The International Journal of Healing and Caring

September 2026 **Volume 26 Number 3**
ISSN:1538-1080 **https://doi.10.78717/ijhc.2026.3.45**

OPINION

Artificial Intelligence in Pain Management and Integrative Care: Helpful Tool or Patient-Safety Risk?

Markus Ploesser MD, LLM[1,2,3], Jose Daniel Lomeli Diaz[4], Lic Psych

[1] Department of Psychiatry & Neuroscience, UC Riverside School of Medicine, 900 University Ave. Riverside, CA 92521, USA
[2] University of British Columbia, Division of Forensic Psychiatry, Department of Psychiatry, Faculty of Medicine, Detwiller Pavilion. 2255 Wesbrook Mall. Vancouver, BC V6T 2A1, Canada
[3] NYU Langone Health, Center for Study & Treatment of Pain, 240 East 38th Street, 14th Floor, New York, NY 10016
[4] Mandl School, The College of Allied Health, 254 W 54 St, New York, NY 10019

Abstract

Artificial intelligence (AI) is increasingly used by patients to obtain health information, interpret symptoms and explore treatment options, raising questions about its safety in pain management and integrative care. This commentary examines recent evidence on general-purpose large language models and purpose-built AI applications used in pain-related contexts. Current studies suggest that contemporary AI systems can provide highly accurate and useful information for relatively bounded tasks, including pain pharmacology, patient education and some aspects of rehabilitation. Newer models generally outperform earlier systems, highlighting the rapidly changing nature of the evidence base. However, high factual accuracy does not necessarily translate into safe individualised treatment advice. Important limitations include incomplete responses, inconsistent adherence to clinical guidelines, difficulty reasoning from incomplete clinical information and failure to account for patient-specific contraindications or other contextual factors. Purpose-built AI rehabilitation and chronic-pain self-management tools show promising results, particularly when operating within clearly defined therapeutic pathways. Evidence from integrative medicine is similarly mixed: AI has demonstrated substantial specialist knowledge in areas such as acupuncture and traditional Chinese medicine, while recommendations concerning dietary supplements appear less consistently reliable. The safety of AI in pain care therefore

depends less on whether AI is inherently “safe” or “dangerous” than on the task it is asked to perform. AI appears most valuable as a tool for education, evidence navigation and supported self-management, whereas diagnosis, medication changes and complex individual treatment decisions continue to require clinical judgement, appropriate safeguards and human oversight.

Keywords: artificial intelligence; chronic pain; pain management; integrative medicine; large language models

Introduction: When Information Becomes Treatment Advice

For many patients, consulting an artificial-intelligence (AI) chatbot about symptoms or treatment now requires considerably less effort than arranging a clinical appointment. A nationally representative West Health–Gallup survey of 5,660 US adults conducted in late 2025 found that one in four respondents had used AI tools or chatbots for health information or advice. Among recent AI health users, 14% reported that they had not seen a healthcare provider they otherwise would have consulted because of information or advice received from AI (West Health Gallup Center on Healthcare in America, 2026). These findings illustrate that AI-generated health information is no longer merely an academic issue: it may influence real healthcare decisions.

Recent evaluations have shown that patient-facing AI systems can provide inappropriate or unsafe recommendations when responding to medical questions, particularly when essential clinical context is incomplete or the system fails to recognise the need for professional assessment (Draelos et al., 2025). At the same time, media coverage has also documented patients who used AI successfully to identify possibilities that had previously been overlooked by healthcare professionals (Flinn, 2025). The emerging picture is therefore less straightforward than the early narrative that large language models (LLMs) simply “hallucinate” medical information.

This distinction matters in pain care. A person living with persistent pain may ask an AI system to explain the cause of neuropathic pain, compare acetaminophen with a non-steroidal anti-inflammatory drug (NSAID), interpret an adverse effect, suggest exercises, evaluate acupuncture or recommend a dietary supplement. These are not equivalent tasks. Explaining established medical information is fundamentally different from deciding whether a particular patient should take a medication, change its dose or delay seeking urgent assessment.

