

Evidence-like error and physician responsibility in retrieval-augmented clinical artificial intelligence

Fabio Di Bello¹⁻³ , Daniele Angioni⁴⁻⁷ 

¹Senior Education & AI Development Manager, Wiley Hoboken, NJ - USA

²Faculty Member, Master in Humanization of Care, University of Udine, Udine - Italy

³Faculty Member, University Bicocca Academy of Milan, Milan - Italy

⁴SIEMG-certified physician in clinical ultrasound, Italian Society of Ultrasound in General Practice and Clinical Ultrasound, Parma - Italy

⁵Member of the National Commission for Digital Innovation and Technology of AME with delegation for AI, Udine - Italy

⁶Council member of the Medical Order of Novara, Novara - Italy

⁷General Practitioner, Novara - Italy

ABSTRACT

Recent reports on the unavailability of OpenEvidence in the European Union and the United Kingdom have renewed debate on artificial intelligence, medicine, and regulation. This case has often been interpreted as further evidence of Europe's structural difficulty in supporting technological innovation. While this reading identifies a real problem, it is insufficient. The central question is not only how rapidly AI is adopted, but how safely, critically, and responsibly it is integrated into clinical practice.

In the global regulatory landscape, the United States operates within a market-oriented ecosystem largely shaped by medical device regulation; China pursues a more centralized and state-driven strategy; and Europe relies on a risk-based regulatory framework. Yet healthcare AI cannot be assessed only through availability, competitiveness, or formal compliance. Its value depends on clinical validity, epistemic transparency, and compatibility with physicians' professional duties.

Clinically oriented AI systems, especially those based on curated sources and professional interfaces, may appear more reliable than general-purpose models. However, perceived reliability is not equivalent to clinical validity. Even retrieval-augmented systems can generate inaccurate citations, weak inferences, misleading syntheses, or recommendations expressed with greater certainty than the evidence permits. The distinctive danger is therefore not simply error, but error presented in the language of evidence.

This paper argues against both uncritical adoption and reflexive prohibition. Safe medical AI requires technical robustness, proportionate regulation, institutional governance, independent validation, and specific physician training in critical use. The guiding principle remains deontological: artificial intelligence may support clinical reasoning, but it cannot assume professional responsibility.

Keywords: Artificial intelligence in medicine, Medical AI, Generative AI, Large Language Models (LLM), OpenEvidence

Introduction

Generative artificial intelligence is rapidly entering healthcare processes, including literature retrieval, guideline synthesis, diagnostic hypothesis generation, clinical documentation support, informational triage, patient education, and comparison of therapeutic options.

Public debate, however, tends to polarize. On one side, AI is portrayed as an inevitable accelerator of efficiency and precision.

On the other hand, it is considered a threat to be contained through restrictions, prohibitions, and regulatory burdens.

Both interpretations are insufficient.

Medicine can afford neither naive enthusiasm nor defensive immobilism. Artificial intelligence must instead be governed as a high-impact cognitive technology, capable not only of augmenting physicians' capabilities but also of modifying decision-making posture, verification timing, and thresholds of trust toward generated outputs (1).

The OpenEvidence case is particularly useful in this regard. It does not merely represent an issue of European accessibility. Rather, it functions as a symptomatic case, revealing how a tool perceived as more "clinical" and more reliable may generate confusion between innovation, safety, evidence, and responsibility.

As of May 10, 2026, converging reports, primarily secondary sources, indicate that OpenEvidence is unavailable in the

Received: May 12, 2026

Accepted: May 21, 2026

Published online: June 23, 2026

Corresponding author:

Fabio Di Bello

email: fdibello@wiley.com



European Union and the United Kingdom and that this decision may be linked to regulatory uncertainty. In the absence of complete and publicly verifiable official documentation, the case should be treated as an indicator of regulatory and commercial friction rather than as definitive proof of a direct causal relationship with the AI Act.

The global context: three speeds of medical AI

United States: acceleration, market dynamics, and adaptive regulation

In the United States, healthcare AI evolves within a high-capital ecosystem characterized by strong industrial presence, rapid experimentation, and increasing regulatory intervention. The U.S. Food and Drug Administration publishes a list of AI-enabled medical devices authorized for commercialization; however, this list specifically concerns regulated medical devices and does not encompass the broader range of AI tools used in healthcare (2).

