

Secțiunea „Rezultate studiu clinic”

DEZVĂLUIREA VULNERABILITĂȚILOR MODELELOR LINGVISTICE MARI ÎN TERAPIA INTENSIVĂ: PERSPECTIVE DIN APLICAȚIILE DE GESTIONARE A SEPSISULUI

Victor Iapăscuță^{1,2}

¹ Universitatea Tehnică a Moldovei, Chișinău, Republica Moldova

² Universitatea de Medicină și Farmacie Nicolae Testemițanu, Chișinău, Republica Moldova

Introducere. Sepsisul rămâne o cauză principală a mortalității în ATI, necesitând decizii rapide, bazate pe dovezi, privind terapia cu antibiotice și respectarea ghidurilor. Modelele lingvistice mari (LLM) integrate în sistemele multi-agent oferă instrumente promițătoare de asistență decizională, însă tendința lor spre halucinații - rezultate neacceptate sau excesive - prezintă riscuri semnificative în contexte cu mize mari. Acest studiu evaluează un sistem bazat pe LLM multi-agent pentru gestionarea sepsisului, evidențiind aceste riscuri pentru a informa aplicații mai sigure în terapia intensivă.

Materiale și metode. Un sistem multi-agent a fost dezvoltat folosind Palmyra-Med 70B LLM, orchestrat prin Crew.ai, cu trei agenți: gestionarea sepsisului, recomandarea antibioticelor și respectarea ghidurilor. Generarea augmentată de recuperare (RAG) a utilizat baze de date vectoriale Chroma populate cu literatură și ghiduri recente despre sepsis. Zece cazuri diverse de sepsis (8 din baza de date MIMIC-IV, 2 din literatura de specialitate) au fost procesate ca viniete. Rezultatele au fost evaluate programatic folosind TruLens (prin GPT-3.5 Turbo) pentru fundamentare, relevanță a răspunsului și relevanță contextuală (cu scor 0-1). Doi intensivști au furnizat evaluări de specialitate pe o scală de la 0 la 10 pentru relevanță și fundamentare. Acordul interrater a fost măsurat cu Kappa Cohen.

Rezultate. Recomandările de antibiotice au obținut scoruri medii mari (0,89-1,0), reflectând o fundamentare puternică a dovezilor. Cu toate acestea, rezultatele privind gestionarea sepsisului și conformitatea au avut scoruri mai mici (0,44-0,94 și, respectiv, 0,50-0,89), cu halucinații frecvente (de exemplu, semne clinice fabricate în cazurile de fasciită necrozantă) și detalii excesive. Evaluările experților au reflectat tendințele programatice, cu Kappa Cohen la 0,622 ($p=0,003$) pentru cazul inițial, dar o variabilitate mai mică în celelalte cazuri. Testarea comparativă a LLM (Palmyra-Med 70B, GPT-3.5 Turbo, GPT-4o Mini) a semnalat halucinații în 20-40% din rezultate per LLM.

Concluzii. Deși sistemele multi-agent bazate pe LLM excelează în sugestiile rapide de antibiotice, halucinațiile compromit fiabilitatea în gestionarea sepsisului, riscând erori clinice în îngrijirea la ATI. Rafinarea RAG și a prompturilor de agent este esențială; studiile viitoare ar trebui să valideze în timp real și să exploreze cadre hibride om-IA pentru a atenua aceste riscuri și a spori încrederea.

Cuvinte cheie: managementul sepsisului, modele lingvistice mari, halucinații, sisteme multi-agent, terapie intensivă.

Studiul a fost finanțat de Agenția Națională pentru Cercetare și Dezvoltare (ANCD) din Moldova în cadrul grantului 25.00208.5007.07/PD pentru proiectul „Metode și algoritmi software inteligenți pentru terapia intensivă. Abordare prin prisma științei complexității”, implementat la Universitatea Tehnică a Moldovei.

Section "Clinical study results"

UNVEILING LARGE LANGUAGE MODEL VULNERABILITIES IN INTENSIVE CARE: INSIGHTS FROM SEPSIS MANAGEMENT APPLICATIONS

Victor Iapăscuță^{1,2}

¹ Technical University of Moldova, Chișinău, Republic of Moldova

² Nicolae Testemițanu University of Medicine and Pharmacy, Chișinău, Republic of Moldova

Introduction. Sepsis remains a leading cause of ICU mortality, demanding rapid, evidence-based decisions for antibiotic therapy and guideline adherence. Large language models (LLMs) integrated into multi-agent systems offer promising decision-support tools, yet their propensity for hallucinations—unsupported or

excessive outputs—poses significant risks in high-stakes settings. This study evaluates a multi-agent LLM-based system for sepsis management, highlighting these risks to inform safer applications in intensive care.

Materials and Methods. A multi-agent system was developed using Palmyra-Med 70B LLM, orchestrated via Crew.ai, with three agents: sepsis management, antibiotic recommendation, and guidelines compliance. Retrieval-augmented generation (RAG) utilized Chroma vector databases populated with recent sepsis literature and guidelines. Ten diverse sepsis cases (8 from MIMIC-IV database, 2 from literature) were processed as vignettes. Outputs were assessed programmatically using TruLens (via GPT-3.5 Turbo) for groundedness, answer relevance, and context relevance (scored 0-1). Two intensivists provided expert evaluations on a 0-10 scale for relevance and groundedness. Interrater agreement was measured with Cohen's Kappa.

Results. Antibiotic recommendations achieved high mean scores (0.89-1.0), reflecting strong evidence grounding. However, sepsis management and compliance outputs scored lower (0.44-0.94 and 0.50-0.89, respectively), with frequent hallucinations (e.g., fabricated clinical signs in necrotizing fasciitis cases) and excessive details. Expert ratings mirrored programmatic trends, with Cohen's Kappa at 0.622 ($p=0.003$) for the initial case but lower variability across others. Comparative LLM testing (Palmyra-Med 70B, GPT-3.5 Turbo, GPT-4o Mini) flagged hallucinations in 20-40% of outputs per LLM.

Conclusions. While LLM-based multi-agent systems excel in rapid antibiotic suggestions, hallucinations compromise reliability in sepsis management, risking clinical errors in ICU care. Refining RAG and agent prompts is essential; future studies should validate in real-time settings and explore hybrid human-AI frameworks to mitigate these risks and enhance trustworthiness.

Keywords: Sepsis management, large language models, hallucinations, multi-agent systems, intensive care.

The study was funded by the National Agency for Research and Development (NARD) of Moldova under grant number 25.00208.5007.07/PD for the project "Intelligent Software Methods and Algorithms for Intensive Care: A Complexity Science Approach," implemented at the Technical University of Moldova.