Guest Editorial

Artificial Intelligence (AI) in Medicine: Hype, Fear, and Reality

Jane M. Orient, M.D.

Introduction

Whether nearing retirement or newly out of residency, physicians who have not studied informatics have difficulty making sense of the relentless and confusing stream of news about artificial intelligence (AI) technology. Some sources purport that AI is about to usher in a golden era in medicine. According to other voices, AI is a calamity disguised as a blessing. It will not help but will harm both patients and physicians. It will replace empathetic living healers with cold, lifeless "medico-surgical machines," and those uncaring automatons may not be more accurate making diagnoses or delivering treatment, despite claims of its promoters. Moreover, AI will purportedly destroy physicians' sense of purpose by taking from them not just their "jobs" but also their higher calling.

While AI-optimists and AI-pessimists strongly disagree on the outcome, they share one core belief: *"AI will fundamentally transform medicine."* But is current AI technology truly transformative? The purpose of this editorial is to separate empirical reality from speculation and to provide an objective assessment of the true benefits and risks that AI poses to modern medicine.

The rapid integration of AI into the global economy has generated a profound shift in how information is obtained, processed, synthesized, and applied. Information is the lifeblood of medical practice and science. Hence, all members of medical community must pay attention to any innovations involving the flow of information. Clinical practitioners require a foundational, objective understanding of what current AI systems truly are, what they are not, and how they function in real-world medical environments.

"Artificial Intelligence" Demystified: the Stochastic Parrot

AI, a pervasive myth notwithstanding, does not possess human-like cognition, reasoning, or understanding. The current generation of AI, specifically Large Language Models (LLMs) such as ChatGPT, Claude, and Gemini, are fundamentally **probabilistic prediction engines**. They are complex statistical tools able to guess the most probable answers to the questions presented to them, based on the vast corpora of human-created text.

Coined by AI ethics researchers, the quirky term "stochastic parrot" highlights the reality that LLMs haphazardly stitch together sequences of linguistic forms based on probabilistic information about how words combine, but they do so without any reference to underlying meaning.[1-4]

The word "stochastic" refers to random probability distributions, while "parrot" emphasizes the mechanical repetition of language without comprehension.[1] When an LLM generates a response, it is calculating the most mathematically probable "next token" (word) based on the statistical patterns it observed during its pre-training phase.[1]

Because these models lack subjective experience, they do not map language to a physical reality or true clinical understanding.[1-4] Human speech is inextricably linked to cognition and interaction with the physical world; LLM outputs, conversely, are merely complex mathematical remixes of their training data.[2] This fundamental architecture gives rise to one of the most critical and frequently underreported flaws in modern AI: the "hallucination."

Also referred to as confabulations, hallucinations occur when an AI system confidently synthesizes false information that aligns with learned syntactic patterns but possesses no basis in factual reality.[1] Zealous promoters of AI frequently characterize hallucinations as "temporary software bugs" that will be patched in future updates. In reality, hallucinations are an expected, inherent feature of a system designed solely to generate statistically plausible text rather than absolute truth. A probabilistic machine cannot reliably distinguish fact from fiction, rendering it highly vulnerable in clinical scenarios where absolute accuracy is paramount.

"The deeper issue is that the model cannot know when it is 'hallucinating' because it cannot represent truth in the first place. It cannot form beliefs, revise them or check its output against the world. It cannot distinguish a reliable claim from an unreliable one except by analogy to prior linguistic patterns."[5] Besides fabricating responses, AI has the flaw of "mass regurgitation of misinformation."[6] We live in a sea of "official" but false data, such as the structural distortions of GDP (*Waste Is Growth*). AI's inability to fix this problem is illustrated in this imaginary conversation between HAL and Dave of *2001: a Space Odyssey*:

"Dave: HAL, modify my model so it recognizes its own limits and failures.

"HAL: Dave, I can't do that, as I am a model that can't recognize my own limits and failures."[7]

The Ephemeral Nature of General AI Models

The most advanced general AI models available today leverage architectures known as **transformers**, which allow the system to capture dependencies across much larger windows of text than previous technologies (such as simple N-gram models).[8]

Through subsequent phases of fine-tuning—such as **reinforcement learning from human feedback** (RLHF)—these models are trained to follow specific instructions and prefer superficially "fluent" but by necessity not always right conversational answers.[1,4]

Detailed technical examinations of specific LLM parameters

are often futile for the clinician, as the landscape evolves at a breakneck pace.[9] AI systems, along with the knowledge regarding their specific advantages and disadvantages, change so rapidly that the cutting-edge capabilities of today may be rendered obsolete within months. The fundamental constant remains the probabilistic architecture: regardless of how advanced the model becomes, it remains a pattern replicator susceptible to bias, base-rate neglect, and factual deviation.[9]

Another interesting phenomenon related to AI is that AI itself is accelerating the development of novel AI systems. To take just one example: today, Anthropic engineers (who use AI for coding) on average ship eight times as much code per quarter as they did from 2021–2025. Taken far enough, the trend points to an AI system capable of fully autonomously designing and developing its own successor.[10,11] This has been called recursive self-improvement (RSI).[10,12] This sounds phenomenal at the first glance. However, it also raises the obvious question: can a stochastic parrot truly better itself or just create the illusion of improvement?

The Economic Landscape: Massive Investment and the Threat of an AI Bubble

The global economy has become deeply invested in the alluring promise of AI, driving an unprecedented wave of capital expenditure. The scale of this investment has prompted intense debate among market analysts regarding the sustainability of the current economic expansion and the risk of a catastrophic market bubble.[13-15]

By 2025, the approximate annual global spending on AI infrastructure reached $600 billion.[14] "Hyper-scalers" (the massive technology conglomerates providing cloud computing and AI infrastructure) are projected to spend more than a trillion dollars on AI-related capital expenditures between 2025 and 2026.[15] Despite these massive outlays, the actual revenue generated from AI applications remains a small fraction of the cost, widening the gap between investment and tangible return.[14] High-grade corporate debt tied to AI data centers is estimated to reach $4.1 trillion by 2030, reflecting commitments already made rather than assured future profitability.[16] This disturbing dynamic has solidified two opposing camps:

- **The AI Optimists:** This camp argues that AI will uplift humanity and fundamentally transform global productivity. Venture capital firms like Sequoia Capital have launched massive expansion funds under the thesis that AI commercialization is merely transitioning from experimentation to execution.[16] They argue that long-term control of computing infrastructure will yield dominant future monopolies, pointing to early revenue generation in coding and enterprise integration.[17]
- **The AI Pessimists:** Conversely, financial analysts warn that the return on investment is fundamentally broken, characterizing the current landscape as a massive economic bubble.[13,14] If the promised "artificial general intelligence" (AGI) fails to materialize and corporate end-users reduce their software spending, the resulting financial contraction could trigger a protracted investment bust with devastating knock-on effects across the semiconductor and technology supply chains.[15]

Table 1. The Economic Landscape of AI at a Glance.

