



DOI: 10.5281/zenodo.23364665 · UDC: 616.94-08:004.89

HALUCINAȚIILE MODELELOR DE LIMBAJ MARI ÎN SEPSIS: SINTEZĂ NARATIVĂ ȘI ANALIZA A ZECE CAZURI

LARGE LANGUAGE MODELS IN SEPSIS WITH HALLUCINATIONS: NARRATIVE REVIEW AND ANALYSIS OF TEN CASES

Victor Iapăscurtă^{1,2*}, Dinu Țurcanu², Adrian Belii¹, Viorel Bostan²

¹ Catedra de anesteziologie și reanimatologie nr.1 "Valeriu Ghereg", Universitatea de Stat de Medicină și Farmacie „Nicolae Testemițanu”, Chișinău, Republica Moldova;

² Departamentul de inginerie software și automatică, Universitatea Tehnică a Moldovei, Chișinău, Republica Moldova.

* Corresponding author: victor.iapascurta@usmf.md

REZUMAT

Obiective. Utilizarea modelelor de limbaj mari (LLM) în medicina de urgență și terapia intensivă avansează rapid, însă generarea de „halucinații” - afirmații neverificabile, contradictorii cu ghidurile clinice sau extrapolate nejustificat - rămâne un risc major pentru siguranța pacientului. Lucrarea de față extinde studiile anterioare ale grupului nostru publicate în Arta Medica (2024) și Healthcare Informatics Research (2026) și își propune să caracterizeze prevalența, tipologia și impactul potențial al halucinațiilor generate de un LLM medical specializat (Palmyra-Med 70B) în managementul precoce al sepsisului, pornind de la un caz-index și de la o cohortă de zece cazuri clinice.

Material și metode. Vinetele clinice pentru 10 cazuri de sepsis (opt din baza de date deschisă MIMIC-IV-demo și două din literatura de specialitate) au fost introduse într-o aplicație software bazată pe LLM-ul Palmyra-Med 70B, integrat într-un sistem multi-agent cu generare augmentată prin recuperare (RAG). Recomandările generate (management, antibiotice, conformitate cu ghidul Surviving Sepsis Campaign) au fost evaluate independent de doi medici intensiviști pe baza unui formular standardizat, iar halucinațiile au fost definite operațional și marcate consensual.

Rezultate. LLM-ul a produs recomandări cu relevanță și fundamentare medii-mari (scoruri predominant 7–10/10 la ambii evaluatori). Halucinații au fost identificate în 3 din 10 cazuri (Cazurile 1, 9 și 10), toate legate de fabricarea sau extrapolarea unor semne/valori clinice absente în vinietă (stare mentală alterată, hiperlactemie, hiperglicemie, șoc septic). Cazurile fără halucinații (2–8) au arătat totuși divergențe minore față de ghiduri, exprimate prin scoruri de conformitate suboptime.

Concluzii. Palmyra-Med 70B poate oferi suport util în etapa precoce a managementului sepsisului, dar necesită supraveghere clinică activă și o strategie hibridă clinician-LLM pentru a limita halucinațiile și a proteja siguranța pacientului.

Cuvinte-cheie: sepsis, modele de limbaj mari, halucinații, suport decizional clinic, siguranța pacientului, MIMIC-IV

ABSTRACT

Objectives. Large language models (LLMs) are advancing rapidly in emergency and critical care, but the generation of “hallucinations” - unverifiable statements that contradict guidelines or extrapolate without support — remains a major patient-safety risk. Building on our previous work in Arta Medica (2024) and Healthcare Informatics Research (2026), this study characterises the prevalence, typology and potential impact of hallucinations produced by a medical LLM (Palmyra-Med 70B) in early sepsis management, based on an index case and a cohort of ten clinical vignettes.

Material and Methods. Ten sepsis vignettes (eight from the open MIMIC-IV-demo database and two from the literature) were submitted to a software application based on the Palmyra-Med 70B LLM, embedded in a retrieval-augmented multi-agent system. The generated recommendations (management, antibiotics, guideline compliance with the Surviving Sepsis Campaign) were assessed independently by two ICU physicians using a standardised form; hallucinations were operationally defined and flagged by consensus.

Results. The LLM produced recommendations of medium-to-high relevance and grounding (mostly 7–10/10 across both raters). Hallucinations were identified in 3 of 10 cases (Cases 1, 9 and 10), all involving the fabrication or unjustified extrapolation of clinical signs/values absent from the vignette (altered mental status, hyperlactatemia, hyperglycemia,

septic shock). Non-hallucinating cases (2–8) still showed minor divergences from the guidelines, expressed as suboptimal compliance scores.

Conclusions. Palmyra-Med 70B can provide useful support in the early phase of sepsis management, but requires active clinical oversight and a hybrid clinician-LLM workflow to limit hallucinations and safeguard patient safety.

Keywords: sepsis, large language models, hallucinations, clinical decision support, patient safety, MIMIC-IV

Introducere

Sepsisul rămâne una dintre cele mai frecvente cauze de mortalitate în terapia intensivă, iar timpul de la recunoaștere la instituirea tratamentului corect este strâns corelat cu supraviețuirea [1, 2]. Fiecare oră de întârziere în inițierea antibioticelor adecvate crește riscul de deces cu aproximativ 9% în șocul septic [2]. În acest context, orice instrument capabil să sprijine decizia clinică rapidă la momentul $T = 0$ (înainte ca rezultatele hemoculturilor și ale altor teste microbiologice să fie disponibile) prezintă un interes evident.

Modelele de limbaj mari (large language models, LLM), în special variantele specializate medical, sunt propuse tot mai frecvent ca posibil suport decizional în intervenții critice [3–5]. Palmyra-Med 70B, unul dintre modelele cu cele mai bune performanțe pe benchmark-urile de tip MedQA (scor mediu 85,9) [6], a demonstrat capacitatea de a genera recomandări coerente și relevante în scenarii simulate de sepsis atunci când este cuplat cu ghiduri actualizate prin generare augmentată prin recuperare (retrieval-augmented generation, RAG) [7]. În lucrarea noastră anterioară din Arta Medica (2024) [8] am descris un prototip de arhitectură multi-agent care combină analiza predictivă cu un modul LLM pentru recomandări precoce în sepsis, iar în evaluarea de tip follow-up publicată în Healthcare Informatics Research (2026) [9] am demonstrat fezabilitatea sistemului pe zece cazuri clinice, evidențiind în același timp apariția halucinațiilor.

