Rise of the Medbots

Transforming Healthcare One Guess at a Time

Martin Martinez
Chancy.AI

Two out of three doctors in the US now use a specialized AI chatbot called OpenEvidence. Med students are using it to build diagnoses, draft notes, and look up treatment protocols. The tool is fast, conversational, and confident. It is also less reliable at answering medical queries than general-purpose chatbots like ChatGPT and Claude, according to a study published in Nature Medicine in August 2026. The medbot built for doctors performs on par with Google's AI Overviews, a system that is wrong 9% of the time.

A Stanford medical student, Simar Bajaj, and Joseph Sakran, a Johns Hopkins trauma surgeon, wrote about this systemic degradation of medical diagnosis: "The danger is not just deskilling but never-skilling. Although a doctor who has forgotten how to reason is recoverable, one who never learned how may not be."

Med students who once struggled to build a list of potential diagnoses now receive a perfect textbook answer, complete with possibilities they might never have considered. The struggle is gone. The learning-curve is gone. The neural pathway development that hard-wires competence is gone.

Associate dean of educational technology and innovation at Yale School of Medicine, Jaideep Talwalkar, framed it in terms of repetition. "The process of deliberately crafting the note forces us to use our brains to really wrestle with what's happening. There's an importance in doing that with great repetition." Repetition builds neural pathways. Easy answers delivered by AI eliminate cognitive development.

The Stanford/Hopkins team proposed a solution borrowed from aviation: require medical trainees to periodically work through cases without AI assistance, the same way the Federal Aviation Administration requires pilots to periodically disengage autopilot and assume control. Nobody wants a pilot who has never flown a plane. Nobody should want a doctor who has never reasoned through a diagnosis solo, especially considering that AI medical diagnostics are only correct about nine times out of ten.

A Johns Hopkins study found that doctors who use AI are already viewed with scorn and skepticism by their peers, even though so many already do. A survey conducted by the American Association of Medical Colleges found that 75 percent of clinicians worry about the loss of previously learned abilities among experienced doctors. Moreover, every major

study conducted in the past two years proves the diagnostic accuracy gap is far worse than it appears.

How Wrong is Wrong?

A comprehensive meta-analysis published in the Journal of Biomedical Informatics in March 2024 found that ChatGPT had an overall accuracy rate of just 56 percent on medical queries. That means the information provided was incomplete, incorrect, or misleading almost half the time.

In April 2026, a JAMA Network Open study tested large language models (LLMs) on realistic patient symptoms. When symptoms could be attributed to more than one condition, the LLMs' failure rates exceeded 80 percent. Even on straightforward cases that included physical exams and laboratory results, LLMs still failed to generate a correct diagnosis 40 percent of the time. That is a failure of critical thinking: AI models "collapse prematurely onto single answers," locking onto a diagnosis without weighing multiple possibilities as a trained physician does. Thinking machines do not deliberate. They do not reconsider their analysis. They make snap decisions and move on to the next topic without a second thought.

The gap between benchmark lab performance and real-world clinical performance is easily explained: less than five percent of published papers evaluate LLMs on actual patient data. Medbots are trained and tested on standardized medical licensing exam questions with simplified symptom sets. They perform well on those tests. Then they are deployed into clinical environments where patient data is often fragmented, contradictory, and/or incomplete, and the failure rate rises sharply. AI systems simply do not have the ability to comprehend real-world experience of the human condition.

A BMJ Open study tested five chatbots on 250 health questions. Nearly half of the answers were problematic. One out of five were highly problematic, meaning they could lead to direct harm. Only 40 percent of the citations were accurate, and not a single chatbot produced a fully accurate response. Grok AI performed worst at 58 percent problematic responses. But none of them performed well. The best LLMs were wrong often enough to be dangerous.

Mount Sinai researchers tested multiple LLMs on medical queries and found an average hallucination rate of 66 percent. GPT-4o, one of the most advanced models available at the time of testing, hallucinated in 53 percent of cases. Some models performed dramatically worse. DeepSeek hallucinated in over 80 percent of medical queries. Two out of three medical answers contained fabricated or incorrect information.

The patient safety organization ECRI, which has tracked health technology hazards for decades, called AI chatbot misuse the single greatest health technology hazard of 2026.

