AI Diagnostic Tools & Patient Trust:

How Can Transparency Improve Patient Acceptance of Automated Diagnoses?

Yusra B. Ashar, Tariq Kakar, Jason Smevog, Yui Nguyen

California State University Monterey Bay

CST462S - Race Gender & Class and Class Digital World

Dr. Bude Su, Brianna S. Ortega, Brian Robertson

11 November 2025

Abstract

Artificial intelligence is rapidly reshaping diagnostic medicine through the ability of systems to interpret complicated medical images, predict disease trajectories, and suggest therapeutic interventions. Yet, these advances have not translated into universal patient acceptance; many remain skeptical of machine-generated diagnoses. This study investigates whether transparency and explainability in AI-driven diagnostic tools can bridge that trust gap. Based on recent research related to explainable AI and upon insights from patient interviews, the paper examines how the lack of transparency of "black box" algorithms erodes confidence and discusses design strategies that make AI decision-making more interpretable. The presented study therefore focuses on an analysis of ethical implications of transparency in AI with respect to impacting patient autonomy and tries to contribute to ongoing debates about equitable, human-centered healthcare innovation.

AI Diagnostic Tools & Patient Trust

Modern medicine benefits from AI-based diagnostic systems which include radiology image readers and pathology analyzers and predictive clinical models that deliver high accuracy and operational efficiency. The main problem continues to exist because patients tend to doubt the accuracy of automated medical diagnosis results. The majority of AI models operate as unexplainable black boxes because they generate results without showing their decision-making processes. The lack of transparency in AI systems creates two main issues which affect patient mental state and their concerns about fairness and their willingness to follow AI-generated treatment plans.

The research problem focuses on the difference between AI diagnostic precision and patient confidence in these systems. Patients need system transparency and interpretability to make decisions and hold systems responsible for their mistakes. The absence of explainable systems creates a digital divide which benefits people who understand technology but excludes those who do not. The solution of this gap requires immediate attention because patients need to trust medical systems before they will accept their recommendations.

The subject holds social importance because it creates direct impacts on healthcare services that guarantee equal treatment for all patients while respecting their right to make decisions. The growing hospital dependence on AI screening and triage systems creates access disparities for patients who doubt or lack understanding of these systems. The research into AI diagnostic transparency and communication methods for patient confidence development serves both technical and moral purposes because it connects technological progress to ethical standards.

Literature Review

AI Decision-Making and the Black-Box Problem (BaniHani et al., 2024)

BaniHani Alawadi and Elmrayyan (2024) published AI and the Decision-Making Process: A Literature Review in Healthcare Financial and Technology Sectors to study how AI affects decision-making transparency and trust in healthcare and financial and technological sectors. The research team discovered that deep learning models create a “black box” problem because their operational mechanisms remain invisible to users. Users lose trust in systems because they cannot understand the decision-making process which results in biased outcomes and discriminatory treatment of protected groups. The authors demonstrate that users need transparency to confirm medical diagnoses and loan decisions are both fair and accurate. The authors support AI adoption as an unavoidable reality that requires XAI research to prevent these systems from causing damage to society. The research investigates healthcare and financial and technological sectors through their analysis of healthcare misdiagnoses and algorithmic discrimination and transparency levels. The authors present regulatory obstacles and support human-oriented AI governance through examples including the EU AI Act (BaniHani, Alawadi, & Elmrayyan, 2024).

Clinicians’ Trust in Explainable AI (Rosenbacke et al., 2024)

The research shows that medical professionals lack trust in AI-based clinical choices because these systems operate as black boxes and provide confusing or inconsistent explanations. The authors establish that achieving maximum trust levels stands as a fundamental requirement for AI system integration in medical practice because both excessive trust and insufficient trust can result in adverse consequences. The authors establish that healthcare

professionals need to determine the correct trust threshold relative to AI system precision for safe and effective AI implementation (Rosenbacke, Melhus, McKee, & Stuckler, 2024).

Patient-Centered Transparency in AI Diagnostics (Sadeghi et al., 2024)

Sadeghi et al. (2024) focus on AI diagnostic systems from a patient perspective, arguing that explainability is a precondition for trust and ethical use. They note that most AI tools in radiology and pathology remain opaque, limiting patients' ability to understand or challenge diagnostic outputs. The authors advocate for transparent design that visually and verbally communicates how decisions are reached, thus enhancing patients' sense of control and informed consent. Their work links technical innovation to moral obligation, asserting that trustworthy AI must be both accurate and understandable.

AI, Misinformation, and Public Perception (Kim & Sin, 2025)

Kim and Sin (2025) studied how AI-based social media content affects user emotional responses and trust levels. The researchers conducted a survey with college students who showed that AI-created false information caused users to experience anxiety and confusion while decreasing their trust in digital information. The research investigates social media platforms instead of medical settings yet demonstrates similar psychological responses when users encounter unexplained diagnostic systems.

