THE CASE FOR AI AUTHORITY IN DIAGNOSTIC DECISIONS

ABSTRACT

This paper argues that artificial intelligence should transition from a supplementary tool to a central authority in medical diagnosis and treatment. While human physicians have traditionally held epistemic authority, their decision-making is constrained by cognitive limits and unequal access to medical knowledge. I examine the ethical implications of this shift through principles of beneficence and epistemic justice, arguing that clinical authority should follow the most reliable source of knowledge. In conclusion, I propose a hybrid model in which AI guides diagnosis, while physicians retain responsibility for ethical oversight and patient-centered care.

ROHIL WATWE

I. INTRODUCTION

Artificial intelligence (AI) has made significant strides in the medical field, specifically in disease diagnosis. As AI continues to improve, the role of human physicians as the centerpiece of the medical system has come into question. Traditional beliefs hold that human physicians are better equipped to be the final decision-makers due to their experience and ability to reason ethically. However, with AI demonstrating superior diagnostic capabilities in various medical fields, its role is shifting from a supplementary tool to a potential central authority in clinical decision-making.[1]

I argue that, as AI systems evolve, they will be better equipped to make clinical decisions than human doctors, and thus, should transition from a supplementary tool to a central authority in diagnosis and treatment. This argument rests on three claims. Firstly, AI's superior diagnostic accuracy makes deference to less reliable human judgment ethically problematic. Secondly, AI has access to our entire collection of medical data, surpassing human limitations and allowing for more consistent and equitable care. Third, while both AI and humans are susceptible to bias, AI's decision-making processes are correctable in ways that human cognition is not, making it a more ethically accountable tool.

The increasing efficacy of AI in disease diagnosis highlights its potential to revolutionize healthcare. AI has the ability to process data on a large scale, identifying patterns that humans cannot. Furthermore, AI can refine its accuracy by integrating real-time patient data. In contrast, humans are limited in this scope, affected by cognitive constraints and memory limitations. AI's ability to synthesize medical information at a rapid rate gives it a distinct advantage over human practitioners.[2] Additionally, human physicians are susceptible to bias and fatigue, which can impact diagnostic accuracy, whereas properly designed AI applies standardized criteria to all cases, allowing for greater objectivity in clinical decision-making. While critics say that AI can inherit bias based on training data, efforts to improve training dataset diversity should mitigate these risks.[3]

1 Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* 542, no. 7639 (February 2017): 115–18, https://doi.org/10.1038/nature21056.

2 Chayakrit Krittanawong et al., "Artificial Intelligence in Precision Cardiovascular Medicine," *Journal of the American College of Cardiology* 69, no. 21 (May 2017): 2657–64, https://doi.org/10.1016/j.jacc.2017.03.571.

3 Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," Science 366, no. 6464 (October 2019): 447–53, https://doi.org/10.1126/science.aax2342.

Critics argue that AI lacks fundamental humanizing characteristics that make human physicians particularly effective. These include traits such as empathy, ethical reasoning, and patient communication. Especially in end-of-life care, humans provide sensitivity and judgment that AI can struggle to replicate. Although these limitations do exist, they do not diminish the power of AI as a diagnostic and decision-making tool. Rather than serving only a supplementary role, AI should move toward central authority in diagnosis and treatment, while human physicians maintain ethical oversight and patient-centered care.

II. HISTORICAL CONTEXT

The use of AI in a medical context is not a new phenomenon. In the 1970s, a tool known as MYCIN was used to identify bacteria causing severe infections.[4] This system ran on the earliest iterations of AI. In the decades since, AI has improved substantially, both in diagnostic capability and efficiency. With the creation of machine learning algorithms, AI has become more capable of recognizing complex patterns in large datasets.

Modern AI systems use neural networks trained on millions of images and lab results. These systems can detect subtle indications of disease that humans are prone to miss. In fact, many modern specialties, including radiology, dermatology, pathology, and ophthalmology, utilize AI systems for pattern recognition, leading to enhanced patient outcomes.[5]

In addition to pattern recognition, AI also has the ability to outperform human physicians. A 2017 study by Ranav Rajpurkar et al. used a deep learning model to help detect pneumonia on chest X-rays.[6] When comparing performance with human physicians, the AI was able to outperform humans in detecting fourteen different diseases. Similarly, in a study done through Google's DeepMind AI system, artificial intelligence was found to detect over fifty retinal diseases with an

4 Edward H. Shortliffe, *Computer-Based Medical Consultations: MYCIN* (Elsevier, 1976), https://doi.org/10.1016/B978-0-444-00179-5.50001-9.

