Journal of Electronic Research and Application, 2026, Volume 10, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Predictive Hegemony and the Interpretive Gap: On the Philosophical Premises and Ethical Boundaries of AI-Assisted Clinical Decision-Making

Yubo Wang

Sendelta International Academy, Shenzhen, Guangdong, China

Abstract: When debating the application boundaries of artificial intelligence (AI) predictive models in clinical medicine, it is clear that high predictive accuracy is desirable, but on its own, does not provide a sufficient condition for clinical application. Drawing on three example, AlphaFold's prediction of protein structure, radiomics' prediction of disease diagnosis and prognosis, and clinical risk scoring models' prediction of morbidity, and engaging with David Hume's empiricist skepticism towards causality, argue that interpretability is an indispensable condition in a discipline that values mechanistic explanation. In order for AI to evolve from a capable recommender to a decision-making machine that begins to develop a sense of individual self, several preconditions need to be fulfilled. Predictions must be falsifiable, minimally grounded in mechanistic knowledge, accompanied by partially explicable decision logics, designed to be fair to populations, and embedded in an error-tolerant architecture that enables correction and rollback. The utility of AI today lies in its ability to dramatically reduce the costs of human trial and error, but should not diminish the doctor's right to make, learn from, and reflect on mistakes as the final accountable link in the chain.

Keywords: AI interpretability; Clinical decision-making; Hume; Machine learning; Predictive models

Online publication: February 27, 2026

1. Introduction

The story of Elizabeth Holmes and Theranos as portrayed in "Bad Blood" makes for a cautionary tale for current debates on artificial intelligence in medicine. Theranos's revolutionary blood-testing technology proved to be a fraud that ultimately collapsed as a scandal. It fell apart not because the technology worked or did not work, but because the operations of the system were protected from verification and opaque to transparency. Lives were put at risk, clinical dangers were examined through microscopes, public trust was in jeopardy. This story serves as a warning to the black box in medicine [1].

Today, predictive models in clinical medicine powered by artificial intelligence enter the medical field with high levels of accuracy and efficiency. Their very success, however, gives rise to a fundamental philosophical and practical quandary: when sufficiently “precise,” should artificial intelligence predictions be allowed to directly decide the fates of patients? How, if at all, can the prediction-decision gap be bridged?

A philosophical analysis is an indispensable tool to address this question. David Hume’s empiricist skepticism towards causality argues that we never perceive the causal connection between two events; we only observe one event repeatedly following another. Because of this, our belief of A causes B was not derived from rational proof, but from habit, expectation, and mental pattern-making. The predictions of artificial intelligence, however brilliant, are in a sense of computational-generated form of correlation. Its constant conjunction is a logically, or perceptual certainty, not a psychological habit. Medicine, by contrast, requires causal explanation and mechanistic understanding. The relationship between these two epistemic approaches lies at the heart of the current debate.

In this study, we focus on three representative examples: AlphaFold in protein structure prediction, radiomics in disease prediction for diagnosis and prognosis, and clinical risk scoring models for measuring morbidity, as well as discussing the dilemma of high prediction but low interpretability [2,3]. Then, based on these analyses, we argue that there are five preconditions for AI to upgrade from the level of recommendation to the level of decision-making: falsifiable predictions; minimally grounded mechanistic knowledge; partially explicable decision logics; population-fairness; and the capability of error tolerance and rollbacks.

2. Hume’s empiricism and the nature of AI prediction

David Hume was one of the most vocal critics of the idea that causation can be understood through reason or a priori analysis. In his empiricism, what we think of as causal necessity is nothing more than the habit of the mind that comes from repeated experience [4]. This study has observed that event A is followed by event B repeatedly, and then we come to believe that A causes B. The belief in that causal relation, however, is merely psychological, and it does not stand up to scrutiny. There is nothing demonstratively necessary about the relation between A and B. When we see one billiard ball strike another and the second moves, we do not perceive causation; we perceive two events in constant conjunction. In Hume’s mind, causation is something we do not perceive in the world, but something was imposed on its mind. Our belief that the first ball caused the second to move is a product of custom and habit, a feeling of expectation imprinted by repeated experience.

