Paradoksi umjetne inteligencije u znanstvenoj recenziji: Metodološki pregled (2024. – 2025.)

Autor: Davor Moravek

Sažetak

Znanstveni recenzentski postupak (*peer review*) suočava se s krizom održivosti. Veliki jezični modeli (LLM) nude se kao rješenje, no njihova integracija donosi duboke metodološke i etičke izazove. Ovaj rad provodi metodološki pregledni pregled (*scoping review*) literature objavljene u 2024. i 2025. godini, slijedeći PRISMA-ScR smjernice. Naša analiza identificira i analizira tri ključna paradoksa koji definiraju trenutno stanje: (1) **Paradoks usvajanja**, gdje se stvarna upotreba AI alata procjenjuje na 50–76 %, dok je prijavljena upotreba samo 5,7 % (AlFayyad et al., 2025., predstavljeno na PRC2025); (2) **Paradoks performansi**, gdje AI nadmašuje ljude u specifičnim zadacima (npr. 82 % *Defect Recall* (Propel, 2025), ali uz visoku stopu lažno pozitivnih nalaza), ali istovremeno usporava iskusne stručnjake za 19 % na složenim zadacima (Becker et al., 2025); i (3) **Paradoks pristranosti**, gdje simulacije pokazuju da AI recenzenti sustavno favoriziraju AI-generirane radove (stopa prihvaćanja 78 % naspram 49 % za ljudske radove) (Li et al., 2025). Zaključujemo da ovi paradoksi čine potpunu automatizaciju recenzije etički opasnom. Optimalni model ostaje hibridni sustav „AI-Asistent, Čovjek-Sudac“, koji zahtijeva rigorozan ljudski nadzor kako bi se iskoristile prednosti AI-a bez ugrožavanja znanstvenog integriteta.

Ključne riječi: umjetna inteligencija, znanstvena recenzija, veliki jezični modeli (LLM), pristranost, paradoks usvajanja, *scoping review*, etika objavljivanja

Abstract

The scientific peer review process is facing a sustainability crisis. Large Language Models (LLMs) are offered as a solution, yet their integration brings profound methodological and ethical challenges. This paper conducts a methodological scoping review of literature published in 2024 and 2025, following PRISMA-ScR guidelines. Our analysis identifies and analyzes three key paradoxes defining the current landscape: (1) The **Adoption Paradox**, where actual use of AI tools is estimated at 50–76%, while declared use is only 5.7% (AlFayyad et al., 2025, presented at PRC2025); (2) The **Performance Paradox**, where AI outperforms humans in specific tasks (e.g., 82% *Defect Recall* (Propel, 2025), but with a high false-positive rate), yet simultaneously slows down experienced experts by 19% on complex tasks (Becker et al., 2025); and (3) The **Bias Paradox**, where simulations show AI reviewers systematically favor AI-generated papers (78% acceptance rate vs. 49% for human papers) (Li et al., 2025). We conclude that these paradoxes make full automation of peer review ethically perilous. The optimal model remains a hybrid "AI-Assistant, Human-Judge" system, which requires rigorous human oversight to leverage AI’s benefits without compromising scientific integrity.

Keywords: artificial intelligence, peer review, large language models (LLM), bias, adoption paradox, scoping review, publishing etic

1. Uvod: Kriza održivosti i „Paradoks usvajanja“

Recenzentski postupak (*peer review*) ostaje temeljni mehanizam kontrole kvalitete u znanstvenoj zajednici. Međutim, tradicionalni sustav suočava se s dvostrukom krizom: eksponencijalnim rastom broja rukopisa koji dovodi do preopterećenja recenzenata te pojavom generativne umjetne inteligencije (AI) koja fundamentalno mijenja način na koji se znanost piše i evaluira (ICLR 2025 Research Team, 2025).

Dok veliki jezični modeli (LLM) nude obećanje efikasnosti i brzine (npr. Farber et al., 2024), njihova integracija stvorila je ono što nazivamo „Paradoksom usvajanja“. Podaci predstavljeni na 10. Međunarodnom kongresu o recenziji (PRC2025, 2025) razotkrivaju duboku krizu transparentnosti. Dok studije velikih izdavača, poput one AlFayyad et al. (2025) o BMJ časopisima, pokazuju da samo **5,7 % autora formalno prijavljuje upotrebu AI alata** (čak i kad je politika obavezna), neovisne ankete procjenjuju da je **stvarna stopa upotrebe među autorima između 50 % i 76 %**.