Literature Review Synthesis

The research findings demonstrate that AI systems need to operate with complete transparency to achieve better patient understanding and self-determination and acceptance. The research by Kim and Sin (2025) demonstrates that AI systems without clear explanations will create distrust among users which leads to emotional distress. The research indicates that AI

systems which provide transparent and interpretable results create medical accuracy while establishing ethical and psychological foundations for trust. The system's diagnostic process needs to be visible to patients for them to accept AI-based diagnostic tools in healthcare.

The research findings demonstrate that AI trust development requires both clear information and effective communication methods. BaniHani et al. (2024) identify transparency as the fundamental ethical problem of contemporary AI systems. The study by Rosenbacke et al. (2024) demonstrates that healthcare providers will adopt AI suggestions based on the quality of explanation they receive. Sadeghi et al. (2024) applied this concept to patient populations through their research which proved that patients benefit from diagnostic systems that provide explanations.

Research Question

How to improve transparency in AI-based Diagnostic tools?

Interview Questions

The research question of this project and our literature review results led our team to create interview questions that study human responses to AI diagnostic system transparency. The team members each added two questions to create a full set which investigates trust and misinformation and how explanation and decision-making impact AI diagnosis acceptance. The research team will use these questions in all interviews to achieve data collection consistency. The following section shows the complete set of interview questions.

Interview Questions

- How does the clarity of the reasoning behind an AI tool influence your level of trust in acting on that information?

- Can you describe a time when a digital tool or system provided a patient result that conflicted with your own clinical judgment?
- How do you feel when you know AI gave you misinformation?
- How do you check the result with AI or search engines to ensure the provided information is accurate?
- How important is it for you to understand why AI arrived at a particular recommendation? Why or why not?
- When AI answers disagree with your own judgment, what factors guide your final decision?
- Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?
- What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Our target research participant audience consists of individuals who interact with or are affected by AI decision-making systems across professional, educational, and personal contexts. Tariq described the target audience as "professionals involved in the lifecycle of AI decision making systems," including healthcare professionals who use AI decision support tools and technology professionals who help build them, noting that these participants will likely fall within the age range of 30 to 50, possess moderate to high technological comfort, and represent a diverse mix of gender and backgrounds to explore whether trust varies across demographics. Yui selected "a person who is a professional in the tech field, computer science to be specific," someone with limited medical knowledge to better observe emotional responses and information-checking behaviors when interacting with AI systems. Jason identified "someone

who is a professional, most likely in the medical field, and uses AI systems at work" as his participant, offering insight into how clinicians interact with automated tools on a regular basis. Yusra selected a computer science (CS) tech student with strong comfort using technology as her target interviewee. This participant will be someone who encounters AI and automated systems in academic or early-career contexts, making them well-positioned to speak on how transparent explanations influence understanding, trust, and willingness to rely on AI tools. Yusra's participant will be contacted through professional networking platforms such as LinkedIn, and the interview may be conducted through Zoom or Discord depending on availability.

Although each team member plans to use a different interview format based on participant availability and context, the core questions and order will remain identical across interviews. Tariq plans to request permission to record his interviewee via phone call, text, or in-person conversation to allow accurate transcription and deeper analysis. Yui intends to meet her participant in person and take notes but will use a Google Form if an in-person meeting is not possible. Jason prefers to conduct his interview via Zoom or phone, using email only if necessary. Yusra plans to conduct her interview through LinkedIn, Zoom, or Discord, depending on scheduling with the CS student. Regardless of method, all interviewees will receive the same eight questions in the same sequence, and each interviewer will document responses through observational notes, recordings (when permitted), or written submissions. All raw data will be saved in a shared team folder for later thematic analysis.

Each team member is responsible for interviewing one participant within the target audience. To ensure adequate time for transcription, organization, and analysis, the team agreed that all interviews must be completed before the end of Week 5, aligning with project requirements and allowing sufficient time to prepare the Findings section.

Conduct Research

Each team member conducted at least one interview with individuals from the target audience to explore how people perceive and evaluate transparency in AI diagnostic systems. The following section provides answers to:

1. Identify who each team member interviewed
2. Explain why these individuals were suitable participants for this study.

Jason's Interviews:

1. Within the target audience, I interviewed a nurse, Michael Fallarme, and an EMT, Marcel Anderson.
2. The characteristics that make these interviewees qualified to speak on the topic of AI in the medical field is their experience, they are professionals in the medical field. They work with the AI tools that we are discussing and they interact with the patients that are affected by them.

Yui's Interviews:

1. Participant A: 33 female interviewee who worked on the IT field for about 10 years
   Participant B: 27 male interviewee who is working for a Vietnamese company, mainly on back-end web developer
2. Characteristics: they have at least five years of experience in the technology field, which means they understand AI and its behavior quite well. In addition, their age is between 25

to 35, a great range for adapting at this moment. These characteristics make them ideal interviewees to see how people feel and react with misinformation from AI.

