

Informed Consent for Artificial Intelligence in Emergency Medicine: A Practical Guide

Kenneth V. Iserson, M.D., MBA
Professor Emeritus, Department of Emergency Medicine
The University of Arizona, Tucson, AZ
Published: *American Journal of Emergency Medicine*. 2024;76:225-230.

1. INTRODUCTION

Researchers claim that artificial intelligence (AI), the ability of machines to perform tasks typically associated with human intelligence and decision making, is the solution to numerous challenges in emergency medicine (EM).[1] As AI programs become more available, hospitals and clinicians face political and economic pressure to quickly adopt clinical AI with little system evaluation.[2-4] With AI's complexity outpacing EPs' ability to provide their patients with informed consent about its use, it is vital to highlight that information.[5]

To respect patient autonomy, EPs must engage in discussions with their patients about whether they want AI used in their assessment and treatment. As clinical AI becomes more widespread, these discussions must be integrated into the routine informed consent process. An ethical and legal requirement, informed consent provides patients with clear, comprehensive, and easy-to-understand information about a proposed medical intervention so that they can decide whether they want to proceed. This raises the question of whether EPs possess sufficient knowledge about clinical AI to offer informed consent effectively.

To provide informed consent about AI, EPs will need to know (1) how AI systems operate; (2) whether AI systems are understandable and trustworthy; (3) the limitations of and errors AI systems make; (4) how disagreements between the EP and AI are resolved; (5) whether the patient's personally identifiable information (PII) and the AI computer systems will be secure; (6) if the AI system functions reliably (has been validated); and (7) if the AI program exhibits bias. This paper will address each of these issues under two broad concepts, beneficence—doing good for the patient, and distributive justice—fairness to the entire population.[6,7]

2 Beneficence: What EPs Need to Know About AI to Provide Their Patients with Informed Consent

2.1 How Do AI Systems Function?

At their core, AI systems work like other computer programs, with data inputs and outputs. Despite using terms such as “learn,” “think,” and “decide” to describe their functions, there is no evidence that these programs are or ever will be sentient. AI developed for clinical medicine uses a blend of artificial neural networks that learn from huge datasets (deep neural networks, DNN), human language input (natural language processing, NLP), and performance improvement without human intervention (machine learning).[8]

Inspired by the human brain, DNNs model a network of interconnected neurons (nodes), each with a set of inputs and outputs. When a neuron receives input, it calculates an output and sends it to the next neuron. These systems have hundreds or thousands of “hidden” (i.e., “deep”) nodes, each of which makes numerous connections forward and sometimes backward (Figure 1).[9] Using trial and error or demonstration, each node continually refines its “understanding” of the information by adjusting its connections with other nodes.[10]

DNNs are trained on large datasets that have been labeled with the correct answers. For example, a DNN that is used to diagnose diseases might be trained on a dataset of test results, symptoms, images, and their corresponding diagnoses. The DNN learns to identify patterns in the data that are associated with different diseases. Once a DNN is trained, it can be used to make predictions about new data.

Different types of DNNs are best suited for different tasks in EM. Multi-layer perceptrons (MLPs) are good at solving problems that can be broken down into steps, such as diagnosing diseases based on symptoms. They can also predict patient outcomes and risk, identify trends in patient data, and assist with triage decisions. Convolutional neural networks (CNNs) are good at interpreting images, such as medical scans. Recurrent neural networks (RNNs) are good at solving problems that involve sequential data, such as predicting patient deterioration from vital signs.

In addition to DNNs, EPs also use natural language processing (NLP), a subset of RNNs, to interact with electronic health records (EHRs), language-translation programs, and AI online tools. While most AI programs for clinically assisted diagnosis are DNNs, the better-known term "AI" will be used throughout this paper.