The central issue is therefore not simply the factual accuracy of AI-generated medical information, but the point at which general information begins to function as individualised treatment advice and may introduce safety risks.

How Well Does AI Answer Pain-Related Questions?

Recent pain-specific evidence is considerably more reassuring than some early discussions of medical AI might suggest.

Stroop and colleagues evaluated ChatGPT-4 on clinical pharmacological questions concerning pain therapy, including drug interactions, doses and contraindications. Healthcare professionals considered 99% of the responses

comprehensible, 96% medically accurate and 93% satisfactory overall (Stroop et al., 2025). These results are important because they directly challenge the assumption that a general-purpose chatbot necessarily provides poor-quality information about analgesic treatment. However, only 63% of the responses in the same study were considered complete, with gaps particularly concerning pharmacokinetics and drug interactions (Stroop et al., 2025). An answer can therefore be medically correct without being sufficiently complete for an individual treatment decision.

Other pain studies show a similarly mixed but increasingly encouraging picture. A 2025 study of commonly asked questions about endodontic pain found that ChatGPT-3.5 produced high-quality and reliable answers and outperformed Gemini on those measures, although ChatGPT's responses required a considerably higher reading level (Aljamani et al., 2025). The fact that this relatively strong performance was obtained with the now-outdated GPT-3.5 model also illustrates the difficulty of extrapolating older research to newer systems.

The most comprehensive recent synthesis in one pain field is a 2026 scoping review of generative AI in low-back-pain management. The review identified 31 studies, including investigations of diagnosis and treatment support, patient education, imaging interpretation, clinical documentation and research applications (Tu et al., 2026). AI-generated advice showed greater agreement with reference recommendations when tasks were relatively simple, prompts supplied more detailed information and newer-generation models were used (Tu et al., 2026).

This last observation deserves particular emphasis. Evaluations of GPT-3.5 or early GPT-4 systems remain historically informative, but they cannot automatically be assumed to describe the performance of systems available several years later. In a large 2026 evaluation of 21 frontier LLMs, newer versions generally outperformed earlier models within the same model families, and reasoning-optimised models performed significantly better than non-reasoning models (Rao et al., 2026). Any commentary on AI safety therefore has an unusually short evidence half-life: the technology may change significantly during the interval between a study being performed and its publication.

Factual Accuracy Does Not Equal Clinical Judgement

The positive findings should not, however, be interpreted as evidence that a chatbot can independently manage pain.

A useful example comes from lumbosacral radicular pain. In a 2025 study comparing ChatGPT-4o, Gemini, Copilot, Claude and Perplexity with clinical-practice-guideline recommendations, none of the tested systems fully reproduced guideline-concordant recommendations, and substantial variation occurred between models. The authors therefore cautioned against unsupervised reliance on these systems for healthcare advice (Rossettini et al., 2025).

The 2026 low-back-pain review similarly identified inconsistency with guidelines or clinicians, hallucinations, poorer accuracy on complex questions, sensitivity to prompt construction and problems with interpretability. Importantly, none of the studies included in that review had yet demonstrated clinical effectiveness of generative AI in real-world low-back-pain care (Tu et al., 2026).

A broader 2026 study helps explain this discrepancy. Rao and colleagues tested 21 contemporary LLMs, including GPT-5 and other frontier systems, across the sequential stages of clinical reasoning. Overall accuracy appeared impressive, ranging approximately from 81% to 90%, yet performance was substantially weaker when the systems had to construct a differential diagnosis from incomplete information (Rao et al., 2026). Failure

rates exceeded 80% for differential-diagnosis tasks in the study, whereas final-diagnosis questions were answered much more successfully (Rao et al., 2026).

This distinction closely resembles what happens when patients consult AI. Once a diagnosis is established, providing information about commonly used treatments for lumbar radiculopathy is a relatively bounded task. By contrast, responding to an undifferentiated symptom presentation requires the system to identify missing information, consider alternative diagnoses, recognise potentially serious causes and determine whether urgent assessment is needed.