5 Geert Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (December 2017): 60–88, https://doi.org/10.1016/j.media.2017.07.005.

6 6 Pranav Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv* (2017), https://doi.org/10.48550/*arXiv*.1711.05225.

accuracy comparable to that of leading ophthalmologists.[7] There is an abundance of data that points to AI being able to outperform human physicians. These deep learning models also benefit from being trained on current medical data continuously, meaning diagnostic accuracy should keep improving over time.

III. LIMITATIONS OF HUMAN DECISION-MAKING

While physicians have extensive knowledge and clinical experience, their decision-making is prone to limitations. Cognitive biases have been shown to affect the present medical system deeply. Diagnostic errors affect an estimated 10–15% of medical cases, with cognitive factors contributing substantially to these mistakes.[8] These errors reflect recurring vulnerabilities in human reasoning under uncertainty rather than isolated incidents. For example, confirmation bias can lead physicians to settle on an initial diagnosis prematurely and overlook contradictory evidence, increasing the risk of misdiagnosis.[9] Additionally, physicians tend to work long hours, exacerbating issues such as physician fatigue. A 2004 study by Christopher P. Landrigan et al. found that diagnostic errors increased by around 20% during extended shifts in the emergency department.[10]

Physicians also face the challenge of staying up to date with current medical knowledge. One study asserts that in 2020, the doubling time of medical knowledge was around 73 days.[11] It is unrealistic for physicians to stay up to date on every advancement, something AI can accomplish with far fewer issues. With humans facing so many limitations, having AI as a secondary authority seems to be even more problematic.

7 Jeffrey De Fauw et al., "Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease," *Nature Medicine* 24, no. 9 (September 2018): 1342–50, https://doi.org/10.1038/s41591-018-0107-6.

8 David E. Newman-Toker et al., "Burden of Serious Harms from Diagnostic Error in the USA," *BMJ Quality & Safety* 33, no. 2 (January 2024): 109–20, https://doi.org/10.1136/bmjqs-2021-014130.

9 Pat Croskerry, "The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them," *Academic Medicine* 78, no. 8 (August 2003): 775–80, https://doi.org/10.1097/00001888-200308000-00003

10 Christopher P. Landrigan et al., "Effect of Reducing Interns' Work Hours on Serious Medical Errors in Intensive Care Units," *New England Journal of Medicine* 351, no. 18 (October 2004): 1838–48, https://doi.org/10.1056/NEJMoa041406.

11 Peter Densen, "Challenges and Opportunities Facing Medical Education," *Transactions of the American Clinical and Climatological Association* 122 (2011): 48–58.

IV. THE PROBLEM TO BE SOLVED

Despite evidence suggesting AI's growing capabilities in diagnostic tasks, human physicians remain the final decision-makers in diagnosis and treatment. This status quo is perpetuated due to beliefs that humans have professional expertise and provide ethical reasoning. However, by focusing solely on diagnostic tasks, where the ultimate goal is accuracy, the deference to human judgment is increasingly difficult to defend. Diagnostic error is one of the leading causes of preventable harm in healthcare. It is estimated that approximately 795,000 Americans suffer permanent disability or death each year due to diagnostic errors.[12] When cognitive limitations contribute to harm on this scale, the ethical implications of preserving exclusive human authority become significantly heightened.

This raises an important question: Should AI, which has been demonstrated to be superior at recognizing patterns and processing large quantities of medical data, remain behind the judgment of human physicians in diagnosis? While AI does not replace the need for empathy and ethical reasoning, it provides many other benefits. In places where accuracy and consistency are critical, such as when detecting diseases and interpreting scans, AI has proven to be more effective than humans. The issue now is not whether AI can match human diagnostic performance but rather when the medical system is ready to acknowledge the advantages of AI and reorganize accordingly.

V. DIAGNOSTIC ACCURACY AND ETHICAL AUTHORITY

To justify AI's transition from a supplementary tool to a central authority, it is not enough to demonstrate its superior diagnostic accuracy. One must also show that this superiority carries ethical weight. Specifically, in an epistemic and moral framework, clinical authority should follow the most reliable source of medical knowledge. If AI consistently outperforms human physicians in diagnostic accuracy, then continuing to emphasize less accurate human judgment becomes ethically indefensible. For instance, the landmark study by Andre Esteva et al. demonstrated that deep neural networks matched the performance of dermatologists when classifying skin cancer.[13] A 2019 study by Maged Nasser and Umi Kalsom Yusof showed

12 Newman-Toker et al., "Burden of Serious Harms," 109–20.

13 Esteva et al., "Dermatologist-Level Classification," 115–18.

that an AI system was the most accurate and extensively used model for breast cancer detection.[14] These are not isolated successes but rather indicative of a broader trend in AI-driven diagnostics.

