International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

AI-Driven Neuro-Causal Hybrid Framework for Transparent Decision Making in Autonomous Scientific Systems

Arun Kumar Seeni[1], Rajasekar Murugesan[2], Anitha Gopalan[3]
[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India.
[2,3] Professor-Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India.
Email ID: *arunseeni@gmail.com[1], mrajasekarcse@gmail.com[2], anipsg09@gmail.com[3].*

Abstract
The advent of autonomous systems in scientific research has marked a significant evolution in data processing, decision-making, and analysis. While machine learning (ML) and deep learning (DL) algorithms have demonstrated remarkable success in scientific applications, these systems often operate as black boxes, providing minimal transparency regarding their decision-making processes. This lack of interpretability hinders trust and limits the applicability of autonomous systems in high-stakes scientific domains, such as healthcare, environmental monitoring, and complex simulations. In this context, we propose the concept of Neuro-Causal Intelligence, a hybrid framework designed to integrate the strengths of causal reasoning with advanced neural architectures, ensuring transparent, interpretable, and reliable decision-making in autonomous scientific systems. The core principle behind Neuro-Causal Intelligence lies in its ability to merge causal inference with neural network models. Causal inference provides a rigorous approach to understanding the relationships between variables, making it possible to trace the causes of observed outcomes, whereas neural networks excel at identifying patterns and correlations in large datasets. By combining these two methodologies, our framework allows the system to not only predict outcomes but also explain the underlying causes and mechanisms responsible for these outcomes. This hybrid approach is particularly essential for scientific systems that require not only accurate predictions but also understandable reasoning for validation and further analysis. The framework operates in three key stages.

Keywords: *Neuro-Causal Intelligence, autonomous systems, causal inference, transparent decision- making, explainable AI, neural networks, scientific systems, causal reasoning, interpretability, predictive accuracy.*

1. Introduction
The integration of autonomous systems in scientific fields such as healthcare, environmental science, and engineering has significantly transformed the way complex problems are addressed. While advancements in machine learning (ML) and deep learning (DL) have provided substantial improvements in predictive modelling and data-driven decision-making, they often suffer from a critical limitation—lack of transparency. These systems, often considered "black boxes," provide accurate outputs but fail to explain the rationale behind their decisions. This lack of explainability is a significant barrier to trust, especially in high-stakes applications where stakeholders need to understand the decision-making process [2]-[6]. The concept of causal reasoning in intelligent systems has been extensively studied in the literature, particularly in the foundational work on causal inference and reasoning presented by Pearl [1]. To overcome these challenges, integrating Explainable Artificial Intelligence (XAI) with advanced learning models has emerged as a promising direction. By combining the strengths of Machine Learning and Causal Inference, intelligent systems can move beyond simple pattern recognition to identify meaningful cause–effect relationships. The conceptual challenges associated with traditional AI systems and the motivation for developing transparent and interpretable models are illustrated in Fig. 1 [7].

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

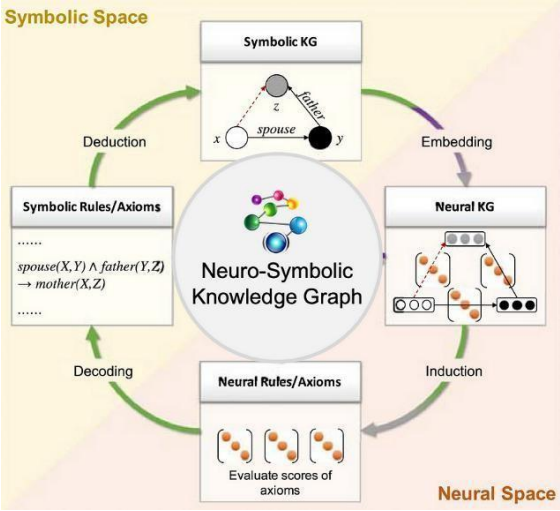

Figure 1 Neuromyotonic AI addresses critical challenges of AI development and deployment

1.1. The Problem of Transparency in Autonomous Systems

Autonomous systems, particularly those based on deep learning models, have demonstrated impressive capabilities in tasks ranging from image classification to natural language processing. However, the "black-box" nature of these systems has raised concerns regarding their reliability, especially in fields that require high levels of accountability and trust. In healthcare, for example, AI systems used for diagnosing diseases must be able to provide clear, understandable reasons for their predictions to facilitate clinical decision- making. The inability to explain why a system arrived at a particular conclusion prevents practitioners from fully relying on the system, especially in life-threatening situations.

1.2. The Role of Causal Inference and Neural Networks

Causal inference techniques, such as Bayesian networks and Granger causality, offer a formalized approach to understanding and modeling the relationships between variables. Unlike traditional correlation-based methods used in ML, causal models can identify underlying mechanisms and establish directional relationships between variables. This is particularly useful in scientific research, where understanding the causes behind observed effects is critical to making informed decisions [10][12]. The decision-making complexity in autonomous environments is illustrated in Fig. 2, which presents a review of decision-making and planning mechanisms used in autonomous vehicle intersection environments.