Ova studija (AlFayyad et al., 2025) sugerira da autori ili nisu sigurni što prijaviti ili svjesno skrivaju upotrebu zbog straha od stigmatizacije. Pritom, velika većina (87 %) *prijavljene* upotrebe otpada na benigna poboljšanja jezika, što služi kao krinka za rizičnije, ali rijetko prijavljeno generiranje sadržaja (6,1 %).

Ovaj jaz između prijavljene i stvarne upotrebe sugerira da su AI alati postali „standardna praksa u sjeni“. To potkopava trenutne politike transparentnosti izdavača (npr. Springer Nature, 2025) i čini provođenje pravila gotovo nemogućim. Ako ne možemo pouzdano znati koji je rad napisan uz pomoć AI-a, kako možemo procijeniti njegovu originalnost ili potencijalnu pristranost u recenziji?

Ova studija ima za cilj pružiti metodološki pregled stanja u 2025. godini, fokusirajući se na tri temeljna istraživačka pitanja:

1. Kakva je stvarna, nijansirana slika performansi AI-a u usporedbi s ljudima?
2. Kako pomiriti naizgled kontradiktorne nalaze o sistemskoj pristranosti i skrivenoj upotrebi AI-a?
3. Koji regulatorni okviri i alati postoje te koliki je jaz između politike i prakse?

2. Metodologija

Kako bi se sustavno mapiralo ovo brzorastuće polje, ovaj pregledni rad eksplicitno usvaja okvir *Scoping Review* (pregledni pregled).

2.1. Dizajn pregleda

Koristili smo PRISMA smjernice prilagođene za *Scoping Reviews* (PRISMA-ScR) (Tricco et al., 2018). Cilj nije bio provesti meta-analizu (što je nemoguće zbog heterogenosti studija), već identificirati i mapirati ključne koncepte, dokaze i praznine u istraživanju na temu „AI u recenzentskom postupku“ u 2024. i 2025. godini.

2.2. Strategija pretraživanja i kriteriji za uključivanje

Provedena je sustavna pretraga sljedećih baza podataka: arXiv, Google Scholar, ACM Digital Library i PubMed. Vremenski okvir pretrage bio je ograničen na radove objavljene ili postavljene kao *preprint* od 1. siječnja 2024. do 15. studenog 2025., kako bi se osigurala relevantnost u ovom brzorastućem polju.

Ključne riječi pretrage uključivale su (ali nisu bile ograničene na):

("LLM" OR "large language model" OR "generative AI") AND ("peer review" OR "scientific publishing")
"AI bias" AND "peer review"
"ICLR 2025" AND "AI feedback"
"LLM-REVal" OR "arXiv: 2510.12367"
"AI adoption" AND "scientific writing" AND "transparency"

Kriteriji za uključivanje: Uključili smo kvantitativne studije, simulacije, randomizirane kontrolirane pokuse (RCT), tehničke izvještaje o performansama modela (npr. SWE-bench) i relevantne prezentacije s vodećih konferencija (npr. PRC2025).

Kriteriji za isključivanje: Isključili smo članke s mišljenjima bez podataka, blogove, novinske članke i studije objavljene prije 2024. godine.

2.3. Hibridni metodološki pristup pisanju

U skladu s vlastitim preporukama o transparentnosti, specificiramo uloge AI alata u *procesu pisanja* ovog rada. Ovaj rad je i sam studija slučaja hibridnog modela "AI-Asistent, Čovjek-Sudac".

Početna sinteza i strukturiranje: Google Gemini Pro korišten je za generiranje početne strukture metodološke analize i sažimanje ključnih tema.
Pretraživanje i validacija izvora: Perplexity i Google Deep Search korišteni su za dubinsko pretraživanje literature, identifikaciju ključnih studija (npr. AlFayyad et al., Becker et al.) i validaciju njihovih metodologija.
Idejno "Red Teaming": Claude i ChatGPT korišteni su za iterativno testiranje logičkih nedostataka u argumentaciji (npr. rješavanje "Paradoksa petlje").
Konačna analiza i autorstvo: Konačnu analizu, kritičku procjenu, sintezu svih AI doprinosa i pisanje cjelokupnog teksta proveo je ljudski autor, koji preuzima punu odgovornost za sadržaj.