The logic of contemporary deep learning is a computational instantiation of this Humean epistemology. At its core, AI is simply finding patterns of constant conjunction within a vast dataset. AlphaFold has found statistically reliable connections between protein sequences and their structures. Radiomics models have found similar connections between imaging features and disease states. What these systems output are extraordinarily fine-grained, highly reliable correlations. When applying it to the Humean epistemology, however, their outputs are predictive habits of pattern recognition rather than discoveries of underlying biophysical or pathophysiological mechanisms (**Figure 1**).

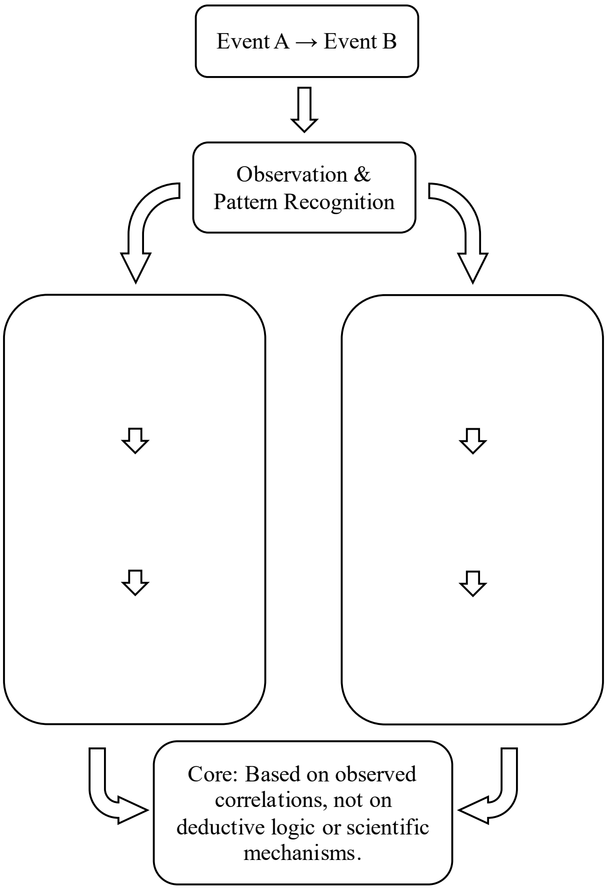
Event A → Event B
Observation & Pattern Recognition
Core: Based on observed correlations, not on deductive logic or scientific mechanisms.

Figure 1. The Humean-AI Epistemic Parallel. Just as Hume holds that constant conjunction produces a psychological expectation of causation, modern AI systems convert statistical conjunctions into predictive outputs. While this correlation-based mode of inference can generate reliable predictions, it operates independently of the causal and mechanistic reasoning on which clinical understanding typically depends.

But medicine cannot be cursorily reduced to a predictive science. In addition to predicting outcomes, medicine is also an effort to understand why things happen. Doctors need mechanistic explanations to support their diagnostic confidence, to choose therapies that target underlying causes, to anticipate side effects that have not yet been observed, and to live up to patients' expectations that medicine is not merely an art of black box prediction. For example, a physician treats a bacterial infection with an antibiotic because of a mechanistic understanding of cellular biology, not merely because antibiotics and recovery have been constantly conjoined in past data. Prediction without explanation would be epistemically incomplete and would erode patients' trust.

And so, when arrived to a fundamental epistemological contradiction, on one side stands AI prediction, which exposes the limits of computational inference by producing, in a Humean register, nothing more than extraordinarily fine-grained correlation, an advanced technological form of constant conjunction. On the other side is medicine, which insists on some degree of causal understanding and mechanistic clarity. The epistemic tension between these two camps is the central question in determining the prospects for AI in medical decision-making: whether, and under what conditions, prediction can ever be a viable substitute for, or complement to, explanation.