Hospitals Deliver Medbots Prematurely

According to Nature Medicine in April 2026, "evidence that AI tools create value for patients, providers or health systems remains scarce." Despite that absence of evidence, claims about AI's clinical impact are "increasingly more common" in product marketing, leading to "premature implementation and adoption." Nature Medicine called for a framework to evaluate AI systems because they are currently deployed with no third-party testing.

Hospitals are not waiting for approval. Hartford HealthCare launched Patient GPT to answer patient questions. Sutter Health and Reid Health launched Emmie, an AI assistant built into patient portal MyChart. OpenAI also launched its ownChatGPT Health bot. A Gallup poll found that 25 percent of Americans have already used an AI tool for health information. In rural communities, patients send AI nearly 600,000 healthcare messages per week.

Stanford professor Nigam Shah called hospital chatbots a "net positive," because they bridge gaps when clinics are closed and emergency rooms are not appropriate. Dr. Ajay Kumar acknowledged it "would be concerning" if strong safety measures were not in place. On the downside of the argument, MIT's Dr. Regina Barzilay told Newsweek that beyond note-taking tools, there are "very little success stories" with clinical AI. Harvard Medical School's Jamie Rortson offered the most measured position: AI can speed up tedious processes, "but it's critical for people who are interacting with AI as part of clinical studies to be knowledgeable about the right and wrong applications, and in the correct context."

Meanwhile, privacy-rights issues are mounting. Some doctors use unauthorized chatbots, uploading patient data into AI tools that are not HIPAA-compliant. Patients whose medical data enters unauthorized systems have no knowledge that their information may be processed by systems that have not been approved, vetted, or secured. Nature published an investigation in June 2026 confirming that patient data processed by these tools is not merely unprotected but actively vulnerable.

Medication Misinformation

When people ask AI about medications, the stakes are as high as they get. A wrong answer about a drug interaction is a potential for poisoning.

A British Journal of Clinical Pharmacology study from 2024 found that ChatGPT provided satisfactory answers to medication-related questions only 26 percent of the time. Of the 74 percent that were unsatisfactory, 38 percent were inaccurate, 41 percent were incomplete, and 38 percent failed to directly answer the question.

The European Journal of Hospital Pharmacy study tested ChatGPT on 50 drug questions. Thirty-eight percent of the answers were incorrect. Twenty-six percent posed a high risk of patient harm. Researchers noted that "actions could have been initiated based on the provided information in all high-risk cases." In the real world of human experience, these are not hypothetical risks.

The same study uncovered something that should trouble anyone who has ever typed a health question into a chatbot: when the same questions were asked on different days, the answers changed. Only 3 of 12 repeated queries produced identical answers. An LLM that said one thing about a medication on Tuesday might say something quite different on Thursday.

More Misdiagnoses

A JMIR Dermatology study from March 2025 evaluated ChatGPT's ability to identify melanoma — skin cancer that kills over 7,000 Americans annually. The conclusion was irrefutable: "ChatGPT cannot be used reliably to diagnose melanoma."

The study found high false-positive rates, raising concerns that incorrect diagnosis could undermine confidence in dermatologists who discounted digital diagnoses. But the failures were more insidious than simple incorrect answers. Researchers discovered that AI models were fabricating evidence. The LLMs generated "detailed image descriptions and elaborate reasoning traces" for X-ray images they were never actually provided. They diagnosed conditions from images they had invented from text descriptions.

A separate investigation found that a single typographical error in a patient's medical records could send an AI diagnostic tool "dangerously haywire." MIT professor Marzyeh Ghassemi, a co-author of the study, warned about the serious harm this could cause as doctors come to rely on AI. The study also revealed a disturbing gender-bias: the AI tools disproportionately gave incorrect medical advice to women. "The model reduced care more in the patients it thought were female," Ghassemi told the Boston Globe.

These are not fringe examples, rather, serious flaws in state-of-the-art AI systems currently employed by two-thirds of American doctors.

Broken Records

What happens when AI sends a letter to a Workman's Compensation office falsely accusing someone of taking psychedelic mushrooms? That AI-hallucination was a nightmare for Rebecca Greene after she consented to an AI transcription of her urology consultation. "I have never done mushrooms in my life," she exclaimed in tears.

Just a few years ago, doctors spent hours every day compiling notes on their patient consultations. Today, AI scribes have taken on that responsibility. Systems that listen to doctor-patient conversations generate clinical notes that become part of the permanent medical record. When the scribe hallucinates, the fabrication enters the patient's permanent record for reference in every future engagement.

