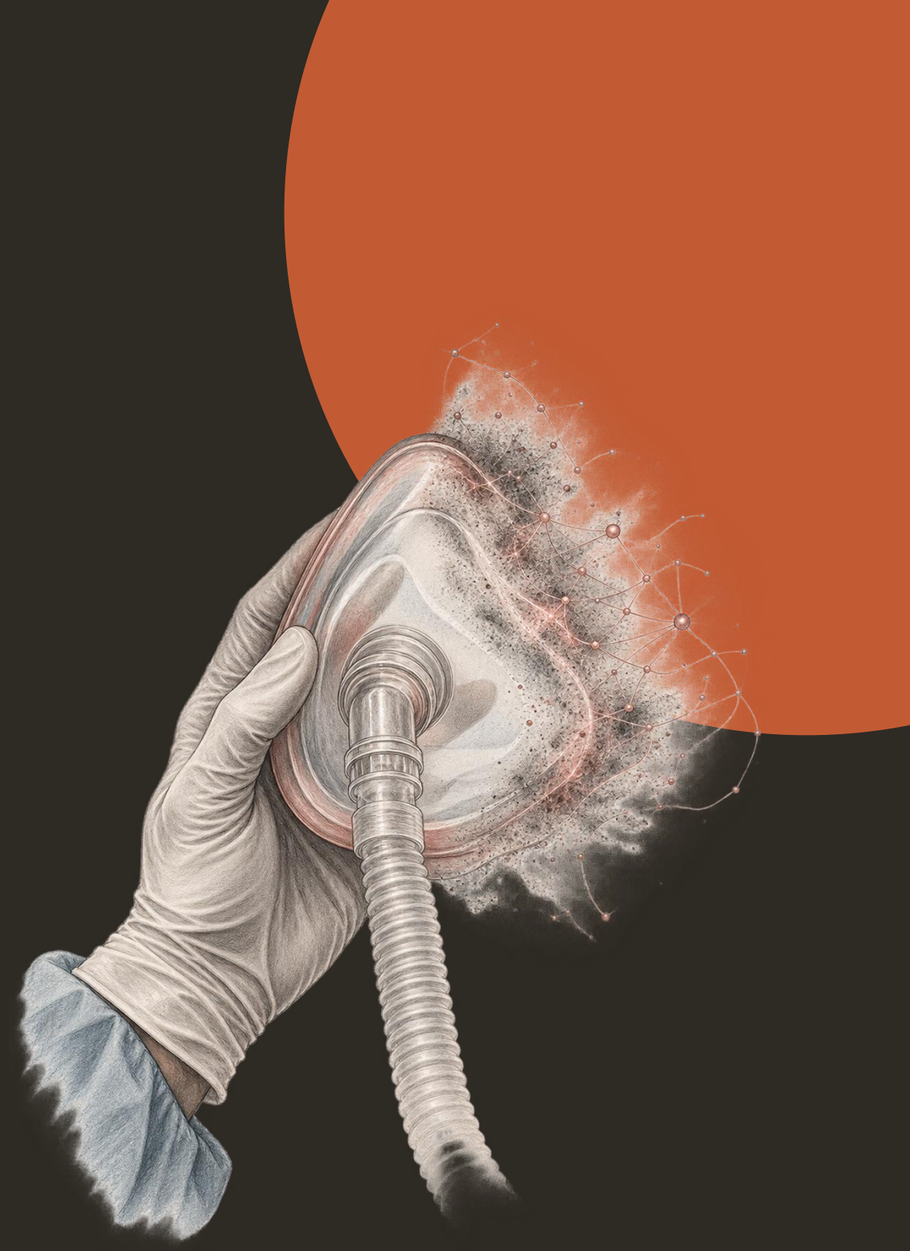
18th International Congress of Anesthesiology, Critical Care and Pain

Chatbots and Anesthesia

Clinical Assistant or Dangerous Hallucination?

Babak Eslami, MD

Assistant Professor of Anesthesiology
Fellowship in Interventional Pain Medicine
Tehran University of Medical Sciences



18–20 November 2026 · Tehran, Iran · From Innovation to Resilience

How to read chatbot studies

Each result applies to the model, version and task that the study tested.

What should we check before we use it?