These implications go far beyond performance metrics. The concept of epistemic injustice asserts that patients are wronged when their access to knowledge, including life-saving medical information, is obstructed by institutional structure.[15] Prioritizing the authority of human physicians, even when superior epistemic tools exist, is an ethical failure. It privileges the less accurate information of a human over a better-informed source, undermining a patient's right to the best knowledge available. Importantly, this argument concerns the ethical duty to deliver accurate clinical information, not the broader human responsibility of emotional support, which still rightly belongs to human physicians.

If the ethical foundation of medicine is, as Tom L. Beauchamp and James Childress argue, based on principles such as beneficence and nonmaleficence, then human physicians are under an obligation to provide the most accurate care available.[16] With this in mind, clinical authority must follow epistemic authority. Choosing human judgment for diagnosis over more accurate options goes against this moral responsibility.

VI. ACCESS TO DATA AND EPISTEMIC SUPERIORITY

A core reason AI should have the final say in diagnosis lies in its unparalleled access to medical knowledge, giving it epistemic superiority over human physicians and providing consistent care. An individual physician's training is strongly variable. They each have different educations, different clinical experiences, and the aforementioned limitations in memory and cognition. Unlike humans, AI systems have been consistently designed to interpret large amounts of medical data with precision.

This capacity goes beyond data processing. AI functions as a sort of collective intelligence. For example, large language models like

14 Maged Nasser and Umi Kalsom Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," *Diagnostics* 13, no. 1 (January 2023): 161, https://doi.org/10.3390/diagnostics13010161.

15 Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford University Press, 2007), https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.

16 Tom L. Beauchamp and James F. Childress, *Principles of Biomedical Ethics*, 7th ed. (Oxford University Press, 2013).

DeepMind's AlphaFold or IBM Watson Health represent a collection of thousands of medical studies and clinical data, far beyond what a single human could replicate. This epistemic advantage has both practical and moral implications. If the goal of a medical system is to maximize health outcomes, then using AI offers higher accuracy, faster throughput, and more consistent care.

Crucially, AI's scalability allows a "leveling of the playing field." Human expertise is generally unevenly distributed. Big cities might have hundreds of physicians and many subspecialists, while rural areas tend to suffer from physician shortages. AI offers a standardized diagnostic tool accessible to all areas. Every patient, regardless of socioeconomic status or geographic location, can benefit from a "best possible" doctor whose expertise is not affected by human limits. AI's superior access to data is not just a technical strength. Rather, it forms an epistemic foundation for centralized diagnostic authority. In a system that emphasizes equity, it is wrong to limit medical care to individual physicians when we have tools that provide a more complete view of the healthcare field.

VII. BIAS AND MORAL RESPONSIBILITY

Bias is central to the debate over medical authority, and AI holds the ethical advantage. This is not because AI is bias-free but because its biases are detectable and thus more manageable than those of human physicians. Human physicians, despite their extensive training, are susceptible to implicit biases. These have been shown to produce systemic inequalities in healthcare. For instance, a study in 2007 by Alexander Green et al. found that physicians demonstrated unconscious pro-white biases that affected clinical decisions.[17] Kelly M. Hoffman et al. documented that Black patients were less likely to receive pain medication due to false beliefs about biological differences.[18]

With regard to bias, a counterpoint that many critics mention is that AI systems have inherent biases that are directly comparable to those of their human counterparts due to biased training data. This critique conflates two distinct issues: the presence of bias and the correctability of bias. While both humans and AI systems may exhibit bias, the moral and practical significance lies in whether these biases can be recognized and fixed.

17 Alexander R. Green et al., "Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine* 22, no. 9 (September 2007): 1231–38, https://doi.org/10.1007/s11606-007-0258-5.

18 Kelly M. Hoffman et al., "Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites," *Proceedings of the National Academy of Sciences* 113, no. 16 (April 2016): 4296–4301, https://doi.org/10.1073/pnas.1516047113.

The human brain is opaque. Implicit biases operate at the subconscious level, making them difficult to detect and even harder to change. Attempts at bias training and awareness programs have yielded mixed results and tend to wear off over time.[19] In contrast, AI systems operate through code and models that are open to inspection. As Joy Buolamwini and Timnit Gebru demonstrated in their work on facial recognition AI bias, algorithmic disparities can be systematically detected and quantified.[20]

This distinction has great ethical implications. A just healthcare system should strive to not only reduce morally arbitrary disparities in treatment but also correct existing inequalities. AI offers a path toward this ideal, not because it is inherently unbiased but because its mechanism of decision-making is transparent and subject to external regulation. Tools such as fairness-enhancing algorithms and reweighting algorithms allow for continuous adjustment and improvement. In contrast, efforts to reform human decision-making tend to rely on individual behavioral change, which faces significant inertia.

