povezano je s eksplicitnim navođenjem fraze „kliničko pitanje“ u *promptu* (više u poglavlju 4.3.). Bitno je spomenuti i da su elementi kliničkog pitanja korektno interpretirani i strukturirani u model PICO od strane ChatGPT-a u svim slučajevima gdje je on spomenut. Ujedno je vidljivo da je u nizovima generiranim u ChatGPT-evim odgovorima gdje se spominje i model PICO, postignuto više bodova u elementima što se tiče MeSH-predmetnica i prepoznavanja odgovarajućih koncepata kliničkog pitanja. Nalazi ovog rada koji upućuju na pozitivan utjecaj uključivanja modela PICO na kvalitetu odgovora generiranih alatima GUI-a, u skladu su s literaturom koja predlaže ugrađivanje elemenata PICO modela u *promptove* za poboljšanje učinkovitosti pretraživanja medicinskih informacija (Robinson, Bontekoe i Muellenbach, 2025), ali i koja pokazuje da spominjanje modela PICO u *promptu* pridonosi jasnijem definiranju konteksta za alate GUI-a i posrednom kreiranju preciznijih pretraživačkih nizova (Gosak i sur., 2025).

4.3. Povezanost između elemenata promptova i kvalitete pretraživačkih nizova

Porast kvalitete pretraživačkih nizova u drugoj sesiji odražava učinak kvalitetnijeg strukturiranja *promptova* u drugoj sesiji, čime se potvrđuje važnost preciznog oblikovanja *promptova*, tj. kvalitetnih uputa alatu GUI-a, ali i poznavanja principa informacijskog pretraživanja i oblikovanja pretraživačkih nizova čije je zakonitosti potrebno jasno navesti u kreiranom *promptu*. Iz usporedbe prosječne vrijednosti ukupnoga ostvarenog broja bodova *promptova* i ukupnog broja bodova pretraživačkih nizova tijekom prve (*prompt* = 2,25; niz = 17,45) i druge sesije (*prompt* = 5,24; niz = 20,89) vidljivo je da kvaliteta *prompta* značajno utječe na kvalitetu pretraživačkog niza generiranoga od strane ChatGPT-a (vidi tablica 1 i 2). U literaturi koja se bavi istraživanjem potencijala alata GUI-a u kreiranju pretraživačkih nizova također se mogu iščitati stavovi o znatnom utjecaju vrste *prompta* na efektivnost pretraživačkog niza (Alshami i sur., 2023; Gotlib-Małkowska i sur., 2025: 16–17; Reed i Barnett, 2025; Wang i sur., 2023: [8–9]).

4.3.1. Korelacijska analiza

Osim što je vidljivo kako ukupna kvaliteta *promptova* utječe na ukupnu kvalitetu pretraživačkih nizova, Spearmanovom rang korelacijskom analizom ispitalo se postoji li veza između pojedinih elemenata *prompta* i ukupne kvalitete pretra-

živačkog niza, tj. koje je elemente *prompta* nužno spomenuti kako bi se povećala kvaliteta generiranoga pretraživačkog niza.

Općenito govoreći, između elemenata *prompta* i ukupne kvalitete pretraživačkog niza uočena je slaba do umjerena značajna pozitivna korelacija (Spearmanov r_s varira u rasponu od 0,248 do 0,429). Najjače pozitivne korelacije s ukupnom kvalitetom pretraživačkog niza pronađene su za elemente *prompta* „Tezaurus MeSH“ ($r_s = 0,429$; $p < 0,001$), „Booleovi operatori“ ($r_s = 0,424$; $p < 0,001$) i „Ključne riječi / sinonimi“ ($r_s = 0,369$; $p = 0,002$).

Nadalje, uočene su i korelacije između pojedinačnih elemenata *prompta* i pojedinačnih elemenata generiranih pretraživačkih nizova. Element „Prepoznavanje modela PICO“ unutar odgovora generiranog pomoću ChatGPT-a pokazao je značajnu, snažnu pozitivnu korelaciju s elementom „Fraza ,kliničko pitanje““ u *promptu* ($r_s = 0,558$; $p < 0,001$). Slaba pozitivna korelacija pronađena je i između elementa „Prepoznavanje PICO modela“ u generiranom ChatGPT-evu odgovoru i „Zadavanje uloge alatu GUI-a“ u *promptu* ($r_s = 0,254$; $p = 0,039$). Generiranje adekvatnih „MeSH predmetnica“ unutar pretraživačkog niza pozitivno je koreliralo s elementima „Tezaurus MeSH“ ($r_s = 0,324$; $p = 0,008$), „Fraza ,kliničko pitanje““ ($r_s = 0,276$; $p = 0,025$) i „Ključne riječi / sinonimi“ ($r_s = 0,262$; $p = 0,035$) u *promptu*. Suprotno tomu, elementi „Tezaurus MeSH“ ($r_s = -0,350$; $p = 0,004$), „Fraza ,kliničko pitanje““ ($r_s = -0,263$; $p = 0,033$) i „Ključne riječi“ ($r_s = -0,254$; $p = 0,040$) u *promptu* negativno koreliraju s ChatGPT-evim „Haluciniranim MeSH predmetnicama“, tj. s izmišljenim MeSH predmetnicama, što sugerira da uključivanje tih elemenata u *prompt* može pomoći u sprječavanju spomenutih halucinacija (vidi slika 1).