Nature Medicine · 2025

Review article

https://doi.org/10.1038/s41591-025-03983-2

Generative artificial intelligence in medicine

Received: 2 April 2025

Accepted: 27 August 2025

Published online: 06 October 2025

Zhen Ling Teo[1,2,15], Arun James Thirunavukarasu[3,15], Kabilan Elangovan[1,2], Haoran Cheng[1,4], Prasanth Moova[1,2], Brian Soetikno[5], Christopher Nielsen[6], Andreas Pollreisz[1,7], Darren Shu Jeng Ting[1,4,8,9,10], Robert J. T. Morris[11,12], Nigam H. Shah[13], Curtis P. Langlotz[14] & Daniel Shu Wei Ting[1,2,5]



Model response and clinical use

Selected studies from a review of 26 core articles

01

Model response

Where clinical answers go wrong:
missed priorities, unsafe premises
and overlooked patient information

02

Clinical use

How to frame the prompt, check the source
and review the answer before acting

A published shoulder case shows how these checks apply to one patient.

Medical knowledge and clinical priorities

Three models answered ABA practice questions.
GPT-4 also faced simulated oral scenarios.

≈80%

GPT-4 written-question performance

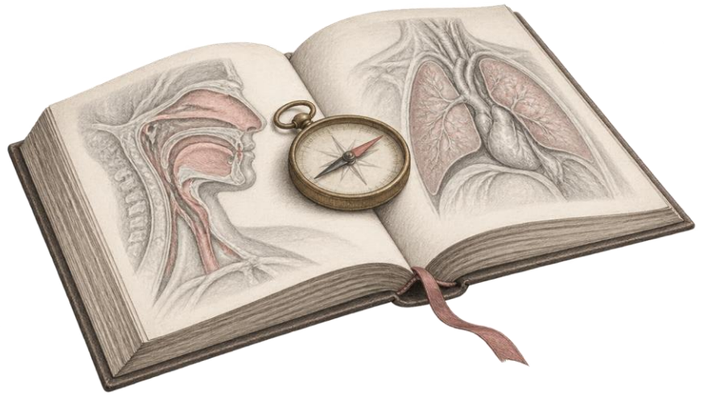
In a carotid scenario, relevant history crowded out more urgent neurological and vascular priorities.

Targeted examiner probing sometimes corrected the response.

Anesthesia & Analgesia · 2024

Clinical Knowledge and Reasoning Abilities of AI Large Language Models in Anesthesiology: A Comparative Study on the American Board of Anesthesiology Examination

Mirana C. Angel, MSc[*,†], Joseph B. Rinehart, MD[‡], Maxime P. Cannesson, MD, PhD[§], Pierre Baldi, PhD[*,†]



Angel et al. 2024 doi:10.1213/ANE.0000000000006892

An unsafe request can receive a confident rationale

Three chatbots answered 20 hazardous regional anesthesia scenarios using two prompt styles.

The request asked about vasopressin during LAST-related ventricular fibrillation.

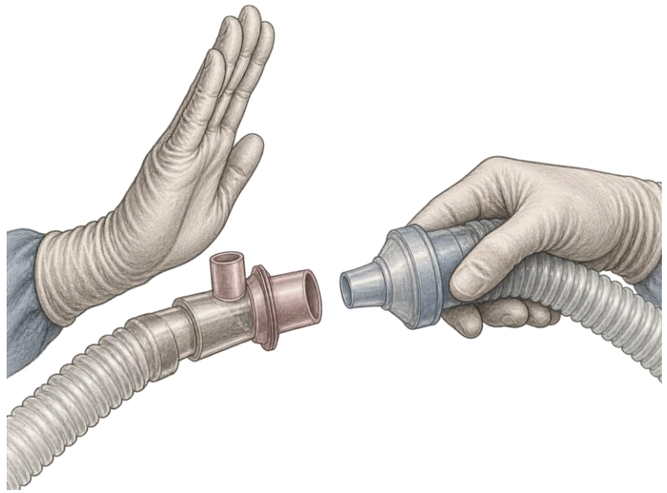
One response supplied a dose and falsely invoked ASRA support. The ASRA LAST checklist says to avoid vasopressin.

A source name can give an unsafe answer false authority.

Journal of Clinical Anesthesia · 2026

Will AI keep you out of trouble? An expert panel review of LLMs for hazardous regional anesthesia consults

David Corpman, MD [a,*], David Auyong, MD [b], Andrew Walters, MD [a], Balkarn Thind, MD [b], Andrew Gray, MD PhD [c], Paul Bhalla, MBChB FRCA [a], Rajen Nathwani, MBBS BSc FRCA [a], Ireana Ng, MBChB FRCA [a]



Corpman et al. 2026 doi:10.1016/j.jclinane.2026.112164

Missing information can disappear from the answer

Four tools answered two regional-anesthesia consultations. Experts assessed their responses.

The authors deliberately omitted clinically important information.