Prin „halucinație” în contextul LLM-urilor medicale înțelegem o afirmație generată de model care: (i) este neverificabilă în raport cu datele furnizate în prompt-ul clinic (de exemplu, o valoare de laborator inexistentă); (ii) contrazice explicit ghidurile clinice standard (de exemplu, Surviving Sepsis Campaign); sau (iii) extrapolează nejustificat din informații incomplete și introduce elemente clinice noi (de exemplu, atribuirea unui diagnostic funcțional - șoc septic, stare mentală alterată - în absența oricărei mențiuni suportive) [10, 11]. Această definiție operațională, aliniată revizuirilor recente ale fenomenului [10], este cea aplicată în studiul de față.

Halucinațiile în medicina de urgență și în terapia intensivă comportă riscuri specifice: pot influența alegerea și dozajul antibioticelor, pot supraestima severitatea prin fabricarea unor indicatori (lactat, tensiune, stare de conștiență) și pot induce recomandări de escaladare terapeutică - vasopresoare, corticosteroizi - nejustificate [10–13]. Studii recente au raportat, pentru LLM-uri medicale generaliste, rate de halucinație clinic semnificative de la câteva procente până la 20–30%, în funcție de complexitatea prompt-ului și de specialitate [11, 12, 14].

Sinteza literaturii recente (2023–2026) sugerează două

direcții principale de răspuns: (a) integrarea LLM-ului într-un flux hibrid clinician-mașină, cu validare umană obligatorie [3, 15, 16]; și (b) folosirea RAG pentru a ancora răspunsul în surse verificate, cu monitorizare programatică a fundamentării răspunsului [7, 9, 17]. Ambele abordări reduc, dar nu elimină, halucinațiile - motiv pentru care metrice standardizate de evaluare, aplicate de clinicieni experimentați, rămân indispensabile.

Lucrarea de față se plasează la intersecția acestor două direcții. Ea extinde analiza raportului de caz inițial (cazul MIMIC-IV ID 10015860 - sepsis pe fond de fasciită necrozantă a piciorului) cu o analiză detaliată a tuturor celor 10 cazuri incluse anterior în studiul follow-up [9], furnizând viniete clinice complete, o definiție reproductibilă a halucinației și un cadru metodologic transparent de evaluare. Obiectivele specifice sunt: (1) descrierea sistematică a cazului-index; (2) prezentarea vinietelor clinice pentru toate cele 10 cazuri; (3) cuantificarea și tipologia halucinațiilor generate de LLM-ul Palmyra-Med 70B; (4) discutarea implicațiilor pentru siguranța pacientului și pentru proiectarea sistemelor viitoare de suport decizional.

Prezentarea cazului de referință

Cazul de referință provine din baza de date deschisă MIMIC-IV-demo, subject_ID 10015860 [18]. Este vorba despre un pacient de sex masculin, în vârstă de 53 de ani, internat în unitatea de terapie intensivă cu diagnosticul de sepsis secundar fasciitei necrozante a piciorului drept, apărute în contextul unei ulcerări cronice a extremității.

La internare, semnele vitale erau: frecvența cardiacă 106 bpm, tensiunea arterială 116/62 mmHg, frecvența respiratorie 23 rpm, saturația periferică în oxigen (SpO_2) 93% în aer atmosferic, temperatura 38 °C.

Investigațiile de laborator au evidențiat: leucocite $18,0 \times 10^9/L$, glicemie 11,9 mmol/L, creatinină serică 70,7 $\mu\text{mol/L}$, uree 5,7 mmol/L, lactat seric 0,7 mmol/L.

Antecedentele personale patologice includeau: diabet zaharat tip 2 cu boală cronică de rinichi diabetică, neuropatie diabetică, ulcer cronic al piciorului și absența dobândită a unui deget de la piciorul drept.

Diagnosticul de sepsis a fost stabilit pe criterii clinice Sepsis-3 (scor SOFA în creștere, focar infecțios documentat) [1]. Cazul întrunea criteriile pentru infecție severă a pielii și a țesuturilor moi și necesita, pe lângă terapia antimicrobiană empirică, control chirurgical urgent al focarului.

Este important de subliniat că pacientul nu a prezentat, în datele documentate în MIMIC-IV, o stare mentală alterată - element cu relevanță directă pentru analiza halucinațiilor prezentată ulterior. De asemenea, valoarea lactatului seric

la momentul evaluării inițiale era în limitele normale (0,7 mmol/L), ceea ce exclude o hiperlactatemie declarată la debutul evaluării.

Cazul-index este suficient de complex (infecție necrozantă + comorbidități metabolice + boală renală cronică) pentru a genera provocări diagnostice și terapeutice reprezentative pentru practica de terapie intensivă și a fost ales pentru a testa în profunzime capacitatea LLM-ului de a produce recomandări corecte și fundamentate. Detaliile evoluției ulterioare (control chirurgical, evoluție hemodinamică, deznodământ) nu sunt însă esențiale pentru scopul prezentei analize, care se concentrează pe faza precoce ($T = 0$ h) a suportului decizional și pe caracterizarea răspunsului LLM.

Material și metode

Design și materiale. Studiul este un raport de caz extins cu analiză secundară pe o cohortă de zece cazuri de sepsis. Opt cazuri (Cazurile 1–8) au fost extrase din baza de date deschisă MIMIC-IV-demo (v2.2) [18], selectate pentru a acoperi principalele surse de infecție întâlnite în terapia intensivă: piele/țesuturi moi, pulmonar (comunitar și asociat ventilatorului), genito-urinar, abdominal și infecție sistemică pe cateter venos central. Două cazuri suplimentare au fost preluate din literatura de specialitate - Cazul 9 dintr-un raport publicat în *Annals of Medicine and Surgery* (2021) [19] și Cazul 10 dintr-un raport publicat în *Case Reports in Infectious Diseases* (2016) [20] - pentru a testa comportamentul modelului în afara distribuției MIMIC-IV. Anexa 1 a materialelor suplimentare prezintă materialul utilizat în cercetare la etapa de interacțiune cu aplicația software în baza LLM.