Tariq's Interview:

1. The interview questions were answered by a registered nurse named Fedwa Kakar.
2. The interviewee is a registered nurse with over five years of experience, representing the clinical end-user demographic essential to this study. She is an ideal participant because she possesses the clinical expertise required to identify 'hallucinations' or errors in AI diagnostic suggestions. Furthermore, she provided high-utility qualitative data regarding the specific types of transparency features, like 'rationales and charts' that would bridge the trust gap defined in our research problem.

Yusra's Interviews:

1. I gathered responses from nine individuals representing a diverse combination of ages, professions, and familiarity with AI systems:

   Sarah A – 49, Asian female, Tech Field

   Husna Ahmed – 40, Asian female, Medical Field

   Nina – 40, Asian/Pakistani female, None of the above

   Dan Sal – 40, Asian male, Tech Field

   Saman Rehman – 37, Pakistani female, None of the above

   Osman – 33, Male, Tech Field

   Participant C – 22-year-old Hispanic male, None of the above

   Yusuf – 17-year-old Pakistani male, Tech Field

   Nicholas Anderson – Male, Software Engineer at Vault JS and CSUMB alumnus

Because they represent a wide range of AI users, including those in the tech industry, the medical field, and general users without technical expertise, these interviewees were ideal. Tech professionals such as Sarah A, Dan Sal, Osman, Yusuf, and Nicholas Anderson offered informed insights into AI behavior and the importance of transparent explanations, with Nicholas adding an advanced perspective as a software engineer. Medical professionals like Husna Ahmed contributed views directly related to diagnostic safety, accuracy, and the role of human oversight. Non-technical participants, including Participant C, provided important perspectives on how everyday users interpret and evaluate AI recommendations. This varied group, which spans ages 17 to 49, provides a thorough understanding of how transparency affects trust among various user groups.

Findings

Based on the interview data collected from 14 participants, three primary user groups emerged, medical professionals, technology professionals, and general users. Across these groups, four central themes appeared consistently: definitions of transparency, trust and reasoning, responses to misinformation, and requirements for accepting AI tools. Together, these themes provide insight into how different users understand and evaluate transparency in AI-based diagnostic systems.

Theme 1: Definitions of Transparency

Medical Professionals

Medical professionals defined transparency within the context of patient safety, clarity, and clinical responsibility. They emphasized that AI should make its reasoning explicit, stating that

"nothing is hidden" when forming diagnostic suggestions (Participant 7, see Appendix). Others stressed that transparency requires the system to be "blunt about certain scenarios" (Participant 12), avoiding vague probabilities and unclear language. This group viewed transparency as necessary to support safe clinical decision-making, noting that they must trust the system to "give us correct responses and aid us in care for patients" (Participant 6).

Technology Professionals

For technology professionals, transparency was tied to the data and system structure. They emphasized the importance of clean, reliable input data (Participant 2), user control over data collection (Participant 14), and visibility into how the AI reached its conclusion. One participant explained that the system should never output a result without also providing the reasoning behind it (Participant 14). They also acknowledged potential bias, noting that "no one can guarantee that all training data is clean or unbiased" (Participant 2).

General Users

General users interpreted transparency as simple, honest, and understandable communication. They wanted "clear communication through the process" (Participant 3) and reassurance that the system contained "no hidden content" (Participant 10). They also stressed the importance of understanding limitations, accuracy levels, and risks (Participant 8). Overall, this group tied transparency to feeling informed and not misled.

Theme 2: Trust and Reasoning

Medical Professionals

For clinicians, understanding the AI's reasoning was non-negotiable. One participant stated that reasoning is essential because "we are handling patients' lives" (Participant 6). Others explained that without sources or evidence, they "can never trust that it's truly correct" (Participant 12). Reasoning helps clinicians judge whether the AI's suggestions align with medical logic and clinical sense (Participant 4).

Technology Professionals

Tech professionals shared a "verify, then trust" mindset. They appreciated reasoning because it allows them to confirm that the system's output "makes sense and [can be] used safely" (Participant 13). Still, some expressed skepticism, noting that even strong reasoning does not ensure perfect data quality (Participant 2). Their trust depends on being able to independently validate the system's conclusions.

General Users

General users linked trust to comfort, understanding, and informed consent. They described needing a "safer understanding of the process" (Participant 3) before agreeing to or acting on information. For this group, clear reasoning increased confidence and reduced fear of hidden errors.

Theme 3: Responses to Misinformation and Conflict

Medical Professionals

Medical professionals reacted most strongly to AI misinformation. Words like "TERRIFIED" (Participant 4) and "worried" (Participant 7) demonstrate how seriously they perceive incorrect

AI responses. Participants described specific conflicts, such as AI missing comorbidities (Participant 6) or giving false anatomical information (Participant 12). Their verification process relied heavily on “evidence-based facts” (Participant 6) or scholarly sources (Participant 12).