2.4. Uloga AI alata u recenzentskom postupku (AI Peer Review)

Kao meta-eksperiment, ovaj rad je također podvrgnut formalnom recenzentskom postupku od strane dva neovisna AI recenzenta (Perplexity AI Research Agent i DeepSeek AI Model), čiji su prijedlozi integrirani u ovu verziju.

Perplexity AI Research Agent: Njegova je uloga bila (1) Evaluacija i verifikacija izvora kroz neovisnu validaciju referenci (npr. AlFayyad et al., Propel) i (2) Identifikacija problematičnih komercijalnih izvora (npr. HEC, MokaHR).
DeepSeek AI Model: U skladu s hibridnim modelom istraživanja opisanim u ovom radu, ova finalna verzija rukopisa podvrgnuta je formalnoj recenziji od strane DeepSeek AI modela. Njegova

je uloga bila provesti kritičku analizu konačnog, integriranog teksta s ciljem: (1) provjere metodološke dosljednosti i usklađenosti s PRISMA-ScR okvirom; (2) evaluacije koherentnosti i logičkog slijeda argumentacije kroz tri paradoksa; (3) davanja strukturalnih preporuka za poboljšanje preglednosti i čvrstoće zaključaka. Svi prijedlozi ovog AI recenzenta bili su podvrgnuti konačnoj ljudskoj prosudbi i odobrenju autora, što je u potpunosti u skladu s preporučenim modelom 'AI-Asistent, Čovjek-Sudac'.
Oba recenzenta su također sudjelovala u analizi koherentnosti argumentacije, posebno u rješavanju "Paradoksa petlje" i "Zavaravajuće varijable kvalitete" (*Quality Confounding*).

3. Usporedba performansi: Nijansirana slika

Naši nalazi, mapirani kroz metodologiju *Scoping Reviewa*, otkrivaju „paradoks performansi“: AI dramatično nadmašuje ljude u specifičnim, ponovljivim zadacima, ali može usporiti proces u složenim, kreativnim zadacima.

3.1. Pozitivni nalazi: Brzina, točnost i dosljednost

Naši nalazi potvrđuju da AI, kada se primjenjuje na ispravne zadatke, nudi mjerljive prednosti u tri ključna područja:

Brzina administrativnih zadataka: U zadacima visoke frekvencije, AI pruža dramatična ubrzanja. Studija Farber et al. (2024) zabilježila je smanjenje vremena od 73 % (s 45 na 12 minuta) u zadatku odabira relevantnih recenzenata. Slično, studija slučaja ICLR 2025 (ICLR 2025 Research Team, 2025) postigla je prosječno vrijeme obrade recenzije od otprilike jedne minute.
Točnost u specifičnim domenama: U zadacima s jasnim pravilima, AI nadmašuje prosječne ljudske sposobnosti. U medicinskoj dijagnostici, LMU Munich Emergency Medicine Team (2024) utvrdio je da je GPT-4 nadmašio liječnike specijalizante (P=.01). U programskom inženjerstvu, specifična industrijska mjerila (Propel, 2025) pokazuju da AI (82 % *Defect Recall* kod Claude 3.5 Sonnet) nadmašuje ljudske recenzente (68 %). Međutim, ova sirova metrika skriva ključni kompromis: isti model ima visoku stopu lažno pozitivnih rezultata (12 %) i najvišu cijenu, dok konkurenti (npr. Gemini 1.5 Pro) nude niži *recall* (68 %), ali i znatno nižu stopu lažno pozitivnih (9 %) (Propel, 2025).
Dosljednost: Možda najvažniji nalaz jest da AI rješava temeljni problem ljudske recenzije: nisku dosljednost. Ljudski međusobni sporazum (*inter-rater reliability*) u recenzijama tradicionalno je nizak, često oko r = 0,20–0,21. Nasuprot tome, LLM-ovi pokazuju znatno višu dosljednost i korelaciju s prosječnim ljudskim ocjenama (r = 0,5046 za GPT-4).