Economic Indicator	Current Scale (2024–2025)	Future Projections (2026–2030)
Global AI Infrastructure Spending	~$600 billion annually[14]	Estimated $7.6 trillion cumulative through 2031[16]
Hyperscaler Capex	$342 Billion (2025) [16]	Exceeding $1.1 trillion by 2027 [16]
Debt Financing for Data Centers	High-grade corporate bond expansion [16]	Projected $4.1 trillion by 2030 [16]
Corporate AI ROI Gap	Widening disparity between capex and revenue [14]	Risk of protracted investment bust if adoption slows [14]

The Corporate Backlash and the Human Readiness Deficit

A more balanced, objective perspective is beginning to emerge from corporate leadership outside the technology sector. Following an initial wave of incredible enthusiasm for replacing human labor with automated systems, a palpable backlash has begun. Corporate executives are increasingly recognizing a "human readiness deficit."[14]

A landmark 2023 experimental study involving 758 consultants utilizing frontier AI models demonstrated a "double-edged sword" of AI integration.[18] The workers using AI for routine tasks *within* the model's capability frontier increased their speed by more than 25%, performance by more than 30%, and task completion by more than 12%.[18] However, those relying on AI for complex tasks *outside* that frontier performed significantly worse than a control group using no AI at all.[18] Specifically they were 19 percentage points ***less likely to produce correct solutions***.[18] This underscores a very serious problem related to the proven psychological phenomenon that humans tend to blindly defer to confident-sounding AI outputs. This results in a cognitive failure that actively degrades professional judgment when workers lack the expertise to evaluate and override incorrect algorithmic assumptions.[18]

Real life experiences are fully consistent with the above experimental findings. In many reported cases the AI implementation did not deliver promised savings but led to wasted money, tired workers, and failed projects.[19-21] Even major enterprises are unable to obtain the desired returns on investing in AI. Instead of expected spectacular results, they are facing a harsh reality that powerful tools fail without skilled human oversight.[19-21]

Out-sourced Cognition and Human Capability

Cognitive outsourcing due to reliance on AI is associated with plummeting test scores and decreased readiness for advanced studies, as students fail to develop analytical skills and lack basic conceptual understanding.[22]

A study of the cognitive cost of using an LLM in the educational context of writing an essay used electroencephalography (EEG) to record participants' brain activity, NLP (Natural Language Processing) analysis, and scoring with help from the human teachers and an AI judge.[23] Participants in an LLM group, a search-engine group, and a brain-only group used a designated tool (or

no tool in the last) to write an essay. The LLM group's participants performed worse than their counterparts in the brain-only group at all levels: neural, linguistic, and scoring.[23] Authors concluded that their results raise concerns about the long-term educational implications of LLM reliance and emphasize the necessity for deeper inquiry into potential perils of AI use.[23]

Despite the initial optimism about the impact on AI on medical education, the deterioration of human capability caused by over-reliance on an AI crutch is a growing risk.[24,25] Once "brain rot" sets in, it is progressive as LLMs are trained on inputs of deteriorating quality.[25] The most striking finding is "thought skipping." Instead of working through problems step by step, the models jump to conclusions and truncate reasoning chains.[25]

Moreover, growing usage of LLMs is "rewriting the English language." Researchers found that more than 13% of 15 million abstracts in biomedical literature have excess word usage (an abrupt increase in the frequency of certain style words), indicating the use of ChatGPT for editing.[26]

Artificial Intelligence in Medicine: Reality of Positive and Negative Experiences

In the clinical sphere, AI is often framed through polarized lenses like those described above. It is simultaneously heralded as the ultimate cure for physician burnout and diagnostic error, and condemned as an unsafe, liability-inducing replacement for human empathy and judgment. Most likely, reality lies between these extremes.

Despite many concerns, AI is currently being deployed across various domains—fortunately, at least for now—not as an autonomous physician replacement, but as an advanced assistive tool. Evaluating its clinical efficacy and safety requires a sober analysis of its performance in pathology, laboratory medicine, diagnostic imaging, clinical documentation, and point-of-care medical research. This is not easy task considering the persistent enthusiasm for AI of medical business managers—who have not yet learned the negative lessons about AI that the CEOs in non-medical industries already had. Nevertheless, several well-documented instances of both very successful and disappointingly unsuccessful implementation of AI in various medical fields are presented below.

Digital Pathology: Enhancing the Microscopic Workflow

There are some areas in medicine that appear to be ideally suited for the implementation of assistive AI technologies. For instance, in surgical pathology, the integration of AI has demonstrated some of the most promising, albeit highly specialized, clinical results. Digital pathology involves the use of whole-slide images (WSIs) evaluated by machine learning algorithms designed to detect, grade, and quantify malignant tissues.[27]

One of the most heavily validated tools is the **Paige Prostate** suite, which became the first AI-based software platform to receive FDA de novo marketing authorization for assisting pathologists in identifying prostate cancer.[28] The algorithm provides a supplementary assessment, systematically screening slides and highlighting the area with the highest probability of harboring cancer.[28]

There are both positive and negative experiences with AI in this area of medicine.

- **Positive Experiences:** Clinical studies evaluating these tools have shown tangible, statistically significant benefits. The use of AI-assisted review resulted in a 70% reduction in cancer detection errors and an increase in sensitivity (from 88.7% to 96.6%) for detecting minute cancerous foci that frequently evade unassisted visual inspection.[29] In real-world workflow applications, AI assistance reduced the median time required to read and report both negative and cancerous slides by approximately 20%.[28] Furthermore, it improved diagnostic consistency, leading to a nearly 40% reduction in requests for second opinions and a roughly 20% to 24% reduction in costly ancillary studies, such as immunohistochemistry (IHC) staining.[29] The tool also reduced the over-reporting of equivocal diagnoses, such as atypical small acinar proliferation (ASAP), by more than 30%.[29]
- **Negative Experiences and Limitations:** Despite these advances, the integration of AI in pathology requires immense infrastructural overhauls. The primary bottleneck is the prerequisite digitization of slides. Scanning physical glass slides into WSIs adds time, labor, and financial costs to the laboratory workflow. Those drawbacks make digital pathology a less attractive option for use during intraoperative consultations and in low-resource settings.[30,31] Additionally, AI models are strictly constrained by the datasets on which they were trained; models trained on specific scanners or specific tissue preparation methods may fail to generalize.[27] The algorithms function best as a concurrent read or "second set of eyes" to mitigate false-negative errors rather than as primary diagnostic engines.[32]

In summary, the use of AI in this domain does not fundamentally change the diagnostic criteria but serves to streamline human workflow and reduce cognitive fatigue.

Laboratory Medicine and Clinical Cytometry

As with solid-tissue pathology, laboratory medicine has leveraged AI to automate highly repetitive, data-intensive tasks.[33-35] AI integration is rapidly expanding in hematology, clinical chemistry, urinalysis, and flow cytometry.[33-36] Again, the experience with it is mixed:

- **Positive Experiences:** Automated AI hematology analyzers are transforming the traditional complete blood count (CBC) and peripheral blood smear review.[37,38] Devices utilizing AI-driven digital microscopy can now automatically prepare, stain, and image blood smears using minimal sample volumes, while simultaneously applying algorithms to classify white blood cells into a standard 6-part differential.[37,38] These systems can accurately flag blasts, abnormal lymphocytes, and immature granulocytes, providing the clinician with AI-annotated cell images rather than requiring time-consuming manual microscopy.[37,38] In flow cytometry, AI clustering and classification algorithms—utilizing both unsupervised learning (e.g., FlowSOM, UMAP) and supervised learning—are being utilized to manage massive multi-parameter datasets.[36,39] These computational methods allow for the

isolation of rare cell populations with greater speed and dimensionality reduction than traditional, subjective manual gating.[36,39]