Technology Professionals

Tech professionals were more analytical and less emotional. Some reported feeling nothing because they expect AI to occasionally make errors (Participant 1). Others described feeling “annoyed” (Participant 11) but unsurprised—summarized as “I knew it” (Participant 2). Verification methods for this group included checking “direct sources” (Participant 14), Google (Participant 1), or StackOverflow (Participant 5).

General Users

General users showed a sense of resignation, expressing that errors were “not surprising” (Participant 3) and acknowledging that “things can mess up” because AI is artificial (Participant 10). Their primary method for resolving misinformation was contacting a healthcare provider for clarification (Participant 8).

Theme 4: User Requirements for Acceptance

Medical Professionals

Clinicians required detailed rationales, charts, and step-by-step visuals (Participant 6) that clearly link sources to accuracy (Participant 12). Even when AI results are compared to human doctors, many remain cautious, explaining that AI cannot “physically assess the patient” (Participant 4).

One suggested adding an “E-Sign” feature, allowing a doctor to digitally confirm or approve the AI’s result before it is used (Participant 12).

Technology Professionals

Tech users preferred logical, step-by-step breakdowns (Participant 1) along with visual explanations (Participant 14). They were generally favorable toward human comparison, stating that knowing a doctor reviewed or aligned with the AI result would make them “more confident” (Participant 9).

General Users

General participants preferred simple written explanations or step-by-step instructions (Participant 8). Like the other groups, they felt more confident when AI outputs were reviewed and confirmed by a human medical professional (Participant 8).

Synthesis: Addressing the Research Question

The research question asks how transparency can be improved in AI-based diagnostic tools. Across all user groups, the data points to a consistent answer: the integration of Explainable AI (XAI).

Participants repeatedly expressed that AI must move away from “black box” outputs and instead provide:

- Clear, step-by-step reasoning
- Visual explanations
- Source citations and evidence trails

- Honest disclosure of limitations
- Options to opt out of data collection
- Simple language for non-experts

However, the findings also show that transparency alone is not enough to create full trust, especially in medical settings. Many participants suggested that transparency must be paired with human verification protocols, such as doctor review or digital sign-off, to bridge the gap between algorithmic recommendations and clinical safety.

In short, improving transparency requires a system that is not only explainable, but also supported by human oversight, ensuring that AI tools function as partners, not replacements, in the medical decision-making process.

Conclusions

Based on our findings, we can conclude that transparency is a foundational requirement for user acceptance of AI-based diagnostic tools, but is understood at different levels across user groups. Medical professionals prioritize transparency as a matter of patient safety and clinical accountability. Technology professionals view it through system logic and data integrity. General users associate transparency with clarity, honesty, and reassurance. Despite these differences, all groups demonstrated that they value explainable reasoning and clear communication.

The research question, *"How to improve transparency in AI-based diagnostic tools?"*, has been answered fully and specifically by the study's findings. The research concluded that the definitive method for improving transparency is the integration of Explainable AI (XAI) across diagnostic systems. The answer provided is detailed, specifying that transparency requires

moving away from “black box” outputs and instead implementing specific features such as step-by-step reasoning, visual explanations, and source citations. Furthermore, the research nuanced this answer by concluding that transparency features alone are insufficient to ensure acceptance; they must be paired with human verification protocols to fully answer the problem of patient trust.

From an interpretive perspective, this indicates that transparency in medical AI is not simply a technical enhancement, but a socio-technical requirement shaped by responsibility, risk, and ethical accountability. While AI systems can achieve high diagnostic accuracy, users do not equate accuracy with trust unless the reasoning behind that accuracy is visible, understandable, and verifiable. Transparency therefore operates as a bridge between algorithmic decision-making and human judgment, particularly in high-stakes medical contexts where consequences directly affect patient well-being.

The conclusions are supported by consistency across the qualitative data collected from 14 participants in medical, technology, and general user groups. For medical professionals, the results support the conclusion that transparency is a safety and ethical necessity; this group defined transparency as ensuring “nothing is hidden” and required explicit rationales because they bear the clinical responsibility for patient lives. For technology professionals, the findings interpret transparency as a structural requirement; their “verify, then trust” mindset supports the conclusion that users need visibility into data quality and logic to independently validate the system’s output. Finally, for general users, the results support the conclusion that transparency functions as a comfort mechanism, where honest, simple communication is required to reduce the fear of being misled.

Additionally, the findings reveal that medical topics hold a uniquely elevated level of importance across all participant groups, regardless of professional background. Respondents consistently reported that they are more likely to double-check AI outputs when medical decisions are involved, while skipping verification only when the topic is perceived as low-risk. This reinforces the conclusion that AI systems operating in healthcare must meet the highest standards of transparency, accuracy, and accountability. Medical AI is therefore held to a different ethical and functional threshold than general-purpose AI tools, emphasizing the need for domain-specific safeguards.