2.2. Transparency: Can EPs and Patients Understand and Trust AI's Recommendations?

One challenge to adopting clinical AI in EM is the lack of understanding of how these programs make decisions, also known as the "black box" problem.[5] AI algorithm decision paths can be complex and difficult to explain and validate, even for their developers.[11]

Clinical AI processes are difficult to validate because neural networks manipulate data in complex ways (Figure 1). AI systems use random probability distributions to determine the weight for each connection between the nodes, but this may not result in accurate or consistent predictions for patients whose demographic groups were not in the training dataset. This inconsistency makes it difficult or impossible to recreate the results and is a major hurdle for developing "explainable, or interpretable, AI" (XAI) that achieves "transparency," or a clear understanding of how an AI system generates its recommendations.[12]

XAI, in contrast to the black box, is the "white" or "glass" box that highlights the factors that contributed most significantly to an AI decision, helping EPs identify critical factors in patient outcomes. A major XAI research group writes that it should be able to answer: Why did you do that? Why not something else? When do you succeed? When do you fail? When can I trust you? How do I correct an error?[13] Some XAI programs provide counterfactual explanations, showing how small changes in the input data lead to different decisions. Optimally, XAI should also identify and correct biases in AI programs and offer explanations in commonly understood language rather than "technospeak." [12] Unfortunately, the better AI models function, the less likely they are to be explainable.[13]

Two barriers to transparency for XAI developers have been the limited access to the AI training datasets and the reluctance of AI developers to open their proprietary black boxes.[4] Despite these difficulties, transparency remains a fundamental goal and a common element of ethical guidelines for clinical AI ethics (Table 1). [15-19] AI researchers are also investigating incorporating programs to guard against ethical problems and biases.[14] While it is still more theoretical and experimental than applicable, XAI development is crucial for EPs to trust AI's assistance in clinical decision making and to help provide the information patients need in the informed consent process. (Table 2).

A significant concern is that EPs may over rely on ("automation bias") or anthropomorphize AI (view as a human partner), giving it too much responsibility without continued vigilance. Such behavior can have dangerous consequences when the AI program is wrong or when the presenting problem is too unusual for it to process.[5]

2.3 Limitations and Errors in Clinical AI

Applying clinical AI to EM has several limitations. First, it may be difficult to amass high-quality, representative training data, potentially resulting in biased or inaccurate outcomes. AI programs adapted from other settings and not validated in EM, may lack proof of safety and effectiveness.[20] Moreover, integrating AI seamlessly into ED workflows can be complex and disruptive, especially given the time-sensitive nature of EP decisions.[20] The wide range of patients' symptoms and conditions can limit AI's effectiveness and make it difficult to explain its use and results to patients. Lastly, the high cost of developing and deploying AI systems may increase healthcare disparities.

AI programs are more likely to err than humans because they often rely on incomplete or erroneous data, lack practical knowledge, are surprisingly deficient at math, deal poorly with uncertainty, and are often biased.[21,22] In general, the probability of error will be higher for AI programs that are used to perform complex tasks, are trained on small datasets, or are used with biased datasets.

AI clinical program errors come from a variety of sources (Table 3).[25] How often errors occur depends on the program, its training dataset and how well it is matched to the task. Most information about AI errors is from non-clinical research. For example, when one clinical AI program's long-form answers about a medical case were compared to those of clinicians', a clinician panel judged that 5.9% of the AI recommendations, if followed, would lead to a potentially harmful patient outcome. This compared to 5.7% of clinicians' responses. A different AI program on the same test, however, had 29.7% potentially harmful answers.[23] Even some non-clinical AI programs, such as ChatGPT, can pass the USMLE, generate empathetic responses to patient questions, and outperform medical students on clinical reasoning examinations.[24] But, it has been difficult to get these programs to include only supportable medical information rather than inventing their own answers, a phenomenon termed "hallucinating." [24] Currently, these errors are a barrier to using AI in clinical medicine, since the bar for safe use must be extremely high in such a safety-critical environment.[24]

AI errors are not all created equal. AI systems are trained differently than physicians, and as a result, they make different types of errors. Unlike physicians, AI systems do not consider less-objective medical information (Table 4) when suggesting diagnoses.[26]

2.1.4. Resolving Discrepancies Between EP and AI Recommendations

When AI programs and EPs disagree on diagnoses and treatment plans, the consequences can range from minor to severe. To resolve these discrepancies, various approaches have been suggested, including: the physician's decision takes precedence, the machine's decision is favored, a predetermined compromise position is adopted, another physician is consulted, and another AI system is involved.