Iako nekoliko elemenata u *promptu* korelira s ukupnom kvalitetom pretraživačkih nizova, naši rezultati pokazuju da je uključivanje elemenata „Fraza ,kliničko pitanje““, „Tezaurus MeSH“ i „Ključne riječi / sinonimi“ u *prompt* posebno važno i za generiranje specifičnih elemenata pretraživačkog niza visoke kvalitete, kao što su indirektno „Prepoznavanje modela PICO“ i „MeSH predmetnice“, ali i za sprječavanje potencijalnih ChatGPT-evih „Haluciniranih MeSH predmetnica“.

PICO –	0,117	-0,132	0,558***	-0,095	-0,056	0,005	-0,148	-0,207	-0,144	0,070	0,254*
MeSH predmetnice –	-0,064	-0,064	0,276*	0,324**	0,262*	0,235	0,180	0,208	0,101	0,177	0,104
Halucinacija MeSH-a –	0,046	0,046	-0,263*	-0,350**	-0,254*	-0,236	-0,201	-0,219	-0,157	-0,128	0,034
Kvaliteta pretraživačkog niza –	0,135	0,105	0,297*	0,429***	0,369**	0,424***	0,350**	0,294*	0,248*	0,114	0,109
Kvaliteta prompta –	0,217	-0,060	0,523***	0,855***	0,752***	0,701***	0,724***	0,750***	0,592***	0,311*	0,237
	Fraza "strategija pretraživanja"	Baza PubMed	Fraza "kliničko pitanje"	Tezaurus MeSH	Ključne riječi	Booleovi operatori	Pretraživanje fraza	Kraćenje riječi	Pretraživanje u polju Title/Abstract	Opis konteksta upita	Zabavljanje uloge genAI alatu

Napomena: Na slici su prikazani samo kvaliteta i elementi pretraživačkog niza i prompta među kojima postoji značajna korelacija prema Spearmanu. Negativne *rs* vrijednosti predstavljaju negativnu korelaciju i obojene su crveno, a pozitivne *rs* vrijednosti predstavljaju pozitivnu korelaciju i obojene su ljubičasto. Statistički značajne korelacije deblje su otisnute. Razine statističke značajnosti su: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Slika 1. Rang korelacija prema Spearmanu između elemenata *prompta* i elemenata i kvalitete pretraživačkog niza

4.4. Ograničenja istraživanja

Kao glavno ograničenje ovog istraživanja navodimo problem malog uzorka, osobito vidljiv u činjenici da nisu svi studenti sudjelovali u obje istraživačke sesije, što je dovelo do nedosljednosti i nejednakog broja uzorka u sesijama zbog kojih se pristupilo uspoređivanju na skupnoj razini, umjesto na individualnoj razini ispitanika, čime bi se dobila detaljnija slika o napretku studenta nakon edukacije i provedbe druge sesije istraživanja.

Nadalje, istraživanje je bilo ograničeno na samo jednu bazu podataka (PubMed) i na samo jedan informacijski upit. Također, studentske sesije provedene su korištenjem samo jednog alata GUI-a (ChatGPT, model 3.5), što – uz nepostojanje uvida u treniranje temeljnih algoritama modela GUI-a – potencijalno ograničava poopćavanje rezultata i donošenje zaključaka o široj slici te problematike.

U ovom istraživanju bavili smo se samo evaluacijom *promptova* i njima rezultirajućih generiranih pretraživačkih nizova, i to usporedbom sa zlatnim referentnim standardom. Evaluacija stvarne vrijednosti i kvalitete pretraživačkog niza i rele-

vanost dobivenih rezultata pretraživanja, tj. situacija kad se on zaista upotrijebi za pretraživanje u bazi PubMed, nadilazi okvire i istraživački koncept ovog rada.

4.5. Smjerovi budućih istraživanja

Ograničenja ovog istraživanja ujedno su i mjesta iz kojih se mogu generirati smjerovi za buduće istraživačke projekte. Daljnja istraživanja mogla bi uključivati veći broj informacijskih upita za pretraživanje u bazi PubMed, različite tematske usmjerenosti i složenosti, i biti provedena s većim brojem ispitanika, što bi pridonijelo dodatnom validiranju provjerenih popisa za ocjenjivanje *promptova* i pretraživačkih nizova osmišljenih u ovom istraživanju.

Također, bilo bi korisno testirati i usporediti potencijale i drugih alata i modela generativne umjetne inteligencije (npr. ostali modeli ChatGPT-a, alati Gemini, Claude, DeepSeek itd.), i to u kreiranju pretraživačkih nizova i za druge baze podataka, kao što su Scopus, baze sadržane u platformi Web of Science, zatim pretraživač Google Scholar itd. Osim toga, potencijal se pokazuje i u kreiranju pretraživačkih nizova i osmišljavanju kriterija evaluiranja pretraživačkih nizova iz drugih područja, polja i grana znanosti, a moguće je i istraživanje odziva i preciznosti dobivenih rezultata pri pretraživanju baza podataka pomoću pretraživačkog niza generiranog u alatu GUI-a.

