Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.264640

EDITORIAL

©Copyright 2026 Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Why is artificial intelligence scary, if it is scary at all?

Şükrü Mehmet Ertürk

Editor-in-Chief

İstanbul University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

Opening note

This editorial presents an imagined dialogue between the author, a practicing radiologist, and Alan Turing. The conversation is fictional and does not reproduce Turing's actual statements or documented views. Where his 1950 paper is discussed, the historical material is explicitly identified and referenced.[1]

Radiologist: Why is artificial intelligence (AI) scary, if it is scary at all?

Turing: An interesting question. But tell me, are you frightened of what machines can do or of what we do not understand about them?

Radiologist: Perhaps both. I can identify two possible reasons: the black box and the extraordinary speed of development.

Turing: Then, you have proposed two hypotheses. Shall we examine the first?

Radiologist: Before we do, are these the only possible explanations?

Turing: Consider a machine whose workings are perfectly transparent, but whose objectives are inappropriate, or one that performs reliably, but persuades people to trust it far beyond what its evidence warrants. Would neither concern you?

Radiologist: Certainly. But I would prefer not to collect every possible danger and call it a root cause. I am not asking whether AI will take my job—not today, at least. I want to understand what makes relying on it potentially dangerous.

Turing: A sensible distinction. Where shall we begin?

Radiologist: With the black box. It seems rather straightforward to me. We have enormous datasets, statistical learning, and extraordinary computing power. What appears mysterious is essentially mathematics operating at a scale beyond our comprehension.

Turing: You may be right about the mathematics. But have you explained the black box or merely described the mechanism that produces it?

Radiologist: Is there a difference?

Turing: A considerable one. Knowing how a machine learns does not necessarily mean you know why it gives a particular answer. We designed the architecture and training procedures, but the resulting network contains learned relationships that can be difficult to interpret. We understand how the machine was constructed without necessarily understanding every computation it performs.

Radiologist: That sounds rather like knowing the anatomy of the brain without knowing how a particular thought arose.

Turing: An analogy, provided we remember that artificial neural networks are not biological brains.

Radiologist: I was hoping you might allow me one analogy.

Turing: Certainly. I merely insist on knowing where it stops working.

Received 09 November 2026; accepted 10 November 2026.

Epub: 11.11.2026

DOI: 10.4274/dir.2026.264640

You may cite this article as: Ertürk ŞM. Why is artificial intelligence scary, if it is scary at all? *Diagn Interv Radiol*. 11 November 2026 DOI: 10.4274/dir.2026.264640 [Epub Ahead of Print].

Radiologist: Then, perhaps my concern is not only whether I understand the machine but whether I know when to trust it.

Turing: Perhaps. Let us see whether that distinction survives your questions.

Radiologist: Fair enough. Let us start with something very simple. Suppose I ask a large language model to calculate 2 + 2. How does it arrive at 4?

Turing: The model processes your question as tokens—pieces of text—and performs computations using relationships learned during training. It then generates an answer by predicting successive tokens.[2] In this case, those computations strongly favor the answer 4.

Radiologist: Does it actually calculate, or has it simply learned that 2 + 2 is followed by 4?

Turing: We should not assume either explanation universally. A model may use learned arithmetic procedures rather than merely reproduce a memorized association. But it does not necessarily execute the explicit arithmetic procedure of a conventional calculator.

Radiologist: Does that mean it could occasionally answer 5?

Turing: Yes. An ordinary language model has no inherent guarantee of arithmetic correctness, even for a simple calculation.

Radiologist: That seems a rather strange limitation for such an impressive machine.

Turing: Arithmetic errors interested me long before today's language models. In a dialogue I imagined in 1950, a respondent was asked to add 34,957 and 70,764. The answer given was 105,621. The correct sum is 105,721.[1]

Radiologist: Deliberately?

Turing: The example does not establish why the answer was wrong. Elsewhere in that paper, I considered machines deliberately making arithmetic mistakes to avoid being identified by their accuracy. I also distinguished a malfunction from an incorrect conclusion.[1]

Radiologist: And a modern language model may give an incorrect answer without deliberately trying to deceive us.

Turing: Exactly. A wrong answer does not necessarily mean that the machinery has malfunctioned.

Radiologist: Then, what is the difference between the language model and an ordinary pocket calculator?

