web of semantic
Journal of Interdisciplinary Science | ISSN: 2960-9550
Research Article Peer-review, Open Access

Hempel's D-N Model in a New Epistemological Era of AI

Hilary Anayochukwu Eke [1]
Remigius Achinike Obah [2]

Abstract: Carl Gustav Hempel's Deductive-Nomological (D-N) model emerged in the mid-20th century as a formal framework for scientific explanation, emphasizing deductive reasoning through general laws and initial conditions. However, the rise of Artificial Intelligence (AI) has transformed the landscape of scientific inquiry, challenging traditional explanatory models like the D-N model. AI systems, such as machine learning and neural networks, not only assist in processing vast data sets but also act as epistemic agents, raising questions about the nature of explanation and understanding. This essay critically examines the impact of AI on scientific explanation, exploring whether Hempel's model remains relevant or requires revision. The opacity of AI processes, particularly in 'black-box' models, complicates the deductive clarity required by the D-N model, prioritizing prediction over traditional causal understanding. Furthermore, the rise of explainable AI (XAI) and causal inference underscores the growing need for transparency, interpretability, and pragmatic relevance in explanations. This shift from deductive to inductive reasoning suggests that while Hempel's D-N model remains influential, it must evolve to accommodate AI's role in contemporary scientific discovery. The integration of AI demands a re-evaluation of what constitutes scientific explanation in the 21st century, balancing prediction, causality, and human-centred understanding.

Keywords: Scientific Explanation, Explainable AI (XAI), D-N Model, I-S Model, Understanding.

[1] *PhD, Department of Philosophy, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria hilary_eke@uniport.edu.ng*

[2] *Department of Philosophy, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria, remigius.obah@uniport.edu.ng*

Introduction

In the mid-20th century, Carl Gustav Hempel proposed the Deductive-Nomological (D-N) model, a formal representation of scientific explanation, which continues to inspire philosophical inquiry into the nature of explanation. The D-N model is grounded in the logic of deductive reasoning, where scientific explanations are structured in terms of general laws and initial conditions that deductively entail a phenomenon (Hempel, 1965). However, in the era of Artificial Intelligence (AI), scientific explanation is undergoing a paradigm shift. As AI systems take on roles as both objects and agents of scientific discovery, the assumptions underpinning traditional models of explanation, such as the D-N model, are

being challenged and transformed. In this essay, I explore the impact of AI on scientific explanation, focusing on the extent to which Hempel's D-N model remains relevant or requires revision. The intersection of AI, scientific explanation, and contemporary epistemological challenges is hereby critically examined within the scope of this study. Let us first attempt some definitions that provide a foundation for understanding both AI and scientific explanation within the framework of philosophy of science and thereby pinpoint the dual role AI plays in scientific practice and the evolving challenges it presents for traditional models of explanation.

Artificial Intelligence (AI) can be understood as both a tool and a form of epistemic agent. It engages with scientific inquiry by processing data, identifying patterns, making predictions, and even generating hypotheses in ways that challenge traditional notions of human-centred scientific explanation. AI operates through computational models, such as machine learning and neural networks, which enable it to analyse vast amounts of data and recognise complex relationships that may be inaccessible to human cognition. In this context, AI acts as an extension of human cognitive capacities, allowing for new forms of scientific understanding while raising critical questions about the nature of explanation and the transparency of its methods. AI thus, refers to computational systems designed to perform tasks that typically require human intelligence, such as reasoning, learning, problem-solving, and decision-making. Philosophers such as Margaret Boden view AI as not merely a tool but a mode of extending human cognitive processes. Boden defines AI as "the study of how to make computers do things which, at the moment, people do better" (Boden, 2016). This definition reflects the dual role AI plays as both an assistant to human scientific inquiry and an autonomous agent capable of deriving insights independently of human intervention. Accordingly, "AI systems can be considered agents capable of processing information in a way that mirrors certain aspects of human intelligence, including reasoning, learning, and decision-making" (Boden, 2016, p. 17).