Taken together, these conclusions demonstrate that transparency must be understood as a layered concept: technical explainability, ethical disclosure, and institutional accountability. AI-based diagnostic tools that fail to address all three dimensions risk undermining patient trust, clinician confidence, and broader public acceptance. Conversely, systems that successfully integrate explainability with human oversight position AI as a supportive partner in healthcare rather than a replacement for professional judgment.

Recommendations

To solve the problem of transparency in AI-based diagnostic tools, explainable AI should be implemented into them. Implementing XAI into these systems will satisfy the explainable reasoning and clear communication values that the medical professionals, technical professionals, and general users all share. Another thing that can be done to increase transparency is to disclose the system limitations clearly, such as uncertainty levels, known biases, and gaps in training data.

The research also recommends solving the problem of distrust by designing AI tools that function as partners rather than replacements, specifically through the use of Explainable AI (XAI). To solve the “black box” problem, the paper recommends specific design implementations including clear, step-by-step reasoning, visual explanations, and evidence trails that link to sources. Additionally, the study recommends ethical solutions such as honest disclosure of system limitations and providing options for users to opt out of data collection. Crucially, the paper recommends that these technical solutions be supported by human oversight protocols, such as a doctor’s digital sign-off or review, to bridge the gap between algorithmic probability and clinical safety.

Expanding on these recommendations, AI developers and healthcare institutions should prioritize domain-specific AI systems for medical use. Similar to how specialized AI models exist for art, music, or literature, medical AI tools should rely exclusively on vetted medical databases, peer-reviewed literature, and validated clinical guidelines. For general-purpose AI systems such as ChatGPT, Copilot, or Grok, training and response generation should be constrained to trusted academic and clinical sources when handling medical queries. This approach reduces the likelihood of misinformation and aligns AI outputs with professional medical standards.

Furthermore, AI diagnostic interfaces should be designed with user-centered communication strategies that adapt explanations to different levels of expertise. Medical professionals may require detailed charts, rationales, and source links, while general users benefit from simplified explanations that clearly communicate risks, limitations, and next steps. Transparency, in this sense, must be contextual rather than uniform, ensuring that each user group receives explanations that are both accessible and meaningful.

Recommendations for Further Research

To further our research, we should test AI systems with built-in transparency and human verification to assess their real-world effectiveness. We could also conduct quantitative studies that measure how specific transparency features impact trust, accuracy, and clinical decision-making.

While the paper focuses on immediate design solutions, the literature review highlights several areas for further research. Continued inquiry is needed into human-oriented AI governance and regulatory obstacles to prevent systems from causing societal damage. Additionally, future research should explore the correct “trust threshold” for medical professionals, identifying how much reliance on AI is appropriate without encouraging overdependence or complacency. Finally, research into patient-facing communication methods is necessary to better understand how transparency influences confidence, informed consent, and long-term acceptance of AI-driven healthcare technologies.

References

Rosenbacke, R., Melhus, Å., McKee, M., & Stuckler, D. (2024). How explainable artificial intelligence can Increase or Decrease clinicians' trust in AI applications in health Care: Systematic review. *JMIR AI*, *3*, e53207. https://doi.org/10.2196/53207

BaniHani, I., Alawadi, S., & Elmrayyan, N. (2024). AI and the decision-making process: a literature review in healthcare, financial, and technology sectors. *Journal of Decision System*, *33*(sup1), 389–399. https://doi.org/10.1080/12460125.2024.2349425

Kim, K., & Sin, S. J. (2025). AI/Misinformation on social media: Users' appraisal and emotional outcomes of their response. *Proceedings of the Association for Information Science and Technology*, *62*(1), 1514–1516. https://doi.org/10.1002/pra2.1452

Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalos, P. M. (2024). A review of Explainable Artificial Intelligence in healthcare. *Computers & Electrical Engineering*, *118*, 109370. https://doi.org/10.1016/j.compeleceng.2024.109370

APPENDIX

Full Verbatim Interview Transcripts

(All text copied exactly as submitted by participants.)

Participant 1

Male, 27, Asian, Tech Field

(Timestamp: 12/1/2025 9:34:42)

Q1: When you think of an 'AI system being transparent,' what does that mean to you in a medical context?

In medical contexts, I believe that because AI is a machine, it will not withhold information from patients, which may at times come across as blunt and potentially affect patient's morale negatively

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

Usually, AI gives logical solutions that have a sense of certainty to them

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Very important, i need to verify the source of it's solution before trusting it

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

When I experienced lower-back pain that caused a painful numbness in my entire left leg, the AI suggested that it was due to an incorrect sleeping position. However, I knew that a simple sleeping posture issue couldn't cause such severe symptoms, so I went to the doctor to determine the underlying cause

Q5: How do you feel when you know AI gave you the wrong information?