3. Case studies–The spectrum of predictive accuracy and interpretability

3.1. AlphaFold–The Humean habit of protein folding

AlphaFold is a revolution in protein structure prediction, highly accurate in ways almost indistinguishable from experimental structure determination [2]. Its predictions are staggering: given the sequence of amino acids that make up a protein, AlphaFold can propose a three-dimensional structure with great reliability. Its success is a triumph of identifying constant conjunctions between amino acid sequences and their three-dimensional configurations.

The interpretive profile, however, is different. From a Humean perspective, AlphaFold's internal process is a black box consists of deep learning layers that even the model's creators cannot read. This value, however, is secured because its output, a predicted 3D structure, is itself a mechanistic hypothesis. Protein biochemists can subject this output to the established causal models of chemistry and biophysics. AlphaFold provides the correlative what; we human provide the causal why.

Consequently, clinical AlphaFold is used as a hypothesis-generating tool that expedites discovery and guides experimental design. It remains a powerful recommender, not a decision-maker.

3.2. Radiomics–The opacity of visual conjunctions

Radiomics takes prediction to the next level by applying it to medical images themselves, extracting huge sets of quantitative features from CT, MRI, PET, and other common modalities [3]. These sets often exceed the capacity of the human visual system; models can use them to predict tumor malignancy, genetic subtypes, or treatment response, often outperforming experienced radiologists. Its predictive power derives from finding constant conjunctions between textural patterns and disease states.

The epistemic problem is interpretability. The model may have reached its conclusion based on correlations that may be unintuitive or unconnected to disease biology. A cluster of pixels might be constantly conjoined with malignancy, but if its biological basis is unknown, the prediction remains a pure Humean correlation. Given this output, clinicians cannot easily explain to patients why the model believes they will develop lung cancer rather than a benign disease. The epistemic gap makes it hard to build trust.

As a result, radiomics is today used as an auxiliary clinician-in-the-loop tool that plays the role of a second reader. The problem of interpretability is blocking its advance toward autonomy in decision-making, such as whether to order a biopsy, as that action requires a causal, explainable justification.

3.3. Clinical risk scoring models–Population level conjunctions

Machine learning has also been applied to electronic health records (EHR) to output risk scores stratifying patients by likelihood of future morbidity events such as heart failure or sepsis. These models are meaningful in their accuracy, and have become a fixture in early-warning beds. It marked an ability to computationally model constant conjunctions between a patient's documented clinical history and future morbidity on the population-level.

While sometimes built on interpretable linear models, models based on deep learning are also black boxed that offer little insight into the mechanism underlying their predictions. Their epistemic limitation is their statistical nature. They predict based on population-level conjunctions, which may not hold for a specific individual's pathophysiology. A high-risk score can be a correlation, not a diagnosis; it indicates that a patient shares features with others who had poor outcomes. It triggers heightened review but cannot specify the causal mechanism at work in this patient.

Among clinical workflows, they act like triage: they trigger an alert, prompt us to review, and suggest that we monitor more closely. They can identify that a patient needs our attention, but they still cannot tell us what to do. They are epistemically authoritative but not decisional authoritative.

4. The five preconditions for moving from recommendation to decision

The arguments above suggest that the use of AI in clinical practice cannot be justified on predictive accuracy alone. If the responsible passage of AI from an advisory role to a decisional role is to be achieved, a sequence of epistemic, ethical, and institutional safeguards must be established. Five preconditions are particularly salient for AI systems to meet in order to bridge the Humean-Mechanistic gap.