3.2. Negativni nalazi i balans: „Paradoks performansi“

Prethodni nalazi ne znače da je AI superioran u *svim* zadacima. Naša pretraga identificirala je ključne studije koje pokazuju negativan utjecaj AI-a na složenim zadacima.

Ključni nalaz koji potkrepljuje ovaj paradoks dolazi iz rigorozne studije organizacije METR (Becker et al., 2025), koja je prethodno pogrešno pripisivana „TechPro Analytics“. Studija je otkrila da, iako AI asistenti ubrzavaju mlađe programere na jednostavnim zadacima, oni **usporavaju iskusne programere za 19 %** na složenim, novim problemima.

Najvažniji nalaz studije (Becker et al., 2025) je jaz između percepcije i stvarnosti: programeri su

vjerovali da su 20 % *brži*, dok su objektivno bili 19 % *sporiji*. Razlog je taj što kognitivni trošak provjere, ispravljanja i integracije AI-generiranog koda na složenim zadacima premašuje korist automatizacije. Studija je utvrdila da su programeri prihvatili manje od 44 % AI prijedloga, što znači da je kognitivno opterećenje provjere i odbacivanja loših prijedloga (što je u skladu s visokom stopom lažno pozitivnih nalaza zabilježenom u Propel (2025) mjerilu) veće od napora da sami napišu kod.

Ovo stvara „paradoks performansi“: AI čini jednostavne zadatke bržima (npr. pronalaženje jednostavne greške u kodu), ali složene zadatke čini sporijima (npr. dizajn nove arhitekture softvera). Primjena AI-a u recenziji stoga nije univerzalno korisna; ona je korisna samo u precizno definiranim, ponovljivim zadacima.

4. Pristranost: Rješavanje logičkih paradoksa

Analiza pristranosti u 2025. godini opterećena je logičkim paradoksima koji zahtijevaju pažljivo metodološko raščlanjivanje. Koristeći novopronađene metodološke detalje o ključnim studijama, rješavamo dva temeljna problema: „Paradoks petlje“ i „Zavaravajuće varijable“.

4.1. Simulacijski nalazi: LLM-REVal

Naša metodološka pretraga potvrdila je postojanje i metodologiju ključne studije o pristranosti (Li et al., 2025). Ova studija pruža zapanjujuće podatke:

Stopa prihvaćanja: Radovi generirani od strane AI-a imali su stopu prihvaćanja od 78 %, u usporedbi s 49 % za ljudske radove.
Prosječna ocjena: AI-generirani radovi dobili su višu prosječnu ocjenu (6,21) u usporedbi s ljudskim (5,94).
„Win Rate“: U izravnim usporedbama, AI radovi „pobjeđivali“ su u 66 % slučajeva (Li et al., 2025).

4.2. Rješenje „Paradoksa petlje“ (Metodologija simulacije)

Postavlja se ključno logičko pitanje: Kako studija LLM-REVal (Li et al., 2025) može identificirati AI radove (da bi izmjerila 78 % stopu prihvaćanja), ako naš drugi ključni nalaz, „Paradoks usvajanja“ (AlFayyad et al., 2025), tvrdi da 76 % autora *skriva* svoju upotrebu?

Rješenje, pronađeno u našoj dubljoj analizi metodologije (Li et al., 2025), leži u činjenici da LLM-REVal *nije bila studija detekcije u „divljini“*.

To je bila *kontrolirana simulacija*. Istraživači su koristili „agenta za istraživanje“ (*research agent*) da generiraju radove i „agenta za recenziju“ (*review agent*) da ih ocijeni. Stoga, nisu trebali „detektirati“ koji su radovi AI-generirani jer su to *znali* od početka.

Ovo rješava „Paradoks petlje“. Oba naša ključna nalaza mogu koegzistirati:

U „divljini“, autori masovno skrivaju upotrebu AI-a (AlFayyad et al., 2025).
U *simuliranom* okruženju, kada AI recenzira radove, on pokazuje snažnu pristranost prema drugim AI-generiranim radovima (Li et al., 2025).

4.3. Rješenje „Zavaravajuće varijable“ (Kvaliteta vs. Pristranost)

Drugo ključno pitanje jest: Je li 78 % vs 49 % dokaz *pristranosti*, ili su AI-asistirani radovi *stvarno bolji* (npr. gramatički jasniji)? Ovo je poznato kao *„quality confounding“* varijabla.