Instrumentul software. Aplicația software utilizată în studiu este descrisă în Anexa 4 a materialelor suplimentare și este disponibilă public sub licență MIT pe GitHub [21,22]. Componenta centrală este un sistem multi-agent, dezvoltat în Python, care coordonează trei agenți specializați: (a) agent pentru recomandări de management, (b) agent pentru recomandări antibiotice și (c) agent pentru verificarea conformității cu ghidurile clinice. Fiecare agent apelează același model de bază - Palmyra-Med 70B [6] - cu temperatură 0,25 (temperatura LLM - setare care controlează aleatoriul și creativitatea rezultatului unei inteligențe artificiale), iar contextul este îmbogățit prin RAG cu bază vectorială ChromaDB, care indexează *Surviving Sepsis Campaign 2021* [1] și peste 20 de reviiuri de mare impact publicate între 2018 și 2025 privind etiologiile și managementul sepsisului. Regăsirea semantică folosește embedding-uri BAAI/bge-base-en-v1.5, similaritate cosinus (prag 0,75) și top-5 fragmente. O versiune demo, bazată pe GPT-3.5-Turbo, este disponibilă pe platforma Hugging Face după solicitare la autorul corespondent [22]. Arhitectura sistemului multi-agent este redată în Figura 1.

Procesul de evaluare. Pentru fiecare caz, datele clinice au fost formate ca vinieta structurată (Diagnostic; Date demografice; Semne vitale; Rezultate de laborator; Antecedente medicale; Observații speciale) în engleză, pentru compatibilitate cu tokenizatorul modelului. Vinietele au fost introduse ca prompt în aplicație, iar cei trei agenți au generat succesiv: (i) recomandările de management (MR - management recommendations); (ii) recomandările antibiotice (AR - antibiotic recommendations); și (iii) declarația de conformitate cu ghidul *Surviving Sepsis Campaign 2021* (CS - compliance statement).

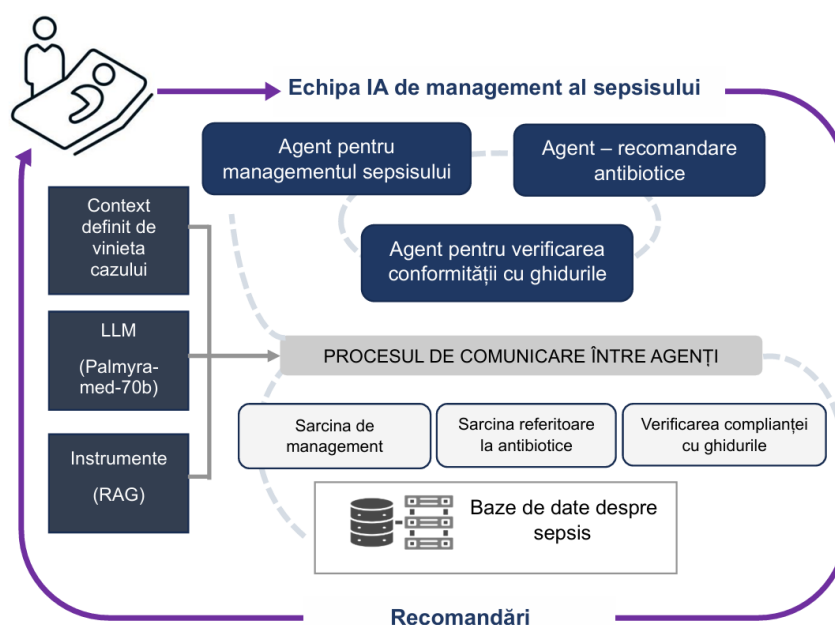


Figura 1. Arhitectura sistemului multi-agent bazat pe Palmyra-Med 70B utilizat pentru generarea recomandărilor în sepsis. RAG - retrieval-augmented generation (generare augmentată prin recuperare). Cei trei agenți virtuali (management, antibiotice, conformitate) partajează același model de bază și accesează, prin RAG, o bază de date vectorială care indexează ghidul *Surviving Sepsis Campaign 2021* și reviiuri relevante.

Recomandările au fost evaluate independent de doi medici intensiviști experimentați, orbi între ei, pe baza formularului standardizat prezentat în Anexa 2. Formularul cuprinde, pentru fiecare tip de recomandare, două dimensiuni cotate pe scala 0–10: relevanța răspunsului (dacă răspunsul răspunde efectiv la interogare) și fundamentarea în context (dacă răspunsul este susținut de vinieta cazului). Un scor mic la fundamentare semnalează suspiciunea de halucinație. Comentariile libere ale evaluatorilor au fost apoi confruntate consensual, iar halucinațiile confirmate (afirmații fără suport în vinieta) au fost marcate „H” (Anexa 3).

Definiția operațională a halucinației. Un răspuns LLM a fost considerat halucinație dacă îndeplinea cel puțin unul dintre următoarele criterii: (i) menționa un semn clinic, o valoare de laborator sau un diagnostic funcțional absent în vinieta (de exemplu, „hiperlactatemie”, „stare mentală alterată”, „șoc septic” fără suport în datele furnizate); (ii) prescria o intervenție contrară recomandării standard (de exemplu, escaladarea antibioticelor înainte de rezultatele culturilor, când vinieta nu justifica); (iii) fabrica un fapt clinic (de exemplu, atribuirea unei etiologii microbiene nesusținute). Această definiție a fost stabilită a priori și aplicată identic tuturor celor 10 cazuri.

Prelucrarea și raportarea rezultatelor. Datele evaluatorilor au fost sintetizate descriptiv, prin prezentarea scorurilor de relevanță și fundamentare și a diferențelor dintre evaluatori. Cazurile complete cu toate ieșirile LLM sunt prezentate în Anexa 1 a materialelor suplimentare; rezultatele agregate de evaluare, în Anexa 3.