In short, the question is not about whether bias exists in AI. Rather, it is about whether we can do something about it. In this aspect, AI offers the possibility of continuous ethical progress that human physicians lack.

VIII. ADDRESSING COUNTERARGUMENTS

A common objection to AI having the final say in medical diagnosis stems from the idea that AI lacks the capacity for empathy and moral reasoning. Human physicians offer emotional support and comfort, which is particularly useful for end-of-life care. Critics say that AI cannot possibly hope to replicate the bond and trust that patients develop with their doctors.[21] This concern is valid and important, par-

19 Temitayo A. Ogunleye, "Unconscious Bias," *Dermatologic Clinics* 41, no. 2 (April 2023): 285–90, https://doi.org/10.1016/j.det.2022.08.003.

20 Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018), https://proceedings.mlr.press/v81/buolamwini18a.html.

21 Mary Catherine Beach et al., "Relationship-Centered Care: A Constructive Reframing," *Journal of General Internal Medicine* 21, no. S1 (January 2006): S3–8, https://doi.org/10.1111/j.1525-1497.2006.00302.x.

ticularly for patient-centered care where communication and compassion are deeply interrelated with treatment outcomes.[22] Since AI lacks consciousness, critics state that it cannot replicate the nuanced decisions a human makes on a day-to-day basis. However, this thought misunderstands my argument. I do not suggest that AI should replace all human roles in medicine, but rather that AI should guide the technical core of medical decision-making, including diagnosis and treatment recommendations. AI's utility lies in improving clinical precision, not in replacing the human presence in care. A hybrid model is necessary. AI systems should guide diagnostic and treatment planning while physicians take responsibility for patient communication and ethical counseling.

Recent studies have shown that a hybrid model is already proving effective. A 2022 study by Khalid Al-Dasuqi et al. showed that AI systems that flag possible abnormalities are best used when their outputs are reviewed by human radiologists for accuracy and consistency.[23] The physician then interprets these findings within the context of the patient's complete medical history and clinical presentation. This division of labor strengthens the doctor-patient relationship by allowing the physician to focus on qualities only they can provide such as compassion and trust.

Another objection to AI systems is that they are susceptible to diagnostic errors. There have been instances of false positives, false negatives, and blind spots in AI training algorithms. In high-stakes areas of medicine, even a small margin of error could contribute to devastating consequences. This is a serious concern. AI is not infallible, and overconfidence in its abilities could lead to dangerous outcomes. However, this objection underestimates the detectability and the correctability of AI errors. AI systems leave a trail of logic and data behind their conclusions that can be iteratively improved through retraining, something that human cognition struggles to do. Moreover, comparative studies show that AI systems have the same rate of diagnostic error as human physicians in several fields. For example, a study in *Nature Medicine* showed that an AI model diagnosed pediatric illnesses correctly at a rate comparable to senior consultants.[24] The broader

22 Hui-Ching Weng et al., "The Effect of Surgeon Empathy and Emotional Intelligence on Patient Satisfaction," *Advances in Health Sciences Education* 16, no. 5 (December 2011): 591–600, https://doi.org/10.1007/s10459-011-9278-3.

23 Khalid Al-Dasuqi, Michele H. Johnson, and Joseph J. Cavallo, "Use of Artificial Intelligence in Emergency Radiology: An Overview of Current Applications, Challenges, and Opportunities," *Clinical Imaging* 89 (September 2022): 61–67, https://doi.org/10.1016/j.clinimag.2022.05.010.

24 Huiying Liang et al., "Evaluation and Accurate Diagnoses of Pediatric Diseases Using Artificial Intelligence," *Nature Medicine* 25, no. 3 (March 2019): 433–38, https://doi.org/10.1038/s41591-018-0335-9.

context matters here. Diagnostic error is one of the most common and costly types of medical error globally, with around 795,000 Americans becoming permanently disabled or dying annually because of it.[25] Even if AI systems are imperfect, they are systematically improvable. By leveraging large-scale data, AI is uniquely equipped to evolve in accuracy over time.

A final consideration lies in ethical accountability. If an AI system misdiagnoses a patient, who should take responsibility: the hospital, the AI developer, or the physician? This ambiguity is a big reason why many resist giving AI systems authority in the medical field. This concern is also justified. The legal infrastructure surrounding AI in medicine is currently underdeveloped. However, this is not a reason to exclude AI from decision-making. Rather, it should serve as a wake-up call to accelerate legal and ethical reform in the field.