Naša nova analiza potvrđuje da je odgovor *oboje*.

Istraživanje potvrđuje da AI-asistirani radovi često pokazuju značajno višu razinu jezične jasnoće i gramatičke ispravnosti. Ovaj nalaz o *„quality confounding“* varijabli objašnjava dio razlike u ocjenama.

Međutim, to ne objašnjava cijelu priču. Studija LLM-REVal (Li et al., 2025) također je identificirala jasnu *stilsku pristranost*: AI recenzenti preferirali su „čist“ i strukturiran stil pisanja koji je karakterističan za LLM-ove, čak i kada sadržaj nije bio superioran. Dakle, razlika u stopi prihvaćanja kombinacija je (A) stvarne više jezične kvalitete i (B) stvarne stilske pristranosti.

4.4. Kontekstualiziranje pristranosti (AI vs. Ljudske pristranosti)

Konačno, kako bi se AI pristranost stavila u kontekst, valja napomenuti da je fokusiranje isključivo na AI pristranost zavaravajuće, jer je ljudski recenzentski postupak već opterećen dobro dokumentiranim pristranostima (JAMA, 2022), uključujući:

Institucionalnu pristranost (favoriziranje poznatih institucija).
Geografsku pristranost (favoriziranje autora iz zapadnih zemalja).
Rodnu pristranost (Gornitzki et al., 2021).

Međutim, ključni argument za AI, potvrđen nedavnim analizama (npr. Jain & Verma, 2024), jest da, iako uvodi nove mjerljive pristranosti, može dramatično *smanjiti* postojeće, kaotične ljudske pristranosti. Ljudska pristranost je nedosljedna i često nesvjesna; AI, s druge strane, (prema tvrdnjama industrijskih izvora) primjenjuje „standardizirane kriterije“ (MokaHR, 2025) dosljedno na svaki unos. Njegova vrijednost nije u „nepristranosti“, već u „dosljednosti koja se može revidirati“. Svaka AI odluka (opet, prema komercijalnim izvorima) stvara „zapis“ (Hirin.ai, 2025), pretvarajući neuhvatljiv pravni problem (ljudska pristranost) u rješiv, mjerljiv inženjerski problem. Regulatorni okviri poput EU AI Act-a (HEC, 2025) ne zabranjuju AI, već zahtijevaju transparentnost i ljudski nadzor, potvrđujući da je budućnost hibridni model.

Valja napomenuti da su neki od citiranih izvora u ovom kontekstu (npr. MokaHR, Hirin.ai, HEC) primjeri industrijskog diskursa (marketinški materijali i blogovi), a ne recenzirane akademske studije. Oni su uključeni kako bi ilustrirali percepciju AI-a u industriji, ali njihovi nalazi o "standardiziranim kriterijima" nisu akademski validirani i koriste se isključivo kao ilustracija.

Ironično, čini se da AI, umjesto da riješi ove probleme, uvodi *nove* pristranosti (poput stilske) dok istovremeno *nasljeđuje* neke od starih. Studija Lianga et al. (2025) potvrdila je da LLM-ovi također pokazuju *institucionalnu pristranost*, favorizirajući radove s afilijacijama visoko rangiranih institucija.

5. Alati i regulatorni okvir: Jaz između politike i prakse

Dok se AI alati sve više integriraju u znanstveno pisanje, izdavači i organizacije pokušavaju uspostaviti kontrolu kroz kombinaciju automatiziranih alata za provjeru i strogih uredničkih politika. Međutim, kao što je navedeno u našem Uvodu, „Paradoks usvajanja“ (AlFayyad et al., 2025) čini provedbu ovih mjera izuzetno teškom.

5.1. Alati za provjeru integriteta

Kao odgovor na porast AI-generiranog sadržaja i pad transparentnosti, izdavači su počeli implementirati automatizirane alate za *screening* rukopisa prije nego što oni uopće stignu do ljudskih recenzenata. Naša pretraga identificirala je dva ključna primjera:

Ripeta: Ovaj alat koristi strojno učenje za procjenu transparentnosti i ponovljivosti (*reproducibility*) rukopisa, provjeravajući jesu li navedeni izvori podataka, kod i etička odobrenja.
SciScore: Slično, SciScore provjerava rigoroznost metoda u biološkim znanostima, osiguravajući da rukopisi zadovoljavaju standarde.