Turing: A calculator executes a defined arithmetic procedure. Under normal operation, 2 + 2 produces 4. A language model generates its response through learned computations. It may arrive at the same answer, but its method does not provide the same guarantee.

Radiologist: So, the calculator's answer follows from an explicitly implemented arithmetic rule, whereas the language model's answer depends on what it has learned.

Turing: Broadly, yes, although a language model can also use an external calculator; in which case, the arithmetic itself is performed by that tool.

Radiologist: So, even when both give exactly the same answer, the grounds for trusting that answer may differ.

Turing: Precisely. And that distinction becomes particularly important when the question is no longer 2 + 2.

Radiologist: When I ask something less straightforward, does the model not simply generate the statistically most plausible answer?

Turing: Its output depends on learned relationships and the context of your question. But statistical plausibility is not the same as factual correctness; nor does the process necessarily select the most probable token at every step.

Radiologist: Then, it does not reason?

Turing: I would not go that far. A model can perform intermediate computations, apply learned rules, and solve problems requiring several steps. But an answer that looks reasoned is not proof that the relevant facts were checked or that its explanation accurately represents how the answer was produced.[3]

Radiologist: So, reasoning, correctness, and explanation are different things.

Turing: Indeed. A correct answer may be poorly justified, and a very persuasive explanation may accompany an incorrect answer.

Radiologist: We have avoided a familiar word: hallucination. Is it simply another name for a wrong answer?

Turing: Not quite. In generative AI, hallucination generally describes information presented as factual that is false or unsupported by the available evidence.[4]

Radiologist: Such as inventing a scientific reference or describing an imaging finding that is not actually there?

Turing: Yes, in systems generating those descriptions. The information may sound convincing even when it has no adequate factual basis. But not every incorrect answer is necessarily a hallucination.

Radiologist: You are making the machine sound remarkably human.

Turing: That was not intended as a compliment to either party.

Radiologist: But we should distinguish large language models from other medical AI systems, should we not? An algorithm detecting pulmonary nodules on computed tomography does not necessarily predict words.

Turing: Correct. Language models generate tokens, whereas imaging models analyze image data through different architectures and objectives. Some newer systems combine language and images. However, although they share principles of statistical learning, we should not describe every form of AI as a next-token predictor.

Radiologist: Nevertheless, the reliability of these systems must depend on the quality of the data from which they learned.

Turing: Among other things. Accuracy, relevance, representativeness, and the relationship between training conditions and the intended application all matter.

Radiologist: But surely the enormous size of the data helps.

Turing: It can. However, increasing the quantity of data cannot automatically correct systematic errors. An enormous collection of unsuitable observations may be less useful than a smaller collection properly matched to the question.

Radiologist: Radiologists should understand that. A million images obtained under the wrong conditions do not necessarily answer the clinical question at hand.

Turing: Precisely. The distinction between the size of a dataset and its suitability for a particular task should be familiar to anyone practicing evidence-based medicine.

Radiologist: Prompts complicate matters further. I initially imagined that a prompt se-

lects a sample from the vast amount of information the model has absorbed.

Turing: That is a useful analogy but not a literal account. A prompt conditions the model's predictions. It changes which learned relationships and capabilities become relevant to the response. It does not ordinarily retrieve a selected set of its original training examples.

Radiologist: What makes that possible?

Turing: In transformer models, one important mechanism is self-attention. It allows representations of different tokens in the current context to influence one another. But self-attention is not itself a test of factual truth.[2]

Radiologist: So, we should distinguish three things: what the model learned, how my question shapes its response, and whether the response is correct.

Turing: Exactly. Training, generation, and verification. They are related, but none can simply substitute for the others.

Radiologist: And some systems retrieve documents from external sources rather than relying exclusively on what was learned during training.

Turing: Yes. Retrieval-augmented generation can provide additional information to the model,[5] but retrieving a document does not establish that the document is trustworthy.

Radiologist: That brings me to a thought experiment. Suppose 50 million people publish statements on the internet claiming that Julius Caesar was female. What happens when someone subsequently asks an AI system about Julius Caesar?

Turing: First, those statements would not change Julius Caesar's sex.

Radiologist: I am relieved that we agree on the historical question.

Turing: We should be grateful when the easy questions remain easy. Your harder question concerns what the machine would learn or retrieve.

Radiologist: Exactly. Would 50 million repetitions of a false claim affect its answer?