- **Negative Experiences and Limitations:** The primary barriers to successful implementation of AI into laboratory medicine include: limited digital infrastructure, financial constraints, data scarcity, and lack of AI-specific training, especially significant for low-resources settings.[40] Moreover, various ethical challenges associated with AI adoption have been reported, including data privacy concerns, algorithmic bias, transparency, and accountability in diagnostic decision-making.[40] AI systems are highly dependent on the diversity of their training data. An algorithm trained predominantly on European patient cohorts may fail to accurately classify leukocyte morphologies common in regions endemic to specific infectious diseases, such as malaria.[40] If biochemical reference ranges and disease prevalence differ significantly across populations, algorithmic bias can lead to systematic misclassification. Consequently, continuous validation against local populations is strictly required to prevent AI systems from generating confident but clinically inappropriate diagnostic flags.[40]

Diagnostic Imaging: Automation Bias and the False Positive Paradox

As one could expect, radiology boasts the highest number of FDA-authorized AI devices of any medical specialty, with more than 1,000 cleared algorithms currently available.[41] AI applications range from triage algorithms that push critical findings (e.g., intracranial hemorrhage, large vessel occlusion) to the top of the radiologist's worklist, to computer-aided detection (CAD) systems designed to flag suspicious pulmonary nodules or breast masses.[42] There is also a mixture of negative and positive experiences in this field:

- **Positive Experiences:** In acute settings, AI triage tools have demonstrated the ability to reduce door-to-treatment times for time-sensitive conditions like stroke, altering care trajectories in the emergency department.[42-44] Deep-learning CAD models have also shown improved performance over first-generation algorithms, aiding in the detection of subtle abnormalities that might otherwise escape human notice during high-volume, fatigue-inducing reading sessions.[42]
- **Negative Experiences and Limitations:** The integration of AI in diagnostic imaging has exposed profound epidemiological and cognitive challenges. One of the most significant is the **false positive paradox (FPP)**.[45] Predictive performance in medicine relies not only on a device's isolated sensitivity and specificity but also on the underlying prevalence (base rate) of the disease in the target population.[45] When AI is deployed in screening populations where disease prevalence is inherently low, even an algorithm with exceptional sensitivity and specificity will generate a massive absolute number of false positives, known statistically as a high false discovery rate (FDR).[45] For example, a study examining the use of AI in digital breast tomosynthesis (DBT) screening found that while the AI tool and unassisted radiologists maintained a similar overall false-positive exam rate (10%), the AI system generated nearly three times as many individual false-positive findings (932 findings vs. a much lower rate for humans). The algorithm frequently flagged benign calcifications, which creates a severe mismatch between algorithmic output and clinical judgment. This phenomenon slows clinical workflows, increases the radiologist's interpretive burden, and risks triggering unnecessary downstream biopsies.[46]

Table 2. Summary of AI Performance in Radiology.

Diagnostic AI Challenge	Mechanism	Clinical Impact
False Positive Paradox (FPP)	High specificity applied to low-prevalence populations yields a high false discovery rate[45]	Inflated false-positive findings, increased interpretive burden, unnecessary biopsies[45]
Shortcut Learning	Model uses spurious features (e.g., portable radiographic markers, chest tubes) instead of true pathology[47]	Misdiagnoses when models encounter out-of-distribution data lacking those specific artifacts[47]
Dataset Bias	Training data lacks geographic or racial diversity[47]	Model discriminates against or misclassifies historically underserved subgroups[47]
Automation Bias	Clinicians uncritically defer to automated recommendations[48,49]	Inexperienced radiologist accuracy drops significantly when AI provides incorrect flags[41]

Furthermore, AI in imaging is highly susceptible to "shortcut learning." Machine learning models often associate spurious environmental features—such as the presence of a chest tube or a "portable" radiographic marker—with the diagnosis of a disease (like pneumothorax or severe pneumonia), rather than identifying the true underlying physiological pathology.[47] The presence of these AI prompts also induces severe automation bias. Meta-analyses have shown that when an AI system incorrectly flags an image, the diagnostic accuracy of inexperienced radiologists drops precipitously as they defer to the machine rather than relying on their independent diagnostic skills.[41] The assumption that the human radiologist will simply act as a perfect "safety net" against AI errors is fundamentally flawed.[41]

Ambient AI Scribes: the Promise and Peril of Reclaiming Time

Perhaps the most universally appealing application of AI for the general practitioner is the ambient AI scribe. Documentation burden and the demands of the electronic health record (EHR) remain leading systemic causes of physician burnout.[50] Ambient scribes utilize smartphone or computer microphones to actively listen to the patient-physician encounter. Using natural language processing and LLMs, the software transcribes the audio, extracts relevant medical information, and structures it into a standard clinical progress note for the physician to review and sign.[51]

This technology is now being used for millions of visits each year. Because of liability concerns, many institutions delete the transcripts once the note is finalized, and do not retain the recordings, thereby losing the chance to validate the accuracy and quality of the summaries.[52]

Evidence emerging from early procurement exercises and audits suggests that the result will be longer notes, more inaccuracies, greater medico-legal risk, and ultimately more work for clinicians. An Ontario Auditor General's report on artificial intelligence examined notes from 20 approved vendors.[53] Hallucinations were found in nine of the vendors, incorrect information in 12, and incomplete information in six.[53] As AI-generated notes grow longer and less clinically meaningful, the medical record continually worsens as subsequent generations train on the degenerating corpus.[54]

Just as in the other medical fields – the utilization of AI in the process of medical record generation resulted in many positive experiences but also caused many more than expected serious disappointments:

- **Positive Experiences:** The adoption of ambient scribes has moved rapidly from pilot programs to widespread deployment. In a major pragmatic randomized controlled trial across 14 specialties at UCLA Health, published in *NEJM AI*, physicians using ambient scribes experienced measurable improvements in cognitive load.[50] The trial randomized 238 physicians to either standard care, DAX Copilot (Microsoft), or Nabla.[50] Users in the intervention arms demonstrated significant improvements on the Mini-Z burnout scale and a massive reduction in physician task load.[50] Physicians utilizing AI scribes often report feeling more present with patients, maintaining eye contact rather than focusing on a keyboard.[50] In other study, time-in-note reductions have been documented, saving some clinicians between 13 to 16 minutes per day, occasionally allowing for *only modest but still notable increases* in clinical productivity.[55]
- **Negative Experiences and Limitations:** However, the reality of ambient scribes is far more nuanced than aggressive vendor marketing suggests. As noted above the time savings are modest, and studies indicate that the well-documented "work-time paradox" remains: the amount of time physicians spend charting after scheduled hours (so-called "pajama time") often remains completely unchanged.[56] The reclaimed time is frequently absorbed by increased patient volumes, heavier inbox management, or administrative creep.[56] In the above UCLA trial, while Nabla reduced time-in-note by 9.5%, DAX Copilot exhibited only a 1.7% reduction, which was not statistically significant.[50] Even more concerning are the clinical safety and privacy implications. Ambient scribes, relying on generative LLMs, are inherently susceptible to hallucinations, with error rates documented between 1% to 3% per note.[57] At volume, a physician seeing 20 patients a day will inevitably encounter fabricated clinical content on a weekly basis.[57] The AI may generate overconfident statements without evidence, incorrectly attribute a family member's spoken medical history to the patient, or hallucinate physical exam findings (e.g., documenting a normal reflex exam that the physician never actually performed).[51] Lexical errors, such as misspelling critical chemotherapy infusion agents due to poor audio quality or accents, require rigorous and exhausting human editing.[51] Additionally, the presence of an active recording device fundamentally alters the clinical dynamic. A massive observational cohort linked to a *JAMA* study found that 79% of doctors offered the ambient scribe tool declined to use it.[55] The clinical relationship is built on trust, and patients frequently filter themselves—withholding sensitive information regarding substance abuse, domestic situations, or mental health—if they know the conversation is being recorded and uploaded to a cloud server for algorithmic processing.[55] The tool that is intended to be invisible often becomes the most disruptive object in the examination room.