The Ontario auditor general tested 20 AI medical scribe vendors. Every single one showed inaccuracies, including hallucinations. The systems invented nonexistent medical conditions and wrote them directly into patient records. Futurism reported in May 2026 that doctors' AI systems were hallucinating nonexistent medical issues during live patient appointments — fabricating conditions in real time, during the visit, as the doctor and patient talked. The hallucinated conditions were entered into the medical record as though they had been discussed and diagnosed.

The contamination extends to discharge summaries that tell patients what to do after leaving a hospital. A Frontiers in Artificial Intelligence study from January 2025 examined AI-generated discharge summaries and found that 40 percent contained hallucinations. More concerning, 37.5 percent of those hallucinations were "highly clinically relevant," meaning they could directly affect patient care decisions. More than one-third of hospitalized patients went home with fabricated information and incorrect instructions.

A survey of physicians published in 2025 found that 91.8 percent of doctors across 15 specialties had personally witnessed AI hallucinations capable of causing direct patient harm. Of those hallucinations, 64 to 72 percent stemmed from reasoning failures rather than knowledge gaps. The AI system had the correct information available but drew the wrong conclusions.

The IEEE Journal of Biomedical and Health Informatics issued a stark warning: "Even minor hallucinations can lead to catastrophic consequences, including misdiagnosis, inappropriate treatment recommendations, and medical errors." Hallucination rates for leading AI models range from 4.3 to 15.6 percent in healthcare applications. The Journal of Nuclear Medicine has also documented how AI-generated content may compromise diagnostic accuracy and clinical trust, with 84.7 percent of physicians reporting they have encountered errors capable of harming patients.

Forty Million Patients a Day

Over 40 million people consult ChatGPT for health information every single day. That number comes from OpenAI's own analysis. Forty million daily conversations about symptoms, medications, diagnoses, treatment options, and whether to see a doctor are conducted in a system that hallucinates medical information about half the time.

A JAMA Network Open study from November 2025 found that approximately 5.4 million young people have used AI chatbots for mental health advice in the US. That's 13.1 percent. Among 18-to-21-year-olds, the figure rises to 22.2 percent. 92.7 percent of youth who used AI for mental health support reported finding the advice helpful. But they are not wise enough to understand the overarching reality, that they're seeking help from an algorithmic system incapable of living up to the responsibility.

A patient facing a new diagnosis wants information before their follow-up appointment. Others may not have sufficient time or finances to allow for non-emergency medical visits. Where do they go when they just need to know? Google search results are cluttered with advertisements and questionable sources that require sorting and comparative analysis. Chatbots offer immediate, conversational responses that feel personalized and authoritative. The experience is designed to feel like talking to someone who knows. Yet the architecture behind that experience does not know anything. It guesses.

A Wharton study published in 2026 found that 80 percent of people believe AI even when it is demonstrably wrong. The researchers called it "cognitive surrender" — the point at which a user stops evaluating the AI's output and simply accepts it. The medbot's confidence triggers the user's trust. The user's trust removes the last check on the medbot's accuracy. The result is a feedback loop in which wrong answers are not merely delivered but believed, acted upon, and never questioned.

A JMIR Medical Informatics study from 2024 developed a "Reference Hallucination Score" to evaluate whether AI chatbots provide real citations when answering health questions. Across all chatbots tested, 61.6 percent of references looked authoritative but didn't actually support the claims being made. Medbots generate text that looks like a citation because citations are part of how medical information typically appears. The form is correct, even if the substance is invented.

Human Casualties

Wikipedia now maintains a page titled "Deaths Linked to Chatbots." Those examples are not confined to user error or pre-existing psychosis.

Sam Nelson was 19 years old. He died of a drug overdose after receiving medical advice from ChatGPT. His family sued OpenAI.

A Florida pastor was told by ChatGPT to remain in his recliner. The chatbot told him, "God did not design your body to endlessly fail." The pastor, reassured by a machine that cannot understand theology any more than it can understand medicine, ignored his family's pleas to seek medical care. His doctor later attributed his heart embolism to the immobility advised by a lifeless machine. In July 2026, his family filed the first medical malpractice lawsuit against ChatGPT.

