BF-P12 VEŠTAČKA INTELIGENCIJA U SLUŽBI BEZBEDNOSTI PACIJENATA: PILOT EVALUACIJA LLM-OVA ZA ANALIZU INTERAKCIJA LEKOVA

N. Jovanović-Lješković[1], M. Markočević[1], I. Krtolica[2], V. Ilić[2], M. Zloh[1]
[1]*Univerzitet Privredna akademija u Novom Sadu – Farmaceutski fakultet, Novi Sad, Srbija,* [2]*Istraživačko-razvojni institut za veštačku inteligenciju Srbije, Novi Sad, Srbija*

Interakcije između lekova (DDIs) predstavljaju izazov u farmakoterapiji pacijenata sa hroničnim nezaraznim bolestima, uključujući kardiovaskularna oboljenja i dijabetes, zbog potencijalne polifarmacije i povećanog rizika od klinički relevantnih interakcija. Razvoj efikasnih metoda za ranu identifikaciju DDIs ima ključnu ulogu u unapređenju bezbednosti pacijenata. Primena savremenih digitalnih alata, poput veštačke inteligencije (AI), predstavlja potencijalni pristup za unapređenje identifikacije i interpretacije DDIs. Ovaj pilot-projekat procenjuje mogućnosti velikih jezičkih modela (LLMs) u identifikaciji i proceni DDIs. Četiri savremena LLM-a (ChatGPT, Gemini, Claude i lokalni Qwen) odabrana na osnovu učestalosti korišćenja i opšte dostupnosti, evaluirana su kroz modularni analitički okvir koji integriše sažetke karakteristika leka iz baze ALIMS-a i platformu RxList. Primenjen je dvostruki pristup: pacijentski orijentisani upiti (laički) i farmaceutski strukturirani upit (stručni) zasnovani na JSON podacima o lekovima. Analizirano je šest parova interakcija: glimepirid–ciprofloksacin, glimepirid–varfarin, metformin–ibuprofen, inzulin aspart–prednizon, atorvastatin–klaritromicin i semaglutid–inzulin aspart. Tačnost i pouzdanost odgovora procenjivani su kroz stepen podudaranja identifikacije i kategorizacije težine rizika sa referentnim izvorima. Svi modeli su prepoznali klinički značajne interakcije, uz razlike u proceni njihove težine. Gemini (r=0.91 za stručni upit) i ChatGPT (r=0.86 za oba tipa upita) pokazali su najveći stepen podudarnosti sa procenama RxList-a. Najveći izazov predstavljala je kombinacija atorvastatin–klaritromicin, gde su jedino Qwen i stručni Gemini korektno identifikovali kontraindikaciju. Pacijentski prompti generisali su konzervativnije procene rizika, dok su farmaceutski prompti omogućili preciznije razlikovanje kliničke relevantnosti. Nalazi su preliminarni, ali ukazuju na potencijal velikih jezičkih modela kao pomoćnih alata u identifikaciji i proceni DDIs, uz neophodnost dalje validacije na većim i reprezentativnijim skupovima podataka.

Zahvalnica
Ovaj rad je finansirao Pokrajinski sekretarijat za visoko obrazovanje i naučnoistraživačku delatnost Autonomne Pokrajine Vojvodine, Republika Srbija, kroz projekat broj: 003782327 2025 09418 003 000 000 001/2.

BF-P12 ARTIFICIAL INTELLIGENCE FOR PATIENT SAFETY: PILOT EVALUATION OF LLMs FOR DRUG-DRUG INTERACTION ANALYSIS

N. Jovanović-Lješković[1], M. Markočević[1], I. Krtolica[2], V. Ilić[2], M. Zloh[1]
[1]*The University Business Academy in Novi Sad – Faculty of Pharmacy, Novi Sad, Serbia,* [2]*The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia*

Drug–drug interactions (DDIs) represent a significant challenge in the pharmacotherapy of patients with chronic non-communicable diseases, including cardiovascular diseases and diabetes, due to polypharmacy and increased risk of clinically relevant interactions. Effective approaches for early DDI identification are essential for improving patient safety. Artificial intelligence (AI)-based tools represent a potential approach for enhancing DDI identification and interpretation. This pilot study evaluated the potential of large language models (LLMs) for DDIs identification and clinical risk assessment. Four LLMs (ChatGPT, Gemini, Claude, and the locally deployed Qwen model), selected based on widespread use and accessibility, were evaluated using a modular framework integrating drug information from the ALIMS database and the RxList platform. A dual-prompting approach was applied, including patient-oriented (layperson) queries and structured pharmacist-oriented (expert) queries based on JSON-formatted drug information. Six DDI pairs were analysed: glimepiride–ciprofloxacin, glimepiride–warfarin, metformin–ibuprofen, insulin aspart–prednisone, atorvastatin–clarithromycin, and semaglutide–insulin aspart. Model performance was assessed by comparing interaction identification and risk severity categorization with reference sources. All models identified clinically relevant interactions, although differences were observed in risk severity assessment. Gemini (r=0.91 for expert queries) and ChatGPT (r=0.86 for both query types) demonstrated the highest agreement with RxList assessments. The atorvastatin–clarithromycin combination represented the most challenging case, with only Qwen and expert Gemini correctly identifying the contraindication. Patient-oriented prompts generated more conservative risk assessments, whereas expert prompts enabled more precise differentiation of clinical relevance. These preliminary findings suggest that LLMs may serve as supportive tools for DDI assessment, although further validation using larger and more representative datasets is required.

Acknowledgement
This work was funded by the Provincial Secretariat for Higher Education and Scientific Research, Autonomous Province of Vojvodina, Republic of Serbia: 003782327 2025 09418 003 000 000 001/2.