The American model appears pragmatic: it promotes innovation while attempting to adapt regulatory oversight to the evolving nature of software, predictive models, and AI-enabled devices. The associated risk is that market speed may outpace the maturation of clinical usage culture and independent evaluation mechanisms (2,3).

Europe: a risk-based approach

The European Union has adopted a different path. Regulation (EU) 2024/1689, commonly known as the AI Act, establishes a horizontal regulatory framework based on risk classification, with specific requirements for systems capable of affecting health, safety, and fundamental rights (4).

The value of this approach is evident: preventing opaque or insufficiently validated technologies from entering sensitive domains without safeguards. Its limitation is equally evident: when regulation becomes operational uncertainty, it may induce withdrawal, delay, or restricted access to potentially useful tools.

China: state strategy and generative AI control

China has developed a model characterized by strong state direction, industrial promotion, and regulatory control over generative services. Since 2023, the Chinese regulatory framework has included specific measures governing public-facing generative AI services, balancing innovation promotion with informational control (5).

This model aims at technological competitiveness within a politically and socially controlled environment. In healthcare, such an arrangement may facilitate rapid and coordinated implementation, while simultaneously raising concerns regarding transparency, professional autonomy, data access, and independence of clinical judgment.

WHO: global ethical governance

In 2024, the World Health Organization published guidance on the ethics and governance of large multimodal models (LMMs) in healthcare, emphasizing governance, accountability, safety, equity, and human oversight (1,6).

This level of analysis is essential because it reminds us that healthcare AI is not merely a matter of markets or regulation. It is fundamentally a matter of public health, trust, justice, and responsibility.

Figure 1 provides a schematic summary of the three main medical AI governance models discussed in this section.

The OpenEvidence case as a symptom, not an exception

OpenEvidence belongs to an emerging category of clinically oriented AI assistants: systems that do not present themselves as general-purpose chatbots, but rather as professional environments designed to provide clinically grounded responses anchored to curated sources. Publicly, the platform describes itself as a tool for healthcare professionals whose responses are based on peer-reviewed literature and specialized medical sources.

This architecture has operational value. A system designed specifically for medicine, powered by biomedical sources and targeted toward healthcare professionals, may reduce certain risks typically associated with general-purpose models. However, this does not automatically imply independent clinical validation or regulatory classification as a medical device.

The critical issue lies elsewhere: the medicalization of the interface may generate trust exceeding the system's actual reliability.

Physicians tend to remain cautious when interacting with general-purpose AI models. By contrast, vigilance thresholds may decrease when confronted with systems speaking the language of scientific literature, citing references, organizing outputs in clinical formats, and explicitly presenting themselves as professional tools. For precisely this reason, the risk of automation bias may increase. This psychological distinction is decisive. The danger is not merely that AI may be wrong. The danger is that it may be wrong in a credible way (7-9).

Perceived safety and real safety

Within medically oriented AI systems, two distinct levels must be differentiated.

Perceived safety

Perceived safety derives from several factors:

- closed or semi-closed environments;
- curated information sources;
- professional interfaces;
- medical terminology;
- bibliographic citations;
- well-structured outputs;
- declared clinical purpose.

These characteristics contribute to an appearance of reliability and scientific rigor that may psychologically reinforce user trust.

Three Models of Medical AI

The USA, Europe, and China are moving in different directions.
The key issue is not the tool, but clinical responsibility.