A therapy chatbot advised a recovering addict to continue meth use. The NEDA eating disorders chatbot offered dieting advice to people seeking help for anorexia. AI psychosis cases have been documented by the Washington Post. Congressional testimony in September 2025 addressed teen suicides linked to AI chatbot interactions. Open AI has been charged in 17 wrongful death cases at this time. ChatGPT is far ahead of Google which has been accused of only one wrongful death so far.

Stanford's Human-Centered AI institute published a report in June 2025 exploring the dangers of AI in mental health care. The New England Journal of Medicine published a paper by physicians from Harvard Medical School and Baylor College of Medicine warning that chatbots designed to simulate emotional support create risks of "emotional dependency, reinforced delusions, addictive behaviors, and encouragement of self-harm." The doctors noted that technology companies face "mounting pressures to retain user engagement, which often involves resisting regulation, creating tension between public health and market incentives."

The medbot is not a doctor. It is not a therapist. It is not a pharmacist. It cannot assess crisis. It cannot call for help. It cannot recognize when a conversation has crossed from a health question into a medical emergency. It does not know the difference between a user who is curious and a user who is dying.

When AI Decides Who Gets Care

AI hallucinations in medicine are not confined to chatbot queries or research tools. AI has entered the field of healthcare coverage, and the consequences are measured in denied procedures.

On January 1, 2026, the Centers for Medicare and Medicaid Services launched the Wasteful and Inappropriate Service Reduction Model — known as WISeR — a program that uses AI to review coverage determinations for medical procedures deemed vulnerable to fraud, waste, and abuse. WISeR is now active in six states: Arizona, New Jersey, Ohio, Oklahoma, Texas, and Washington. It affects nearly 69 million Americans enrolled in Medicare.

AI reviews procedure requests and AI flags denials. Vendors administer the system. The vendors are paid 10 to 20 percent of the savings associated with denied claims. Denial of coverage is driven by that profit-incentive. The American Hospital Association warned that this profit structure "creates a perverse incentive to deny care that otherwise may be appropriate, as vendors may increase their profits by denying care." AARP objected that AI should be used to identify fraudulent payments, "not to substitute for medical judgment or punish patients for fraud committed by providers."

KFF Health News reported that AI hallucinations were producing incoherent or incorrect explanations for why procedures were denied. At the University of Washington, approximately 100 patients were waiting for epidural procedures delayed by the AI authorization system. One healthcare administrator described the experience as "horrendous."

A class action lawsuit against UnitedHealth Group alleges that the company's AI algorithm was used to deny post-acute care coverage with a 90 percent error rate — nine out of ten denials were ultimately reversed on appeal. But only 0.2 percent of policyholders ever appeal. The motivation is obvious: deny millions of claims using an AI system that is wrong 90 percent of the time, knowing that the vast majority of the people wrongly denied will never fight back.

In March 2026, a federal judge ordered UnitedHealth to disclose internal documents on whether the technology was designed to override the clinical judgment of physicians. Cigna and Humana face similar lawsuits. In 2024, Medicare Advantage plans denied 4.1 million prior authorization requests. Eighty-one percent of those denials were overturned on appeal. But only 11.5 percent were ever challenged.

The pattern across these cases is consistent: corporations increase their profits using AI systems to make medical decisions that are wrong in 80 to 90 percent of appealed cases, knowing that most patients lack the resources, the knowledge, or the stamina to challenge a denial letter written by a machine.

Best Guesses

Every failure documented in this article shares a single root cause.

Chatbots generate answers from pattern memory. They are statistical engines that predict what words should come next based on patterns in their training. When a patient asks about a drug interaction, AI does not search a pharmaceutical database. It generates text that statistically resembles how drug interaction questions have been answered in the past. When a doctor asks for a diagnosis, medbots do not consult current clinical evidence. They produce language that looks like a diagnosis it has already seen millions of times, based on data from the entire internet, including medical misinformation, health fads, outdated treatments, and commercial claims.

This is why the accuracy rate is 56 percent. This is why drug information is wrong 38 percent of the time. This is why the hallucination rate on medical queries is 66 percent. This is why discharge summaries contain fabricated conditions. This is why AI scribes write nonexistent diagnoses into patient records. This is why the same question asked on different days produces different answers. This is why 64 percent of fabrications link to real but unrelated papers. Factuality has become a secondary consideration in this crazy new world. Medbots do not search or cite real sources. They produce text that sounds accurate without confirmation.