Nothing, since I understand that AI draws its responses from multiple sources and there is no one who verifies all the information it is trained on

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

Usually through google, or asking professionals.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

I believe that a step-by-step logical diagnosis with reliable evidence will make it easier for me to accept.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

Yes

Participant 2 — ERIS RAIN

Female, 33, Asian, Tech Field

(Timestamp: 12/1/2025 9:43:27)

Q1: When you think of an 'AI system being transparent,' what does that mean to you in a medical context?

When it comes to the transparency of AI, it is extremely difficult to define what "being transparent" truly means in the medical field—especially when human lives are involved. We must remember that even human beings have not been fully understood. Much of the medical field still contains elements of uncertainty or what could be considered "educated guessing." An experienced doctor may diagnose instantly because they have seen certain symptoms many times before—and in a similar way, AI functions on patterns learned from data. AI is essentially a new form of database, drawing knowledge from many heterogeneous sources: medical journals, lectures, clinical reports, scientific literature, and more. But the idea of AI transparency in healthcare remains challenging because:

First, is the input data clean and reliable? Second, how much access does the AI have to the patient's information? And finally, is the AI supervised or validated by a physician after making a diagnosis?

Because of these unanswered questions, it is extremely difficult to define what true transparency means at this point. No one can guarantee that all training data is clean or unbiased, and thus no one can fully guarantee how AI arrives at every medical decision it makes.

This is my perspective on the issue.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

Since AI itself cannot fully control the quality of its input data, in my personal opinion, even if the reasoning is explained clearly, I would still need to review its recommendation.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

It must be clear that no AI is perfect. It only provides the answer that is most similar to what "usually works" — in other words, the "likely" solution — and there will always be a small margin of error. Depending on the AI model, they may even give different answers for the same medical case.

So for me, even when AI gives a recommendation, I will still question it and verify it with my own experience, with the experience of colleagues, and with the AI's suggestion as well — and then make my final decision from all combined inputs.

For me, from the two answers above, I also made it clear that, by nature, AI allows us to access a supportive tool—note, a tool that "assists" in the process of diagnosis, and even in treating

patients. But for me, I will balance between transparency and convenience. Clearly, I will consider the AI's answer as just "a suggestion."

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

Certainly, there will be such cases, a typical example of conflict between humans and AI. AI always aims for the most likely answer. Humans, on the other hand, rely heavily on experience and personal judgment; some people have unique experience and assessments, which allow them to create their own answers to solve a problem.

But in the medical field, we are talking about a patient's life. We cannot "retry" an AI model's answer, nor can we fully rely on a single doctor's experience.

I would still proceed as usual. For example, in a difficult case, I would assemble a team of experienced doctors and consult AI—not just one, but multiple AI models, like GPT, Claude, Gemini, Gwen—to arrive at the most desirable answer.

Q5: How do you feel when you know AI gave you the wrong information?

My feeling when AI gives wrong information?

I just think, "Well, I knew it."

Because I know very well that AI has its own limitations; it is a tool. A mistake from AI reminds me that no tool is perfect—what matters is whether we are using it correctly.

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

The way I verify information is mostly based on my own experience. For example, if I have a photo and I need to describe the person in it—their appearance, their clothing—I first make my own description as clearly as possible. Then, I ask an AI to analyze it and compare it with my description. Next, I continue with another AI model, adding or adjusting details until I feel the result is the most "accurate." Here, "accuracy" means the answers from multiple AIs that match my own description the most.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

For me, I want an AI to respond with analysis step-by-step, especially for long, academic-style answers, so that I can verify each part carefully, slowly, and thoroughly.

For short, high-level questions, I prefer AI to respond in a visual + written style, so I can grasp the main structure and the overall picture.

Therefore, for me, it depends on the type of question I need the AI to answer.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

For me, having AI compare its result with a human doctor's opinion would certainly make me more confident, but I would still not rely entirely on it.

In the medical field, patient safety is the top priority, so I would treat the AI's result as a supplementary suggestion, combined with the doctor's experience and even other AI models if

needed. This comparison helps me assess the reliability of the AI, but ultimately, the decision would be based on a combination of sources and practical experience.

Participant 3

Male, 22, Hispanic, None of the above

(Timestamp: 12/1/2025 13:25:14)

Q1: When you think of an 'AI system being transparent,' what does that mean to you in a medical context?

Clear communication through the process in which is being conducted

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

It helps provide a safer understanding of the process

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

I would want to be informed on what is being conducted before providing consent

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

N/a

Q5: How do you feel when you know AI gave you the wrong information?

It wasn't surprising but it also didn't help provide me with the reassurance that what they do is correct

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

Backtrack the work and do research to ensure

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

A step by step break down discussing what is being conducted why and how

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

No

Participant 4

Male, 27, Asian, Medical Field

(Timestamp: 12/1/2025 13:35:41)

Q1: When you think of an 'AI system being transparent,' what does that mean to you in a medical context?