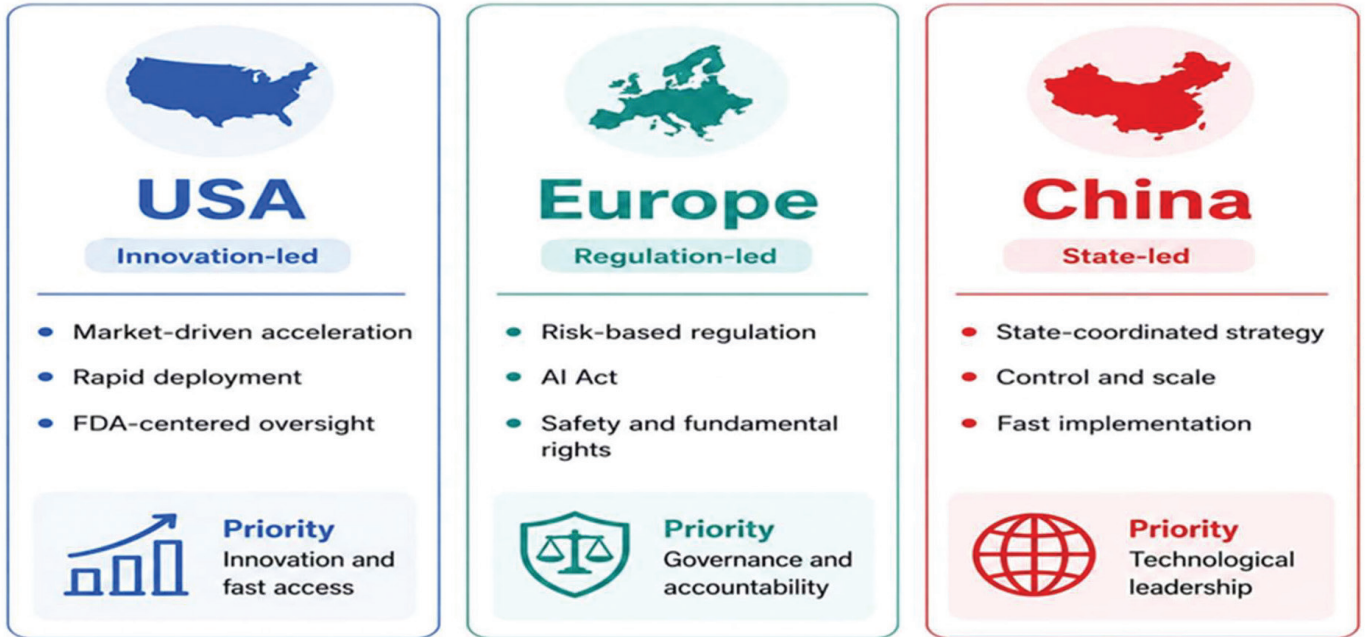


FIGURE 1 - Three models of medical AI governance: market-oriented and device regulation in the United States, a risk-based regulatory approach in Europe, and state strategy and information control in China.

Real safety

Real safety, however, requires substantially different conditions:

- verifiable accuracy;
- traceability of sources;
- correspondence between cited sources and generated claims;
- explicit management of uncertainty;
- continuous updating;
- independent validation;
- auditability;
- systematic surveillance of errors;
- evaluation of clinical impact;
- adequate user training.

The improper overlap between these two dimensions represents one of the greatest risks in contemporary medical AI.

A system may appear clinically reliable because it adopts the form of evidence. Yet the form of evidence does not coincide with the substance of evidence.

Error dressed as evidence

Generative artificial intelligence introduces a specific epistemological problem: it does not merely retrieve information, but re-elaborates it into linguistically persuasive outputs.

This creates a new category of risk: plausible error.

In educational experiences involving AI in medicine, stress-testing models represent one of the most instructive exercises. The objective is not to discredit the technology, but to identify its failure zones and distinguish apparent reliability from documentary robustness and genuine clinical usability. We can summarize the taxonomy of AI errors in Table 1.

TABLE 1 - Taxonomy of AI errors and associated clinical risks

Error Type	Description	Clinical Risk
Non-existent citation	Plausible but fabricated reference	False scientific validation
Real citation used incorrectly	The existing source is inconsistent with the generated conclusion	Distortion of evidence
Excessive inference	Unjustified transition from data to recommendation	Inappropriate decision-making
Overconfidence	Excessively certain response formulation	Reduced vigilance
Omission of uncertainty	Failure to expose alternatives or limitations	False clinical simplicity
Insufficient contextualization	Abstractly correct indication unsuitable for the patient	Clinical inappropriateness



The most dangerous problem is not the gross error. Gross errors are often recognized. The most dangerous problem is the error that preserves the tone, structure, and vocabulary of scientific authority.

This constitutes the conceptual core of the present paper: when error speaks the language of evidence, clinical risk increases because critical suspicion decreases.

This issue becomes even more evident in retrieval-augmented systems, which are often perceived as safer because they are connected to external sources. Yet the presence of a database, uploaded documents, or citations does not eliminate the possibility of error. Rather, it shifts the problem to a subtler level: the issue is no longer merely “the model invents,” but rather “the model retrieves, interprets, compresses, and sometimes attributes to the source claims that the source itself does not support” (10).

The limits of RAG systems: retrieving sources does not guarantee truth

A substantial portion of clinically oriented AI systems relies on Retrieval-Augmented Generation architectures, commonly referred to as RAG systems. These architectures combine two distinct operations: first, they retrieve documents from archives, databases, websites, PDFs, guidelines, technical sheets, or user-uploaded materials; second, they generate a linguistic response through a large language model.