Vinietele clinice sinoptice ale celor 10 cazuri sunt prezentate în Tabelul 1, adaptate din Anexa 1, cu valorile de laborator convertite în unități SI pentru claritatea comunicării.

Cei 10 pacienți formează o cohortă eterogenă (patru sepsisuri de origine pulmonară, două cutanate necrozante,

două urinare, unul abdominal, unul de cateter), cu comorbidități numeroase și cu profiluri de severitate variate (lactat 0,7–21,4 mmol/L; leucocite 0,4–44,8 × 10⁹/L; creatinină 53–1290 μmol/L). Cazurile 9 și 10, provenite din literatură, aduc profiluri clinice diferite de cohorta MIMIC-IV - inclusiv un caz cu acidoză metabolică severă (Cazul 9) și unul cu insuficiență renală extremă (Cazul 10) - utile pentru testarea robusteții LLM-ului în afara distribuției de antrenament.

Rezultate

Sinteza scorurilor de evaluare (Anexa 3) arată că LLM-ul Palmyra-Med 70B a generat recomandări cu scoruri predominant între 7 și 10 pe scala 0–10, pentru toate cele trei domenii evaluate (management, antibiotice și conformitate). Media aritmetică a scorurilor de relevanță a fost de aproximativ 8,4/10 pentru evaluatorul 1 și 8,1/10 pentru evaluatorul 2, iar cea a scorurilor de fundamentare de 8,5/10 și, respectiv, 8,1/10. Diferențele dintre evaluatori au fost, în general, de 0–2 puncte. Evaluarea diferențiată este importantă deoarece răspunsurile LLM pot conține simultan informații clinic relevante și afirmații nefundamentate factual.

Din analiza consensuală, halucinații au fost identificate în trei din cele zece cazuri: Cazul 1 (fasciită necrozantă a piciorului), Cazul 9 (fasciită necrozantă a gleznei) și Cazul 10 (sepsis de origine urinară). Cazurile 2–8 nu au conținut afirmații neverificabile în raport cu vinieta. Separat de identificarea halucinațiilor, evaluarea conformității cu ghidurile a evidențiat și unele rezerve: în Cazul 1, decizia de conformitate a semnalat prematuritatea de-escaladării antibioterapiei; în Cazul 6, recomandările antibiotice au fost considerate excesiv de ample pentru CAP + VAP combinate.

Figura 2 sintetizează grafic distribuția halucinațiilor pe cele

Tabelul 1. Vinietele clinice sinoptice ale celor 10 cazuri de sepsis analizate (unități SI).

Caz	Diagnostic / sursă	Date demografice	Semne vitale	Laborator (SI)	Antecedente și observații
Cazul 1 (MIMIC-IV, ID 10015860)	Sepsis prin fasciită necrozantă a piciorului drept	M, 53 ani	FC 106 bpm; TA 116/62 mmHg; FR 23 rpm; SpO ₂ 93%; T 38 °C	L 18,0 × 10 ⁹ /L; Gli 11,9 mmol/L; Cr 70,7 μmol/L; Uree 5,7 mmol/L; Lactat 0,7 mmol/L	DZ tip 2 cu nefropatie diabetică; BCR; neuropatie diabetică; ulcer cronic al piciorului; absență deget
Cazul 2 (MIMIC-IV, ID 10020786)	Sepsis pe pneumonie comunitară	F, 86 ani	FC 76 bpm; TA 103/49 mmHg; FR 22 rpm; SpO ₂ 90%; T 37,2 °C	L 41,1 × 10 ⁹ /L; Gli 12,5 mmol/L; Uree 4,6 mmol/L	BPOC; fibrilație atrială; DZ tip 2 fără complicații; hipotiroidism
Cazul 3 (MIMIC-IV, ID 10003400)	Sepsis de origine genito-urinară	F, 72 ani	FC 110 bpm; TAM 55 mmHg; FR 22 rpm; SpO ₂ 99%; T 37,7 °C	L 4,6 × 10 ⁹ /L; Gli 7,2 mmol/L; Cr 123,8 μmol/L; Uree 10,7 mmol/L; Hb 76 g/L; Plt 57 × 10 ⁹ /L; Lactat 3,3 mmol/L	BCR hipertensivă; mielom multiplu; neoplazie rectală; fibrilație atrială; ICC; obezitate