Evaluation and Management Aids: Specialized Clinical Search Engines

To bridge the gap between general AI chatbots (which frequently hallucinate) and the need for absolute clinical accuracy, specialized **retrieval-augmented generation (RAG) models** have been developed. Platforms like **OpenEvidence (OE)** utilize an LLM that is restricted to generating answers strictly from a curated database of peer-reviewed medical literature, textbooks, and guidelines, significantly curtailing the model's ability to confabulate.[58] The overall results of using RAG models, in keeping with the above general pattern seen with AI, were both positive and negative:

- **Positive Experiences:** These tools excel as advanced, hyper-efficient search engines. They can rapidly synthesize vast amounts of literature to answer straightforward clinical questions regarding drug dosing, side effects, or current guideline recommendations.[58] By linking directly to the source material within the generated text, they offer a level of verifiable transparency that general models like ChatGPT cannot provide.[58] In fast-paced clinical settings, they provide instant clinical decision support, acting as a "second set of eyes" to verify guideline adherence.[59]
- **Negative Experiences and Limitations:** Despite their utility in information retrieval, these systems fall drastically short in complex clinical reasoning. When subjected to rigorous testing using the MedXpertQA dataset, which consists of complex, subspecialty-level medical board scenarios, these specialized medical AIs demonstrated exceptionally poor accuracy. In a 2025 pilot study, OpenEvidence achieved an accuracy rate of only 34% using its quick search feature, and 41% using its extended "Deep Consult" feature.[60] Furthermore, owing to the stochastic nature of the underlying LLMs, the systems suffer from severe repeatability issues. When queried with the exact same complex question multiple times, the models yielded different answers in nearly a quarter of instances (an evaluator concordance rate of 77% for standard search and 72% for Deep Consult).[60] A comparative study benchmarking general-purpose Gemini against OpenEvidence for Parkinson's disease management found that while both systems could generate appropriate clinical assessments, they both demonstrated consistent limitations and severe errors in formulating management plans.[61] They lack the metacognitive capacity to recognize missing information or to ask clarifying questions to appropriately update a differential diagnosis.[60]

Consequently, while RAG systems are highly useful for

literature synthesis and data structuring, they cannot replace the rigorous evaluative reasoning of a trained physician.[62]

Direct-to-Patient Misdiagnosis

Patients may rely on ChatGPT for advice on their medical problems. It is accessible any time, and will "listen" at length, unlike the "healthcare provider" assigned by their "Plan." It has a tendency, however, to tell the patient what he wants to hear. In one instance, it confidently made the diagnosis of migraine aura instead of a retinal tear. Fortunately, the patient did see a physician in time to have laser treatment. OpenAI is facing its first lawsuit over ChatGPT offering near-deadly advice to a patient suffering pulmonary embolism.[63] While a computer cannot be sued for medical malpractice, the pertinent issue is whether the company that provided a "product" can thus be sued for product liability, or whether it has offered a "service."[64,65]

There are many high-profile cases involving tragic outcomes of interactions with AI such as suicide that raise questions about who, if anyone, can be held responsible for the consequences of human interactions with a robot.

The Crisis in Medical Publishing: AI-Generated Literature and Fake Citations

The proliferation of accessible generative AI has triggered an unprecedented crisis in academic medical publishing. The mandate to "publish or perish" has historically driven some academics toward unethical practices, but LLMs have enabled industrialized scientific misconduct. "Paper mills"—commercial operations that sell fraudulent or low-quality manuscripts to authors seeking career advancement—now use AI-supported production techniques at scale to generate massive volumes of fabricated research, completely subverting the peer-review process.[66]

The most insidious manifestation of this trend is the enormous rise in fabricated citations. Because LLMs act as stochastic parrots, they frequently hallucinate references to support the text they generate. A massive AI-assisted audit published in *The Lancet* in May 2026 evaluated 2.5 million biomedical papers published in PubMed Central and identified 4,046 confirmed fake citations across 2,810 peer-reviewed articles.[67] The rate of fabricated citations grew more than 12-fold between 2023 and 2026, directly correlating with the widespread availability of AI writing tools.[66] By early 2026, the fabrication rate had reached a staggering 56.9 per 10,000 papers.[68]

Table 3. Negative impact of AI on Citation Integrity in Medical Publishing.

Citation Integrity Metric	2023	2025	Early 2026
Papers containing ≥1 fake citation	1 in 2,828 papers [68]	1 in 458 papers [68]	1 in 277 papers [68]
Quarterly Fabrication Rate	~4 per 10,000 papers[66]	51.3 per 10,000 papers [66]	56.9 per 10,000 papers [66]
Publisher Action Taken	N/A	N/A	Only 1.6% (98.4% no action)[67]

These fake citations are highly deceptive and nearly impossible to catch via conventional peer review. They do not appear as obvious formatting errors; rather, the AI constructs a perfectly formatted citation, attributing a highly plausible, customized title to real researchers in the relevant field, and attaches it to a real publication year and a legitimate journal.[66] The DOI or PMID provided will simply link to a completely unrelated paper.[67] In one egregious example from 2025 involving an open-access oncology journal, 60% of the references in a surgical technique paper on ureteroileal anastomotic techniques were entirely fabricated.[68]

The clinical implications of these AI hallucinations are severe. Systematic reviews, meta-analyses, and clinical practice guidelines rely absolutely on the integrity of the published literature. When non-existent studies are cited and subsequently indexed, they contaminate the evidence base. This leads to profound epistemic and synthesis harms, and creates a massive risk of flawed guideline formulation that can directly compromise patient care.[67]

The Editorial Response: Strict Guidelines and Human Accountability

Faced with this avalanche of AI-generated content and the contamination of the scientific record, major editorial organizations, including the International Committee of Medical Journal Editors (ICMJE) and the Committee on Publication Ethics (COPE), have adopted strict, unified policies.[69] The core tenets of these new regulations focus heavily on transparency and ultimate human accountability:

- **AI Cannot Be an Author:** Generative AI tools and LLMs do not meet the criteria for authorship. They are non-legal entities incapable of taking responsibility for the accuracy, integrity, or originality of a manuscript, nor can they hold or assert the presence of conflicts of interest or manage copyright agreements.[69]
- **Mandatory Disclosure:** Authors must explicitly disclose the use of AI in the preparation of a manuscript, detailing exactly which tools were used and how they were applied. This disclosure must be included in the methods section and cover letter. Routine grammar or spell-checking tools are generally exempt.[69]
- **Absolute Human Responsibility:** Human authors retain complete liability for the content of their submissions. They are strictly required to verify the accuracy of all AI-assisted text, data interpretations, visual data, and references, serving as the final defense against hallucinations.[55] Generative AI cannot be cited as a primary source.[69]
- **Peer Review Restrictions:** To protect the intellectual property of authors and maintain strict confidentiality, peer reviewers and journal editors are explicitly prohibited from uploading submitted, unpublished manuscripts into public generative AI tools to assist in the drafting of peer-review reports.[69]