4.1. **Falsifiability**

AI predictions must be subject to rigorous testing and, where possible, refutation within real clinical settings. Performance demonstrated only on retrospective datasets is not enough; AI predictions should be tested on prospective studies and preferably randomized controlled trials. Only when a system's predictions can survive the possibility of being proven wrong can they approach a correlation toward real scientific credibility.

4.2. Minimal mechanistic knowledge

This study does not require AI to provide a complete causal explanation; but its predictions should not be completely divorced from contemporary biomedical knowledge. At the very least, its predictions should be compatible with established mechanisms of disease. If a deep learning model on radiology images were to predict malignancy, for example, the image features identified by the model should not be entirely without relation to proven pathological mechanisms. This anchoring in mechanism, however partial, protects AI from spurious correlations.

4.3. Articulable rationale

Decisions must be accompanied by reasons that clinicians and patients can understand, even if only in a limited sense. Techniques such as explainable AI, where relevant image regions are highlighted for the clinician, are examples of articulable rationales. Where full explainability is not possible, uncertainty should at least be quantifiable, for example through confidence scores, so that medical judgment can weigh prediction against risk.

4.4. Population fairness

Predictive models must be rigorously tested for bias across gender, race, age, and other relevant population characteristics. A correlation that holds for one demographic may not hold for another, for example, at equivalent algorithm-predicted risk scores, Black patients exhibit a significantly greater burden of illness compared to their White counterparts [5]. Therefore, it is imperative that AI systems are designed not to perpetuate or exacerbate healthcare disparities. Technical fixes in training data are not enough; there must also be institutional commitment to rigorously monitor outcomes across diverse populations.

4.5. Reversibility and human-in-the-loop

Finally, AI must be embedded in error-tolerant systems where there is a circuit breaker that guarantees reversibility when a prediction is suspected or shown to be erroneous. When AI recommends a treatment that turns out to be wrong, there must be a clinical professional who can immediately reclaim decision authority and take remedial action. Human oversight must be constant; otherwise, no AI-driven recommendation should be capable of irreversible harm without clinical intervention.

5. Conclusion

Hume's philosophy reminds us that even the most miraculous correlations generated by artificial intelligence are nothing more than super-correlations, a digital habit of the mind. Across the three case studies, another pattern is also evident: the need for interpretability rises in direct proportion to the clinical consequences at stake. When the consequences are minimal, correlations will do; when it is a question of life and death, explanation is required.

This perspective is completely reversed, destining Theranos's failure, when it treated its technology as an indubitable black box, protected from transparency, validation, and criticism. The responsible path for AI is the opposite: to build systems that are open to falsification, interpretation, and revision. The value of AI is to lower the cost of the trial and error that produces success, filtering choice and eliminating absurd options, without lowering the physician's right, and responsibility, to engage in the trial and error and reflective judgment [6].

The way forward is not substitution but symbiosis. A sustainable model of collaboration between human and machine will consist of a Humean AI, generating hidden correlations and precise predictive hypotheses, used by a mechanistic medicine that grounds these in causal understanding, ethical reflection and global clinical reasoning. Only then will the power of predictive technology be realized, while the core of the humanistic practice of medicine is protected and the safety of patients guaranteed.

Disclosure statement

The author declares no conflict of interest.

References

[1] Rudin C, 2019, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature Machine Intelligence, 1: 206–215.

[2] Jumper J, Evans R, Pritzel A, et al., 2021, Highly Accurate Protein Structure Prediction with AlphaFold. Nature, 596: 583–589.

[3] Gillies R, Kinahan P, Hricak H, 2016, Radiomics: Images Are More Than Pictures, They Are Data. Radiology, 278(2): 563–577.

[4] Hume D, 2000, An Enquiry Concerning Human Understanding. Oxford University Press.

[5] Obermeyer Z, Powers B, Vogeli C, et al., 2019, Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science, 366(6464): 447–453.

[6] Topol E, 2019, High-Performance Medicine: The Convergence of Human and Artificial Intelligence. Nature Medicine, 25: 44–56.