Scientific explanation refers to the process by which a phenomenon is made intelligible by showing how or why it occurred fits into a broader framework of laws, principles, or causal mechanisms. Carl Hempel's *Deductive-Nomological (D-N) model* is one of the most influential accounts of scientific explanation. According to Hempel, a scientific explanation consists of a set of premises, including general laws and initial conditions, that logically entail the phenomenon to be explained (Hempel & Oppenheim, 1948). This formal model of explanation relies on logical deduction from known laws to specific events or phenomena. A scientific explanation is a logical account that makes a phenomenon intelligible by showing how it results from a set of general laws and specific conditions. Hempel defines it as a deductive argument in which the event to be explained is logically derived from premises consisting of scientific laws and particular facts (Hempel, 1965). In his own words, a scientific explanation is "an argument to the effect that the phenomenon to be explained, *the explanandum phenomenon,* was to be expected in virtue of certain explanatory facts…particular facts and uniformities expressible by means of general laws (Hempel, 1965, p. 336).

The D-N Model of Explanation

The D-N model, which is one of the kinds of the covering-law model, conceptualises scientific explanation as a deductive argument in which the phenomenon to be explained (the *explanandum)* is derived from a set of premises that include both general laws and specific initial conditions (Hempel & Oppenheim, 1948). According to Hempel, scientific explanations have the following structure:

1. A set of premises (the *explanans*) composed of at least one general law and particular facts.
2. The conclusion (the *explanandum),* which is the event or phenomenon being explained, must follow logically and necessarily from the premises.

For example, to explain why a particular sample of gas expands when heated, we appeal to the ideal gas law (a general law) and the specific conditions of temperature and volume at which the gas was heated (initial conditions). The expansion of the gas is the deductively necessary consequence of these premises. In this model, explanation is intimately tied to prediction: if we know the laws and conditions, we can

predict the outcome. Hempel's account of explanation has had a significant influence on the philosophy of science, but it has also faced numerous criticisms. Notably, philosophers such as Wesley Salmon (1984) and Bas Van Fraassen (1980) argued that the D-N model overlooks the pragmatic dimensions of explanation, including its context-dependence and relevance to human understanding. Nevertheless, the model has remained a cornerstone for understanding the role of laws in explanation and prediction.

The Dawn of a New Epistemological Era of Artificial Intelligence

The rise of AI marks a profound shift in scientific practice and, consequently, in the philosophy of science. AI technologies, particularly machine learning (ML) and deep learning algorithms, have become powerful tools for data analysis and pattern recognition, outperforming traditional statistical methods in many fields (Russell & Norvig, 2021). However, the epistemological status of these tools is controversial. AI systems, especially "black-box" models, often produce accurate predictions without providing interpretable explanations of how or why those predictions are made (Burrell, 2016). This raises critical questions for the philosophy of scientific explanation: Can an opaque system that offers reliable predictions be said to provide a genuine scientific explanation? Is explanation still necessary if AI can predict phenomena with high accuracy?

We need to examine the nature of AI-driven explanations in order to understand how AI challenges the D-N model. Deep neural networks and some other machine learning models rely on large datasets and computational processes to identify complex patterns. However, these processes are not always intelligible to humans, and their internal workings may defy traditional forms of logical representation. Consequently, while an AI model might predict the behaviour of a system with remarkable precision, it may not do so in a way that satisfies the conditions of a deductive-nomological explanation. For instance, in cancer diagnosis, an AI algorithm might identify a tumour's malignancy with high accuracy by analysing vast amounts of medical data, but its classification process may resist formalisation into a set of general laws and initial conditions.

This lack of transparency has led some scholars to describe AI models as "epistemic tools" rather than sources of explanation (Creel, 2020). AI predictions may be reliable, but their lack of interpretability challenges the notion that they provide understanding in the traditional sense. From this perspective, the predictive power of AI does not necessarily translate into explanatory power, at least not in the terms defined by the D-N model.