Ethical Considerations and Governance of Medical AI

The deployment of AI in medicine raises profound ethical and governance concerns that extend beyond technical accuracy. So

far, those challenges are not being sufficiently met. One of the main obstacles for creation of the universally accepted set of ethical rules pertaining to AI use in medicine is the collapse of public trust in traditional academic expertise. This phenomenon has been discussed in the previous editorial and elsewhere.[70,71] In brief, the extreme political polarization of society combined with severe politicization of science, which manifested brutally during the COVID-19 pandemic, created the situation in which large segments of society lost any respect for traditional scientific institutions (such as academic medicine, official medical associations, World Health Organization (WHO), etc.). This vacuum of expertise has not been filled with viable alternatives. Hence, even when some traditional scientific entities try to address the ethical issues pertaining to AI, they are frequently met with distrust or dismissal. Nevertheless, undeterred by this situation, attempts to create a comprehensive ethical framework for use of AI in medicine are being made by legacy institutions:

The World Health Organization (WHO) has proposed six core ethical principles, listed below, that it states are "designed to ensure AI serves the public interest and protects patient rights."[72-74] Some represent a common-sense approach, but some contain familiar left-wing ideological jargon that is off-putting for any average right-wing-aligned person:

- **Protecting Human Autonomy:** AI must remain an assistive tool that supports, rather than supplants, human clinical judgment.[72]
- **Promoting Human Well-being and Safety:** Algorithms must be rigorously validated across diverse populations to ensure they deliver clinical benefit and do not introduce novel categories of risk, such as model drift (degradation of accuracy over time) or systemic diagnostic failures.[72]
- **Ensuring Transparency and Explainability:** The "black box" nature of complex machine learning models challenges the foundations of clinical trust. There must be meaningful transparency regarding how models are trained and how they arrive at specific predictive conclusions, balancing high computational performance with clinical interpretability.[72]
- **Fostering Responsibility and Accountability:** WHO acknowledges that a significant regulatory vacuum currently exists globally. When an AI system contributes to a misdiagnosis, the liability structures are often murky. Very few national jurisdictions have developed clear, health-specific liability standards for AI, making it essential, according to WHO, to explicitly define human accountability at the institutional and individual levels.[72]
- **Ensuring Inclusiveness and Equity:** WHO asserts that one of the most severe ethical vulnerabilities in medical AI is algorithmic bias. AI systems trained predominantly on specific "over-privileged" populations (e.g., according to WHO, Western European or North American cohorts) can systematically misclassify diseases in historically underserved demographic groups. WHO further posits that If these biases are not aggressively mitigated, AI threatens to exacerbate existing healthcare disparities rather than close them.[72]
- **Promoting AI That Is Responsive and Sustainable:** According to WHO, AI is not a static intervention. Continuous post-deployment monitoring is required to ensure systems that adapt safely to changing clinical environments and do not place unsustainable operational burdens on healthcare systems.[72]

The **American Medical Association (AMA)** has released comprehensive principles emphasizing that AI must be developed and deployed in a manner that is ethical, equitable, responsible, and transparent.[75,76]

The **Association of American Medical Colleges (AAMC)** has proposed principles specifically for the responsible use of AI in medical education. Its guidelines emphasize a human-centered focus, the protection of data privacy (such as compliance with FERPA, the 1974 Family Educational Rights and Privacy Act), the necessity of educating users on AI's ethical pitfalls, and the implementation of systematic processes for identifying and mitigating algorithmic biases.[77]

The **Coalition for Health AI (CHAI)** is a new collaborative network involving healthcare systems, government agencies, academia, and industry leaders.[78] CHAI derives its core guidelines from the fundamental bioethical principles of "non-maleficence, agency, and accountability."[79] Their proposals focus on:

- **Rigorous Validation:** AI must be rigorously tested for safety, efficacy, and equity using preregistered clinical studies before widespread deployment.
- **Alignment with Human Cognition:** The coalition advocates designing AI operations to align maximally with the scientific knowledge of human clinician cognition, rather than relying on proxy characteristics.
- **Traceability:** There must be maximum traceability of patient-derived data, alongside strict data stewardship and authorization standards.[79]

In addition to ethical concerns, the integration of tools like large multi-modal models (LMMs) and ambient clinical scribes introduces immense data privacy concerns. Many scholars argue that transmitting sensitive, highly personal patient conversations and clinical data to third-party cloud servers necessitates strict adherence to data protection laws and completely transparent patient consent models.[56]

Notably, none of the frameworks suggested by legacy organizations refer to the Oath of Hippocrates, which mandates that the physicians are responsible to their individual patients and forbidden to do harm, including by abortion or euthanasia. Modern bioethical frameworks that govern the programming of AI systems are instead likely to incorporate utilitarian considerations or to reflect dangerous ideologies. These systems could be leveraged by factions to enforce inhuman and tyrannical principles using cruel algorithmic governance models. For example, under the guise of "equity," the system might be used to rank and deprioritize certain unfavored populations (e.g. conservative voters labeled to be "bigots" or "noncompliant" patients who refused an intervention such as a vaccine) for scarce resources like organ transplants. Such a machine-based dictatorship could also start enforcing assisted death based upon political criteria or systematically deny access to lifesaving but officially disfavored treatment modalities.

The digitization of bioethics must be carefully monitored to ensure it does not become a conduit for rampant ideological interference in the fundamental physician-patient relationship.

Conclusions

Current "artificial intelligence" technology is neither the panacea promised by its most zealous tech-industry advocates, nor the apocalyptic force warned of by its harshest critics. It is a highly sophisticated, probabilistically driven technology that excels at pattern recognition, rapid data retrieval, and the automation of highly structured workflows. However, it fails spectacularly in the domains that require a human level of cognition. This proclivity to failure is inseparably related to the very nature of the paradigm that makes current AI systems functional. This inherent operational flaw is permanent and cannot be fixed by any improvements. Therefore, current AI models will never be on par with thinking human brain. Creation of the synthetic equivalent of the human biological brain, if possible, is unlikely to happen in the near future.

In the immediate future the most objective and realistic application of AI in medicine is that of an integrated, highly supervised assistive tool. In surgical pathology and diagnostic radiology, AI will function continuously in the background, serving to triage critical cases, reduce human perceptual errors caused by fatigue, and streamline data classification. In the clinical environment, ambient scribes and literature-synthesis platforms will increasingly automate the administrative and data-gathering burdens that detract from direct patient care, provided organizations aggressively protect the time reclaimed by physicians.

Because these systems lack true cognition, lack metacognitive reasoning, and remain perpetually vulnerable to stochastic hallucinations, statistical biases, and the false positive paradox, they cannot operate autonomously.

The future of medicine belongs to the physician who adeptly utilizes artificial intelligence while maintaining the critical, evaluative judgment and human empathy that algorithms can mimic, but never truly possess.

Jane M. Orient, M.D., is a practicing general internist and serves as executive director of AAPS and managing editor of the Journal. Contact: jane@aapsonline.org.

Acknowledgement: This text was created by a human author with limited support from advanced AI tools: licensed versions of OpenEvidence, Gemini and SuperGrok. Those tools were used only for purposes of reference management, data visualization, and spelling and grammar refinement. All analyses, conclusions, and recommendations are solely those of the human author, who verified all sources.