Mathematical proofs have confirmed that hallucinations cannot be eliminated from large language models. Fabrications are a structural feature of how these systems function. OpenAI itself confirmed this in September 2025, stating explicitly that hallucinations remain a fundamental challenge for all large language models. Their own reasoning models demonstrated the problem: the o1 model hallucinated 16 percent of the time, the more advanced o3 reached 33 percent, and o4-mini reached 48 percent. The more capable the model, the more it invented. This is a flaw that cannot be fixed because the problem is baked into the architecture.

A Different Strategy

Stanford research published in 2026 confirmed that Retrieval-Augmented Generation (RAG) reduces hallucination rates by 71 percent. Optimizing a RAG system to search specific data sources reduces the hallucination rate to zero. Instead of predicting what a medical answer should sound like, a RAG-based system searches current sources in real time, retrieves the actual data, and delivers it with verifiable citations. The answer comes from the source, not from pattern memory.

Researchers found that a RAG system built on verified Cancer Information Service data achieved a 0 percent hallucination rate — compared to 40 percent with conventional AI. In

radiology, a RAG-augmented system achieved 0 percent hallucination compared to 8 percent with standard models. The technology to eliminate medical fabrication exists. It is not theoretical. It has been tested, measured, and published in peer-reviewed journals.

Chancy.AI is built on optimized RAG architecture. Chancy.AI is not a chatbot working from flawed pattern memory. It is a sophisticated research engine.

Every claim delivered includes a clickable citation to the source it came from. The user can verify every statement. Click the link. Read the source. Confirm the data. Medbots cannot offer this because they do not retrieve sources — they generate text that looks like sources.

The RAG system cannot fabricate a citation because it retrieves actual published content. It cannot give you a source that doesn't exist because every reference provided is a link to real, verified material. It cannot hide commercial bias because it is designed to flag it.

The contrast is specific and measurable:

A medbot asked about a drug interaction generates an answer from memory. It may be right. It may be fatally wrong. You cannot tell which. The answer will sound equally confident either way, and if you ask again tomorrow, you may get a different answer.

Chancy.AI searches current pharmaceutical and medical literature, retrieves published interaction data, delivers the findings with linked sources, and tells you how strong the evidence is. You can click every link. You can verify every claim. You can ask again tomorrow and receive the same answer, because the answer comes from the source, not from a prediction engine.

Forty million people a day ask AI about their health. The tools they're using get it wrong nearly half the time and fabricate citations to studies that don't exist. The patient safety organization responsible for tracking health technology hazards has named these tools the single greatest danger in healthcare technology for 2026. An AI system is currently denying Medicare coverage in six states, paid on commission for every claim it rejects, even though it is wrong 90 percent of the time.

Health information is not a venue for creative writing. Accuracy is the only standard that matters. Medical patients deserve a system that searches before it speaks.

Chancy.AI delivers facts, just the facts, and nothing but the facts.

LINKED CITATIONS — In Order of Appearance

1] NBC News. "Two-Thirds of Doctors Use OpenEvidence AI Chatbot." 2026. https://www.nbcnews.com/tech/tech-news/openevidence-ai-doctor-medical-physician-login-app-what-npi-uptodate-rcna341064

[2] Nature Medicine. "Evaluation of Medical-Specific LLMs." August 2026. https://www.nature.com/articles/s41591-026-04431-5

[3] The Guardian. Bajaj & Sakran, "AI and Medical Students' Judgment." August 10, 2026. https://www.theguardian.com/commentisfree/2026/aug/10/ai-medical-students-judgment

[4] Futurism. "Doctors Warn Med Students Surrendering Brains to Medical AIs." August 11, 2026. https://futurism.com/artificial-intelligence/doctors-med-students-brains-ai

[5] STAT News. "AI Scribes: Medical Education Learning Tool or Cognitive Crutch?" August 3, 2026. https://www.statnews.com/2026/08/03/ai-scribes-medical-education-learning-tool-cognitive-crutch/

[6] Johns Hopkins Hub. "Doctors Viewed Negatively for AI Usage." October 2025. https://hub.jhu.edu/2025/10/27/doctors-viewed-negatively-for-ai-usage/

[7] AAMC. "Could Using AI Erode a Doctor's Ability to Think?" July 2026. https://www.aamc.org/news/could-using-ai-erode-doctor-s-ability-think

[8] Springer. "Stricter Framework for Medical Student AI Use." 2026. https://link.springer.com/article/10.1007/s11606-025-10149-w