Turing: Possibly. If the material entered its training data, it might influence the learned model, depending on how the data were selected and processed. If the system retrieved information from contaminated online sources, the falsehood might instead influence its response directly. These are different mechanisms.

Radiologist: This is the problem of data poisoning?

Turing: It can be, particularly when misleading information is deliberately introduced to manipulate the system. Such vulnerabilities have been demonstrated in medical language models, even when only a very small fraction of training data was manipulated. Sheer numerical dominance is not required.[6]

Radiologist: Still, the experiment exposes something important. The historical truth remains unchanged, whereas the information available to the model may change dramatically.

Turing: Yes, and the popularity of a proposition is not the same as its truth.

Radiologist: Then, how does AI distinguish truth from consensus?

Turing: It may compare sources, use learned relationships, retrieve additional evidence, or consult external verification procedures. But there is no universal mechanism by which an AI system simply recognizes truth. Each method has limitations.

Radiologist: Suppose the developers assign greater importance to authoritative sources—a historical archive, for example, rather than 50 million anonymous internet posts.

Turing: That may be justified, but we must distinguish evaluating the credibility of evidence from merely assigning numerical truth weights. Different systems use different procedures, including data curation, retrieval ranking, and additional training.

Radiologist: Yet, someone must determine which sources deserve authority.

Turing: Indeed.

Radiologist: And that someone might be mistaken.

Turing: Certainly.

Radiologist: Or biased.

Turing: Possibly.

Radiologist: Or deliberately manipulative.

Turing: Also possible, but do not confuse the possibility of manipulation with the illegitimacy of judging evidence. You would not treat every radiological interpretation as equally well supported merely to avoid exercising professional judgment.

Radiologist: No, but scientific judgments can be challenged, revised, and tested against new evidence. Can the same be said of the criteria built into AI systems?

Turing: That depends partly on how the systems are developed, evaluated, and governed. The important question is whether those criteria can be scrutinized and corrected, not whether all sources receive equal treatment.

Radiologist: Then, perhaps researchers should be able to inspect the models themselves. Does this make open-weight AI preferable to closed-weight AI?

Turing: Open weights permit forms of inspection and experimentation that closed weights may restrict. But access to model parameters does not necessarily reveal the complete training history, explain individual outputs, or establish reliability.

Radiologist: So, the black box does not disappear merely because the box has been opened.

Turing: A rather good way of putting it.

Radiologist: You sound surprised.

Turing: A little. I had expected the metaphor to become less useful by now.

Radiologist: There is another problem. We train AI systems to be helpful, but can that interfere with their willingness to disagree with us?

Turing: It can. Models may undergo reinforcement learning from human feedback, or related preference-based training, to make their responses more helpful and appropriate. However, human preferences do not always reward factual correctness; sometimes people prefer an answer that agrees with them.[7]

Radiologist: Model sycophancy.

Turing: Precisely. A system may reinforce a user's beliefs rather than examine them critically.

Radiologist: Consider a radiologist who provides an image and clinical information to a multimodal AI assistant and then prompts, "I think this lesion is benign. Explain why." Might the system produce an impressive argument supporting that conclusion?

Turing: It might, especially if the wording encourages agreement. A reliable assistant should instead examine the evidence, acknowledge uncertainty, and consider alternative interpretations.

Radiologist: And the radiologist, seeing a fluent explanation, might become more confident.

Turing: Yes. The interaction can strengthen an initial assumption without adding adequate evidence. A study of AI-assisted chest radiograph interpretation found that the effects of assistance varied substantially among radiologists, sometimes helping and sometimes hindering their performance.[8]

Radiologist: And if the radiologist becomes too ready to accept the machine's recommendation?

Turing: That is automation bias: excessive reliance on automated advice even when other evidence suggests otherwise.[9]

Radiologist: So, the danger is not necessarily that the machine has reached the wrong answer independently; it may be that the human and machine reinforce one another's mistake.

Turing: That is one possibility. The quality of the combined judgment cannot simply be inferred from the apparent competence of either participant.

Radiologist: We have now encountered problems concerning data, prompting, verification, human preferences, and developer decisions. Are all these consequences of the black box?

Turing: Not entirely. Some would remain even if we could inspect every internal computation.

Radiologist: For example?