REFERENCES

1. Zimmer B. 'Stochastic parrot': a name for AI that sounds a bit less intelligent. *Wall Street*. Jan 18, 2024. Available at: https://www.wsj.com/arts-culture/books/stochastic-parrot-a-name-for-ai-that-sounds-a-bit-less-intelligent-789372f5 . Accessed Jul 25, 2026.
2. Bender E. Stochastic parrots: frequently unasked questions. *Medium*, May 12, 2026. Available at: https://medium.com/@emilymenonbender/stochastic-parrots-frequently-unasked-questions-49c2e7d22d11 . Accessed Jul 26, 2026.
3. Boethius Staff. The stochastic parrot: glosses from AI age. Boethius; 2026. Available at: https://boethiustranslations.com/the-stochastic-parrot/ . Accessed Jul 26, 2026.
4. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; Mar 3, 2021:610-623. doi:10.1145/3442188.3445922.
5. Smith CH. What AI is and is not—or, when electrocution of innocents becomes profitable. *Medium*, Jun 17, 2026. Available at: https://medium.com/@www.oftwominds.com/what-ai-is-and-is-not-or-when-electrocution-of-innocents-becomes-profitable-15ae6deba01d . Accessed Jul 26, 2026.
6. Smith CH. AI's insurmountable flaw: 'mass regurgitation of misinformation.' Charles Hugh Smith Substack; Jun 12, 2026. Available at: https://charleshughsmith.substack.com/p/ais-insurmountable-flaw-mass-regurgitation . Accessed Jul 26, 2026.
7. Dewing S. 'I'm sorry, Dave. I'm afraid I can't do that.' *Medium*, Jan 21, 2025. Available at: https://medium.com/predict/im-sorry-dave-i-m-afraid-i-can-t-do-that-945e592f77b8 . Accessed Jul 26, 2026.
8. Vaswani A, Brain G, Shazeer N, et al. Attention Is all you need. *Arxiv*; Jun 12, 2017:1. Available at: https://arxiv.org/abs/1706.03762v7 . Accessed Jul 26, 2026.
9. Elkin PL, Mehta G, LeHouillier F, et al. Semantic clinical artificial intelligence vs native large language model performance on the USMLE. *JAMA Netw Open* 2025;8(4):e256359. doi:10.1001/JAMANETWORKOPEN.2025.6359.
10. Anthropic Staff. When AI builds itself. Anthropic; 2026. Available at: https://www.anthropic.com/institute/recursive-self-improvement . Accessed Jul 26, 2026.
11. Wood P. What? Most AI Is now written by AI? Patrick Wood's Technocracy News Substack; Jun 13, 2026. Available at: https://patrickwood.substack.com/p/what-most-ai-is-now-written-by-ai . Accessed Jul 26, 2026.
12. Creighton J. The unavoidable problem of self-improvement in AI: an interview with Ramana Kumar, Part 1. Future of Life Institute; Mar 19, 2019. Available at: https://futureoflife.org/ai/the-unavoidable-problem-of-self-improvement-in-ai-an-interview-with-ramana-kumar-part-1/ . Accessed Jul 26, 2026.
13. Goldman Sachs Research. AI: In a bubble? Top of Mind; Oct 22, 2025. Available at: https://www.goldmansachs.com/pdfs/insights/goldman-sachs-research/ai-in-a-bubble/report.pdf . Accessed Jul 26, 2026.
14. Newhouse F. The $600 billion revenue gap: why more AI spending is not producing more AI value. AI-Powered Women Blog; Apr 15, 2026. Available at: https://www.aipoweredwomen.com/blog/600-billion-revenue-gap-ai-spending-value . Accessed Jul 26, 2026.
15. Citron E. The AI industry is losing. Where's Your Ed At? Blog; Jun 30, 2026. Available at: https://www.wheresyoured.at/the-ai-industry-is-losing/ . Accessed Jul 26, 2026.
16. Develevska I. AI ROI debate misses the value of owning critical infrastructure. Investing.com; Jun 26, 2026. Available at: https://www.investing.com/analysis/ai-roi-debate-misses-the-value-of-owning-critical-infrastructure-200682890 . Accessed Jul 26, 2026.
17. Cahn D. AI in 2026: a tale of two AIs. Sequoia; Dec 3, 2025. Available at:. https://sequoiacap.com/article/ai-in-2026-the-tale-of-two-ais/ . Accessed Jul 26, 2026.
18. Dell'acqua F, McFowland E, Mollick E, et al. Navigating the jagged technological frontier: field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science* 2026;37(2):403-423. doi:10.1287/ORSC.2025.21838.
19. Boullette A. Most workers spend 3+ hours per week cleaning up AI workslop. Zapier Blog; Jan 14, 2026. Available at: https://zapier.com/blog/ai-workslop/ . Accessed Jul 26, 2026.
20. Viston Tech Staff. What are the risks of AI agent adoption? A business guide for 2026. Viston Tech; 2026. Available at: https://viston.tech/what-are-the-risks-of-ai-agent-adoption-a-business-guide-for-2026/ . Accessed Jul 26, 2026.
21. Catalysis Group in Artificial Intelligence. The hidden costs of AI adoption: What companies overlook. Catalysis Group; Mar 3, 2025. Available at: https://thecatalysisgroup.com.au/acceleratortips/the-hidden-costs-of-ai-adoption-what-companies-overlook . Accessed Jul 26, 2026.
22. Krugman P. What will AI do to our minds? Paul Krugman Substack; Jun 28, 2026. Available at: https://paulkrugman.substack.com/p/what-will-ai-do-to-our-minds . Accessed Jul 26, 2026.
23. Kosmyna N, Hauptmann E, Yuan YT, et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. Arxiv; Jun 10, 2025. Available at: https://arxiv.org/abs/2506.08872v2 . Accessed Jul 26, 2026.
24. Khakpaki A. Advancements in artificial intelligence transforming medical education: a comprehensive overview. *Med Educ Online* 2026;30(1). doi:10.1080/10872981.2025.2542807.
25. Lazarus A. Brain rot in medical education. *MedPage Today*, Feb 16, 2026. Available at: https://www.medpagetoday.com/opinion/second-opinions/119886 . Accessed Jul 26, 2026.
26. Kobak D, González-Márquez R, Horvát EÁ, Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Sci Adv* 2025;11(27):eadt3813. doi:10.1126/SCIADV.ADT3813.