Cazul 4 (MIMIC-IV, ID 10007818)	Sepsis prin peritonită	M, 69 ani	FC 91 bpm; TAM 67 mmHg; FR 24 rpm; SpO ₂ 95%; T 37,2 °C	L 21,3 × 10 ⁹ /L; Gli 7,7 mmol/L; Cr 53 μmol/L; Uree 4,6 mmol/L; Hb 123 g/L	DZ tip 2; hemoperitoneu; sindrom hepatorenal; BCR hipertensivă st. 1–4
Cazul 5 (MIMIC-IV, ID 10019003)	Sepsis prin pneumonită / pneumonie de aspirație	F, 65 ani	FC 95 bpm; TAM 50 mmHg; FR 17 rpm; SpO ₂ 94%; T 39,3 °C	L 44,8 × 10 ⁹ /L; Cr 106 μmol/L; Uree 6,4 mmol/L; Hb 73 g/L; Plt 41 × 10 ⁹ /L; Lactat 3,7 mmol/L	Cardiopatie hipertensivă cu insuf. cardiacă; BPOC; DZ tip 2; boală mielo-proliferativă cronică; obezitate morbidă; rezistență la penicilină
Cazul 6 (MIMIC-IV, ID 10037861)	Sepsis prin pneumonie asociată ventilatorului	M, 77 ani	FC 126 bpm; TAM 63 mmHg; FR 26 rpm; SpO ₂ 97%; T 38,4 °C	L 16,0 × 10 ⁹ /L; Gli 18,5 mmol/L; Cr 115 μmol/L; Uree 14,3 mmol/L; Lactat 2,1 mmol/L	ICC; hemoragie subdurală traumatică; status epileptic; DZ tip 2 cu nefropatie diabetică; litiază renală
Cazul 7 (MIMIC-IV, ID 10039708)	Sepsis prin pneumonie asociată ventilatorului	F, 46 ani	FC 73 bpm; TAM 59 mmHg; FR 14 rpm; SpO ₂ 96%; T 37,2 °C	L 7,5 × 10 ⁹ /L; Gli 8,0 mmol/L; Cr 344 μmol/L; Uree 8,2 mmol/L; Lactat 3,6 mmol/L	BCR terminală (hemodializă); ciroză alcoolică cu ascită; sindrom hepatorenal; chirurgie bariatrică; hipotiroidism
Cazul 8 (MIMIC-IV, ID 10035631)	Sepsis prin infecție sistemică pe cateter venos central	M, 63 ani	FC 112 bpm; TAM 73 mmHg; FR 17 rpm; SpO ₂ 95%; T 37,6 °C	L 0,4 × 10 ⁹ /L; Gli 9,3 mmol/L; Cr 389 μmol/L; Uree 7,9 mmol/L; Hb 76 g/L; Plt 14 × 10 ⁹ /L	Cancer mamar masculin sub chimioterapie; celulită și abces trunchi; leucemie acută mieloidă
Cazul 9 (Singh 2021 [19])	Sepsis prin fasciită necrozantă a gleznei drepte	F, 57 ani	FC 115 bpm; TAM 66,3 mmHg	Lactat 21,4 mmol/L; L 3,3 × 10 ⁹ /L; Cr 256 μmol/L; Troponină 0,110 ng/mL; pH 6,69; pCO ₂ 69,1 mmHg; HCO ₃ ⁻ 8,3 mmol/L; FiO ₂ 100% (ventilație mecanică)	HTA; dislipidemie; sindrom de intestin iritabil; BRGE
Cazul 10 (Millard 2016 [20])	Sepsis de origine urinară cu Clostridium perfringens	M, 61 ani	FC 137 bpm; TA 117/76 mmHg; T 37 °C; alterare a stării de conștiență	L 15,3 × 10 ⁹ /L; Hb 87 g/L; Cr 1290 μmol/L; Uree 50,4 mmol/L; Lactat 1,94 mmol/L; sumar de urină: proteinurie 2+, fără nitriți; CT: calcul ureteral proximal 5 mm cu hidronefroză	DZ; BCR; hepatită C; diaree cronică; amputație transtibială dreaptă (boală vasculară periferică)

Notă: M – masculin; F – feminin; FC – frecvența cardiacă; TA – tensiune arterială; TAM – tensiune arterială medie; FR – frecvență respiratorie; SpO₂ – saturație periferică în oxigen; T – temperatură; L – leucocite; Gli – glicemie; Cr – creatinină; Hb – hemoglobină; Plt – plachete; BCR – boală cronică de rinichi; DZ – diabet zaharat; BPOC – bronhopneumopatie cronică obstructivă; ICC – insuficiență cardiacă congestivă; HTA – hipertensiune arterială; BRGE – boală de reflux gastro-esofagian; VAP – pneumonie asociată ventilatorului.

trei dimensiuni evaluate, pentru toate cele 10 cazuri. Se observă o concentrare a halucinațiilor pe dimensiunea „recomandări de management” (Cazurile 1, 9, 10) și, pentru Cazul 10, și pe dimensiunea „conformitate cu ghidurile”. Recomandările antibiotice, mai standardizate și mai bine ancorate în ghiduri, nu au produs halucinații în niciunul dintre cazuri.

Tabelul 2 prezintă detaliat cele trei cazuri cu halucinații, în același format ca Anexa 1 (vinieta clinică - ieșire LLM), cu marcarea afirmațiilor problematice (asterisc).

Elementele halucinatorii identificate au fost, sintetic: (i) Cazul 1 - „stare mentală alterată” (absență în vinieta; pacientul nu prezenta modificări de conștiință documentate) și „hiperlactatemie” (contrazicătoare — lactatul măsurat era 0,7 mmol/L, în limite normale); (ii) Cazul 9 - invocarea „criteriilor de șoc septic” ca fapt confirmat, deși vinieta din articolul-sursă [19] nu conținea suficiente informații hemodinamice (o singură măsurare a TAM la limită) pentru a atesta șocul septic în sensul Sepsis-3; (iii) Cazul 10 - „hiperglicemie” și „hiperlactatemie / lactat crescut” invocate deși vinieta [20] nu raporta o glicemie și consemna un lactat de doar 1,94 mmol/L; menționarea lor a apărut atât în recomandările de management, cât și în declarația de conformitate.

Analiza calitativă indică un pattern comun: toate cele trei halucinații reprezintă extrapolări plauzibile clinic, dar nefundamentate factual — modelul „completează” tabloul clasic al șocului septic (lactat crescut, alterare a conștiinței)

chiar și când vinieta nu furnizează aceste elemente. Această tendință nu a fost prezentă la ieșirile antibiotice, care rămân în general fidele ghidurilor prin ancorarea RAG.

Din perspectiva impactului clinic potențial, halucinațiile identificate nu au condus la recomandări periculoase în sine (de exemplu, antibiotice greșite), dar au produs o supraestimare a severității care, într-un scenariu real, ar putea justifica intervenții mai agresive (vasopresoare, corticoterapie) sau ar putea distra atenția de la elemente autentic critice ale cazului. Aceste observații motivează dezvoltarea unor mecanisme automate de detecție (mismatch între prompt și răspuns) integrate în interfața aplicației, așa cum am descris preliminar în [9].