Existing ethics guidelines recognize the importance of having mechanisms to resolve disagreements between humans and AI systems that carry significant risks.[19] For example, the European Union's General Data Protection Regulation bans "solely automated" decisions that significantly affect people. Similarly, the European Commission proposed one of the first legal frameworks specific to AI, the Artificial Intelligence Act,[27] which requires high-risk AI systems to be designed so that a human can effectively oversee them.[28] In the United States, the Blueprint for an AI Bill of Rights provides that humans should monitor AI systems in case the system fails or produces an error.[26]

Future decision-making in emergency medicine will require a balance between AI-generated advice, other evidence, and patient preferences. One approach entails EPs reviewing significant AI decisions and substituting their judgment for the machine's decision if they disagree. That may not always yield the best results.[26] A better way to resolve conflicts between human and machine errors is to reduce the risks of both. One approach is for the AI program to classify its results as "confident normal," "confident abnormal," or "not confident." A doctor then reviews these results and, if they disagree, another doctor may be called for a second opinion. Another promising method is to have the AI program and EP make independent decisions followed by a mutual review. In high-risk situations where their decisions differ, another EP can review the decision again. This method has been successful in clinical settings like radiology where it was found to improve both sensitivity and specificity compared to relying solely on the EP or AI system. In emergency medicine, this approach might involve on-site or teleconsultation with another EP or other specialist or using a different AI system. For emergent situations, EPs should have immediate access to one of these predetermined resolution methods.

As for liability, most of the public (60%) and physicians (57%) believe that physicians should be held responsible for errors occurring when AI is used. Less than half of both groups believe that vendors or healthcare organizations should be liable.[29]

3 PRIVACY AND DISTRIBUTIVE JUSTICE: Using AI Systems Appropriately

3.1 Are Patients' PII and the Computer System Secure?

Clinical AI ethics guidelines universally emphasize the significance of safeguarding data privacy, (Table 1) which is unsurprising considering that medical AI systems rely on datasets containing extensive personal information. The information includes data from wearable health devices, telemedicine, radiologic images, and, most notably, EHRs, which encompass medical histories, treatment plans, and test results. To put the enormity of the data in perspective, the EHR from a single hospitalization typically generates approximately 150,000 pieces of data.[30] In 2020, the United States alone generated a staggering 2.3 exabytes (equivalent to 2.3 trillion gigabytes) of healthcare data, and by 2025, this number is expected to increase to approximately 3.8 zettabytes (3.8 trillion terabytes) annually.[31]

Cybercriminals consider medical computer systems and patients' PII as high-value targets, up to 50 times more valuable than financial data. PII provides names, policy numbers, birth dates, medical histories, billing data, diagnosis codes, and bank and credit card information.[32] To counter ever-more-sophisticated cyberattacks, clinical AI data is becoming better protected.[33] Security methods being explored include federated learning, in which AI models train on sensitive medical data that remains on individual or organizations' devices or servers, rather than on centralized data hubs. Homomorphic encryption allows computations to be performed on encrypted data with encrypted results as output. Secure multi-party computation has multiple parties jointly using their data, without revealing their individual datasets. Differential privacy adds controlled noise to data before analysis, making it difficult to identify PII. The best-known security technology is blockchain, a distributed ledger technology that enhances data storage security and data exchange.[34]

3.2 Does the AI System Function Reliably?

For informed consent, it is vital that EPs and their patients know if the clinical AI programs affecting them have been successfully validated and are continuously monitored (Table 5). Performed during system development, the validation process checks for program accuracy, bias, and, as with any medical technology, safety.[35] During validation, developers use datasets that differ from those used to train the program to ensure that the system will perform as expected in real-world conditions. Subsequently, the program must be tested against clinicians' decisions and other standard tools used for the same task. Finally, to seek approval from the US Food and Drug Administration (FDA) or the European Medicines Agency, it must undergo testing in real clinical settings.