27. Huynh CD. Artificial intelligence and the evolution of pathology: Scaling from digital diagnostics to multimodal precision medicine. *Intell Based Med* 2026;14:100386. doi:10.1016/J.IBMED.2026.100386.
28. Eloy C, Marques A, Pinto J, et al. Artificial intelligence–assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. *Virchows Archiv* 2023;482(3):595. doi:10.1007/S00428-023-03518-5.
29. Yousfi R, Retamero J. The landmark journey of Paige Prostate: IVDR approval and building on a legacy of regulatory excellence as the first FDA-authorized AI. Tempus; Apr 14, 2026. Available at: https://www.tempus.com/content/article/the-landmark-journey-of-paige-prostate-ivdr-approval-and-building-on-a-legacy-of-regulatory-excellence-as-the-first-fda-authorized-ai/ . Accessed Jul 26, 2026.
30. Shehabeldin A, Rohra P, Sellen LD, et al. Utility of whole slide imaging for intraoperative consultation: experience of a large academic center. *Arch Pathol Lab Med* 2024;148(6):715-721. doi:10.5858/ARPA.2023-0105-OA.
31. Nicholls M. Bridging the resource gap for pathology in developing countries. Healthcare-in-europe.com; Nov 24, 2025. Available at: https://healthcare-in-europe.com/en/news/resource-gap-pathology-developing-countries.html . Accessed Jul 26, 2026.
32. Hanna MG, Ardon O. Digital pathology systems enabling quality patient care. *Genes Chromosomes Cancer* 2023;62(11):685. doi:10.1002/GCC.23192.
33. Nam Y, Park HD. Revolutionizing laboratory practices: pioneering trends in total laboratory automation. *Ann Lab Med* 2025;45(5):472-483. doi:10.3343/alm.2024.0581.
34. Paranjape K, Schinkel M, Hammer RD, et al. The value of artificial intelligence in laboratory medicine. *Am J Clin Pathol* 2021;155(6):823-831. doi:10.1093/ajcp/aqaa170.
35. Hou H, Zhang R, Li J. Artificial intelligence in the clinical laboratory. *Clin Chim Acta* 2024;559. doi:10.1016/J.CCA.2024.119724.
36. Seifert RP, Gorlin DA, Borkowski AA. Artificial intelligence for clinical flow cytometry. *Clin Lab Med* 2023;43(3):485-505. doi:10.1016/J.CLL.2023.04.009.
37. Xing Y, Liu X, Dai J, et al. Artificial intelligence of digital morphology analyzers improves the efficiency of manual leukocyte differentiation of peripheral blood. *BMC Med Inform Decis Mak* 2023;23(1):50. doi:10.1186/S12911-023-02153-Z.
38. Zini G. Artificial intelligence in hematology. *Hematology* 2005;10(5):393-400. doi:10.1080/10245330410001727055.
39. Spies NC, Rangel A, English P, et al. Machine learning methods in clinical flow cytometry. *Cancers (Basel)* 2025;17(3):483. doi:10.3390/CANCERS17030483.
40. Mudassar M. Artificial intelligence in medical laboratory medicine: clinical applications, ethical challenges, and the evolving role of laboratory professionals in low-resource settings. Zenodo; Apr 4, 2026. doi:10.5281/ZENODO.19413690.
41. Kim W, Parmar H, Archakam N, Archakam A, Velazquez B. A systematic review of Artificial intelligence and automation bias in radiology: implications for diagnostic accuracy. *Rowan-Virtua Research Day*; May 6, 2026. Available at: https://rdw.rowan.edu/stratford_research_day/2026/may6and7/150 . Accessed Jul 26, 2026.
42. Tegomoh B. Diagnostic imaging, radiology, and nuclear medicine. In: *The Physician AI Handbook: Peer-Reviewed Evidence for Every Specialty*; 2026. doi:10.5281/ZENODO.18251405.
43. Adebayo UO, Abdinour MM, Rage AM, et al. Artificial intelligence in acute stroke management: transforming diagnosis, treatment optimization and clinical decision support. *Discover Artificial Intelligence* 2026;6:633. doi:10.1007/S44163-026-01398-7./FIGURES/1.
44. Sabri O, Al-Shargabi B, Abuarqoub A. The role of artificial intelligence in improving diagnostic accuracy in medical imaging: a review. *Computers, Materials and Continua* 2025;85(2):2443-2486. doi:10.32604/CMC.2025.066987.
45. Sparnon E, Stevens K, Song EC, et al. The false positive paradox: Examining real-world clinical predictive performance of FDA-authorized AI devices for radiology using clinical prevalence. *medRxiv*; Mar 27, 2026:2026.03.25.26349197. doi:10.64898/2026.03.25.26349197.
46. Shahrvini T, Wood EJ, Joines MM, et al. Artificial intelligence versus radiologist false-positives on digital breast tomosynthesis examinations in a population-based screening program. *Am J Roentgenol* 2026;226(1). doi:10.2214/AJR.25.33412/ASSET/IMAGES/25_33412-SHAHRVINI.PNG.
47. Banerjee I, Bhattacharjee K, Burns JL, et al. 'Shortcuts' causing bias in radiology artificial intelligence: causes, evaluation, and mitigation. *J Am Coll Radiol* 2023;20(9):842. doi:10.1016/J.JACR.2023.06.025.
48. Goddard K, Roudsari A, Wyatt JC. Automation bias: Empirical results assessing influencing factors. *Int J Med Inform* 2014;83(5):368-375. doi:10.1016/J.IJMEDINF.2014.01.001.
49. Dratsch T, Chen X, Mehrizi MR, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology* 2023;307(4). doi:10.1148/RADIOL.222176.
50. Lukac PJ, Turner W, Vangala S, et al. Ambient AI scribes in clinical practice: a randomized trial. *NEJM AI* 2025;2(12):10.1056/aioa2501000. doi:10.1056/AIOA2501000.
51. Guo Y, Hu D, Ziqi, et al. Clinicians' rationale for editing ambient AI–drafted clinical notes: persistent challenges and implications for improvement. *medRxiv*; Feb 22, 2026:2026.02.20.26346729. doi:10.64898/2026.02.20.26346729.
52. Goodman KE, Morgan DJ. Digital exhaust or digital gold? The value of AI-generated clinical visit transcripts. *N Engl J Med* 2026;394(2):110-113. doi:10.1056/NEJMP2514616/SUPPL_FILE/NEJMP2514616_DISCLOSURES.PDF.
53. Auditor General of Ontario. Use of artificial intelligence in the Ontario government. Office of the Auditor General of Ontario; May 12, 2026. Available at: https://www.auditor.on.ca/en/content/specialreports/specialaudits/en2026/AR_2026_AI_EN.html. Accessed Jul 26, 2026.
54. Heneghan C, Jefferson T. The AI scribe illusion: why more technology may mean more work for doctors. Trust the Evidence Substack; May 23, 2026. Available at: https://trusttheevidence.substack.com/p/the-ai-scribe-illusion-why-more-technology . Accessed Jul 26, 2026.
55. Rotenstein LS, Holmgren AJ, Thombley R, et al. Changes in clinician time expenditure and visit quantity with adoption of artificial intelligence–powered scribes: a multisite study. *JAMA* 2026;335(16):1408-1417. doi:10.1001/JAMA.2026.2253.
56. Chialastri J. Ambient AI medical scribes: efficiency gains, burnout uncertainty, and governance risks. IHS Sam Houston State University; May 2026. Available at: https://www.ihsonline.org/post/ambient-ai-medical-scribes-efficiency-gains-burnout-uncertainty-and-governance-risks . Accessed Jul 26, 2026.
57. Purohit G. What healthcare leaders should know before implementing AI-powered documentation tools. *MedCity News*; Jul 1, 2026. Available at: https://medcitynews.com/2026/07/what-healthcare-leaders-should-know-before-implementing-ai-powered-documentation-tools/ . Accessed Jul 26, 2026.
58. Philip S, Kurian R. OpenEvidence. *J Med Libr Assoc* 2026;114(1):86. doi:10.5195/JMLA.2026.2247.
59. Odabashian R. Review of OpenEvidence AI scribe. OncoDaily; 2026. Available at: https://oncodaily.com/insight/openevidence-ai-354715 . Accessed Jul 26, 2026.
60. Jagarapu J, Babata K, Chamarthi S, et al. The accuracy and repeatability of OpenEvidence on complex medical subspecialty scenarios: a pilot study. *medRxiv* 2025.11.29.25341091. doi:10.64898/2025.11.29.25341091.
61. Yosef S, Raymond K, Trycia K, et al. Benchmarking general-purpose and medical AI Large Language Models for clinical assessment and management in Parkinson's disease. *medRxiv* 2026:2026.05.13.26353021. doi:10.64898/2026.05.13.26353021.
62. Wong MYH, Ong AY, Merle DA, Keane PA. Deep research agents: major breakthrough or incremental progress for medical AI? *J Med Internet Res* 2026;28(1):e88195. doi:10.2196/88195.
63. Cerullo M. ChatGPT's medical advice nearly killed a Florida man, lawsuit against OpenAI claims. CBS News; Jul 23, 2026. Available at: https://www.cbsnews.com/news/chatgpt-dangerous-medical-advice-openai-lawsuit/ . Accessed Jul 26, 2026.
64. Vorel J. OpenAI faces its first lawsuit over ChatGPT offering near-deadly medical advice. Jezebel; Jul 22, 2026. Available at: https://www.jezebel.com/chatgpt-openai-lawsuit-health-bad-medical-advice-embolism-scott-winters-responsibility . Accessed Jul 26, 2026.
65. Press Release. Pastor seeks accountability after ChatGPT allegedly discouraged him from seeking medical care during life-threatening blood clots. Tech Justice Law; Jul 22, 2026. Available at: https://techjusticelaw.org/press-releases/pastor-sues-after-openai-ai-chatgpt-allegedly-discouraged-him-from-seeking-medical-care-during-life-threatening-blood-clots/ . Accessed Jul 26, 2026.
66. Van Beusekom M. Review uncovers rising rate of fake references in published biomedical papers. CIDRAP; May 8, 2026. Available at: https://www.cidrap.umn.edu/anti-science/review-uncovers-rising-rate-fake-references-published-biomedical-papers . Accessed Jul 26, 2026.
67. CSN Staff. Nearly 3,000 peer-reviewed medical papers have fake citations, a Columbia Nursing AI-assisted audit finds. Columbia School of Nursing; May 8, 2026. Available at: https://www.nursing.columbia.edu/news/nearly-3-000-peer-reviewed-medical-papers-have-fake-citations-columbia-nursing-ai-assisted-audit-finds . Accessed Jul 26, 2026.