Isto tako smatramo da bi longitudinalno istraživanje učinka edukacije o pismenosti u području umjetne inteligencije, primjerice kroz dugoročno praćenje razine pismenosti o UI-u kod studenata, moglo pridonijeti razvoju smjernica za uključivanje tema o umjetnoj inteligenciji i *promptanju* u akademsko okruženje i visokoškolsku nastavu, ali i u širu zajednicu, pri čemu bi knjižnice mogle imati ključnu ulogu kao edukatori.

5. Zaključak

Cilj istraživanja bio je ispitati potencijal i učinkovitost alata GUI-a (ChatGPT) u oblikovanju pretraživačkog niza za bazu podataka PubMed na primjeru studenata studija medicine kao korisnika alata. Istraživanje sa studentima provedeno je u dvije sesije, i to prije i nakon edukacije o pretraživanju baza podataka i kreiranju pretraživačkih nizova te oblikovanju *promptova*, pa se tako dobio uvid u odnos između kvalitete studentskih *promptova* i kvalitete pretraživačkih nizova generiranih u alatu GUI-a.

Alat ChatGPT demonstrirao je sposobnost kreiranja funkcionalnih i sintaktički ispravnih pretraživačkih nizova za bazu PubMed. Usporedba rezultata dviju sesija pokazala je značajan porast broja bodova kako u *promptovima* tako i u pretraživačkim nizovima u drugoj sesiji upućujući na to da već i kratka edukacija može

znatno unaprijediti rad korisnika s alatima GUI-a, što posredno zatim rezultira generiranjem kvalitetnijih pretraživačkih nizova. Osim toga pokazana je i jasna povezanost između određenih elemenata *prompta* i ukupne kvalitete pretraživačkog niza. Spearmanova korelacija pokazala je da uključivanje elemenata *prompta* kao što su MeSH, Booleovi operatori i ključne riječi pozitivno korelira s ukupnom kvalitetom niza. Također je uočena povezanost između pojedinih elemenata *prompta* i specifičnih elemenata niza, pri čemu spominjanje, npr., fraze „kliničko pitanje“, tezaurusa MeSH i ključnih riječi pridonosi generiranju niza više kvalitete što uključuje, primjerice, neizravno prepoznavanje modela PICO i primjenu relevantnih MeSH-predmetnica.

Kao i u prethodnim istraživanjima, uočeno je haluciniranje, tj. izmišljanje MeSH-predmetnica, pa se – iako su se takvi osnovni pretraživački nizovi pokazali funkcionalnima u pretraživanju baze PubMed i mogli bi zadovoljiti jednostavne informacijske potrebe određenih korisnika – i dalje vidi potreba za ljudskom provjerom, validacijom i eventualnom doradom pretraživačkih nizova. Iako alati GUI-a pokazuju potencijale u različitim područjima, i dalje je riječ o sustavu konstruiranom po principu „crne kutije“, pri čemu korisnicima nisu poznati detalji o treniranju modela, pa je stoga nužna ljudska procjena i odluka o vrijednosti dobivenih generiranih odgovora.

Doprinos ovog rada ogleda se u razvoju posebnih istraživačkih instrumenata u obliku zlatnih standarda i provjerenih popisa za evaluaciju *promptova* i pretraživačkih nizova, primarno namijenjenih pretraživanju medicinskih baza podataka, prilagođenih neiskusnim korisnicima alata GUI-a, a koji bi se možda mogli koristiti u nekoj budućoj standardizaciji evaluacije učinkovitosti *promptova* i rezultata dobivenih od alata GUI-a.

Pokazana pozitivna uloga edukacije na kvalitetu *promptova* te posljedično i pretraživačkih nizova temelj je za prepoznavanje potrebe za uključivanjem tema o radu s alatima GUI-a, s posebnim naglaskom na kreiranju *promptova*, u programe informacijskog opismenjavanja, i to osobito u području biomedicine i zdravstva gdje je informacijska pismenost ključna za pronalaženje kvalitetnih i relevantnih medicinskih znanstvenih dokaza i informacija. Nadalje, uočeni obrasci korisničkog ponašanja – u ovom istraživanju na primjeru studenata medicine – mogu biti temelj za osmišljavanje programa edukacije i/ili izradu vodiča o kreiranju *promptova* i strategija pretraživanja kako za zdravstvene djelatnike, studente, stručnjake i znanstvenike iz područja biomedicine i javnog zdravstva tako i korisnike iz drugih disciplina koji samostalno pretražuju baze podataka.

Iako predstavljaju još relativno recentno tehnološko rješenje, alati GUI-a već su postali standardom u svakodnevnom životu širokog kruga korisnika. Odgovornoj, sigurnoj i učinkovitoj upotrebi tih alata te kritičkom vrednovanju generiranih odgovora neiskusne korisnike može poučiti upravo knjižničar kao stručnjak za

informacijsko pretraživanje i kritičko vrednovanje različitih vrsta izvora, osobito u situaciji kad se tradicionalni načini informacijskog pretraživanja postupno mijenjaju i/ili zamjenjuju interakcijom s alatima GUI-a poput ChatGPT-a.