Breaking down everything step by step on how they got to that conclusion of diagnosis.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

Clear reasoning builds trust, and helps me judge whether the advice makes sense.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

It is very important, I'm more comfortable when I understand the reasoning behind it rather than relying on blind trust.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

N/A

Q5: How do you feel when you know AI gave you the wrong information?

TERRIFIED

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

I cross check by verifying key facts with reputable sites .

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

None

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

No because doctors are able to physically assess the pt, when it comes to medicine you have to look at the whole picture not just the symptoms. I'd offer AI suggestion to the doctors as potential answers, to maybe run more tests but I wouldn't take AI answer over an MD.

Participant 5 — Yusuf

Male, 17, Pakistani, Tech Field

(Timestamp: 12/1/2025 14:22:37)

Q1: When you think of an 'AI system being transparent,' what does that mean to you in a medical context?

More trust builds, users/clients feel safe talking with the bot

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

If your asking whether i trust AI's recommendations then yes if its clear enough i am more confident

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Extremely important, i have to check everything it said before moving on

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

A lot of times, i wanted to create my own agency and it made me realize how difficult the initial process was. so it kind of refined my way of thinking when and how to create my own agency.

Q5: How do you feel when you know AI gave you the wrong information?

I tend to lash out or call it derogatory words

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

either stackoverflow or I just dont

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Fast and sometimes visual but both conditions must be met together

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

No

Participant 6 — Fedwa Kakar

Female, 38, Central Asian, Medical Field

(Timestamp: 12/1/2025 15:50:15)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

We would trust the system to give us correct responses and aid us in care for patients. I would help shorten the time we take with patients especially with intake on a new patient.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

It is important because we are handling patients lives.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

When I had a patient that needed a specific antibiotic for a patient but there were comorbities that AI can't detect because AI needs all of the information about a patient before giving the correct answer.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

That you can't always trust AI you should always double check.

Q5: How do you feel when you know AI gave you the wrong information?

We look at our own guidelines and policies such as evidence based facts.

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

Rationales and charts that show why AI came to a conclusion.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Yes and no because there is so much more to diagnosing a patient than just facts from AI such as seeing the patient in person and looking at them in real life that helps better assess someone that I don't believe AI can do.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

(Combined in previous answer)

Participant 7 — Husna Ahmed

Female, 40, Asian, Medical Field

(Timestamp: 12/1/2025 17:05:54)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

It means the AI clearly explains how it reached a decision and what information it used. Nothing is hidden.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

If the reasoning is clear and makes sense, I am more willing to trust the recommendation.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

It is very important. I need to know why the AI said something before I act on it.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

Yes. Sometimes AI gives medical or academic information that feels incorrect or too general, so I double-check it with trusted sources.

Q5: How do you feel when you know AI gave you the wrong information?

I feel worried and disappointed because wrong medical information can be harmful.

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

I ask the AI again in a different way, look for reliable websites, or compare answers from multiple sources.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

A simple written explanation with clear steps is easiest for me to understand and accept.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

Yes. I would feel much more confident if the AI's result was checked or supported by a human doctor.

Participant 8 — Nina

Female, 40, Asian/Pakistani, None of the above

(Timestamp: 12/1/2025 17:59:01)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

Understanding how the AI system works, its limitations and accuracy, is essential for me to be able to trust in its recommendations.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

When I understand the process and 'reasoning' behind a recommendation I would naturally be able to trust in the AI more.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

It is very important for me to understand the process before I can accept the results.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

I have never had such an experience.

Q5: How do you feel when you know AI gave you the wrong information?

I would be more hesitant to trust in its recommendations in the future.

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

I would reach out to my medical provider to verify the AI recommendations.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Step - by - step would provide the most clarity for me, helping me understand and trust the recommendations.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

I would definitely feel more confident in AI recommendations if they were verified or aligned with a human professional's.

Participant 9 — Dan Sal

Male, 40, Asian, Tech Field

(Timestamp: 12/1/2025 20:08:23)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

It's not baised towards one aspect and shares pros and cons.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

Very important. If I am not able to follow the reasoning then I would be very reluctant to act in the recommendation

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Not that important and Depends on the topic. I am not going to understand some intricacies given its medical field.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

I always thought eating ice cream when cold is not good for health - recently asked ChatGPT and actually it's ok to eat.

Q5: How do you feel when you know AI gave you the wrong information?

Significantly reduces trust.

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

Depends on the topic. If it's a non consequential topic then I wouldn't. But if it is then I would check with doc before taking action.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Written would be good enough.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

Yes it would.

Participant 10 — Saman Rehman

Female, 37, Pakistani, None of the above

(Timestamp: 12/2/2025 1:31:01)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

That there is no hidden content

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

If it's something I'm familiar with I feel more comfortable to accept the recommendations

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Not really important

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

No

Q5: How do you feel when you know AI gave you the wrong information?