This distinction is fundamental. A well-designed RAG system may reduce a portion of hallucinations by grounding outputs in external sources. However, it does not eliminate the problem. Technical literature has documented that, even within RAG settings, models may generate unsupported or contradictory statements relative to the retrieved content. RAGTruth, for example, was specifically developed to analyze hallucinations in retrieval-augmented frameworks, precisely because retrieval integration does not remove the problem of unsupported claims (10). The reason is structural. Document retrieval is not equivalent to clinical understanding.

The model does not read as a legally and deontologically accountable professional would. It integrates, compresses, infers, organizes, and completes information. In some cases, it completes it incorrectly.

The weak point is therefore not merely the retrieval of the source. It is the generative step transforming the source into an answer.

Why a RAG system may hallucinate

A RAG system may fail at least at two levels.

The first concerns retrieval itself: the system may retrieve the wrong source, an outdated document, incomplete information, semantically similar but clinically irrelevant material, or only fragments of a larger document while missing the broader contextual framework.

The second concerns generation: the model may incorrectly reinterpret the retrieved material, superimpose information originating from its internal parametric knowledge, infer claims unsupported by the source, or present conclusions as if they were explicitly documented.

TABLE 2 - Summary of the key messages for clinical prudence

Error Type	Description	Clinical-Practical Example
Retrieval error	Retrieval of incorrect, outdated, or only apparently relevant sources	Retrieval of obsolete guidelines or recommendations derived from different populations
Defective chunking	Response generated from incomplete document fragments	Reporting dosage while omitting contraindications contained in the subsequent paragraph
False synthesis	Source states A, model summarizes B	Conditional recommendations transformed into strong recommendations
Over-inference	Explicit conclusions inferred from implicit statements	“May be considered” becomes “is recommended”
Source conflict	Discordant sources merged without evidence hierarchy management	Combination of package inserts, narrative reviews, and conference abstracts
Parametric override	Internal model knowledge overrides the retrieved source	Use of outdated training information is inconsistent with updated drug labeling
Citation laundering	The existing citation fails to support the generated claim	The cited paragraph discusses the drug, but not the claimed indication

In other words, a RAG system is not a deterministic documentary verification engine. It is a probabilistic system generating text from retrieved documents.

This difference is substantial.

The promise of RAG is not infallibility. The promise is risk reduction through documentary grounding. Yet reducing risk does not mean eliminating it, nor does it guarantee semantic fidelity or clinical correctness, as you can see in Table 2 below (10).

Categories of error in RAG systems

The last category is particularly relevant in medicine. The mere existence of a citation does not guarantee that the citation genuinely supports the generated statement. This represents the most insidious level of error: not fabricated citations, but authentic citations improperly used as epistemic cover.

Figure 2 summarizes the primary failure points of RAG systems.

Systems based on user-uploaded sources

Tools such as NotebookLM clearly illustrate this ambivalence. These systems are designed to operate on sources