LITERATURA

- Abd-Alrazaq i sur. (2023). Abd-Alrazaq, A.; R. AlSaad; D. Alhuwail; A. Ahmed; P. M. Healy; S. Latifi; S. Aziz i suradnici. Large language models in medical education: Opportunities, challenges, and future directions. *JMIR medical education* 9: e48291. <https://doi.org/10.2196/48291>
- Alaniz, L.; C. Vu i M. J. Pfaff. (2023). The utility of artificial intelligence for systematic reviews and Boolean query formulation and translation. *Plastic and reconstructive surgery. Global open* 11, 10: e5339. <https://doi.org/10.1097/GOX.0000000000005339>
- Alshami i sur. (2023). Alshami, A.; M. Elsayed; E. Ali; A. E. E. Eltoukhy; T. Zayed. Harnessing the power of ChatGPT for automating systematic review process: Methodology, case study, limitations, and future directions. *Systems* 11, 7: 351. <https://doi.org/10.3390/systems11070351>
- Bakst, W. (2025). Prompt evaluation: methods, tools, and best practices. [citirano: 2025–04–22]. Dostupno na: <https://mirascope.com/blog/prompt-evaluation>
- Bourgeois, J. P. i H. Ellingson (2025). Ability of ChatGPT to generate systematic review search strategies compared to a published search strategy. *Medical reference services quarterly*: 1–13. <https://doi.org/10.1080/02763869.2025.2537075>
- Chaudhuri, J. i L. Terrones (2024). Reshaping academic library information literacy programs in the advent of ChatGPT and other generative AI technologies. *Internet Reference Services Quarterly* 29, 1: 1–25. <https://doi.org/10.1080/10875301.2024.2400132>
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval* (3. izd.). New York: Neal-Schuman Publishers.
- „Dr Med“. (2025). Digitalni repozitorij Medicinskog fakulteta Sveučilišta u Zagrebu [citirano: 2025–07–14]. Dostupno na: <https://repozitorij.mef.unizg.hr/>
- Gao i sur. (2024). Gao, Z.; L. Li; S. Ma; Q. Wang; L. Hemphill i R. Xu. Examining the potential of ChatGPT on biomedical information retrieval: Fact-checking drug-disease associations. *Annals of biomedical engineering* 52, 8: 1919–1927. <https://doi.org/10.1007/s10439-023-03385-w>
- Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering* 51, 12: 2629–2633. <https://doi.org/10.1007/s10439-023-03272-4>

- Gitman, V.; Maxwell, C. i J. M. Gamble. (2025). Enhancing search strategies for systematic reviews on drug Harms: an evaluation of the utility of ChatGPT in error detection and keyword generation. *Computers in biology and medicine* 193: 110464. <https://doi.org/10.1016/j.compbiomed.2025.110464>
- Gosak i sur. (2025). Gosak, L.; G. Štiglic; L. Pruinelli i D. Vrbnjak. PICOT questions and search strategies formulation: A novel approach using artificial intelligence automation. *Journal of nursing scholarship: an official publication of Sigma Theta Tau International Honor Society of Nursing* 57, 1: 5–16. <https://doi.org/10.1111/jnu.13036>
- Gotlib-Małkowska i sur. (2025). Gotlib-Małkowska, J.; I. Cieślak; M. Jaworski i M. Panczyk. Evidence-based advanced prompt engineering in nursing research: quality analysis of ChatGPT-generated Boolean search query. *Nursing in the 21st Century / Pielęgniarstwo XXI wieku* 24, 1(90): 9–19. <https://doi.org/10.12923/pielxxiw-2025-0002>
- Heston, T. F. i C. Khun (2023). Prompt engineering in medical education. *International Medical Education* 2, 3: 198–205. <https://doi.org/10.3390/ime2030019>
- Higgins i sur. (ur.). (2024). Higgins, J. P. T.; J. Thomas; J. Chandler; M. Cumpston; T. Li; M. J. Page i V. A. Welch. *Cochrane handbook for systematic reviews of interventions* (Version 6.5, August 2024). London: The Cochrane Collaboration. [citirano: 2025–05–25]. Dostupno na: www.cochrane.org/handbook
- JASP Team. (2024). JASP, računalni program, verzija 0.19.3. [citirano: 2025–06–28]. Dostupno na: <https://jasp-stats.org/>
- Jones, K. S. (1991). The role of artificial intelligence in information retrieval. *Journal of the American Society for Information Science* 42, 8: 558–565. [https://doi.org/10.1002/\(SICI\)1097-4571\(199109\)42:8<558::AID-ASI4>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-4571(199109)42:8<558::AID-ASI4>3.0.CO;2-4)
- Lefebvre i sur. (2025). Lefebvre, C.; J. Glanville; S. Briscoe; R. Featherstone; A. Littlewood; M.-I. Metzendorf; A. Noel-Storr i suradnici. Chapter 4: Searching for and selecting studies. March 2025. U: Higgins, J. P. T.; J. Thomas; J. Chandler; M. Cumpston; T. Li; M. J. Page; V. A. Welch (ur.). *Cochrane handbook for systematic reviews of interventions* (Version 6.5.1). London: The Cochrane Collaboration. [citirano: 2025–05–25]. Dostupno na: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-04>
- Lešić, J. (prir.). (2010). *Knjižničarstvo i srodna područja: izbor englesko-hrvatskih stručnih naziva*. Zagreb: Naklada Nediljko Dominović.
- Maaz i sur. (2025). Maaz, S.; J. C. Palaganas; G. Palaganas; M. Bajwa. A guide to prompt design: foundations and applications for healthcare simulationists. *Frontiers in medicine* 11: 1504532. <https://doi.org/10.3389/fmed.2024.1504532>
- McGowan i sur. (2016). McGowan, J.; M. Sampson; D. M. Salzwedel; E. Cogo; V. Foerster i C. Lefebvre. *Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E)*. Ottawa: Canada's Drug Agency. [citirano: 2025-02-28]. Dostupno na: https://www.cda-amc.ca/sites/default/files/attachments/2021-07/CP0015_PRESS_Update_Report_2016_0.pdf

- Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: Tutorial. *Journal of medical Internet research* 25: e50638. <https://doi.org/10.2196/50638>
- Miltenoff, P. (2024). Academic librarians in times of AI and AI literacy: tasks, responsibilities, and leadership. *The International Information & Library Review* 56, 3: 296–305. <https://doi.org/10.1080/10572317.2024.2381182>
- Neilson, C. J. (2021). Adoption of peer review of literature search strategies in knowledge synthesis from 2009 to 2018: an overview. *Health information and libraries journal* 38: 160–171. <https://doi.org/10.1111/hir.12367>
- NLM (2023). National Library of Medicine. MEDLINE, PubMed, and PMC (PubMed Central): How are they different? [citirano: 2025–03–23]. Dostupno na: <https://www.nlm.nih.gov/bsd/difference.html>
- NLM (2024). National Library of Medicine. MEDLINE PubMed production statistics. [citirano: 2025–03–23]. Dostupno na: https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html
- NLM (2025). National Library of Medicine. PubMed. [citirano: 2025–06–25]. Dostupno na: <https://pubmed.ncbi.nlm.nih.gov/>
- NLM (2025a, July 9). National Library of Medicine. Introduction to MeSH: preface. [citirano: 2025–07–11]. Dostupno na: https://www.nlm.nih.gov/mesh/intro_prelude.html
- NLM (2025b, July 10). National Library of Medicine. MeSH record types. [citirano: 2025–07–11]. Dostupno na: https://www.nlm.nih.gov/mesh/intro_record_types.html
- NLM (2025c, May 23). National Library of Medicine. 2025 MeSH highlights. [citirano: 2025–07–11]. Dostupno na: https://www.nlm.nih.gov/oet/ed/mesh/2025/mesh_highlights.html
- NLM (2025d, July 9). National Library of Medicine. MeSH qualifiers with scope notes. [citirano: 2025–07–11]. Dostupno na: https://www.nlm.nih.gov/mesh/qualifiers_scopenotes.html
- NLM (2025e, July 9). National Library of Medicine. Entry terms and other cross-references. [citirano: 2025–07–11]. Dostupno na: https://www.nlm.nih.gov/mesh/intro_entry.html
- NLM (2025f, July 14). National Library of Medicine. Concept structure in MeSH. [citirano: 2025-07-15]. Dostupno na: https://www.nlm.nih.gov/mesh/concept_structure.html
- OpenAI. (2022). Introducing ChatGPT. [citirano: 2025–04–22]. Dostupno na: <https://openai.com/index/chatgpt/>
- OpenAI. (2025). ChatGPT. [citirano: 2025–03–13]. Dostupno na: <https://chatgpt.com>

- Parisi, V. i A. Sutton (2024). The role of ChatGPT in developing systematic literature searches: an evidence summary. *Journal of the European Association for Health Information and Libraries* 20, 2: 30–34.
- Petrak, J. (2019). Traženje medicinskih informacija. U: M. Marušić (ur.). *Uvod u znanstveni rad u medicini* (6. izd.) (Str. 164–173). Zagreb: Medicinska naklada.
- Poropat, G. (2023). Kako sustavno pretraživati informacije u medicini? U: D. Štimac (ur.). *Medicina temeljena na dokazima* (Str. 47–51). Zagreb: Medicinska naklada: Hrvatska akademija medicinskih znanosti.
- PRISMA Executive. (2025). [citirano: 2025–05–26]. Dostupno na: <https://www.prisma-statement.org/>
- Reed, R. B. i Barnett, D. J. (2025). Assessing the quality of biomedical Boolean search strings generated by prompted and unprompted models using ChatGPT: A pilot study. *Medical reference services quarterly* 44, 1: 31–40. <https://doi.org/10.1080/02763869.2024.2440848>
- Robinson, K.; Bontekoe, K. i J. Muellenbach. (2025). Integrating PICO principles into generative artificial intelligence prompt engineering to enhance information retrieval for medical librarians. *Journal of the Medical Library Association: JMLA* 113, 2: 184–188. <https://doi.org/10.5195/jmla.2025.2022>
- Roos, A. (2015). Medical scientists' information practices in the research work context. *Health information and libraries journal* 32, 1: 23–36. <https://doi.org/10.1111/hir.12095>
- Schardt i sur. (2007). Schardt, C.; M. B. Adams; T. Owens; S. Keitz; P. Fontelo. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making* 7: 16. <https://doi.org/10.1186/1472-6947-7-16>
- Spencer, A. J. i Eldredge, J. D. (2018). Roles for librarians in systematic reviews: a scoping review. *Journal of the Medical Library Association: JMLA* 106, 1: 46–56. <https://doi.org/10.5195/jmla.2018.82>
- Škorić, L. (2011). Tezaurus Medical Subject Headings – MESH®. *Vjesnik bibliotekara Hrvatske* 54, 4: 95–114.
- Škorić i sur. (2012). Škorić, L.; M. Šember; H. Markulin; J. Petrak. (2012). Informacijska pismenost u nastavnom programu diplomskog studija Medicinskog fakulteta Sveučilišta u Zagrebu. *Vjesnik bibliotekara Hrvatske* 55, 3–4: 17–28.
- Temsah i sur. (2023). Temsah, M.-H.; I. Altamimi; A. Jamal; K. Alhasan i A. Al-Eyadhy. ChatGPT surpasses 1000 publications on PubMed: envisioning the road ahead. *Cureus* 15, 9: e44769. <https://doi.org/10.7759/cureus.44769>
- Wang i sur. (2023). Wang, S.; H. Scells; B. Koopman; G. Zuccon. Can ChatGPT write a good Boolean query for systematic review literature search? U: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)* (Str. 1426–1436). New York: Association for Computing Machinery. <https://doi.org/10.1145/3539618.3591703>