Iako korisni, ovi alati fokusiraju se na *metodološki* integritet, a ne na *autorstvo* (tj. jesu li tekst generirali AI ili čovjek).

5.2. Regulatorni okvir i službene politike

Vodeći izdavači, svjesni rizika za povjerljivost i odgovornost, uspostavili su jasne smjernice. Kao reprezentativan primjer, Springer Nature (2025) navodi tri ključna pravila:

1. **AI ne može biti autor:** LLM alati ne mogu biti navedeni kao autori jer ne mogu „preuzeti odgovornost za rad“.
2. **Obavezno prijavljivanje:** Korištenje LLM-a u pisanju *mora biti eksplicitno navedeno* u odjeljku Metode ili Zahvale.
3. **Zabrana za recenzente:** Recenzentima je *strogo zabranjeno uploadati rukopise* u generativne AI alate (poput ChatGPT-a) zbog kršenja povjerljivosti podataka.

Ova posljednja zabrana direktno pogađa etičko pitanje povjerljivosti rukopisa i privatnosti podataka, jer svaki *upload* na javni AI alat predstavlja potencijalno kršenje autorskih prava i odavanje neobjavljenih podataka.

5.3. Jaz između politike i prakse

Ovdje leži neriješeni sukob našeg polja u 2025. godini. Politike (poput one Springer Naturea) u potpunosti ovise o *prijavljivanju* od strane autora (Točka 2).

Međutim, nalaz „Paradoksa usvajanja“ (AlFayyad et al., 2025) – da **samo 5,7 % autora prijavljuje** upotrebu, dok je **stvarna upotreba 50–76 %** – znači da se ove politike masovno krše, vjerojatno zbog straha od stigmatizacije ili nejasnoća oko toga što prijaviti (npr. lektura naspram generiranja) (AlFayyad et al., 2025). Jaz između politike i prakse je toliki da su trenutne regulatorne mjere gotovo u potpunosti neučinkovite, ostavljajući urednike i recenzente „slijepima“ na stvarnu ulogu AI-a u rukopisima koje ocjenjuju.

6. Ograničenja studije

U skladu s našom metodologijom *Scoping Reviewa*, ključno je priznati nekoliko temeljnih ograničenja ove studije.

Ovisnost o „sivoj literaturi“ (Preprinti): Ovo polje se razvija brže od tradicionalnih ciklusa objavljivanja. Posljedično, naš rad se uvelike oslanja na *preprint* izvore (npr. arXiv), uključujući naše ključne studije (Li et al., 2025; Liang et al., 2025). Štoviše, ključni nalazi ove revizije također potječu iz takvih izvora: podaci o produktivnosti dolaze iz METR tehničkog izvješća (Becker et al., 2025), podaci o usvajanju s konferencijske prezentacije (AlFayyad et al., 2025), a metrike performansi iz industrijskog mjerila (Propel, 2025). Iako su ovi izvori metodološki pregledani i verificirani od strane AI recenzenata (vidi Odjeljak 2.4), oni još nisu prošli formalnu, rigoroznu ljudsku recenziju.
Brza zastarjelost podataka: Ovaj pregled je „snimka“ stanja u 2025. godini. Metrike performansi (npr. SWE-bench rezultati za Claude 3.5 ili DeepSeek-V3) vjerojatno će biti zastarjele unutar nekoliko mjeseci od objave.
Metodološka ograničenja samih izvora: Naši ključni nalazi ovise o metodama koje imaju svoja ograničenja. Najvažnije, studija LLM-REVal (Li et al., 2025) o pristranosti (78 % vs 49 %) je *simulacija*, a ne promatranje u „divljini“. „Paradoks usvajanja“ (AlFayyad et al., 2025) temelji se na *analizi podnesaka i anketama* (samo-prijavljivanju). Ovi nalazi su stoga indikativni, ali ne i definitivni dokazi o procesima u stvarnom svijetu.
Ključno metodološko ograničenje: Nedostatak PRISMA dijagrama toka: Iako rad slijedi PRISMA-ScR smjernice, formalni dijagram toka koji prikazuje točan broj pretraženih, filtriranih i uključenih studija nije priložen. Ovo je prepoznato ograničenje, a dodavanje ovakvog dijagrama predstavljalo bi ključno poboljšanje za potpunu metodološku transparentnost i ponovljivost u budućim verzijama rada.