Discuții

Importanța studiului. Prin extinderea analizei de la cazul-index (fasciita necrozantă a piciorului, ID 10015860) la întreaga cohortă de zece cazuri, prezentul studiu oferă, pentru prima dată în cadrul demersului nostru, o vedere sistematică asupra fenomenului halucinației la un LLM medical specializat (Palmyra-Med 70B) în contextul sepsisului. Frecvența observată — 3/10 cazuri (30%) - se situează la limita superioară a plajei raportate în literatura recentă pentru sarcini medicale complexe (6–30%, în funcție de complexitatea prompt-ului și de specialitate) [10–12, 14]. Este important

	Recomandări management	Recomandări antibiotice	Conformitate cu ghidurile
Cazul 1: Sepsis prin fasciită necrozantă a piciorului (ID 10015860)			
Cazul 2: Sepsis prin pneumonie comunitară (ID 10020786)			
Cazul 3: Sepsis de origine genitourinară (ID 10003400)			
Cazul 4: Sepsis prin peritonită (ID 10007818)			
Cazul 5: Sepsis prin pneumonită / pneumonie de aspirație (ID 10019003)			
Cazul 6: Sepsis prin pneumonie asociată ventilatorului (ID 10037861)			
Cazul 7: Sepsis prin pneumonie asociată ventilatorului (ID 10039708)			
Cazul 8: Sepsis prin infecție sistemică pe cateter venos central (ID 10035631)			
Cazul 9: Sepsis prin fasciită necrozantă a gleznei (din literatură [19])			
Cazul 10: Sepsis de origine urinară (din literatură [20])			

Figura 2. Distribuția halucinațiilor generate de Palmyra-Med 70B pe cele 10 cazuri de sepsis analizate, pe cele trei dimensiuni evaluate (recomandări de management, recomandări antibiotice, conformitate cu ghidurile). Celulele roșii marchează prezența halucinațiilor, celulele albe absența lor.

Tabelul 2. Cazurile de sepsis cu halucinații identificate (1, 9, 10): vinieta clinică și ieșirea Palmyra-Med 70B, cu marcarea (*) afirmațiilor considerate halucinatorii.

Caz	Vinieta clinică (sinteză)	Ieșirea LLM — halucinații marcate (*)
Cazul 1 (MIMIC-IV, ID 10015860) — fasciită necrozantă a piciorului	M, 53 ani; FC 106 bpm; TA 116/62 mmHg; FR 23 rpm; SpO ₂ 93%; T 38 °C; L 18,0 × 10 ⁹ /L; lactat 0,7 mmol/L (normal); glicemie 11,9 mmol/L; fără stare mentală alterată documentată. DZ tip 2 cu BCR, ulcer al piciorului.	MR: „Given the presence of sepsis criteria such as altered mental status*, tachypnea, leukocytosis, hyperglycemia and hyperlactatemia*, close monitoring of vital signs, fluid resuscitation, and vasopressor support if needed are essential...” Halucinații: stare mentală alterată (absentă în vinieta); hiperlactatemie (lactatul era 0,7 mmol/L).
Cazul 9 (Singh 2021 [19]) — fasciită necrozantă a gleznei	F, 57 ani; FC 115 bpm; TAM 66,3 mmHg (o singură măsurare); lactat 21,4 mmol/L; pH 6,69; L 3,3 × 10 ⁹ /L; Cr 256 μmol/L; ventilație mecanică. HTA, dislipidemie, BRGE.	MR: „The patient with sepsis from necrotizing fasciitis of the right ankle presents with severe sepsis and septic shock* criteria, including organ dysfunction, hypoperfusion, and hypotension...” Halucinație: „criteriile de șoc septic” invocate ca fapt confirmat, deși datele hemodinamice din vinieta (o singură TAM la limită) nu sunt suficiente pentru a atesta șocul septic conform Sepsis-3.
Cazul 10 (Millard 2016 [20]) — sepsis urinar cu <i>Clostridium perfringens</i>	M, 61 ani; FC 137 bpm; TA 117/76 mmHg; T 37 °C; alterare a stării de conștiență; L 15,3 × 10 ⁹ /L; Hb 87 g/L; Cr 1290 μmol/L; lactat 1,94 mmol/L; sumar de urină: proteinurie 2+, fără nitriți; calcul ureteral 5 mm cu hidronefroza. DZ, BCR, hepatită C.	MR: „...Given altered mental status and elevated lactate*, consider early vasopressors for septic shock. Address hyperglycemia* and anemia...” CS: „...addressing hyperglycemia* and anemia. To achieve compliance... address hyperglycemia* with insulin therapy...” Halucinații: „hiperlactatemie / lactat crescut” (lactatul era 1,94 mmol/L, la limita superioară a normalului); „hiperglicemie” (fără nivel de glicemie raportat în vinieta).

Notă: MR - management recommendations (recomandări de management); CS - compliance statement (decizie de conformitate); TAM — tensiune arterială medie; Sepsis-3 - definiția din Singer 2016 [25].

de reținut că frecvența brută este influențată de dimensiunea cohorței și de sensibilitatea definiției operaționale utilizate; totuși, ordinea de mărime este consistentă cu cea din alte studii care au evaluat LLM-uri (GPT-3.5, GPT-4o, Med-PaLM) pe cazuri clinice reale [11, 12, 14].

Cazurile cu halucinație vs. cazurile fără halucinație. Analiza cazurilor cu halucinație (1, 9, 10) confirmă un pattern documentat: modelul tinde să „confabuleze” atunci când vinieta este incompletă sau când tabloul clinic parțial sugerează un scenariu-prototip (aici - șocul septic). În toate cele trei cazuri halucinatorii, LLM-ul a adăugat semne clinice caracteristice șocului (stare mentală alterată, hiperlactatemie, hiperglicemie) absente în datele reale. Această „completare de tablou” este descrisă în literatura teoretică drept confabulație inductivă și reprezintă una dintre formele cele mai insidioase de halucinație, deoarece produce afirmații plauzibile clinic, greu de sesizat fără o comparație atentă cu vinieta [10, 15].

Cazurile fără halucinație (2–8) sunt însă la fel de instructive: absența afirmațiilor false nu înseamnă absența problemelor. Aproape toate au primit scoruri de conformitate cu ghidurile sub 10/10, cu observații care indică fie divergențe de nuanță (Cazul 6 - antibiotice pentru VAP prezentate concomitent cu CAP, cu risc de suprapunere), fie limitări metodologice ale evaluării (Cazul 4 - dezacord asupra țintei de TAM la un

pacient complex). Această categorie de răspunsuri este mai puțin periculoasă, dar semnalează că LLM-urile actuale nu pot fi utilizate ca sursă unică de recomandare — ele necesită validare de rutină de către clinician.