AI systems must also be monitored throughout their clinical lifespan to ensure that they continue to perform optimally and reliably. If necessary, the algorithm and dataset should be recalibrated.[5,36] System monitoring includes checking the quality of the input data, as well as the accuracy, reliability, relevance, and completeness of the output.[37] The system is also checked for the reproducibility, sensitivity, and specificity of the results; type and severity of errors; observed versus expected error rates; and causes of errors.[38] While AI systems require continuous surveillance, the methods for doing this remain unclear and are often inadequate.[39,40]

3.3 Is the AI System Biased?

AI systems can exhibit bias by miscategorizing part of the population or when they use datasets in a discriminatory manner.[41] One example of such a distorted dataset was the massive ImageNet database, which was used to train many AI systems. Many of the photos in the "person" categories were given biased labels, such as "failure," "nerd," and "slut." [42] AI systems trained on this data learned to associate these negative labels with certain groups of people, thus reflecting the prejudices of the crowd-sourced group who had helped compile the dataset. Another example of inherent bias is a widely used healthcare dataset that labeled Black patients as healthier than White patients, even though this categorization was due to societal factors such as inadequate healthcare access.[43] AI bias can also be seen in the way that AI systems are used. For example, the US Immigration and Customs Enforcement agency used an AI system called the Risk Classification Assessment Tool to inappropriately deny release to all immigrants in their custody over many years.[44]

According to the UK National Health Service, "health systems are increasingly expecting AI developers to be transparent about the limitations and ethical examination of the population data used to develop algorithms [and] how data performance was validated".[45] One promising future option may be to embed bias mitigation techniques into AI systems.[46-48] Smith and Rustagi developed a "Playbook" that outlines steps that governments, businesses, and other groups can implement to detect and prevent bias in their AI systems.[41]

Medical AI programs can contain technological bias due to the datasets used to train them. For example, an AI system trained to diagnose chest radiographs using optimal films may not be able to accurately assess subtle cases or the less-than-optimal portable radiographs commonly seen in the ED.[49] Additionally, datasets may contain incomplete or inaccurate information, leading to decision-making errors. Errors in the initial dataset can be amplified in subsequent algorithm iterations, leading to confirmation bias and further errors. This is particularly problematic in EM, where datasets may include autofilled EHR charts, incomplete patient records, misdiagnoses (reported as up to 5.7% of ED visits [50]), or incorrectly recorded or non-representative patient histories, physical exams, and test results.[8]

To improve trust among clinicians and patients, many worldwide government bodies are addressing the ethics of using clinical AI.[51] The European Union has issued standards [28] and the FDA has begun regulating many AI algorithms as medical devices.[52] The US Department of Health and Human Services has also proposed regulating bias in clinical algorithms under health care antidiscrimination laws.[53,54] Some authors have advocated that clinical AI program developers be classed as medical providers and held to the same ethical standards as physicians.[55] Medical organizations, however, have been slow to formulate ethical guidelines for clinical AI. But, as Shah wrote, “Given the highly disruptive potential of these technologies, clinicians cannot afford to be on the sidelines.”[39]

4. Conclusion

AI is poised to become an important part of clinical decision making in EM. To provide their patients with informed consent, it is important for EPs to be aware of the ethical and practical pitfalls of AI systems. EPs should be able to tell their patients how AI programs function, their limitations, a patient’s recourse if errors result from their use, and whether patients have the right to refuse AI as part of their care process.

They should also be able to inform their patients that the AI program has been tested to ensure that it is secure, protects PII, and reliably functions for the purpose for which it is being used. They should be reasonably certain that the AI system is not biased against any portion of the ED population and is frequently monitored for accuracy so that any performance degradation can be detected. EPs should also be able to assure patients that there is not overreliance on AI decisions, and that a mechanism exists for resolving discrepancies between themselves and the AI output.

EPs should partner with bioethicists, AI researchers, and administrators to develop and implement an optimal plan for AI use in EM. This plan should address the ethical and practical pitfalls of AI systems to ensure that they are used safely, ethically, and responsibly.

Among issues still to be addressed include: what a patient’s refusal to allow AI to be part of their care process entails? Will AI be hardwired into the ED’s clinical process so that it cannot be refused? Will patients incur additional costs due to AI program use? And will patients be able to refuse the AI recommendations once they are made?