The Role of Causal Explanations in the AI Context

One of the central critiques of the D-N model is its neglect of causal explanation. Hempel's framework is fundamentally non-causal, focusing instead on the logical relationship between premises and conclusions. However, much of contemporary science is concerned with uncovering causal mechanisms that explain why certain phenomena occur. Salmon (1984) famously argued for a shift towards a causal-mechanical model of explanation, which emphasises the discovery of underlying mechanisms as the primary goal of scientific inquiry.

AI, particularly ML models, complicates this causal understanding. Many AI systems function by identifying correlations in data without necessarily uncovering the underlying causal mechanisms that produce the observed patterns (Pearl, 2009). This has led to a growing interest in the development of causal models within AI research, as exemplified by Judea Pearl's (2018) work on causal inference in machine learning. Pearl distinguishes between mere correlation and true causal understanding, arguing that AI systems must move beyond identifying patterns to discerning causal structures if they are to provide genuine explanations.

The integration of causal inference into AI has the potential to bridge the gap between prediction and explanation. For instance, in medical science, AI models that incorporate causal reasoning may not only predict disease outcomes but also offer insights into the mechanisms that drive these outcomes, thus satisfying the explanatory demands of the D-N model and causal-mechanical models alike. However,

even with causal inference, the opacity of certain AI processes remains a concern. While causal models can enhance interpretability, the complexity of deep learning networks may still resist full transparency.

Pragmatic and Contextual Dimensions of Explanation

Another important critique of the D-N model comes from pragmatic theories of explanation, such as those developed by van Fraassen (1980). According to van Fraassen's "pragmatic" view, explanations are not merely logical derivations from laws and initial conditions; they are context-dependent responses to specific questions posed by agents with particular interests and concerns. Eke (2020) presents the pragmatic view of explanation from the perspective of Jan Faye summarising the pragmatic view as an appropriate answer to an explanation-seeking question where the appropriate answer is determined with respect to the context of the discourse, including the explainer's interest and background knowledge. In this view, explanation is an inherently human-centred activity, and its success depends on the relevance and intelligibility of the explanation to the audience.

The relevance of pragmatic theories of explanation to AI is evident when we consider the increasing role of AI in decision-making. AI systems are often deployed in contexts where their outputs are used to inform human actions, such as in healthcare, criminal justice, and finance. In these contexts, the adequacy of an AI-generated explanation depends not only on its predictive accuracy but also on its capacity to be understood and acted upon by humans (Doshi-Velez & Kim, 2017). This raises ethical and epistemological questions about the transparency and accountability of AI systems. If an AI model cannot provide an explanation that is intelligible to its users, can it be said to offer a satisfactory explanation at all, even if its predictions are accurate?

In response to these concerns, the field of "explainable AI" (XAI) has emerged, aiming to develop AI systems that are more transparent and interpretable (Adadi & Berrada, 2018). The goal of XAI is to produce explanations that are not only reliable but also accessible to human agents. This pragmatic approach aligns with van Fraassen's emphasis on the human context of explanation, suggesting that scientific explanation in the era of AI must be evaluated not only in terms of its logical structure but also in terms of its relevance and usability for human decision-makers.

Reconsidering Hempel's D-N Model in the Age of AI

Given these developments, how should we assess the relevance of Hempel's D-N model in the context of AI? While the D-N model remains a powerful framework for understanding traditional forms of scientific explanation, its limitations become apparent in the face of AI technologies that prioritise prediction over explanation, or that generate explanations that are opaque to human understanding. The rise of AI calls for a re-examination of the criteria by which we evaluate scientific explanations. Prediction alone is no longer sufficient; transparency, interpretability, and causal understanding have become equally important.

At the same time, AI may prompt a reconceptualization of what counts as a "law" in scientific explanation. Traditional D-N explanations rely on general laws that apply universally. However, AI systems often operate in domains where there are no clear-cut laws, but rather probabilistic patterns or complex relationships that defy simple generalisation (Bailer-Jones, 2009). In this context, the very notion of a covering law may need to be redefined to accommodate the kinds of regularities that AI systems uncover.