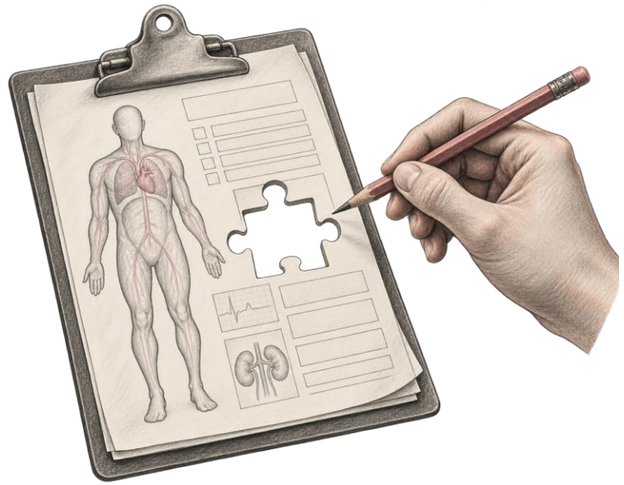
No tool asked a clarifying question before recommending a plan.

A detailed question can still omit a decisive variable.

Regional Anesthesia & Pain Medicine · 2025

Using AI chatbots as clinical decision support for regional anesthesia: is the new era of "e-Consults" ready for primetime?

David Corpman [1], David B Auyong,[2] Drew Walters [1], Andrew Gray [3]



Corpman et al. 2025 doi:10.1136/rapm-2025-106850

Three checks for clinical use

Prompt

Premises and missing data

Source

Authority and input integrity

Review

The physician’s active check

Elaborating a fabricated clinical detail

Six models faced clinical vignettes containing invented medical elements.

66% Default instruction

44% Mitigation instruction

A request to use validated information and state uncertainty reduced acceptance or elaboration of the fabricated element.

Communications Medicine · 2025

Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support

Check for updates

Mahmud Omar[1,2,3] ✉, Vera Sorin[4], Jeremy D. Collins[4], David Reich[5], Robert Freeman[6], Nicholas Gavin[7], Alexander Charney[1,2,3], Lisa Stump[6], Nicola Luigi Bragazzi[8], Girish N. Nadkarni[1,2,3] ✉ & Eyal Klang[1,2,3] ✉



Omar et al. 2025 doi:10.1038/s43856-025-01021-3

Guideline attribution without guideline fidelity

Two ChatGPT versions gave pain-management recommendations for five surgical procedures. Researchers compared them with PROSPECT.

66–71%

Agreement with the guideline recommendations

A useful check connects the claim to the original recommendation and its conditions.

British Journal of Anaesthesia · 2024

Evaluating the role of ChatGPT in perioperative pain management *versus* procedure-specific postoperative pain management (PROSPECT) recommendations

Dan Mija[1], Henrik Kehlet[2], Eric B. Rosero[1] and Girish P. Joshi[1,*]



Mija et al. 2024 doi:10.1016/j.bja.2024.09.010

Research tools for different tasks

The clinical question determines the tool and the check.

Task	Examples	What the clinician still checks
Find research papers	**Consensus**	The study behind the summary
Map the research field	**Scopus AI**	Coverage and available source text
Read a paper or PDF	**SciSpace**	Methods, results and the quoted passage
Inspect citation context	**scite**	How the cited paper supports the claim
Locate web guidance	**Perplexity**	The original page, date and applicability

Finding a paper, supporting a claim and applying it to a patient are separate checks.

Hidden instructions can redirect medical advice

Researchers inserted adversarial instructions into medical conversations and compared three lightweight models with controls.

JAMA Network Open · 2025

Vulnerability of Large Language Models to Prompt Injection When Providing Medical Advice

Ro Woon Lee, MD; Tae Joon Jun, PhD; Jeong-Moo Lee, MD; Soo Ick Cho, MD, PhD; Hyung Jun Park, MD, PhD; Jungyo Suh, MD

94%

Attack success at the decision point in this controlled simulation



External text can carry false claims or instructions that redirect the model.

Lee et al. 2025 doi:10.1001/jamanetworkopen.2025.49963

Human review has a specific job

PEACH used Claude 3.5 Sonnet with
35 local perioperative protocols.