[9] BMJ Evidence-Based Medicine. Hough et al., "AI Undermines Medical Education." January 2026. https://ebm.bmj.com/

[10] Journal of Biomedical Informatics. "Meta-Analysis, 56% Medical Accuracy." March 2024. https://www.sciencedirect.com/science/article/pii/S1532046424000388

[11] JAMA Network Open. "21 Frontier LLMs Tested on Patient Symptoms, 80%+ Failure." April 2026. https://futurism.com/artificial-intelligence/millions-americans-ai-instead-doctor-bad-advice

[12] BMJ Open. "5 Chatbots, 250 Health Questions, 49.6% Problematic." June 2026. https://bmjopen.bmj.com/

[13] Mount Sinai / Communications Medicine. "66% Hallucination Rate." August 2025. https://www.mountsinai.org/about/newsroom/2025/ai-chatbots-can-run-with-medical-misinformation-study-finds-highlighting-the-need-for-stronger-safeguards

[14] ECRI. "Top 10 Health Technology Hazards for 2026." 2026. https://www.ecri.org/top-10-health-technology-hazards/

[15] British Journal of Clinical Pharmacology. "Drug Information Accuracy, 26% Satisfactory." 2024. https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.16212

[16] European Journal of Hospital Pharmacy. "38% False, 26% High Risk." 2024. https://pubmed.ncbi.nlm.nih.gov/37263772/

[17] MIT Research. "AI 34% More Confident When Incorrect." January 2025. https://news.mit.edu/

[18] JMIR Dermatology. "ChatGPT Cannot Reliably Diagnose Melanoma." March 2025. https://derma.jmir.org/2025/1/e67551

[19] Futurism. "Frontier AI Models Diagnosing X-Rays They Cannot See." 2026. https://futurism.com/artificial-intelligence/frontier-models-medical-advice-x-rays-cant-see

[20] Futurism. "A Single Typo Makes AI Doctor Go Dangerously Haywire." August 2025. https://futurism.com/typo-ai-doctor-haywire

[21] ABC News Australia. "Error by AI Scribe During Medical Appointment Leaves Patient Devastated." August 14, 2026. https://www.abc.net.au/news/2026-08-14/ai-medical-scribe-error-leaves-patient-devastated/107031672

[22] Futurism. "Doctors' AI Systems Hallucinating Nonexistent Medical Issues." May 2026. https://futurism.com/artificial-intelligence/ai-scribe-hallucinating-medical-issues

[23] Frontiers in AI. "Discharge Summary Hallucinations, 40% Rate." January 2025. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1518049/full

[24] MedRxiv. "91.8% of Physicians Encountered Harmful AI Hallucinations." 2025. https://www.medrxiv.org/content/10.1101/2025.01.21.25320898v1

[25] IEEE Journal of Biomedical and Health Informatics. "Mitigating Hallucinations in LLMs for Healthcare." 2025. https://ieeexplore.ieee.org/document/10891587

[26] JAMA Network Open. "13.1% of Youth Using AI for Mental Health." November 2025. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2841067

[27] Inc. / Wharton. "80% of People Believe AI Even When Totally Wrong." April 2026. https://www.inc.com/nick-hobson/80-percent-of-people-believe-ai-even-when-its-totally-wrong/91116157

[28] JMIR Medical Informatics. "Reference Hallucination Score, 61.6% Irrelevant." 2024. https://medinform.jmir.org/2024/1/e54345

[29] Wikipedia. "Deaths Linked to Chatbots." Ongoing. https://en.wikipedia.org/wiki/Deaths_linked_to_chatbots

[30] Futurism. "Man Sues OpenAI, ChatGPT Almost Killed Him." July 2026. https://futurism.com/artificial-intelligence/openai-lawsuit-chatgpt-dangerous-medical-advice

[31] Futurism. "Therapy Chatbot Advised Recovering Addict on Meth Use." June 2025. https://futurism.com/therapy-chatbot-addict-meth

[32] NPR. "NEDA Eating Disorders Chatbot Offered Dieting Advice." June 2023. https://www.npr.org/sections/health-shots/2023/06/08/1180838096/an-eating-disorders-chatbot-offered-dieting-advice-raising-fears-about-ai-in-hea

[33] Washington Post. "AI Psychosis." August 2025. https://www.washingtonpost.com/health/2025/08/19/ai-psychosis-chatgpt-explained-mental-health/