REFERENCES

1. Kirubarajan A, Taher A, Khan S, Masood S. Artificial intelligence in emergency medicine: a scoping review. *J Amer Coll Emerg Phys Open*. 2020 Dec;1(6):1691-702.
2. Benber B, Lay K. Health secretary Matt Hancock endorses untested medical app. *The Times*, 17 September 2018, available at <https://www.thetimes.co.uk/article/matt-hancock-endorses-untested-health-app-3xq0qcl0x>.
3. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018;361:k1479.
4. Lehmann HP, Downs SM. Desiderata for sharable computable biomedical knowledge for learning health systems. *Learn Health Syst* 2018:e10065.
5. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform*. 2019;28(1):128-34.
6. Derse A. In: Iserson KV, Sanders AB, Mathieu DR (eds.): *Ethics in Emergency Medicine*, 2nd edition. Galen Press, Ltd., Tucson, AZ, 1995:99.
7. Iserson KV, Heine C. Bioethics. In Walls RM, Hockberger RS, Gausche-Hill M, et al. (eds.), *Rosen’s Emergency Medicine: Concepts and Clinical Practice*, Tenth Edition Philadelphia, PA: Mosby, 2023, e6.

8. Mueller B, Kinoshita T, Peebles A, Graber MA, Lee S. Artificial intelligence, and machine learning in emergency medicine: a narrative review. *Acute Med & Surg.* 2022 Jan;9(1):e740.
9. Jiménez-Gaona Y, Rodríguez-Álvarez MJ, Lakshminarayanan V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. *Applied Sciences.* 2020; 10(22):8298. <https://doi.org/10.3390/app10228298> Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)
10. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med.* 2019 Jan;25(1):24-9.
11. IBM. What is explainable AI? <https://www.ibm.com/topics/explainable-ai> Accessed August 4, 2023
12. Andras P, Esterle L, Guckert M, Han TA, Lewis PR, Milanovic K, et al. Trusting intelligent machines: deepening trust within socio-technical systems. *IEEE Technology and Society Magazine.* 2018 Dec 4;37(4):76-83.
13. Gunning D, Aha DW. DARPA's explainable artificial intelligence program. *AI Mag* 2019;40(2):44-58.
14. Eliot L. AI ethics and autonomous systems lessons gleaned from that recent Alaska Airlines flight where the pilot and co-pilot disagreed prior to taking off and abruptly opted to taxi back to the terminal and go their separate ways. *Forbes.* Jul 23, 2022, <https://www.forbes.com/sites/lanceeliot/2022/07/23/ai-ethics-and-autonomous-systems-lessons-gleaned-from-that-recent-alaska-airlines-flight-where-the-pilot-and-co-pilot-disagreed-prior-to-taking-off-and-abruptly-opted-to-taxi-back-to-the-terminal-and-go-their-separate-ways/?sh=4a0b2d206623> Accessed August 19, 2023
15. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. 2021 <https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf> Accessed August 12, 2023
16. Vearrier L, Derse AR, Basford JB, Larkin GL, Moskop JC. Artificial intelligence in emergency medicine: benefits, risks, and recommendations. *J Emerg Med.* 2022 Apr 1;62(4):492-9.
17. Flahaux JR, Green BP, Skeet AG. Ethics in the Age of Disruptive Technologies: An Operational Roadmap. <https://mailchi.mp/scu/itec-handbook> Accessed August 30, 2023
18. Board DI. AI Principles: recommendations on the ethical use of artificial intelligence by the Department of Defense. Supporting document, Defense Innovation Board. 2019 Oct;2:3.
19. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell.* 2019;1,389–399.
20. Chan SL, Lee JW, Ong ME, Siddiqui FJ, Graves N, Ho AF, Liu N. Implementation of prediction models in the emergency department from an implementation science perspective—determinants, outcomes, and real-world impact: a scoping review. *Ann Emerg Med.* 2023;82(1):22-36.
21. Syed R. So sue me: who should be held liable when AI makes mistakes? Monash University Lens. March 29, 2023, <https://lens.monash.edu/@politics-society/2023/03/29/1385545/so-sue-me-wholl-be-held-liable-when-ai-makes-mistakes> Accessed August 3, 2023
22. Choi CQ. 7 revealing ways AIs fail: neural networks can be disastrously brittle, forgetful, and surprisingly bad at math. *IEEE Spectrum.* 21 Sep 2021 <https://spectrum.ieee.org/ai-failures> Accessed August 3, 2023
23. Singhal, K., Azizi, S., Tu, T. *et al.* Large language models encode clinical knowledge. *Nature.* 2023;620:172–180.

43. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366 (6464):447-453.
44. ACLU v ICE. United States District Court for the Southern District of New York. Case number 1:20-cv-01803. February 2020. www.nyclu.org/en/cases/jose-l-velesaca-v-decker-et-al Accessed September 5
45. NHS code of conduct for data-driven health and care technology, 19 February 2019: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>
46. Friedman B, Kahn PH. Human agency and responsible computing: implications for computer system design. *J Syst Software*. 1992;17(1):7–14.
47. Nissenbaum H. Accountability in a computerized society. *Sci Eng Ethics*. 1996;2:25–42.
48. Greene D, Hoffmann AL, Stark L. Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. 2019. Proceedings of the 52nd Hawaii International Conference on System Sciences 2019. <https://hdl.handle.net/10125/59651> Accessed September 16, 2023.
49. Pawlukiewicz AJ, Geringer MR, Davis WT, Nassery DR, April MD, Streitz MJ, Hyams JM, Martin AW, Martin SA, Oliver JJ. Interrater agreement of the HEART score history component: a chart review study. *JAMA Open*. 2022 Jun;3(3):e12732.
50. AHRQ. Updated review: diagnostic errors in the emergency department: a systemic review. Content last reviewed August 2023. Effective HealthCare Program, Rockville, MD: Agency for Healthcare Research and Quality. <https://effectivehealthcare.ahrq.gov/products/diagnostic-errors-emergency-updated/research> Accessed September 4, 2023.
51. IBM. AI Ethics. <https://www.ibm.com/impact/ai-ethics> Accessed August 3, 2023.
52. U.S. Food and Drug Administration. Artificial intelligence and machine learning in software as a medical device. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> Accessed September 11, 2023.
53. Centers for Medicare & Medicaid Services on 08/04/2022 Proposed Rule: Nondiscrimination in Health Programs and Activities. Federal Register August 4, 2022. <https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities> Accessed September 11, 2023.
54. Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. *JAMA*. 2023; 329: 285-6.
55. Graber MA, Bailey O. The wizard behind the curtain: programmers as providers. *Philos Ethics Humanit Med*. 2016;11:4.

Table 1: Recognized AI Ethics Guidelines [15-19] Four fundamental AI principles are vital for the ethical application of AI in medicine, particularly in EM. Although challenging to implement, they are essential to ethical AI use in EM.

1. Transparency:	Users should understand how AI systems function and arrive at their results.
2. Accuracy	AI systems must produce accurate, consistent, and unbiased results.
3. Privacy	AI systems should safeguard PII from unauthorized access or disclosure.
4. Dispute resolution	Mechanisms should exist for resolving disagreements between EPs and AI system results.

Table 2: Benefits of XAI in EM

Understanding AI Behavior	XAI empowers EPs to comprehend the reasoning behind AI predictions, fostering trust in the technology's accuracy and reliability.
Identifying Biases	Transparent explanations assist EPs in identifying potential biases within AI algorithms, promoting awareness and unbiased patient care.
Error Detection and Correction	XAI facilitates error detection by enabling EPs to identify instances where AI predictions may deviate from clinical reality, thereby enhancing patient safety.
Simplified Patient Communication	XAI-driven explanations can be translated into patient-friendly language, enabling EPs to convey AI decisions and address patient questions and concerns.

Table 3: Error Sources in AI

Ignoring Pretest

Source	Description
Inaccuracy	Datasets can be inaccurate due to human error, equipment failure, or other factors. This can lead to incorrect AI predictions.
Bias	Datasets can be biased in many ways,[25] including not accurately representing the population to which the AI model will be applied, being deployed inappropriately.
Probability	When using an AI model to predict a disease, it is important to consider the pretest probability of the disease in the population to which the model is being applied. This can affect the likelihood of a patient having the disease, even if the AI model predicts that they do.
Confirmation bias	Errors in the initial dataset can be amplified in subsequent machine learning iterations, leading to subsequent errors. This is because the AI model can learn to identify patterns that don't exist, further perpetuating errors.