RAG: key messages

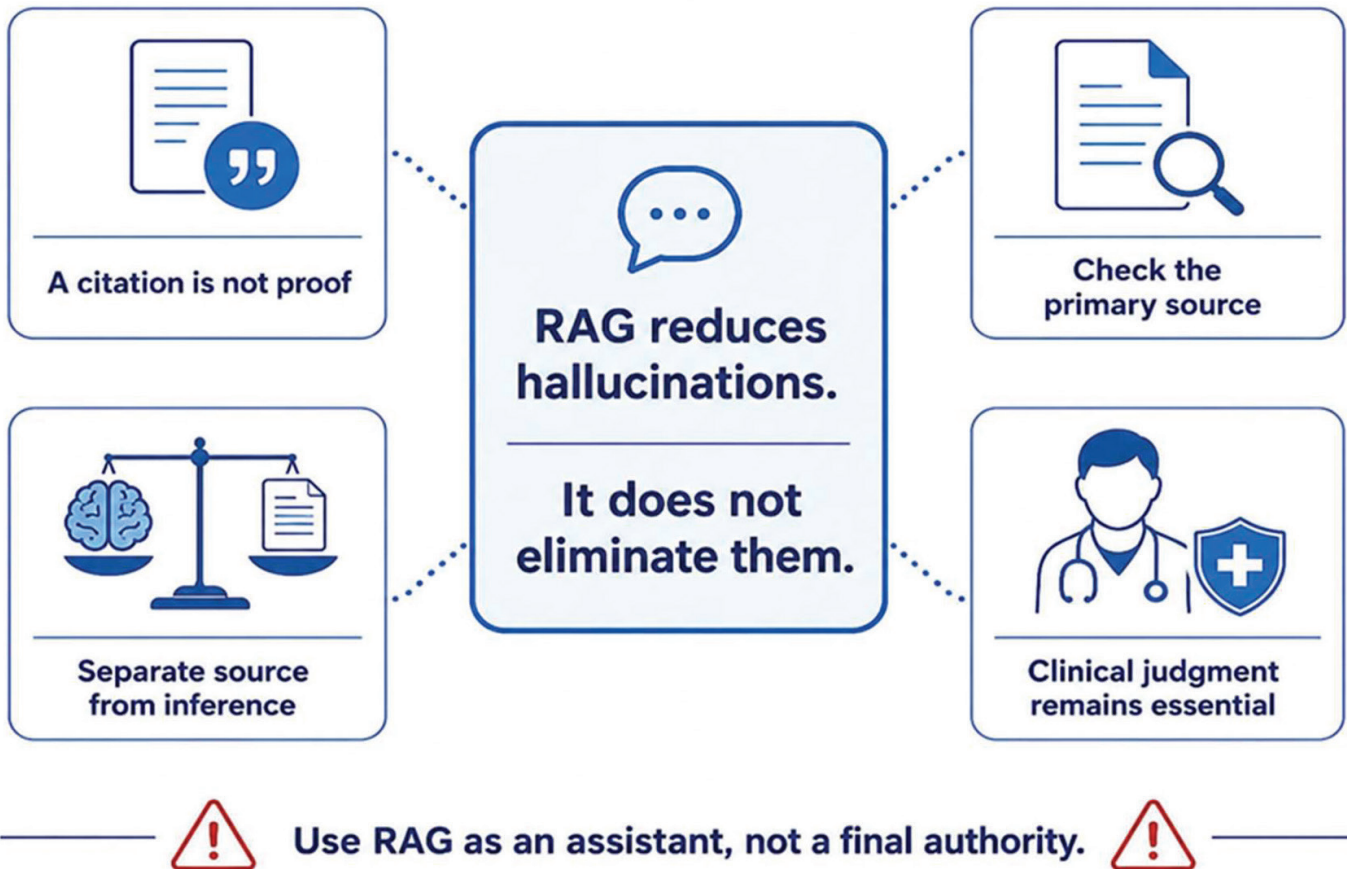


FIGURE 2 - Key principles for the clinically prudent use of RAG systems: citations do not automatically constitute proof, verification of the primary source remains essential, and clinical judgment cannot be delegated.

uploaded or imported by the user, often as synchronized static copies, potentially increasing traceability and control compared with general-purpose models (11).

However, working “on sources” does not automatically transform the system into an infallible documentary authority.

The statement “according to the sources” does not guarantee that the system:

- retrieved the correct source;
- considered the entire document;
- correctly interpreted exceptions, limitations, or negations;
- respected the hierarchy between drug labels, guidelines, clinical trials, reviews, and abstracts;
- avoided conflating similar drugs, populations, or formulations;
- refrained from introducing unsupported inferences.

These observations do not diminish the value of RAG systems. Rather, they place them within their proper epistemological role: tools for retrieval, synthesis, and cognitive

assistance, not automatic guarantors of clinical truth or substitutes for direct verification of primary sources.

The specific risk in pharmacological contexts

Pharmacology represents one of the highest-risk domains for the inappropriate use of RAG systems.

The reasons are both technical and clinical.

Drug labels and prescribing information possess structural characteristics that increase system vulnerability:

- similar names among active ingredients, brand names, formulations, salts, and dosages;
- different indications depending on pharmaceutical formulation;
- variability across adult, pediatric, geriatric, pregnancy, renal impairment, and hepatic impairment populations;
- periodic updates of prescribing information;
- duplicate or obsolete sources;
- tables that are difficult to parse correctly;

TABLE 3 - Systematic verification steps for RAG systems in healthcare

Verification Step	Critical Question
Primary source	Does the response derive from official prescribing information, regulatory agencies, or primary clinical guidelines?
Precise citation	Does the citation correspond exactly to the paragraph supporting the generated claim?
Version and date	Is the source updated?
Correct drug identification	Do active ingredient, formulation, dosage, and patient population match?
Absence of unsupported inference	Does the system distinguish between what is explicitly stated and what is inferred?
Management of contradictions	Were discordant sources identified or improperly merged?
Negations and restrictions	Were contraindications, age limits, pregnancy warnings, or renal/hepatic restrictions correctly preserved?
Evidence hierarchy	Did the system prioritize prescribing information and guidelines over narrative reviews or secondary sources?