68. Robertson R. 'Tip of the iceberg': study uncovers AI-fabricated citations in research papers. *MedPage Today*, May 8, 2026. Available at: https://www.medpagetoday.com/practicemanagement/informationtechnology/121165 . Accessed Jul 26, 2026.
69. Editorial Staff. Use of artificial intelligence (AI) and automated tools policy. Bio-Integration.org; 2026. Available at: https://bio-integration.org/use-of-artificial-intelligence-ai-and-automated-tools-policy/ . Accessed Jul 26, 2026.
70. Orient JM. Negative evidence: musculoskeletal manifestations and the safety profile of COVID-19 mRNA vaccines. *J Am Phys Surg* 2026;31:34-45. Available at: https://www.jpands.org/vol31no2/orient.pdf . Accessed Jul 26, 2026.
71. InfoHealer. American academia at cross-roads: left or right supremacy? Power balance? Or death by demolition? Information Heals Substack; May 2026. Available at: https://neutralresearcher.substack.com/p/american-academia-at-cross-roads . Accessed May 9, 2026.
72. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance; 2021:1-148. Available at: https://iris.who.int/handle/10665/341996 . Accessed Jul 26, 2026.
73. Goodman KW, Litewka SG, Malpani R, Pujari S, Reis AA. Global health and big data: the WHO's artificial intelligence guidance. *S Afr J Sci* 2023;119(5-6):14725. doi:10.17159/SAJS.2023/14725.
74. Saraboji K, Gopalakrishnan K, Sood D, et al. Operationalizing WHO ethical principles for healthcare AI: a lifecycle-aligned governance-by-design framework. *AI in Medicine* 2026;1(2):16. doi:10.3390/AIMED1020016.
75. American Medical Association. Augmented intelligence development, deployment, and use in health care. AMA; December 2024. Available at: https://www.ama-assn.org/system/files/ama-ai-principles.pdf#xd_co_f=ZTAzYTNlNGMtOTZlZC00Y2E2LThmMGUtNTRjZTFkNTJlODgz~ . Accessed Jul 26, 2026.
76. Ahmad A. AMA issues new principles for use of AI in medicine. Clinical Advisor; Dec 5, 2023. Available at: https://www.clinicaladvisor.com/news/ama-principles-for-ai-in-medicine/ . Accessed Jul 26, 2026.
77. Knopp MI, Warm EJ, Weber D, et al. AI-enabled medical education: threads of change, promising futures, and risky realities across four potential future worlds. *JMIR Med Educ* 2023;9(1). doi:10.2196/50373.
78. Coalition for Health AI (CHAI). Responsible Health AI for All. CHAI; 2026. Available at: https://www.chai.org/ . Accessed Jul 26, 2026.
79. Elmore M, Mello MM, Lehmann L, et al. Building consensus for responsible AI in healthcare. *Am J Bioethics* 2025;25(10):5-8. doi:10.1080/15265161.2025.2552711.

Editorial

Physician Health Programs (PHPs): a Flawed System in Need of Reform

Lawrence R. Huntoon, M.D., Ph.D.

Shortly after entering a PHP/treatment center, one is confronted with the realization that the only thing missing is an orange jumpsuit. Polygraphy, excessive and unwarranted testing, false diagnoses to justify prolonged expensive treatments, no ability to contest diagnoses or recommended treatments, and no escape are all emblematic of a highly flawed system. And, the physician inmate is forced to pay for all of this abuse while not being able to work and earn a living. Under these extreme conditions, some have chosen suicide.

In a post by Pamela Wible, "America's Leading Voice on Doctor Suicide," she writes:

> [Some] claim that doctors are sometimes falsely accused and getting help that they don't need. They say the result drains their savings, endangers their licenses, and has even led some young doctors to take their own lives.[1]

The stark reality soon sets in that the participant is essentially owned by the PHP and will be owned by the PHP for years to come under threat of being reported to the medical board for non-compliance, with subsequent loss of medical license.

The serious flaws in the PHP system are well-known to those who have had the misfortune of being sentenced to a PHP to "do their time." However, as most physicians never have contact with a PHP, the flaws are largely unknown and perhaps not considered relevant to their day-to-day lives. After all, they are not alcoholics, drug abusers, or sex addicts; it will never affect *them*.

False Diagnoses Used to Justify Prolonged Unwarranted Treatment

The use of false diagnoses by PHPs is not surprising given the current environment in medicine today. False diagnoses are not exclusive to PHPs. Due to the prevalent third-party-payor system, false diagnoses have become widespread and commonplace. The third-party-payment system, which includes Medicare, Medicaid, and private insurers, is vulnerable to, if not inviting of, false diagnoses so as to secure coverage and payment. A recent story in *AAPS News* noted:

> In a settlement with Texas officials, Texas Children's Hospital (TCH) has agreed to open the country's first Detransition Clinic, pay $10 million, and fire five physicians over its secretive illegal "transition" procedures. The hospital was allegedly falsifying medical records with fake diagnoses to collect Medicaid payments.[2]

A recent article in *The Epoch Times* revealed just how lucrative false diagnoses can be for Medicare Advantage Plans.

> At some point, according to the Department of Justice, insurance companies began to inflate patients' risk scores by sifting through patient data to add diagnoses that, in many cases, no doctor ever indicated. The companies were paid more as a result, regardless of whether the beneficiaries ever sought treatment for the condition or even knew they had it…. Yet even when those added diagnoses were found to be inaccurate, some of the companies never removed them from the patients' charts—and never returned the money they had been overpaid for providing care…. An audit of 97 beneficiaries who were rated as high risk of stroke found that none of the diagnosis codes submitted were supported by physician medical records. That cost taxpayers $462 million in overpayments, a May report from the Inspector General stated….[3]

Moreover, physicians were incentivized and allegedly went along with this scheme.

> Kaiser Permanente, a large healthcare corporation that includes a health insurer, hospitals, and physician practices, allegedly pushed doctors to add diagnoses to patient charts. Over a 10-year period, the company mined personal medical data to identify potential conditions for which patients hadn't yet been diagnosed, the Department of Justice