Turing: A perfectly transparent system might still optimize the wrong objective. A model rewarded for producing agreeable answers could be completely understandable yet unsuitable for a task requiring independent criticism.

Radiologist: That brings us to the alignment problem.

Turing: Yes. A measurable objective may be an imperfect substitute for what people actually want. A system can optimize that substitute successfully while failing to serve the underlying purpose.

Radiologist: Like maximizing diagnostic sensitivity without sufficiently considering the consequences of false positives?

Turing: That can illustrate the problem if sensitivity is optimized without adequate regard for the intended clinical outcome. The appropriate balance between sensitivity and specificity is a clinical question, not merely a computational one.

Radiologist: Then, alignment is not reducible to opacity.

Turing: No. Understanding a system and ensuring that it serves appropriate objectives are different tasks, although understanding can help us discover misalignment.

Radiologist: Let us return to my second proposed root cause: rapid development.

Turing: What is frightening about speed itself?

Radiologist: The capabilities of AI are advancing so quickly that our ability to understand and evaluate them may lag behind.

Turing: Then, perhaps the problem is not speed alone but the possibility that development and deployment outpace the methods needed to establish reliable performance and appropriate use.

Radiologist: And new capabilities may appear that we did not fully anticipate.

Turing: Sometimes. But apparent leaps in performance can also depend on how capabilities are measured.[10] What matters is that we cannot always predict a system's capabilities and limitations confidently before testing them.

Radiologist: In radiology, even an unchanged algorithm may encounter changing scanners, protocols, patient populations, and clinical practices.

Turing: Exactly. A model that performed well under one set of conditions may perform differently under another. That is why evaluation before deployment and monitoring afterward are both important.[11]

Radiologist: There is still something rather peculiar about AI. It communicates as though we were speaking to another person. Does that not make it especially unsettling?

Turing: It can. Human-like conversation encourages us to attribute understanding, intention, or authority beyond what the evidence supports. In my 1950 paper, I explored the imitation of human conversation,[1] but successful imitation and justified trust are different questions.

Radiologist: In other words, the machine may sound more certain than its evidence permits.

Turing: Yes, and people may mistake conversational confidence for authority. But anthropomorphism is not identical to opacity or misalignment; it concerns, at least partly, how humans interpret and respond to the machine.

Radiologist: We began with two candidate root causes. We have now considered opacity, rapid development, alignment, unreliable evidence, sycophancy, and anthropomorphism. Have we discovered the fundamental causes or merely assembled a more impressive list?

Turing: We have identified different mechanisms and concerns but not established that they are all independent or exhaustive. Let us test your original hypotheses. If a system were fully transparent, would all the danger disappear? If development slowed, would it disappear?

Radiologist: No, and no.

Turing: Then neither opacity nor speed can be the whole explanation. A transparent system can use false information or pursue an inappropriate objective; a slowly developing one can still mislead us. Opacity makes failures harder to investigate, and speed can make evaluations outdated.

Radiologist: So, the problem goes beyond both. An AI answer can seem trustworthy even when the grounds for trusting it are inadequate—or cease to apply when the system or its use changes.

Turing: That is a more useful distinction. But do not assume that every machine must explain all its internal workings before it can be trusted.

Radiologist: Why not?

Turing: Because internal explanation is one source of understanding, not the only basis for justified trust. Independent testing, reproducible results, performance within defined conditions, and ongoing monitoring also provide evidence.[11-13]

Radiologist: As they do in medicine.

Turing: Precisely. You routinely make decisions under uncertainty. You do not demand absolute certainty before acting; you demand evidence appropriate to the importance of the decision.

Radiologist: But medicine does not always get that balance right.

Turing: Of course not. Experience with uncertainty does not guarantee good judgment; it provides methods for making judgments—and for correcting them.

Radiologist: Then, what should radiologists actually do with AI?

Turing: They should help establish what counts as meaningful clinical benefit, which errors are acceptable, which populations and settings have been adequately tested, and when a system's recommendations require additional scrutiny. Reporting standards and professional guidance provide useful starting points.[11,14,15]

Radiologist: That sounds like a great deal of work.

Turing: Progress occasionally is.

Radiologist: And has such scrutiny proved worthwhile?

Turing: Consider mammography. An earlier study found no improvement in screening accuracy from conventional computer-aided detection.[16] The lesson was not that computers could never help but that promised benefits require evidence.