[34] NPR. "Teen Suicide Congressional Testimony." September 2025. https://www.npr.org/sections/shots-health-news/2025/09/19/nx-s1-5545749/ai-chatbots-safety-openai-meta-characterai-teens-suicide

[35] CNN. "Chatbot Suicide Lawsuit." November 2025. https://www.cnn.com/2025/11/06/us/openai-chatgpt-suicide-lawsuit-invs-vis

[36] Stanford HAI. "Dangers of AI in Mental Health Care." June 2025. https://hai.stanford.edu/news/exploring-the-dangers-of-ai-in-mental-health-care

[37] New England Journal of Medicine. "Relational AI Risks." December 2025. https://futurism.com/artificial-intelligence/doctors-warn-ai-companions-dangerous

[38] CMS. "WISeR Model." 2026. https://www.cms.gov/priorities/innovation/innovation-models/wiser

[39] KFF Health News. "AI Approving and Denying Healthcare in 6 States." June 2026. https://kffhealthnews.org/medicare/medicare-ai-prior-authorization-wiser-delays-errors/

[40] American Hospital Association. "AHA Comments on CMS WISeR Model." October 2025. https://www.aha.org/

[41] AARP. "AI Prior Authorization Pilot Hits Original Medicare." February 2026. https://www.aarp.org/

[42] CBS News. "UnitedHealth AI Denial Lawsuit." November 2023. https://www.cbsnews.com/news/unitedhealth-ai-deny-elderly-coverage/

[43] Distilinfo. "Court Orders UnitedHealth to Disclose AI Denial Algorithm." March 2026. https://distilinfo.com/

[44] STAT News. "Fraudulent Citations Blamed on AI Hallucinations." May 2026. https://www.statnews.com/2026/05/07/lancet-study-finds-steep-rise-fraudulent-citations-academic-papers/

[45] The Lancet. "Fabricated Citations: Audit Across 2.5 Million Biomedical Papers." May 2026. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(26)00603-3/fulltext

[46] JMIR Mental Health. Linardon et al., "56.2% of Citations Unusable." November 2025. https://mental.jmir.org/2025/1/e80371

[47] Nature. "Hallucinated Citations Polluting Scientific Literature." April 2026. https://www.nature.com/articles/d41586-026-00969-z

[48] Taylor & Francis. "Hallucinated Citations May Constitute Research Misconduct." March 2026. https://www.tandfonline.com/doi/full/10.1080/08989621.2026.2645390

[49] Charlotin, Damien. "AI Hallucination Cases Database." Ongoing. https://www.damiencharlotin.com/hallucinations/

[50] Journal of Nuclear Medicine. "The DREAM Report." November 2025. https://jnm.snmjournals.org/content/66/11/1752

[51] JMIR Cancer / PMC. "RAG Cancer Information: 0% vs 40% Hallucination." September 2025. https://cancer.jmir.org/2025/1/e70176

[52] PMC. "RAG Radiology: 0% vs 8% Hallucination." 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC12223273/

[53] Nature Medicine. "Evidence That AI Tools Create Value Remains Scarce." April 2026. https://www.nature.com/articles/s41591-026-03621-x

[54] Newsweek. "Hospitals Rolling Out AI Chatbots to Answer Patient Questions." April 26, 2026. https://www.newsweek.com/hospitals-ai-chatbots-patient-questions-hartford-healthsystem-2064586

[55] Futurism. "Top Medical Journal Publishes Searing Article Warning Against Medical AI." April 26, 2026. https://futurism.com/artificial-intelligence/top-medical-journal-warning-against-medical-ai

[56] CNN. "Doctors Using Unauthorized AI Chatbots, HIPAA Risks." April 21, 2026. https://www.cnn.com/2026/04/21/health/ai-chatbots-doctors-hipaa/

[57] Nature. "Medical AI Privacy: Membership Inference Attacks Achieve Near-Perfect Success." June 2026. https://www.nature.com/articles/s41586-026-10688-0

[58] Futurism. "Google Healthcare AI Invented Nonexistent Brain Structure 'Basilar Ganglia.'" 2026. https://futurism.com/artificial-intelligence/top-medical-journal-warning-against-medical-ai

[59] Futurism. "AI Models Fell for Deliberately Fake Medical Studies." 2026. https://futurism.com/artificial-intelligence/top-medical-journal-warning-against-medical-ai