Human clinicians do not normally make pain-treatment decisions based on the symptom of pain alone. Age, onset, location, neurological findings, trauma, comorbidities, pregnancy, kidney and liver function, current medication, previous treatments and numerous other variables may alter the clinical interpretation and the safety of a proposed intervention. AI only knows those factors if they are supplied, inferred correctly or obtained through appropriate questioning.

The safety issue may therefore arise even when every individual sentence generated by the AI is technically correct. A recommendation that NSAIDs can relieve musculoskeletal pain is correct in general, but recommending them to an individual without recognising a relevant contraindication is a different matter.

Purpose-Built AI Systems in Pain Management

It is equally important not to treat all "AI pain management" as synonymous with public chatbots. Purpose-built AI systems can operate within a defined therapeutic pathway, restrict the information on which recommendations are based, collect structured patient information and incorporate predetermined safety rules. This may make them fundamentally different from open-ended consumer chatbots.

Evidence for AI-assisted musculoskeletal rehabilitation is already emerging. A 2025 systematic review and network meta-analysis included 33 randomised controlled trials evaluating 13 categories of AI-assisted rehabilitation, including AI feedback systems, exergaming, telerehabilitation and robotic approaches (Luo et al., 2025). The analysis found significant improvements in pain, functional outcomes and range of motion, although the relative effects differed considerably by intervention and longer-term evidence remains necessary (Luo et al., 2025).

In another study, physiotherapists assessed exercises recommended by an AI system for people with musculoskeletal disorders and showed 90.5% agreement that the exercises posed no risk; only 0.08% were considered potentially capable of increasing pain (Griefahn et al., 2024). Such findings suggest that appropriately designed AI systems can individualise some aspects of rehabilitation without necessarily introducing substantial additional risk.

AI-supported services are also moving beyond research settings. An NHS pilot in England allowed more than 2,500 patients with back pain and other musculoskeletal problems to access an AI-powered physiotherapy service during a 12-week period in 2025 (Thomas, 2025). The service was reported to reduce the back-pain waiting list by 55% and save 856 hours of clinician time per month, although professional organisations also warned that digital care must not disadvantage people unable or unwilling to access treatment in this way (Thomas, 2025).

AI-powered tools are also being developed specifically for chronic-pain self-management. Paindrainer, for example, uses patient-entered information to provide individualised coaching based on approaches including pacing, physical activity and sleep management (Olofsson et al., 2026). A previous prospective

multicentre single-arm study cited by the investigators found clinically important improvement in physical function for almost three quarters of 41 participants, although the uncontrolled design means efficacy cannot be established with the same confidence as in a randomised trial (Barreveld et al., 2023; Olofsson et al., 2026).

These examples suggest that the safest application of AI in pain management may not necessarily be the most conversational or apparently intelligent system. A narrower system with clearly defined limits, validated treatment components and appropriate escalation pathways may ultimately be safer than a general-purpose system capable of discussing almost anything.

AI in Integrative Medicine: Opportunities and Challenges

Pain care frequently extends beyond medication. Exercise, pacing, psychological interventions, sleep management, mindfulness and other self-management strategies can all form part of multimodal chronic-pain care, while many patients also explore acupuncture, herbal products or dietary supplements. Digital and AI-supported tools designed to support these broader aspects of chronic-pain self-management have already been tested, with generally promising but still mixed and methodologically limited results (Guebey et al., 2026).

AI therefore offers an intriguing possibility for integrative care: one conversational system can potentially discuss conventional pharmacology, behavioural interventions, rehabilitation and complementary approaches within the same interaction.

Recent evidence concerning acupuncture is surprisingly positive. Liu and colleagues compared seven publicly available LLMs with three professional acupuncturists using a complex real-world case and had 28 acupuncture experts from China, South Korea and the United States assess their recommendations (Y. Liu et al., 2025). The LLMs performed comparably with professional acupuncturists on Western diagnosis, while performance varied across traditional Chinese medicine diagnosis, acupoint selection, needling technique and herbal medicine recommendations. GPT-4o and several other models showed particularly strong alignment with experts for some TCM tasks (Y. Liu et al., 2025).