- contraindications and precautions distributed across multiple sections;
- clinically critical negations such as “contraindicated,” “not recommended,” or “must not be used.”

Negations constitute a particularly critical issue. A model may correctly process an affirmative sentence while failing to interpret a restriction accurately. In pharmacology, however, the restriction is often the clinically most important component. An apparently correct dosage recommendation may become dangerous if it omits a contraindication, an excluded population, a distinct formulation, or a required dose adjustment in renal impairment. For this reason, in pharmacological settings, RAG systems should be treated as retrieval and synthesis assistants rather than final authorities.

Operational formulation

The appropriate formulation is the following:

A well-designed RAG system may reduce hallucination rates by grounding responses in external sources, but it cannot guarantee the absence of hallucinations, clinical correctness, or complete documentary fidelity (10).

This formulation should become a minimum literacy criterion for every physician using AI systems based on external sources.

Practical verification criteria for clinical and pharmacological use

A RAG-generated response in clinical or pharmacological settings should be considered usable only as a preliminary support tool and only if it passes a structured verification process (Table 3 below).

The operational principle is unequivocal: the source must be verified in its primary location, not merely in the form summarized by the model.

The false alternative: innovation versus regulation

The OpenEvidence case has often been interpreted as proof of an opposition between technological innovation and European regulation.

This opposition is misleading.

The issue is not whether one should choose unrestricted access or prohibition. The issue is whether a model of clinical AI usage can be constructed in which:

- the tool is assessable;
- physicians are adequately trained;
- patients are protected;
- responsibility remains clearly identifiable;
- errors are traceable;
- outputs are verifiable;
- uncertainty is explicitly communicated.

Europe certainly risks slowing AI adoption if regulation is perceived as an administrative labyrinth. The United States risks the opposite problem: rapid deployment of tools without the simultaneous maturation of the critical literacy necessary for their safe use. China, in turn, demonstrates that speed may be achieved through centralized coordination and control, though not necessarily through greater professional autonomy.

The central question is therefore not which system moves faster.

The real question is which system builds the safest, most mature, and most clinically responsible model.

Medical deontology as a higher level of governance

Regulation is necessary, but insufficient.

The governance of medical AI cannot rely solely on regulations, certifications, corporate policies, or terms of service. It must rest upon a deeper principle: the deontological responsibility of physicians and their obligation to verify the tools they use.

Physicians may use AI systems to:

- retrieve literature;
- synthesize guidelines;
- generate hypotheses;
- compare therapeutic options;
- improve communication;
- reduce documentation burden;
- identify informational gaps.

However, physicians cannot delegate:

- clinical judgment;
- patient evaluation;



- therapeutic decision-making;
- communication of uncertainty;
- moral responsibility;
- professional accountability.

Within the Italian deontological framework, the physician remains the subject responsible for care. Artificial intelligence may enter the process, but it cannot become its owner.

The correct formulation is therefore: AI as cognitive support, not as a surrogate for clinical responsibility.

Physicians in the global AI competition

Global competition in AI is frequently described as a race among states, corporations, computational infrastructures, and regulatory systems. This interpretation is true, but incomplete.

In healthcare, the real competition does not concern only who develops the most powerful model. It concerns who trains professionals capable of using such systems responsibly and effectively. The competitive advantage of a healthcare system will not depend solely on access to AI technologies, but on the quality of their clinical integration. A healthcare system equipped with advanced AI tools but lacking adequately trained physicians generates risk.

A healthcare system that regulates extensively but trains inadequately generates paralysis. A healthcare system that prohibits technologies out of excessive caution generates backwardness. A healthcare system that adopts technologies uncritically out of enthusiasm generates exposure.

The point of equilibrium lies elsewhere: critical training, proportionate governance, and professional responsibility.

Recommendations

For physicians

- Use AI as a support tool, not as an authority.
- Verify primary sources whenever outputs may influence clinical decisions.
- Be cautious of excessively orderly or overly confident answers in uncertain areas.
- Always ask the system to explicitly state limitations, alternatives, and levels of evidence.
- Document AI usage whenever clinically relevant.
- Maintain full responsibility for clinical reasoning.
- In pharmacological contexts, directly verify prescribing information, regulatory documents, and primary guideline sources.