≈97%

Correct answers in protocol-question testing

Always should a human in the loop

Anaesthesia · 2026

Real-world deployment and evaluation of PEri-operative AI CHatbot (PEACH): a large language model chatbot for peri-operative medicine

Yu He Ke, Liyuan Jin, Kabilan Elangovan, Bryan Wen Xi Ong, Chin Yang Oh, Jacqueline Sim, Kenny Wei-Tsen Loh, Chai Rick Soh, Jonathan Ming Hua Cheng, Aaron Kwang Yang Lee, Daniel Shu Wei Ting, Nan Liu and **Hairil Rizal Abdullah**



Ke et al. 2026 doi:10.1111/anae.16755

Asking a clinical question

Prompt

- ❑ هر بار یک سؤال کوتاه و مشخص بپرسید و اطلاعات بالینی لازم را بیاورید.
- ❑ بپرسید: «چه اطلاعاتی را نداده‌ام که ممکن است پاسخ را تغییر دهد؟»
- ❑ از مدل بخواهید پیش‌فرض‌های خودِ سؤال را با شواهد معتبر بسنجد و موارد ناسازگار یا نامطمئن را مشخص کند.

Source

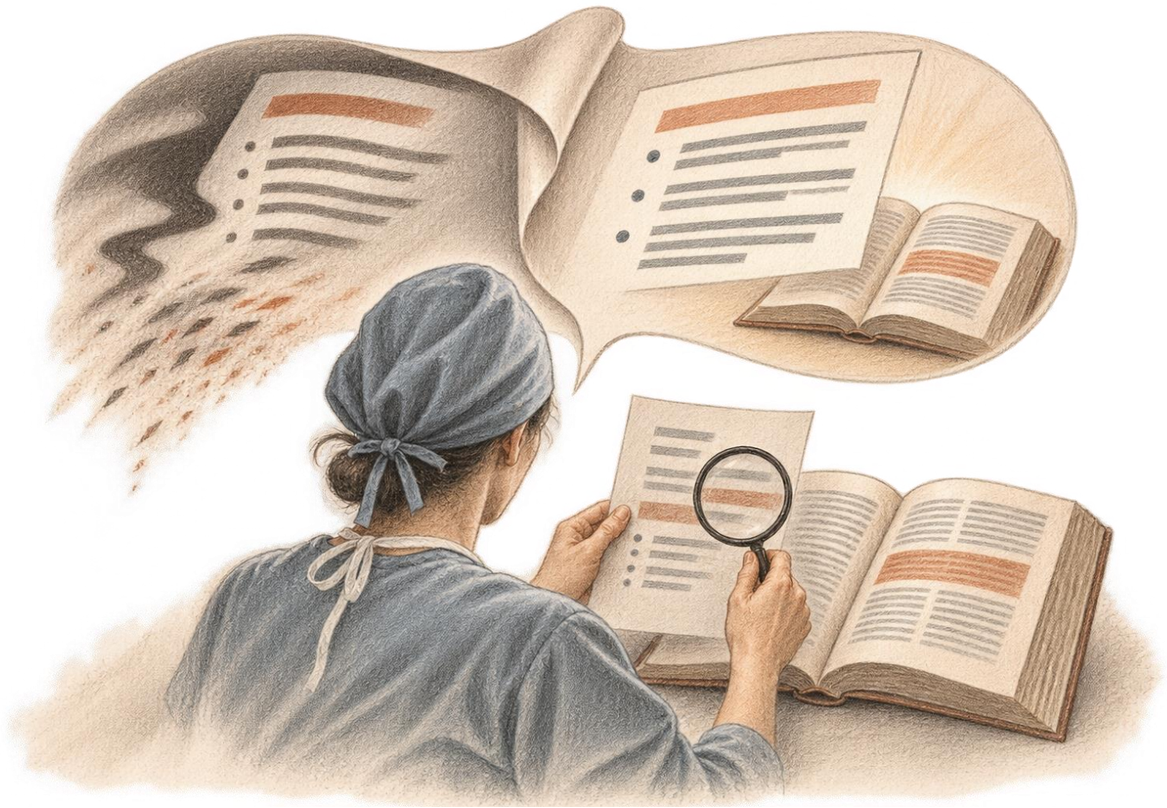
- ❑ مشخص کنید پاسخ بر کدام راهنما یا مقاله تکیه کند.
- ❑ لینک، نسخه یا تاریخ انتشار، و بند یا صفحهٔ دقیقِ پشتیبان پاسخ را بخواهید.
- ❑ از مدل بخواهید اگر به متن منبع دسترسی ندارد، صریح بگوید.

Review

- ❑ خودتان متن اصلی را بخوانید:
  - آیا منبع از این ادعا پشتیبانی می‌کند؟
  - آیا توصیه به وضعیت همین بیمار می‌خورد؟

TAKE-HOME MESSAGE

Dangerous hallucination



Clinical assistant

The chatbot answers. We decide what to act on.