This does not demonstrate that AI can independently practise acupuncture or traditional medicine. The evaluation concerned a single structured case and expert ratings of generated recommendations rather than treatment outcomes in actual patients (Y. Liu et al., 2025). Nevertheless, it demonstrates that contemporary general-purpose models can reproduce a substantial amount of specialist integrative-medicine knowledge.

Dietary supplements expose a more difficult problem. A 2026 study compared GPT-4 and GPT-4o answers concerning 30 dietary supplements and six diseases with an evidence-based reference framework (M. Liu et al., 2026). Overall accuracy was approximately 57%; responses were generally cautious, but ambiguity and occasional irrelevant or incorrect information remained, leading the authors to conclude that there was a moderate risk of misinformation (M. Liu et al., 2026).

This discrepancy may reflect a fundamental challenge for AI in integrative medicine. The literature includes interventions supported by strong evidence, interventions supported by mixed or condition-specific evidence, treatments with mainly traditional or observational support, mechanistic findings that have not translated into clinical benefit and commercially promoted treatments for which meaningful evidence is sparse. An AI system must distinguish the strength and clinical relevance of the evidence, rather than retrieving statements only.

The problem is therefore not necessarily that an AI will invent a fictitious herb or treatment. A subtler and potentially more common problem is that it may describe preliminary evidence accurately but communicate it in a way that makes efficacy appear better established than it is.

Balancing Benefits and Risks in AI-Assisted Pain Care

It would be difficult to defend the statement that contemporary AI systems routinely provide false information about pain treatment. GPT-4 answers to pain-pharmacology questions have been rated 96% medically accurate, and recent studies show useful performance in pain education, rehabilitation and even specialist areas such as acupuncture (Y. Liu et al., 2025; Luo et al., 2025; Stroop et al., 2025).

Equally, it would be difficult to conclude that high factual accuracy makes autonomous AI treatment advice safe. The same pain-pharmacology study that found 96% accuracy found only 63% completeness (Stroop et al., 2025). Chatbots have failed to match clinical guidelines consistently for radicular pain, and the newest review of generative AI in low-back-pain management continues to identify hallucination, inconsistency and difficulties with complex questions (Rossettini et al., 2025; Tu et al., 2026). The distinction between information accuracy and decision safety is therefore central.

A chatbot that explains the usual benefits and adverse effects of an analgesic is functioning as an information resource. A chatbot that tells a particular patient to start, stop, combine or change the dose of medications has moved into individualised clinical decision-making. Likewise, explaining evidence concerning acupuncture is different from determining that acupuncture is appropriate for an unexplained new pain syndrome before serious pathology has been excluded.

The risk probably exists on a continuum. It is relatively low when AI explains established information in a bounded context and substantially higher when the system must infer a diagnosis, recognise missing information, evaluate contraindications or make decisions under clinical uncertainty.

Toward Safer AI-Assisted Pain Management

The appropriate response may therefore be neither rejection nor uncritical adoption. First, patients should be encouraged to distinguish general information from individual treatment advice. AI may be extremely useful for explaining terminology, comparing broad treatment approaches, preparing questions for clinicians and summarising evidence, but these functions should not automatically confer authority to prescribe or alter treatment.

Second, medication questions deserve a higher threshold for verification. The consequences of incomplete information concerning doses, interactions or contraindications are potentially greater than those of an incomplete explanation of pain physiology. The finding that GPT-4 responses were frequently accurate but less consistently complete illustrates why apparently excellent performance may still be insufficient for unsupervised medication management (Stroop et al., 2025).

Third, AI recommendations concerning supplements, herbal products and other integrative interventions should make the strength of evidence explicit. Recent evidence showing only moderate accuracy for dietary-supplement questions indicates that this remains an area requiring particular caution (M. Liu et al., 2026).