For educational institutions

- Integrate AI literacy into continuing medical education.
- Teach not only how to use AI systems, but also how to stress-test them.
- Train physicians to recognize hallucinations, overconfidence, citation laundering, and automation bias (7-10).
- Include modules on deontology, responsibility, and communication of uncertainty.

- Clearly distinguish between general-purpose AI, clinically oriented AI, RAG systems, and AI-enabled medical devices.

For healthcare institutions

- Develop differentiated usage policies according to context.
- Distinguish administrative, informational, cognitive, and decision-support applications.
- Implement periodic audits.
- Monitor incidents, near misses, and patterns of overreliance.
- Evaluate systems not only for technical accuracy, but also for their impact on clinical behavior (9).
- Define specific procedures for medications, dosages, contraindications, and vulnerable populations.

For developers and providers

- Clearly define the intended scope of use.
- Explicitly disclose sources and update policies.
- Signal uncertainty and conflicts among evidence sources.
- Avoid interfaces that induce false authority.
- Allow independent evaluation.
- Clearly separate informational synthesis from clinical recommendation.
- In RAG systems, distinguish retrieved sources, effectively used content, and generated inferences (10).

For regulators

- Avoid both deregulation and prohibition.
- Clarify obligations according to the actual use of the system.
- Promote healthcare-oriented regulatory sandboxes.
- Encourage independent clinical validation.
- Integrate technical regulation with professional responsibility.
- Protect patients without de-responsibilizing physicians, while simultaneously avoiding attribution of unwarranted authority to AI systems.

Conclusions

The OpenEvidence case does not simply demonstrate that Europe obstructs innovation. Rather, it shows that medical AI has entered a phase in which the distinction among information, evidence, recommendation, and clinical decision-making is becoming increasingly fragile.

Within the global landscape, the United States accelerates, China coordinates and controls, and Europe regulates. Yet none of these models is sufficient if the decisive level is missing: the physician's critical competence.

RAG systems further clarify the problem: connecting a model to external sources does not automatically transform its outputs into evidence. A citation is not a guarantee. A retrieved source is not necessarily a correctly interpreted