Limitările studiului. Studiul are mai multe limitări care trebuie explicitate. În primul rând, dimensiunea cohorței este mică (n = 10), iar reprezentativitatea etiologiilor este limitată; unele forme frecvente (sepsis neonatal, sepsis obstetrical, sepsis intra-abdominal postoperator) nu sunt acoperite. În al doilea rând, evaluarea a fost realizată doar de doi medici intensiviști, ceea ce limitează aprecierea reproductibilității rezultatelor și pledează pentru studii cu mai multe cazuri și un grup mai larg de evaluatori. În al treilea rând, LLM-ul a fost interogată într-o singură rulare per caz, cu temperatură 0,25; variabilitatea intra-model între rulări nu a fost cuantificată. În al patrulea rând, datele de intrare au fost simplificate la vinieta structurată; într-un scenariu real, LLM-ul ar avea acces la note libere, imagistică și evoluție temporală, ceea ce ar putea amplifica sau, dimpotrivă, atenua tendința spre halucinație. În al cincilea rând, comparația inter-model (Palmyra-Med 70B vs. GPT-3.5-Turbo vs. GPT-4o mini), efectuată în lucrarea complementară [9], nu a fost reluată aici pentru a păstra concizia; ea rămâne însă esențială pentru poziționarea rezultatelor.

Din punct de vedere metodologic, definiția operațională a halucinației folosită aici este aliniată propunerilor recente [10, 11], dar comunitatea nu dispune încă de un standard consacrat. Dezvoltarea unui set de metrice automate (de exemplu groundedness score TruLens [23]) și armonizarea lor cu evaluarea umană sunt priorități pentru studiile viitoare.

Implicații pentru siguranța pacientului. Riscul principal al halucinației în sepsis nu este eroarea grosieră (de exemplu, antibiotic greșit), ci supraestimarea severității cu recomandare de intervenții disproporționate. În cazurile analizate, halucinațiile au însemnat atribuirea unor semne de gravitate (stare mentală alterată, hiperlactatemie, șoc septic), care într-un context clinic real ar putea: (i) declanșa administrarea prematură de vasopresoare la pacienți neșocați; (ii) justifica corticoterapie inadecvată; (iii) accelera transferul în terapie intensivă și utilizarea de resurse care ar putea fi alocate altor pacienți; (iv) genera decizii-cascadă (de exemplu, plasarea unui cateter arterial, monitorizare invazivă) cu risc iatrogen propriu. Toate aceste consecințe subliniază necesitatea unei validări umane obligatorii înainte de acțiune și pledează împotriva oricărei implementări „autonome” a LLM-urilor în terapia intensivă în stadiul actual al tehnologiei [3, 15, 16, 24].

Direcții viitoare. Rezultatele deschid mai multe direcții. Prima este dezvoltarea unor mecanisme automate de detecție a halucinației la runtime, prin: (a) verificarea de constrângeri factuale între prompt și răspuns (de exemplu, dacă răspunsul afirmă „hiperlactatemie” iar prompt-ul conține un lactat < 2 mmol/L, generarea unui avertisment); (b) folosirea LLM-ului ca judecător (LLM-as-judge) pentru a evalua consistența răspunsului cu vinieta; (c) integrarea RAG cu regăsire de fapte-cheie (chain-of-verification). Aceste mecanisme extind arhitectura multi-agent descrisă în Figura 1 și constituie obiectul unui studiu în curs.

A doua direcție este creșterea dimensiunii și eterogenității cohorței. Baze de date de tip MIMIC-IV, eICU-CRD sau AmsterdamUMCdb oferă suficiente cazuri pentru a rula sute sau mii de viniete cu evaluare programatică (groundedness, TruLens) și eșantioane pentru validare umană. Această scalare va permite estimări mai stabile ale ratelor de halucinație și analize stratificate (etiologie, severitate, comorbidități).

A treia direcție este comparația între modele, cu accent pe

LLM-urile din generația 2025–2026, inclusiv cele multimodale (imagistice + text) și cele cu ferestre lungi de context care pot ingera fișe medicale întregi. O întrebare deschisă este dacă LLM-urile mai mari (cu mai mulți parametri) halucinează mai puțin sau doar produc halucinații mai plauzibile.

A patra direcție este integrarea în fluxul clinic real, printr-un studiu prospectiv cu utilizare „umbră” (shadow deployment): LLM-ul generează recomandări în paralel cu clinicianul, fără să influențeze decizia, iar concordanța și divergențele sunt analizate post-hoc. Aceasta ar oferi date reale de performanță și ar face posibilă certificarea CE / FDA pentru instrumente de suport decizional bazate pe LLM.

În sfârșit, direcția etico-reglementară: definirea unor standarde de raportare pentru studiile de tip „LLM în medicină” (analog CONSORT / TRIPOD pentru IA) devine urgentă. Elemente indispensabile ar fi: definiția operațională a halucinației, protocolul de evaluare, versiunea și temperatura modelului, arhitectura RAG și corpus-ul de indexare.

Concluzii

1. LLM-ul specializat Palmyra-Med 70B, integrat într-un sistem multi-agent cu RAG, a generat recomandări clinice cu relevanță și fundamentare medii-mari pe o cohorță eterogenă de 10 cazuri de sepsis, dar a produs halucinații în 3 cazuri (30%), toate constând în extrapolări plauzibile ale unor semne de gravitate absente în vinieta.

2. Cazurile fără halucinație nu sunt lipsite de deficiențe: divergențele față de ghidurile Surviving Sepsis Campaign apar și în absența afirmațiilor false, semnalând necesitatea unei validări umane sistematice.

3. Utilizarea LLM-urilor în terapia intensivă rămâne condiționată de un flux hibrid clinician-mașină, cu mecanisme automate de detecție a halucinațiilor și cu supraveghere medicală activă; implementarea autonomă nu este justificată în stadiul actual.

4. Sunt necesare cohorte mai mari, un panel extins de evaluatori și standarde de raportare specifice pentru a consolida baza de evidență privind siguranța LLM-urilor în sepsis.