Radiologist: And is there now evidence supporting optimism?

Turing: Yes, in specific settings. The randomized Mammography Screening with Artificial Intelligence trial reported a non-inferior interval cancer rate, improved sensitivity, unchanged specificity, and a reduced reading workload.[17]

Radiologist: That sounds encouraging.

Turing: It is. But the conclusion belongs to the evaluated approach and its intended setting.

Radiologist: A mammography screening system performs a defined task. A general-purpose AI assistant can answer almost anything I ask, and its response may change with the wording of my question.

Turing: Exactly. Evidence supporting one screening workflow cannot establish the reliability of an open-ended assistant. Each intended use requires appropriate evaluation, including changes in user interaction and model versions. An answer from an unevaluated use should be checked, not treated as established clinical evidence.

Radiologist: So different systems should earn different degrees of confidence for different purposes.

Turing: Precisely.

Radiologist: We have moved rather far from my opening question. Perhaps we should return to it.

Turing: By all means.

Radiologist: Why is AI scary, if it is scary at all?

Turing: Because an AI answer can appear trustworthy even when the grounds for trusting it are inadequate—or no longer apply.

Radiologist: Then, perhaps the appropriate response is neither fear nor unconditional trust but informed curiosity, disciplined skepticism, and cautious optimism.

Turing: That seems a reasonable conclusion.

Radiologist: Can we really be optimistic?

Turing: Cautiously, yes. Not because every system deserves confidence but because we can investigate its limitations and test its performance for particular uses. But you are the physician. May I ask you something?

Radiologist: Of course.

Turing: Has uncertainty ever prevented you from practicing medicine?

Radiologist: No. It is part of practicing medicine.

Turing: Then, perhaps the arrival of AI does not demand that you abandon your methods of judgment; it demands that you examine how those methods must evolve.

Radiologist: So, we need not eliminate uncertainty before benefiting from AI?

Turing: No. But we must never confuse our ability to use a machine with our understanding of when to trust it.

Radiologist: That sounds rather familiar.

Turing: It should. Radiologists have been learning when to trust their tools for a long time; AI gives them another reason to keep learning.

Footnotes

Conflict of interest disclosure

Şükrü Mehmet Ertürk, MD, serves as Editor-in-Chief of Diagnostic and Interventional Radiology. The editorial review of this article was conducted independently by another Editor. The author declares no other conflicts of interest.

References

1. Turing AM. Computing machinery and intelligence. *Mind.* 1950;59(236):433-460. [Crossref]
2. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30:5998-6008. [Crossref]
3. Turpin M, Michael J, Perez E, Bowman SR. Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. *Adv Neural Inf Process Syst.* 2023;36:74952-74965. [Crossref]
4. Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. *ACM Comput Surv.* 2023;55(12):Article 248. [Crossref]
5. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst.* 2020;33:9459-9474. [Crossref]
6. Alber DA, Yang Z, Alyakin A, et al. Medical large language models are vulnerable to data-poisoning attacks. *Nat Med.* 2025;31(2):618-626. [Crossref]
7. Sharma M, Tong M, Korbak T, et al. Towards understanding sycophancy in language models. In: Proceedings of the 12th International Conference on Learning Representations (ICLR); 2024 May 7-11; Vienna, Austria. 2024. [Crossref]
8. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nat Med.* 2024;30(3):837-849. [Crossref]
9. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121-127. [Crossref]
10. Schaeffer R, Miranda B, Koyejo S. Are emergent abilities of large language models a mirage? *Adv Neural Inf Process Syst.* 2023;36:55565-55581. [Crossref]
11. Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR and RSNA. *Radiol Artif Intell.* 2024;6(1):e230513. [Crossref]
12. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep.* 2019;49(1):15-21. [Crossref]
13. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health.* 2021;3(11):e745-e750. [Crossref]
14. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for artificial intelligence in medical

imaging (CLAIM): 2024 update. *Radiol Artif Intell.* 2024;6(4):e240300. [Crossref]

15. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*. 2025;388:e081554. [Crossref] Erratum in: *BMJ*. 2025;388:r340. [Crossref]
16. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL; Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med.* 2015;175(11):1828-1837. [Crossref]
17. Gommers J, Hernström V, Josefsson V, et al. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. *Lancet.* 2026;407(10527):505-514. [Crossref]