Fourth, purpose-built AI treatment systems should be judged separately from general chatbots. Clinical validation, clearly defined scope, evidence-based treatment content, appropriate safety

escalation and human oversight may matter more than the apparent sophistication of the conversational interface.

Finally, evaluation needs to be continuous. The 2026 comparison of frontier models demonstrated measurable improvements across successive generations, while the low-back-pain review similarly observed better agreement with newer systems (Rao et al., 2026; Tu et al., 2026). Research performed on yesterday's chatbot should therefore inform caution, but should not automatically be used to characterise tomorrow's technology.

Conclusion

AI is already changing pain care, although probably not in the way suggested by either its strongest advocates or its strongest critics. Current systems can provide accurate, accessible and useful information and are increasingly being incorporated into rehabilitation and chronic-pain self-management. In some bounded tasks, their performance is impressive.

The remaining danger lies less in the proposition that AI is usually wrong than in the fact that an occasional error, omission or unjustified inference may occur at precisely the point where a patient assumes that a fluent answer represents an individual clinical judgement.

Pain management and integrative medicine provide an important testing ground because care often requires long-term navigation among medications, exercise, behavioural approaches, complementary therapies and self-management. AI is uniquely suited to helping patients navigate this complexity. At the same time, this complexity makes incomplete context and uncertain evidence particularly consequential.

The most useful future role for AI may therefore be as an increasingly knowledgeable partner in information and self-management rather than an autonomous prescriber or diagnostician. The central question is not whether AI should participate in pain care, it already does, but how clearly we define the boundaries within which its considerable capabilities can be used safely.

References

Aljamani, S., Hassona, Y., Fansa, H. A., M. Saadeh, H., & Dafi Jamani, K. (2025). Evaluating Large Language Models in Addressing Patient Questions on Endodontic Pain: A Comparative Analysis of Accessible Chatbots. *Journal of Endodontics*, *51*(11), 1617–1624. https://doi.org/10.1016/j.joen.2025.04.015

Barreveld, A. M., Rosén Klement, M. L., Cheung, S., Axelsson, U., Basem, J. I., Reddy, A. S., Borrebaeck, C. A., & Mehta, N. (2023). An artificial intelligence-powered, patient-centric digital tool for self-management of chronic pain: A prospective, multicenter clinical trial. *Pain Medicine*, *24*(9), 1100–1110.

Draelos, R. L., Afreen, S., Blasko, B., Brazile, T. L., Chase, N., Desai, D. P., Evert, J., Gardner, H. L., Herrmann, L., House, A. V., Kass, S., Kavan, M., Khemani, K., Koire, A., McDonald, L. M., Rabeeah, Z., & Shah, A. (2025). *Large language models provide unsafe answers to patient-posed medical questions* (arXiv:2507.18905). arXiv. https://doi.org/10.48550/arXiv.2507.18905

Flinn, R. (2025, July 10). Dr. ChatGPT Will See You Now. *Wired*. https://www.wired.com/story/dr-chatgpt-will-see-you-now-artificial-intelligence-llms-openai-health-diagnoses/

Griefahn, A., Zalpour, C., & Luedtke, K. (2024). Identifying the risk of exercises, recommended by an artificial intelligence for patients with musculoskeletal disorders. *Scientific Reports*, *14*(1), 14472. https://doi.org/10.1038/s41598-024-65016-1

Guebey, J., Gosetto, L., Rehberg, B., Lovis, C., Ehrler, F., & Molinard-Chenu, A. (2026). A Chronic Pain Self-Management Mobile App (Dolodoc): Cross-Sectional Acceptability Study. *JMIR Human Factors*, *13*(1), e77163. https://doi.org/10.2196/77163

Liu, M., Okuhara, T., Shirabe, R., Nishiie, Y., Chang, X., Okada, H., & Kiuchi, T. (2026). Validity and reliability of ChatGPT's responses on dietary supplements in Japan: A quality assessment and content analysis. *PEC Innovation*, *8*, 100461. https://doi.org/10.1016/j.pecinn.2026.100461