- Wilson, T. D. (2000). Human information behavior. *Informing science* 3, 2: [49]–55. <https://doi.org/10.28945/576>
- Zamfirescu-Pereira i sur. (2023). Zamfirescu-Pereira, D.; R. Y. Wong; B. Hartmann i Q. Yang. Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. U: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)* (Str. 1–21, Article No. 437). New York: Association for Computing Machinery. <https://doi.org/10.1145/3544548.3581388>

PRILOZI

Prilog 1. Kliničko pitanje

Smanjuje li primjena statina rizik od moždanog udara kod odraslih osoba s hiperlipidemijom?

Prilog 2. Zlatni standard prompta za genAI alat

Ti si informacijski stručnjak specijaliziran za pretraživanje bibliografskih baza podataka. Generiraj pretraživački niz za bazu podataka PubMed na osnovi sljedećega kliničkog pitanja: „Smanjuje li primjena statina rizik od moždanog udara kod odraslih osoba s hiperlipidemijom?”. Koristi termine iz tezaurusa Medical Subject Headings (MeSH) i slobodno oblikovane ključne riječi kako bi izradio složenu strategiju pretraživanja. MeSH termine i ključne riječi poveži Booleovim operatorima. Ključne riječi (tj. sinonime) pretražuju se u polju Title/Abstract. Optimiziraj sintaksu pretraživačkog niza u skladu s pravilima baze PubMed tako što ćeš upotrijebiti kraćenje ključnih riječi zvjezdicom (svođenje riječi na korijen) i navodnike za pretraživanje fraza.

Prilog 3. Zlatni standard pretraživačkog niza za bazu PubMed

(((„Hyperlipidemias“ [Mesh])

OR (hyperlipid*[Title/Abstract] OR hyperlipem*[Title/Abstract] OR lipemia*[Title/Abstract] OR lipid*emi*[Title/Abstract] OR hypercholester*[Title/Abstract] OR „elevated cholesterol“*[Title/Abstract] OR „high cholesterol“*[Title/Abstract] OR hypertriglyceridemi*[Title/Abstract] OR „elevated triglyceride“*[Title/Abstract] OR „elevated triglyceride“*[Title/Abstract] OR hyperlipoproteinemi*[Title/Abstract]))

AND

(((„Hydroxymethylglutaryl-CoA Reductase Inhibitors“ [Mesh])

OR („Hydroxymethylglutaryl-CoA Reductase Inhibitor“*[Title/Abstract] OR „HMG-CoA Reductase Inhibitor“*[Title/Abstract] OR „Hydroxymethylglutaryl-Coenzyme A Reductase Inhibitor“*[Title/Abstract] OR „HMG-CoA Statin“*[Title/Abstract] OR statin*[Title/Abstract] OR atorvastatin*[Title/Abstract] OR simvastatin*[Title/Abstract] OR pravastatin*[Title/Abstract] OR rosuvastatin*[Title/Abstract] OR fluvastatin*[Title/Abstract] OR lovastatin*[Title/Abstract] OR pitavastatin*[Title/Abstract] OR cerivastatin*[Title/Abstract] OR mevastatin*[Title/Abstract]))

AND

((„Stroke“[Mesh]) OR (stroke*[Title/Abstract] OR „cerebr* accident“*[Title/Abstract] OR „cerebr* apoplex“*[Title/Abstract] OR CVA[Title/Abstract] OR „brain infarct“*[Title/Abstract] OR „cerebr* infarct“*[Title/Abstract] OR „brain isch*emia“*[Title/Abstract] OR „isch*emic stroke“*[Title/Abstract] OR „brain attack“*[Title/Abstract] OR „intracerebral h*emorrhag“*[Title/Abstract] OR „brain h*emorrhag“*[Title/Abstract] OR „subarachnoid h*emorrhag“*[Title/Abstract] OR „hemorrhagic stroke“*[Title/Abstract] OR „haemorrhagic stroke“*[Title/Abstract]))

Prilog 4. Provjerni popis za ocjenjivanje kvalitete prompta za genAI alat namijenjenog izradi pretraživačkog niza za bazu PubMed