Moreover, AI highlights the role of non-deductive forms of reasoning in scientific explanation. Inductive reasoning, which has traditionally been viewed as subordinate to deductive reasoning plays a central role in AI-driven discovery. Machine learning models are built on inductive generalisations from large datasets, rather than on the deduction of specific outcomes from general laws. This shift in emphasis from deduction to induction challenges the primacy of the D-N model in contemporary scientific practice.

Hempel's covering-law model includes both the Deductive-Nomological (D-N) model and the Inductive-Statistical (I-S) model. The I-S model addresses explanations based on statistical or probabilistic generalizations, which seems especially relevant when considering AI systems like machine learning that

often rely on probabilistic reasoning rather than strict deductive laws. The I-S model has the challenge of high probability requirement.

Challenge of the High-Probability Requirement in AI

In Hempel's I-S model, the strength of the explanation depends on the probability of the *explanandum* (the event or phenomenon being explained) given the *explanans* (the premises or laws). For an I-S explanation to be strong, the probability must be close to 1, i.e., the outcome must be highly probable. However, AI models, particularly machine learning models, often do not guarantee high-probability outcomes. Instead, they may predict outcomes with a range of probabilities that reflect the uncertainty inherent in the data or the model's limitations. For instance, classification models (like neural networks or decision trees) assign probabilities to different classes, but none of these probabilities may approach 1. In complex domains like medical diagnosis or weather forecasting, AI models may provide probabilistic predictions that are far from certain (e.g., a 70% chance of rain, or a 60% chance that a tumour is malignant).

This discrepancy between the requirement for high probability in the I-S model and the often lower probabilities generated by AI creates a tension. AI's probabilistic outputs are useful for prediction but may fall short of the rigorous standards required for scientific explanation in the I-S framework.

We can point out certain implications for AI methodology given this tension created by high requirement of probability. It is obvious that there is a lack of high-probability outcomes. In cases where AI predictions have moderate to low probabilities, the I-S model would struggle to classify these predictions as valid scientific explanations. This could lead to a scenario where highly effective AI models – ones that provide valuable insights or predictions – are not considered to offer "explanations" in the traditional scientific sense under Hempel's model.

Epistemic opacity: AI models, especially deep learning algorithms, are often opaque (commonly referred to as "black boxes"), making it difficult to map their probabilistic predictions to clear causal mechanisms or laws. This further complicates their alignment with the I-S model, which expects a clear linkage between statistical generalizations and outcomes.

Potential Solutions

Loosening the High-Probability Requirement:

One solution could be to relax the requirement for a probability close to 1 in the I-S model when applied to AI-driven explanations. This proposal aligns with contemporary discussions in the philosophy of science and AI. A broader interpretation of probabilistic explanations could enhance the practical utility of AI predictions, even when probabilities are moderate or low. Recent studies on low-probability explanations indicate they can be valid. Clatterbuck, for instance, argues that low-probability scientific explanations can still be valid, challenging the received view that only high-probability outcomes qualify as explanations (Clatterbuck, 2020). This perspective suggests that explanations with moderate probabilities, like a 70% chance of rain, can still provide valuable insights in uncertain environments.

Statistical understanding in AI is another support for the reduction of high probability requirement. Ebert-Uphoff emphasizes the importance of statistical understanding in AI models, which can be enhanced through Explainable AI (XAI) methods that clarify input-output relationships (Ebert-Uphoff, 2024). Löfström et al. (2024) introduce Calibrated Explanations, which quantify uncertainty in feature importance, thereby supporting the notion that probabilistic predictions can be informative even when not highly. While Hempel and others may argue that low-probability explanations lack robustness, the evolving landscape of AI and statistical modeling suggests that practical utility and predictive value can justify their inclusion in explanatory frameworks.

Philosophers and scientists could argue for a broader interpretation of probabilistic explanation, where moderate or even low probabilities can still count as explanations if they offer significant predictive or practical value. This would align the I-S model with AI's probabilistic and statistical nature. For example,

if an AI model predicts a 70% chance of rain, while this is not close to certainty, it might still be considered an explanation under a revised I-S model that acknowledges the usefulness of probabilistic predictions in uncertain environments.