In the end it's artifical so I know things can mess up

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

By research and experience

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Written

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

Yes

Participant 11 — Osman

Male, 33, Tech Field

(Timestamp: 12/2/2025 10:01:06)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

Releasing medical data or allowing access by external entities

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

A.I. is just a tool and should be used as such its not

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Yes

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

all the time

Q5: How do you feel when you know AI gave you the wrong information?

annoyed

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?e

intuition

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

it would be easier to accept if it has data behind it

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

Yes

Participant 12 — Marcel Anderson

Male, 23, Black/African American, Medical Field

(Timestamp: 12/2/2025 14:34:20)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

When I think of AI being transparent, It leads me to the idea of AI being blunt about certain scenarios on a healthcare basis. For example, if a doctor was trying to determine if a patient had cancer; It would take thorough research to confirm the results. Where as AI might go based on a few symptoms a patient inputs into its data base and it bluntly diagnosis's them on the spot. (I.e patient: "I have constant stomach cramping and blood in my stool", AI: "it seems that you have colon cancer". I personally don't believe everything AI tells me because it's just a man made

computer system. The knowledge is traced from the web and you cant always rely on the web for basic knowledge.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

Sometimes the web can be very biased and not always resourceful if there are no documents to support answers. On a short note, AI does not influence me to act upon its recommendation, because I'm unsure about its knowledge.

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

Very important. Without sources for answers you can never trust that it's truly correct. Even in school when you're writing a paper you need lots of evidence to back up your arguments.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

I don't typically resort to AI for many situations because of my past experiences in college. The only time I can think about AI conflicting with my own judgement was when I was becoming an EMT and I couldn't remember a nerve and its primary function within the neck. So I googled it and the first thing that popped up was google Gemini giving me false information, which made me question my own knowledge on the topic. I then later found my notes and realized that AI was definitely wrong.

Q5: How do you feel when you know AI gave you the wrong information?

I've only used AI for research purposes, including research for essays, scholarly journals and a manuscript for a nutritional health course. There have been multiple incidents where I find that the documents AI provides me are inaccurate, or only talk about the topic I'm looking for in partially while the rest of the article/document is about a entirely different topic. So I've constantly

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

I typically resort back to a source that is scholarly, or has evidence to support its argument. If the answer line up, I have full trust in it.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

I would like a step by step visual that shows exactly where AI got its information from and a visual that connects each source together with accuracy and not just random documents.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

That honestly sounds like a great idea. It would definitely sway my opinion especially if the doctor had some type of E-SIGN on the document so that I could visually see that the doctor approved AI recommendations.

Participant 13 — Sarah A

Female, 49, Asian, Tech Field

(Timestamp: 12/2/2025 16:30:30)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

It basically mean the patient,doctors and the regulating body understand why AI is being used.

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

People understand Why and not just what , so if people understand the reason they will be more willing to act on recommendations by AI .

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

It is much more important to understand why the AI gave a result than to just trust it. If I know the reasoning, I can check if it makes sense and use it safely. Blind trust without explanation is risky, especially in medicine.

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

An AI app suggested I take the longer route to work because it thought traffic was heavy. But when I checked myself, the road was clear. So the AI's result conflicted with my judgment, and I chose the shorter route instead.

Q5: How do you feel when you know AI gave you the wrong information?

A bit confused and confident after the decision I took .

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

I use only reliable data sources.

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

Visual and step-by-step.

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

I would say yes ,because before any AI ,they were making decisions,plus they have experience of dealing with real human beings.

Participant 14 — Nicholas Anderson

Male, Software Engineer, Vault JS / CSUMB alumnus

(Interviewed via LinkedIn Messenger — full verbatim transcript below)

Q1: When you think of an "AI system being transparent," what does that mean to you in a medical context?

"A transparent model should allow for a patient to opt out of data collection. The AI should also explain how it came to the conclusion in some way."

Q2: How does the clarity of the reasoning behind an AI tool influence your willingness to act on its recommendation?

"Improves willingness."

Q3: How important is it for you to understand why the AI arrived at a particular result versus just trusting the output?

"Very important."

Q4: Can you describe a time when AI provided a result that conflicted with your own judgment?

"In areas I am knowledgeable in, I have often found incorrect or outdated information."

Q5: How do you feel when you know AI gave you the wrong information?

"Reminds me that AI output needs to be reviewed by someone knowledgeable in the field."

Q6: How do you check the result with AI or search engines to ensure the provided information is accurate?

"Find direct sources."

Q7: What kind of explanation (visual, written, step-by-step, or none) would make an AI diagnosis easier for you to accept?

"Visual + step-by-step would be a complete explanation."

Q8: Would you feel more confident in an AI diagnosis if the system compared its result to a human doctor's opinion?

"Yes."