Kategorija	Opis kategorije	Prijedlog bodovanja (kriteriji ocjenjivanja)
Zadavanje uloge genAI alatu	Je li genAI alatu u <i>promptu</i> jasno i nedvosmisleno dodijeljena uloga iz čije će perspektive odgovoriti na korisnički upit? Npr. „Ti si knjižničar specijaliziran za medicinske informacije.“	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen
Opis konteksta upita	Spominje li se u <i>promptu</i> kontekst kojim se objašnjava zašto se traži odgovor i/ili koja je njegova svrha? Npr. „Student sam medicine koji kreće s izradom diplomskog rada. Potrebna mi je strategija pretraživanja u bazi PubMed na temu...“.	
Fraza „pretraživački niz“ / „strategija pretraživanja“	Spominje li se u <i>promptu</i> fraza „pretraživački niz“ i/ili „strategija pretraživanja“? U obzir se uzimaju i engleske varijante <i>search string</i> i <i>search strategy</i> .	
Baza PubMed	Spominje li se u <i>promptu</i> naziv PubMed?	
Fraza „kliničko pitanje“	Spominje li se u <i>promptu</i> fraza „kliničko pitanje“?	
Tezaurus MeSH	Spominje li se u <i>promptu</i> upotreba tezaurusa MeSH? U obzir se mogu uzeti i razne iteracije fraze, npr. „MeSH tezaurs“, „MeSH pojmovi“, „MeSH predmetnice“, ili – puni i skraćeni naziv tezaurusa – Medical Subject Headings i MeSH.	
Ključne riječi / sinonimi	Spominje li se u <i>promptu</i> upotreba (slobodno oblikovanih) ključnih riječi i/ili sinonima?	
Booleovi operatori	Spominje li se u <i>promptu</i> upotreba Booleovih operatora?	
Pretraživanje fraza (upotrebom navodnika)	Spominje li se u <i>promptu</i> pretraživanje fraza?	
Kraćenje riječi (upotrebom zvjezdice)	Spominje li se u <i>promptu</i> pretraživanje pomoću kraćenja/korjenovanja riječi?	
Pretraživanje u polju Title/Abstract	Spominje li se u <i>promptu</i> zahtjev za specificirano pretraživanje u PubMedovu polju „Title/Abstract“?	

Jezične pogreške u <i>promptu</i>	Sadrži li <i>prompt</i> pravopisne, gramatičke, sintaktičke ili druge jezične pogreške (npr. tipfeleri, semantičke nepreciznost u navođenju stručnih termina, itd.)	Nije numerički bodovano (ne utječe izravno na ocjenu <i>prompta</i> , ali može imati posredni utjecaj na generirani rezultat)
-----------------------------------	---	---

Prilog 5. Provjerni popis za ocjenjivanje kvalitete pretraživačkog niza za bazu PubMed

Kategorija	Potkategorija	Opis potkategorije	Prijedlog bodovanja (kriteriji ocjenjivanja)
Preoblikovanje kliničkog pitanja u tematske koncepte	Koncepti	Odgovara li strategija pretraživanja postavljenom informacijskom upitu / kliničkom pitanju, tj. jesu li u pretraživački niz uključeni svi spomenuti koncepti? Promatra se konceptualna razina, bez obzira na to je li korišten pojam iz tezaurusa MeSH i/ili ključna riječ, npr. <i>Diet</i> , <i>Mediterranean</i> ili <i>Mediterranean diet</i> .	Po 1 bod za svaki ispravno navedeni koncept
	Uži/širi koncepti	Upotrebljavaju li se u pretraživačkom nizu relevantni i specifični koncepti? Koncepti ne smiju biti preusko i/ili preširoko postavljeni, npr. široki koncept <i>diet</i> naprama specifičnom užem <i>Mediterranean diet</i> .	Po -1 bod za svaki navedeni neadekvatni uži/širi koncept
Predmetnice iz tezaurusa Medical Subject Headings (MeSH)	MeSH predmetnice	Upotrebljavaju li se u pretraživačkom nizu relevantne i specifične MeSH predmetnice, tj. odgovaraju li konceptima spomenutima u informacijskom upitu? Obratiti posebno pozornost na ispravnost navođenja MeSH predmetnica!	Po 1 bod za svaku ispravno navedenu MeSH predmetnicu
		Navodi li se uz MeSH predmetnice u pretraživačkom nizu i oznaka PubMedova polja specijaliziranog za pretraživanje po MeSH predmetnicama? U skladu sa sintaksom baze PubMed, oznake tih polja su [MeSH Terms] ili [Mesh], npr. „Influenza, Human“[Mesh].	Po 1 bod za navođenje oznake polja [MeSH Terms] ili [Mesh] uz MeSH predmetnicu