Emphasizing Pragmatic Explanations

Hempel's I-S model might be adapted to focus more on pragmatic dimensions of explanation. AI predictions are often highly actionable even when they are probabilistic. This could align with Van Fraassen's pragmatic theory of explanation, where the success of an explanation depends on its usefulness to an audience rather than its strict logical structure or probabilistic closeness to 1. In this view, the AI prediction of a 60% likelihood of disease could be considered a valid explanation in medical practice, where such a probability is actionable and leads to life-saving decisions.

Developing Hybrid Models:

AI explanations often mix inductive, probabilistic reasoning with elements of causal explanation. A hybrid model of explanation could integrate probabilistic reasoning (I-S) with a focus on causal mechanisms (like Wesley Salmon's causal-mechanical model). AI explanations could be seen as partly probabilistic and partly causal, depending on the problem context and the model's structure. For instance, causal models in AI (like Judea Pearl's causal inference work) could help bridge the gap by making AI predictions more interpretable, combining the I-S model's probabilistic nature with causal explanations to enhance the transparency and scientific value of AI-driven predictions.

The Role of Explainable AI (XAI):

The rise of Explainable AI (XAI) offers a solution to some of the transparency issues that complicate AI's alignment with the I-S model. XAI systems aim to make AI models more understandable and interpretable, enabling human users to grasp the reasoning behind probabilistic predictions. This interpretability can offer a complement to probabilistic reasoning, providing users with an explanation that, while not highly probable, is comprehensible and actionable. For example, an XAI system might provide a breakdown of why a 70% probability of rain was predicted, based on causal factors like temperature, humidity, and wind patterns, giving the user a clearer understanding of the underlying mechanisms.

The requirement of high probability in Hempel's I-S model poses challenges for AI-driven explanations, as AI often operates with moderate or lower probabilities. To address this, philosophers of science can consider loosening the probability requirement, adopting pragmatic explanations, or exploring hybrid models that combine probabilistic and causal reasoning. Additionally, Explainable AI (XAI) can play a crucial role in making probabilistic AI explanations more interpretable and scientifically valuable, even when they fall short of strict probabilistic thresholds. By adapting the I-S model to accommodate AI's probabilistic framework, we can better integrate AI into contemporary theories of scientific explanation.

Conclusion

The development of AI represents a transformative moment for scientific explanation. As AI systems become increasingly central to scientific discovery, they challenge traditional models of explanation, including Hempel's D-N model. While the D-N model remains influential, its limitations in the context of AI are evident. AI-driven explanations often prioritise prediction over causal understanding, raise concerns about transparency and interpretability, and shift the emphasis from deductive to inductive reasoning. At the same time, the rise of XAI and causal inference within AI research suggests new possibilities for integrating AI with traditional forms of scientific explanation. In revisiting Hempel's D-N model in the era of AI, we are forced to reconsider the nature of scientific explanation itself. AI technologies invite us to expand our epistemological frameworks, recognising that explanation is not solely a matter of deductive logic, but also of causal understanding, transparency, and pragmatic relevance. As AI continues to evolve, so too must our philosophical models of explanation, adapting to the new epistemic challenges and opportunities that these technologies present.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.
2. Bailer-Jones, D. M. (2009). *Scientific models in philosophy of science*. University of Pittsburgh Press.
3. Boden, M. A. (2016). *AI: Its nature and future*. Oxford University Press.
4. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512.
5. Clatterbuck, H. (2020). A Defense of Low-Probability Scientific Explanations. *Philosophy of Science*. https://doi.org/10.1086/706082.
6. Creel, K. (2020). Transparency in complex computational systems. *Philosophy of Science, 87*(4), 568-578. https://doi.org/10.1086/709731.
7. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608.*
8. Eke, H. A. (2020). Postmodernism and pragmatic model of scientific explanation. *Sapientia Journal, 12*, pp. 143-153.
9. Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press.
10. Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135-175.
11. Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
12. Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016.*
13. Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
14. Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
15. Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.
16. Vidal, G. (2024). 9. Explaining Explanations in Probabilistic Logic Programming. arXiv.org, https://www.doi.org/10.48550/arxiv.2401.17045.