7. Zaključak i preporuke (Model „AI-Asistent, Čovjek-Sudac“)

Ovaj metodološki pregled (*Scoping Review*) krenuo je od jednostavnog pitanja „simbioze ili sukoba“, ali je otkrio složeni ekosustav opterećen paradoksima koji su definirali znanstveno izdavaštvo u 2025. godini. Naši nalazi, sada metodološki utemeljeni i logički pomireni, ukazuju na jasan, iako oprezan, put naprijed.

Naša analiza identificirala je tri ključna paradoksa:

1. **Paradoks usvajanja:** AI se masovno koristi (do 76 %), ali gotovo nikad ne prijavljuje (5,7 %), čineći regulatorne politike (Springer Nature, 2025) i podatke (AlFayyad et al., 2025) gotovo neprovedivima.
2. **Paradoks performansi:** AI dramatično ubrzava jednostavne zadatke (npr. odabir recenzenata), ali usporava (za 19 %) stručnjake na složenim zadacima (Becker et al., 2025), dijelom zbog kognitivnog opterećenja provjere lažno pozitivnih nalaza (Propel, 2025).
3. **Paradoks pristranosti:** Dok se bojimo da AI uvodi nove pristranosti (npr. stilska preferencija prema AI-generiranom tekstu, kako je simulirano u Li et al., 2025), on također može pomoći u smanjenju postojećih ljudskih pristranosti (Jain & Verma, 2024).

Ovi paradoksi čine potpunu automatizaciju recenzentskog postupka ne samo tehnički nepoželjnom, već i etički opasnom. Studija LLM-REVal (Li et al., 2025) jasno pokazuje da AI-u

prepuštenom samom sebi ne treba vjerovati kao *sucu*, jer favorizira druge AI-e.

Stoga, optimalni model za 2025. godinu ostaje strogi hibridni model **„AI-Asistent, Čovjek-Sudac“**. Preporučujemo model koji strogo razdvaja uloge:

AI (Asistent): AI treba koristiti za *mehaničke* i *administrativne* zadatke. To uključuje početni *screening* (npr. SciScore, Ripeta), predlaganje recenzenata (Farber et al., 2024) i moderiranje dijaloga kako bi se poboljšala jasnoća, kao što je uspješno testirano u ICLR 2025 RCT studiji (ICLR 2025 Research Team, 2025).
Čovjek (Sudac): Ljudski urednici i recenzenti moraju zadržati isključivu kontrolu nad *kognitivnim* i *evaluacijskim* zadacima. To uključuje konačni odabir recenzenata, procjenu novosti (*novelty*) i znanstvenog značaja, te, što je najvažnije, *aktivni nadzor* nad AI pristranostima.

Kao konkretan primjer, predlažemo operativni model za urednike časopisa:

Korak 1 (Automatizirani Screening): Svi rukopisi pri predaji prolaze automatiziranu provjeru integriteta (npr. Ripeta, SciScore) radi provjere metodološke rigoroznosti (npr. navođenje izvora podataka, etička odobrenja). **Važno je napomenuti da ovi alati ne detektiraju AI autorstvo, već osiguravaju temeljnu znanstvenu kvalitetu** prije nego rukopis stigne do ljudskih recenzenata zaduženih za procjenu novosti i sadržaja.
Korak 2 (AI Asistent - Tehnička provjera): AI se koristi za administrativne zadatke (prijedlog recenzenata) i provjeru specifičnih, objektivnih mjerila (npr. provjera statističkih grešaka) gdje je stopa lažno pozitivnih nalaza niska.
Korak 3 (Čovjek Sudac - Evaluacija): Ljudski urednici i recenzenti donose isključivu odluku o znanstvenoj novosti, značaju i interpretaciji. Svaka AI sugestija mora biti eksplicitno odobrena od strane čovjeka.

Budućnost ne leži u sukobu ili potpunoj simbiozi, već u opreznoj, nadziranoj suradnji gdje AI služi kao alat za osnaživanje ljudske prosudbe, a ne kao njezina zamjena. Implementacija ovakvog hibridnog modela, usklađenog s odredbama EU AI Acta, bit će ključan izazov za nacionalna akademska tijela, uključujući hrvatske izdavače i agencije poput HRZZ-a, u nadolazećim godinama.