Liu, Y., Yuan, Y., Yan, K., Li, Y., Sacca, V., Hodges, S., Cannistra, M., Jeong, P., Wu, J., & Kong, J. (2025). Evaluating the role of large language models in traditional Chinese medicine diagnosis and treatment recommendations. *Npj Digital Medicine*, *8*(1), 466. https://doi.org/10.1038/s41746-025-01845-2

Luo, Z., Wang, Y., Zhang, T., & Wang, J. (2025). Effectiveness of AI-assisted rehabilitation for musculoskeletal disorders: A network meta-analysis of pain, range of motion, and functional outcomes. *Frontiers in Bioengineering and Biotechnology*, *13*. https://doi.org/10.3389/fbioe.2025.1660524

Olofsson, S., Rosén Klement, M. L., Barreveld, A. M., Mehta, N., Persson, U., & Borrebaeck, C. A. (2026). Cost savings and quality-of-life gains associated with the use of an AI-powered digital tool for self-management of chronic pain based on a multicenter cohort study. *Pain Medicine*, pnag061.

Rao, A. S., Esmail, K. P., Lee, R. S., Jiang, S., Arraiza Carlo, B., Gill, J., Khanna, P., Kalmowitz, E., Montagnese, B., Heydari, K., Jiao, Q., Bott, E., Nguyen, D., Wang, G., Hood, M., Landman, A. B., & Succi, M. D. (2026). Large Language Model Performance and Clinical Reasoning Tasks. *JAMA Network Open*, *9*(4), e264003. https://doi.org/10.1001/jamanetworkopen.2026.4003

Rossettini, G., Bargeri, S., Cook, C., Guida, S., Palese, A., Rodeghiero, L., Pillastrini, P., Turolla, A., Castellini, G., & Gianola, S. (2025). Accuracy of ChatGPT-3.5, ChatGPT-4o, Copilot, Gemini, Claude, and Perplexity in advising on lumbosacral radicular pain against clinical practice guidelines: Cross-sectional study. *Frontiers in Digital Health*, *7*.

https://doi.org/10.3389/fdgth.2025.1574287

Stroop, A., Stroop, T., Zawy Alsofy, S., Wegner, M., Nakamura, M., & Stroop, R. (2025). Assessing GPT-4's accuracy in answering clinical pharmacological questions on pain therapy. *British Journal of Clinical Pharmacology*, *91*(8), 2294–2303. https://doi.org/10.1002/bcp.70036

Thomas, T. (2025, July 31). First NHS AI-run physio clinic in England halves back-pain waiting list. *The Guardian*. https://www.theguardian.com/society/2025/jul/31/nhs-first-ai-run-physio-clinic-in-england-halves-back-pain-waiting-list

Tu, R., French, S., Searant, I., Chung, J. W., Hancock, M., Magrabi, F., & Downie, A. (2026). Use of Generative Artificial Intelligence in the Management of Low Back Pain: A Scoping Review. *Journal of Medical Systems*, *50*(1), 75. https://doi.org/10.1007/s10916-026-02406-0

West Health Gallup Center on Healthcare in America. (2026). *Millions of Americans Now Consult AI Before, After, and Sometimes Instead of, Seeing a Doctor*. West Health. https://westhealth.org/news/millions-of-americans-now-consult-ai-before-after-and-sometimes-instead-of-seeing-a-doctor/

Bio:

Dr. Markus Ploesser, MD, JD as the chair of the journal's board of directors. Dr. Ploesser is trained in Psychiatry, Forensic Psychiatry, Addiction, Pain Management and Integrative Medicine. In addition, he is a member of the New York State, DC and Berlin Bar associations. Speaking 4 languages, he has published many research articles and continues his leadership in the integrative medicine community.

Email: markus.ploesser@nyulangone.org

The International Journal of Healing and Caring
31907 E Davis Ranch Rd.
Marana, AZ 85658
Email: ijhcjournal@earthlink.net Website: http://www.ijhc.org