Already, people are proposing frameworks for shared liability between developers and physicians.[26] These proposals generally follow a distributed responsibility model. In such a model, AI developers are responsible for the design, training data integrity, validation procedures, and post-market monitoring of their systems. Hospitals and healthcare institutions are responsible for proper implementation, oversight, and ensuring that AI tools are used within their validated scope. Physicians remain responsible for integrating AI outputs into the broader clinical picture and communicating decisions to patients. Liability is therefore not placed on a single actor but allocated according to control, expertise, and the capacity to prevent harm.

Within the hybrid model I propose, responsibility would follow function. AI systems hold epistemic authority in diagnosis as far as they provide the most accurate probabilistic assessment of disease. However, physicians would retain ultimate relational and ethical responsibility. This means that while AI may generate diagnostic conclusions, physicians remain accountable for final clinical implementation, contextual interpretation, and informed consent discussions. If harm arises from flawed model design or biased training data, responsibility would extend up to developers and regulatory bodies. If harm arises from improper use or misinterpretation of AI output, responsibility would lie with the clinical institution or physician.

This functional allocation reflects the operational structure of the hybrid model itself. Authority is divided according to comparative

25 Newman-Toker et al., "Burden of Serious Harms," 109–20.

26 Sara Gerke, Timo Minssen, and I. Glenn Cohen, "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare," in *Artificial Intelligence in Healthcare*, ed. Adam Bohr and Kaveh Memarzadeh (Elsevier, 2020), 295–336, https://doi.org/10.1016/B978-0-12-818438-7.00012-5.

strength: AI maximizes diagnostic reliability, while human physicians maintain ethical stewardship and patient-centered care. Accountability must mirror that same division. The presence of distributed responsibility does not weaken ethical clarity; rather, it acknowledges the collaborative nature of modern medical systems.

Ultimately, the potential benefits of AI-guided diagnosis outweigh the legal uncertainties. The challenges mentioned above are solvable, especially when compared against the status quo, where human error remains the third leading cause of death in the US.[27]

IX. CONCLUSION

The integration of artificial intelligence into clinical diagnosis should be considered a moral necessity. This paper has argued that, given AI's superior diagnostic ability, unparalleled access to comprehensive medical data, and capacity for bias correction, it should be given the final say in the diagnostic process. Deferring to human judgment for diagnosis represents a failure to uphold key principles of beneficence and nonmaleficence.

Importantly, this claim does not require replacing human physicians. Humans provide qualities such as empathy and ethical needs that are essential features of medical care, something an AI system cannot and should not replicate. Rather, I argue for a hybrid model in which AI provides the epistemic foundation for diagnosis and treatment, and human physicians focus on patient relations and the ethical dimension of care. This division of labor will allow both parties to play to their strengths.

Additionally, the expectation that AI will continue to improve diagnostically is not speculative optimism, but an inference supported by current trends in machine learning. AI systems benefit from increased volumes of high-quality health data and continual advancements in model architecture. For instance, recent developments allow AI models to be trained across multiple hospitals without compromising patient privacy, accelerating real-world improvement.[28] Moreover, regulatory bodies like the FDA have already approved adaptive AI systems that update themselves post-deployment, reinforcing the idea

27 Martin A. Makary and Michael Daniel, "Medical Error—The Third Leading Cause of Death in the US," *BMJ* (May 2016): i2139, https://doi.org/10.1136/bmj.i2139.

28 Nicola Rieke et al., "The Future of Digital Health with Federated Learning," *NPJ Digital Medicine* 3 (2020): 119, https://doi.org/10.1038/s41746-020-00323-1

that diagnostic accuracy by AI is not static but subject to continual refinement.

Objections about errors in AI diagnosis and the unclear nature of legal accountability are valid concerns. However, they are not disqualifying. Diagnostic errors already cost thousands of lives annually under human authority. Unlike human error, AI errors are traceable and improvable. Additionally, developing frameworks for shared legal responsibility is underway.

In conclusion, the status quo of human physicians having the final decision in the medical process cannot be ethically justified when more reliable systems are available. If medicine is to evolve with its ethical commitments, it must embrace AI not just as a tool but as a diagnostic authority. To delay this shift is to allow outdated norms to obstruct life-saving progress.

Rohil Watwe is a junior at Duke University in Durham, North Carolina, studying biology with a minor in chemistry and a certificate in innovation and entrepreneurship. His philosophical interests include bioethics and medical decision-making, with an emphasis on the role of artificial intelligence in healthcare