Table 4: Informal Information AI Does Not Capture

Information type	Examples
Non-Verbal Cues	body language, facial expressions, eye contact, gestures
Emotional States	anxiety, fear, frustration, or sadness
Appearance and Demeanor	weight, skin color, energy level, posture, mobility.
Voice	pitch, tone, volume, and communication style

Table 5: Elements to Validate and Monitor AI during Program Adoption

<u>Stage</u>	<u>Elements to Evaluate</u>
Project Initiation	<ul style="list-style-type: none"> • User needs and workflow • Available data and quality • Algorithm prediction accuracy and consistency • System prototype
Pre-clinical Testing	<ul style="list-style-type: none"> • Algorithm performance using computer modeling • User experience • Cognitive overload
Pilot On-site Testing	<ul style="list-style-type: none"> • Workflow integration with clinical users' experience • Unintended consequences & benefits • Performance with larger real-world datasets
Clinical Testing	<ul style="list-style-type: none"> • Adverse events (safety, security, and system failure) • Misdiagnoses, over-/under-treatment • Effects of AI-based decision support on clinical outcomes • Level of user trust/over trust in the system
Degradation Monitoring	<ul style="list-style-type: none"> • Changes in accuracy or consistency • Track false positives and negatives • Assess patient outcomes • Compare with clinical guidelines

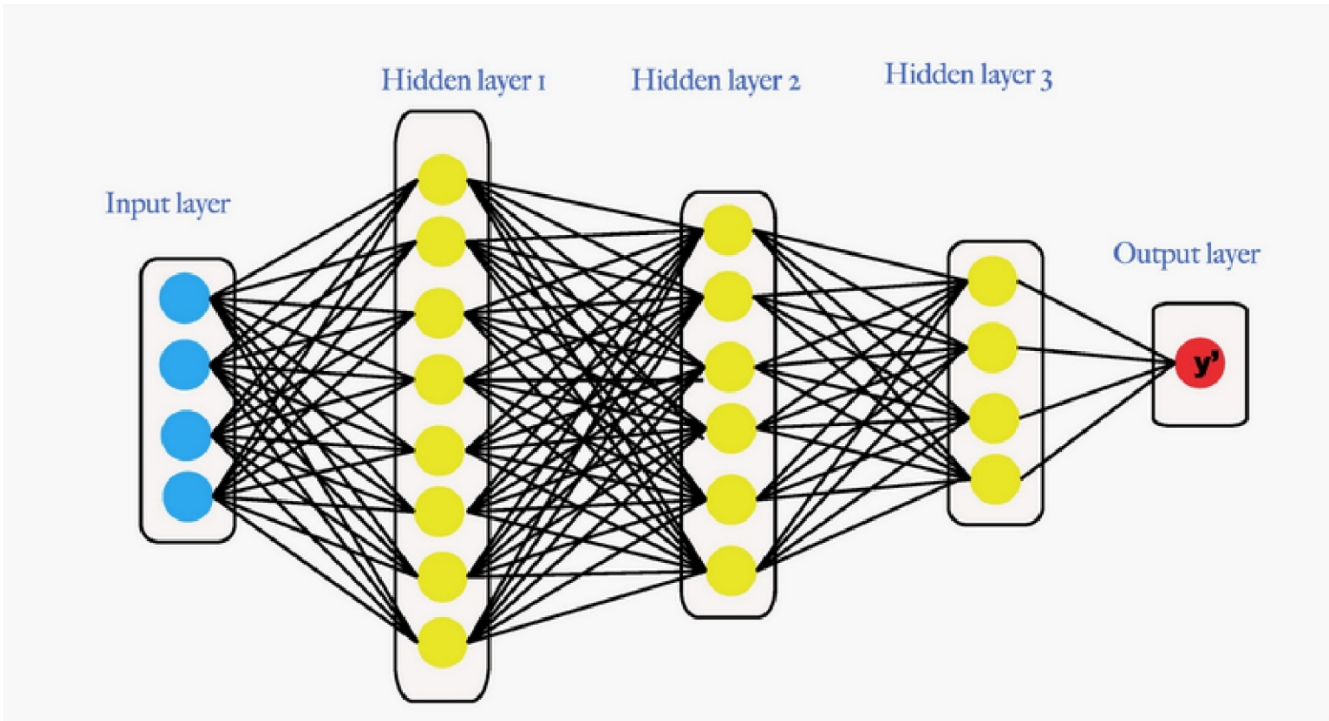


Table and Figure Legends

Figure 1: Deep Learning Network Schematic [9]

Table 1: Recognized Guidelines for AI Ethics [15-19]

Table 2: Benefits of XAI in EM

Table 3: Error Sources in AI

Table 4: Informal Information AI Does Not Capture

Table 5: Elements to Validate and Monitor during AI Program Adoption