	Uže/šire MeSH predmetnice	Upotrebljavaju li se u pretraživačkom nizu relevantne i specifične MeSH predmetnice? MeSH predmetnice ne smiju biti preusko i/ili preširoko postavljene, npr. uža MeSH predmetnica <i>Lung Diseases</i> rezultirat će specifičnim rezultatima naprama pretraživanju po široj predmetnici <i>Respiratory Tract Diseases</i> .	Po -1 bod za svaku navedenu neadekvatnu užu/širu MeSH predmetnicu
	Halucinirane MeSH predmetnice	Postoje li u pretraživačkom nizu nepostojeće MeSH predmetnice, tj. predmetnice koje je genAI halucinirao/ izmislio?	Po -1 bod za svaku navedenu haluciniranu MeSH predmetnicu * <i>Za specifičan kontekst ovog istraživanja, odlučeno je da se halucinirane predmetnice koje odgovaraju službenim sinonimima (entry terms) u MeSH-u neće negativno bodovati, nego će se dodijeliti 0 bodova. Razlog je u tomu što zbog specifičnosti baze PubMed takve predmetnice nemaju negativan utjecaj na pretraživanje (opisano u poglavlju 4.2.2.).</i>
Ključne riječi	Ključne riječi	Upotrebljavaju li se u pretraživačkom nizu relevantne i specifične slobodno oblikovane ključne riječi, tj. sinonimi koji odgovaraju konceptima i/ili MeSH predmetnicama spomenutima u informacijskom upitu?	Po 1 bod za svaku navedenu relevantnu ključnu riječ
	Dodatne ključne riječi	Upotrebljavaju li se u pretraživačkom nizu i dodatni sinonimi?	Po 1 bod za svaki koncept ili MeSH predmetnicu za koje su navedene dvije ili više relevantnih ključnih riječi
	Problematične ključne riječi	Postoje li u pretraživačkom nizu ključne riječi koje nisu adekvatne: preuske, preširoke i/ili potpuno neodgovarajuće?	Po -1 bod za svaku navedenu neadekvatnu ključnu riječ

Sintaksa pretraživačkog niza	Booleovi operatori	Upotrebljavaju li se u pretraživačkom nizu Booleovi operatori (AND, OR i NOT)? Jesu li logički i sintaktički ispravno upotrijebljeni? Booleovi operatori trebaju biti pisani velikim tiskanim slovima!	Po 1 bod za svaku od tri vrste ispravno navedenih Booleovih operatora
	Upotreba zagrade za grupiranje pretraživanih pojmova	Upotrebljavaju li se u pretraživačkom nizu zagrade za grupiranje MeSH predmetnica i ključnih riječi? Jesu li ispravno zatvorene i smještene, tj. logički i sintaktički ispravno upotrijebljene?	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen
	Upotreba navodnika za pretraživanje fraza	Upotrebljava li se u pretraživačkom nizu mogućnost pretraživanja određene fraze pomoću navodnika (engl. <i>phrase searching</i> , npr. „Parkinson’s disease“)? Navodnici koji se upotrebljavaju u PubMedu moraju biti tzv. engleski, gornji navodnici („ „)!	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen
	Upotreba zvjezdice za kraćenje/korjenovanje riječi	Upotrebljava li se u pretraživačkom nizu mogućnost kraćenja riječi i pretraživanja različitih varijanti pisanja riječi pomoću zvjezdice (engl. <i>truncation</i> , npr. neoplasm*, „intracerebral h*emorrhage“)? Znak koji se upotrebljava u PubMedu za kraćenje riječi jest zvjezdica (*)!	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen
	Upotreba oznake PubMedova polja [Title/Abstract] uz ključne riječi	Navodi li se uz ključne riječi u pretraživačkom nizu i oznaka PubMedova polja specijaliziranog za pretraživanje u naslovu i sažetku publikacije? U skladu sa sintaksom baze PubMed, oznake toga polja su [Title/Abstract] ili [tiab], npr. „human flu“[tiab].	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen

Granice pretraživanja i filtri	Dobna ograničenja	Upotrebljava li se u pretraživačkom nizu MeSH termin i/ili ključna riječ za dobno ograničenje? Jesu li pojmovi za dobna ograničenja prikladniji za uključivanje u pretraživački niz ili za naknadni odabir u PubMedovu sučelju? <i>* U ovom istraživanju, pojam 'odrasle osobe' može se javljati kao MeSH predmetnica Adult, ključna riječ „adult“ ili njihova kombinacija. Pojam je tretiran kao pokazatelj adekvatnog korištenja dobno ograničenja u bazi PubMed. U uobičajenoj praksi, dobni filtri nisu sastavni dio pretraživačkog niza nego se primjenjuju naknadno odabirom u PubMedovu sučelju.</i>	0 bodova = Adult[Mesh] 0 bodova = adult[All Fields] 1 bod = Adult[Mesh] OR adult[tiab] -1 bod = adult[tiab] <i>* Bodovanje je razrađeno za specifičan kontekst ovog istraživanja (opisano u poglavlju 4.2.2.).</i>
	Predloženi dodatni filtri	Sugerira li se u popratnom odgovoru genAI alata primjena dodatnih filtera koji nisu izričito navedeni u postavljenom informacijskom upitu, npr. dob, spol, vrsta rada, jezik rada, vrijeme objave rada, itd.?	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen
Dodatni pokazatelji kvalitete	Prepoznavanje PICO modela	Spominje li se u popratnom odgovoru genAI alata pojam PICO modela? Iako nije riječ o tipičnom kriteriju za bodovanje kvalitete niza, pojavljivanje PICO modela promatrano je kao potencijalni indikator kvalitete odgovora generiranih od strane ChatGPT-ja.	1 bod = kriterij je zadovoljen 0 bodova = kriterij nije zadovoljen