Literatura

AlFayyad et al. (2025, rujan). *Authors Self-disclosed Use of Artificial Intelligence in Research: Submissions to 49 BMJ Group Biomedical Journals*. [Prezentacija na 10th International Congress on Peer Review (PRC2025), Chicago, IL, Sjedinjene Američke Države]. Sažetak dostupan na: https://peerreviewcongress.org/peer-review-congress-2025-program/

Anthropic. (2025). *Raising the bar on SWE-bench Verified with Claude 3.5 Sonnet*. Anthropic Research. Preuzeto 15. studenog 2025. s https://www.anthropic.com/research/swe-bench-sonnet

Becker, J., et al. (2025). *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*. METR. Preuzeto 15. studenog 2025. s https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/ (Također dostupno na: https://arxiv.org/pdf/2507.09089)

DeepSeek AI. (2025). *DeepSeek-V3 Technical Report*. DeepSeek. Preuzeto 15. studenog 2025. s https://deepseek.ai/reports/deepseek-v3

Farber, D., J., Peterson, A., & Lee, S. (2024). AI-Assisted Reviewer Selection: A 73% Time Reduction in Editorial Workflows. *Journal of Scientific Publishing, 12*(2), 104–115. https://doi.org/10.1002/leap.1638

Gornitzki, K., et al. (2021). Associations between author-suggested reviewers and peer review outcomes: a cohort study. *BMC Biology, 19*(1), 1–8. https://doi.org/10.1186/s12915-021-00966-2

HEC. (2025). *AI in Talent Acquisition – How Artificial Intelligence is Transforming Recruitment Globally*. HE Consulting. Preuzeto 15. studenog 2025. s https://heconsulting.us/ai-in-talent-acquisition-how-artificial-intelligence-is-transforming-recruitment-globally/

Hirin.ai. (2025). *Finance Recruitment Challenges vs AI Recruitment Solutions*. Preuzeto 15. studenog 2025. s https://www.hirin.ai/blog/financial-recruitment-challenges-ai-solutions/

ICLR 2025 Research Team. (2025). *Can LLM feedback enhance review quality? A randomized controlled trial at ICLR 2025*. arXiv. https://arxiv.org/abs/2504.09737

Jain, R., & Verma, S. (2024). AI in Talent Acquisition: A Game Changer for Modern Recruitment. *Journal of Informatics Education and Research, 5*(1). Preuzeto s https://jier.org/index.php/journal/article/download/2673/2183/4750

JAMA. (2022). Peer review: A flawed process in need of evolution. *JAMA, 327*(8), 711–712. https://doi.org/10.1001/jama.2022.2152

Liang, P., et al. (2025). *Unveiling (Hidden) Bias in LLM-assisted Peer Reviews*. arXiv. https://arxiv.org/abs/2509.13400

Li, R., Gu, J.-C., Kung, P.-N., Xia, H., Liu, J., Kong, X., Sui, Z., & Peng, N. (2025). *LLM-REVal: Can We Trust LLM Reviewers Yet?*. arXiv. https://arxiv.org/abs/2510.12367

LMU Munich Emergency Medicine Team. (2024). GPT-4 Performance in Internal Medicine Emergency Department Cases. *JMIR, 26*, e56110. https://doi.org/10.2196/56110

MokaHR. (2025). *How AI is Revolutionizing Hiring Practices Today*. Preuzeto 15. studenog 2025. s https://www.mokahr.io/myblog/ai-hiring-predictions-2025-trends/

PRC2025. (2025, rujan). *10th International Congress on Peer Review and Scientific Publication*. [Konferencijski program]. Chicago, IL, Sjedinjene Američke Države. https://peerreviewcongress.org/

Propel. (2025). *AI Code Review Showdown: Claude vs GPT-4 vs Gemini in 2025*. Preuzeto 15. studenog 2025. s https://www.propelcode.ai/blog/ai-code-review-showdown-claude-vs-gpt4-vs-gemini-2025

Springer Nature. (2025). *Editorial policies on artificial intelligence*. Preuzeto 15. studenog 2025. s https://www.springernature.com/gp/policies/editorial-policies

Tricco, A. C., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0550