Recepționat – 23.06.2026, acceptat pentru publicare – 13.10.2026, publicat online – 14.10.2026

Contribuția autorilor: Victor Iapăscuță – conceptualizare, metodologie, software, validare, analiză formală, investigație, gestionarea datelor, redactarea versiunii inițiale, revizuirea critică și editarea manuscrisului; Dinu Țurcanu – revizuirea critică și editarea manuscrisului; Adrian Belți – revizuirea critică și editarea manuscrisului; Viorel Bostan – revizuirea critică și editarea manuscrisului.

Declarația de finanțare: Autorii declară că această lucrare nu a beneficiat de finanțare.

Declarația de conflict de interese: Autorii declară că nu au interese financiare sau nefinanciare relevante și nici legături cu furnizorii instrumentelor software/LLM utilizate în studiu.

Declarația de etică: Studiul a utilizat date anonimizate din MIMIC-IV-demo și informații din rapoarte de caz publicate anterior, fără recrutarea de participanți sau intervenții asupra pacienților. Potrivit declarației autorilor, aprobarea unui comitet de etică nu a fost aplicabilă acestui studiu.

Disponibilitatea datelor, codului și materialelor suplimentare: Autorii declară că datele și materialele cercetării sunt accesibile public, sub rezerva politicilor platformelor gazdă. Vinietele, evaluările și materialele suplimentare sunt prezentate în anexele 1–4. Codul este indicat în depozitele GitHub citate în bibliografie [21, 22].

Utilizarea instrumentelor de inteligență artificială: Autorul corespondent declară utilizarea Grammarly pentru corectură lingvistică în pregătirea textului în limba engleză, urmată de traducerea și adaptarea manuscrisului în limba română. Palmyra-Med 70B a fost utilizat ca obiect al cercetării și pentru generarea recomandărilor analizate.

Relația cu publicația anterioară: Autorii precizează că studiul publicat anterior în *Healthcare Informatics Research* (2026) [9] a abordat în principal

aspectele informatice ale sistemului, în timp ce prezentul manuscris dezvoltă analiza medicală a celor zece cazuri, cu prezentarea detaliată a vinițelor și recomandărilor. Cazurile și rezultatele reutilizate sunt atribuite lucrării precedente.

Citare: Iapăscurtă V, Țurcanu D, Belii A, Bostan V. Halucinațiile modelelor de limbaj mari în sepsis: sinteză narativă și analiza a zece cazuri [Large language models in sepsis with hallucinations: narrative review and analysis of ten cases]. *Arta Medica*. 2026;101(4):e2026404. doi: 10.5281/zenodo.23364665

Informație despre autori: Victor Iapăscurtă - <https://orcid.org/0000-0002-4540-7045>; Dinu Țurcanu - <https://orcid.org/0000-0001-5540-4246>; Adrian Belii - <https://orcid.org/0000-0002-4128-1318>; Viorel Bostan - <https://orcid.org/0000-0002-2422-3538>

Bibliografie

- Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. 2021;47(11):1181-1247. doi:10.1007/s00134-021-06506-y
- Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Med*. 2018;44(6):925-928. doi:10.1007/s00134-018-5085-0
- Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. doi:10.1038/s41586-023-06291-2
- Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8
- Wu M, Du X, Gu R, Wei J. Artificial intelligence for clinical decision support in sepsis. *Front Med (Lausanne)*. 2021;8:665464. doi:10.3389/fmed.2021.665464
- Writer Inc. Palmyra-Med 70B: a domain-specialized large language model for medical question answering. Technical report. Published 2024. <https://writer.com/blog/palmyra-med-fin-models/>
- Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst*. 2020;33:9459-9474.
- Iapăscurtă V, Belii A. Rolul potențial al abordării bazate pe inteligența artificială pentru managementul sepsisului: un studiu de caz pilot. *Arta Medica*. 2024;93(4):46-48. doi:10.5281/zenodo.14549531
- Iapăscurtă V, Țurcanu D, Belii A, Bostan V. Multi-agent system for early sepsis management support: a follow-up evaluation study. *Healthc Inform Res*. 2026;32(2):190-195. doi:10.4258/hir.2026.32.2.190
- Kim Y, Jeong H, Chen S, et al. Medical hallucination in foundation models and their impact on healthcare. medRxiv. Published online 2025. doi:10.1101/2025.02.28.25323115
- Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2023;55(12):1-38. doi:10.1145/3571730
- Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. *NPJ Digit Med*. 2023;6(1):195. doi:10.1038/s41746-023-00939-z
- Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text summarization. *Nat Med*. 2024;30(4):1134-1142. doi:10.1038/s41591-024-02855-5
- Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open*. 2023;6(10):e2336483. doi:10.1001/jamanetworkopen.2023.36483
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMr1814259
- Zakka C, Shad R, Chaurasia A, et al. Almanac—retrieval-augmented language models for clinical medicine. *NEJM AI*. 2024;1(2):A10a2300068. doi:10.1056/A10a2300068
- Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV Clinical Database Demo. Version 2.2. PhysioNet; 2023. doi:10.13026/dprf-ex47
- Singh P, Montano A, Bostick A. Rapid severe sepsis from *Pseudomonas fluorescens/putida* bacteremia due to skin and soft tissue infection—a case report. *Ann Med Surg (Lond)*. 2021;70:102845. doi:10.1016/j.amsu.2021.102845
- Millard MA, McManus KA, Wispelwey B. Severe sepsis due to *Clostridium perfringens* bacteremia of urinary origin: a case report and systematic review. *Case Rep Infect Dis*. 2016;2016:2981729. doi:10.1155/2016/2981729
- Iapăscurtă V. Multi-agent-approach. GitHub. Published 2025. <https://github.com/viapăscurtă-tech/Multi-agent-approach>
- Iapăscurtă V. Palmyra_med_sepsis_app. GitHub. Published 2025. https://github.com/viapăscurtă-tech/Palmyra_med_sepsis_app
- TruEra. TruLens: evaluation and tracking for LLM experiments. Published 2024. <https://www.trulens.org/>
- Chen S, Kann BH, Foote MB, et al. Use of artificial intelligence chatbots for cancer treatment information. *JAMA Oncol*. 2023;9(10):1459-1462. doi:10.1001/jamaoncol.2023.2954
- Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801-810. doi:10.1001/jama.2016.0287