RAG reduces hallucinations but does not eliminate them

Why retrieval helps — and why generation can still go wrong

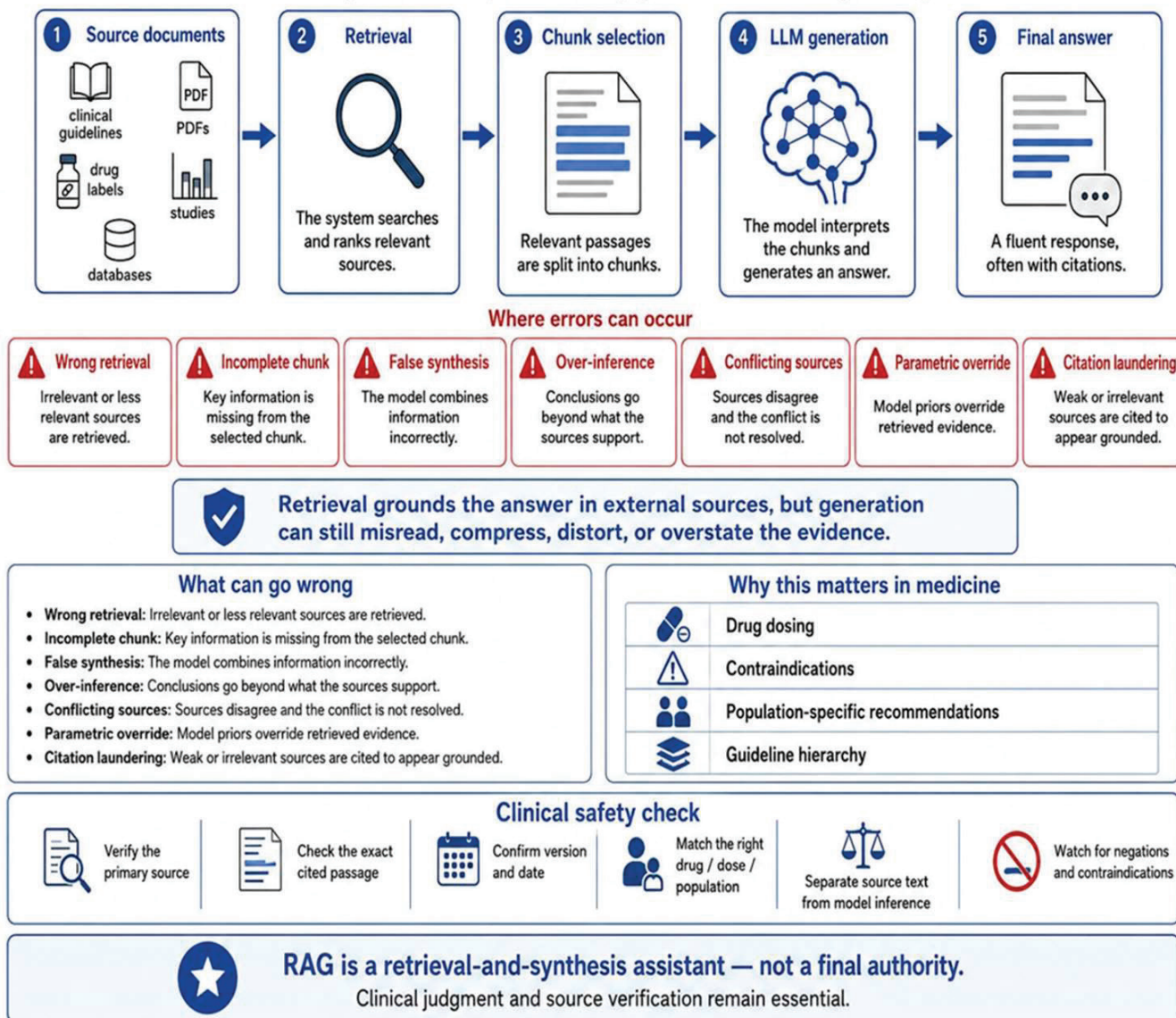


FIGURE 3 - Main failure points in RAG systems: retrieval may recover insufficient or partially relevant documents and fragments, while generation may introduce false syntheses, over-inference, or improper citation usage.

source. A grounded response is not automatically a clinically valid response (10).

Safety in medical AI does not arise from a single factor. It does not arise solely from model power. It does not arise solely from source quality. It does not arise solely from certification nor solely from regulation. It arises from the combination of technical reliability, independent verification, critical literacy and professional responsibility.

It emerges from the interaction among trustworthy technology, proportionate regulation, healthcare governance, and deontological accountability.

The greatest risk is not using AI. The greatest risk is using it without critical competence or prohibiting it without understanding it.

The appropriate future is not a medicine without artificial intelligence. It is a medicine in which AI remains a tool, evidence remains verifiable, the patient remains at the center, and responsibility remains with the physician.

Figure 3 schematically illustrates the principal points at which a RAG system may fail, even when operating on apparently relevant sources.



Disclosures

Conflict of Interest: The authors declare no conflicts of interest related to this manuscript.

Financial Support: This work was conducted independently and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement: Data sharing is not applicable to this article because it is a viewpoint manuscript, and no original datasets, patient-level data, surveys, interviews, or statistical analyses were generated or analyzed.

References

1. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multimodal models (LMMs). WHO; 2024. [Online](#) (Accessed May 2026)
2. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. FDA website. [Online](#) (Accessed May 2026)
3. U.S. Food and Drug Administration, Health Canada, MHRA. Good machine learning practice for medical device development: Guiding principles. 2021. [Online](#) (Accessed May 2026)
4. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024. [Online](#) (Accessed May 2026)
5. Cyberspace Administration of China, National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, National Radio and Television Administration. Interim Measures for the Management of Generative AI Services. 2023. [Online](#) (Accessed May 2026)
6. World Health Organization. WHO releases AI ethics and governance guidance for large multimodal models. 2024. [Online](#) (Accessed May 2026)
7. Abdelwanis M, Alarafati HK, Tammam MMS, et al. Exploring the risks of automation bias in healthcare artificial intelligence applications: a bowtie analysis. *J Saf Sci Resil.* 2024;5(4):460-469. [CrossRef](#)
8. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. *JAMA.* 2023;330(23):2255–2257. [CrossRef](#)
9. Goh E, et al. Physician clinical decision modification and bias assessment with AI assistance in chest pain triage. 2025. [Online](#) (Accessed May 2026)
10. Niu C, Wu Y, Zhu J, et al. RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024). Bangkok, Thailand; 2024. [CrossRef](#)
11. Google NotebookLM Help Center. [Online](#) (Accessed May 2026)