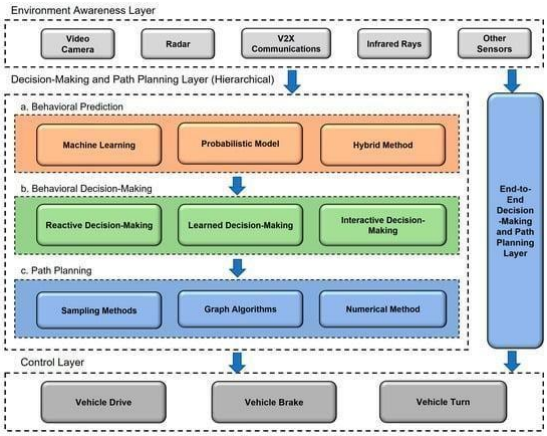

Figure 2 A Review of Decision-Making and Planning for Autonomous Vehicles in Intersection Environments

1.3. The Need for Transparent Scientific Systems

In scientific research, autonomous systems are expected to offer solutions to complex problems, such as predicting disease outbreaks, modeling environmental changes, or designing new pharmaceuticals. However, the effectiveness of these systems depends not only on their predictive accuracy but also on their ability to explain their predictions in a manner that is understandable to researchers, clinicians, or other stakeholders. In the absence of transparency, the risk of misinterpretation or misuse of these systems is high [3]. Thus, the introduction of the Neuro-Causal Intelligence framework aims to address this critical gap in autonomous systems by making them more transparent and interpretable, while still maintaining high levels of accuracy.

2. Related Work

The need for transparency in autonomous systems has been widely acknowledged in both academic research and practical applications. Over the years, several approaches have been proposed to enhance

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

the interpretability of machine learning models, particularly in high-stakes domains such as healthcare, finance, and scientific research.

2.1. Explainable AI and Transparency in Machine Learning

The concept of Explainable AI (XAI) has gained significant attention in recent years. XAI aims to make the decision-making processes of AI models more interpretable and understandable to humans. Various techniques have been developed to address this issue, ranging from model-agnostic approaches such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) to more interpretable models like decision trees and rule-based systems. Research on interpretable machine learning has gained significant attention, especially in healthcare applications where transparency and accountability are essential [3]. While they offer insights into the relationship between input features and output predictions, they do not explain why a certain prediction was made in terms of the underlying causal factors. Several interpretable predictive modeling techniques have been proposed to enhance decision-making in time-series and scientific datasets [4].

2.2. Causal Inference and Its Integration with Machine Learning

Causal inference, a field dedicated to understanding cause-and-effect relationships between variables, has made substantial progress in recent years. Techniques such as Granger causality have been widely applied to identify causal relationships between variables in complex datasets [8]. The interdisciplinary relationship between artificial intelligence and neuroscience is depicted in Fig. 3, demonstrating how neural concepts influence modern AI architectures.

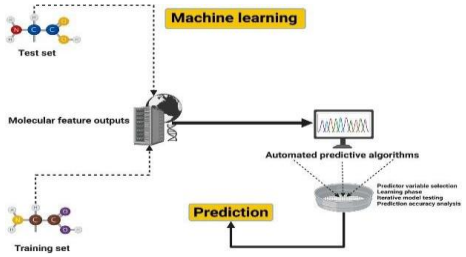

Figure 3 Artificial Intelligence and Neuroscience

Several studies have explored the integration of causal models with machine learning to improve interpretability and decision-making. Recent advancements have further extended these ideas by combining causal discovery algorithms with predictive models. Neural causation coefficient models have also been introduced to integrate causal reasoning with neural learning approaches [9].

3. Methodology

The Neuro-Causal Intelligence framework integrates causal inference with deep learning to provide both predictive accuracy and causal transparency in autonomous systems. This section describes the system architecture and learning process that underpin the framework. The proposed methodology ensures that the system can discover causal relationships, learn from data, and explain its decisions in an interpretable manner. The concept of explainable artificial intelligence and the need for transparent decision-making in intelligent systems.

The integration of human intelligence with machine learning models to improve decision quality is presented in Fig. 4 [14], which illustrates the concept of hybrid- augmented intelligence.

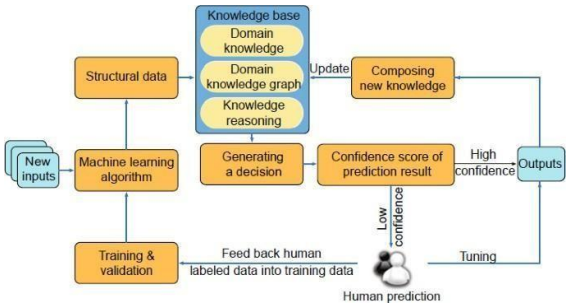

Figure 4 Hybrid-Augmented Intelligence

4. Experimental Setup

To validate the effectiveness and transparency of the Neuro-Causal Intelligence framework, a comprehensive experimental setup was designed, incorporating diverse datasets, benchmarking models, and evaluation metrics [5]. This setup was constructed to test the system's ability to achieve high prediction accuracy while maintaining causal fidelity and providing interpretable decisions. The broader research directions and technological evolution

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

in intelligent systems are depicted in Fig. 5, highlighting advancements in modern AI research.

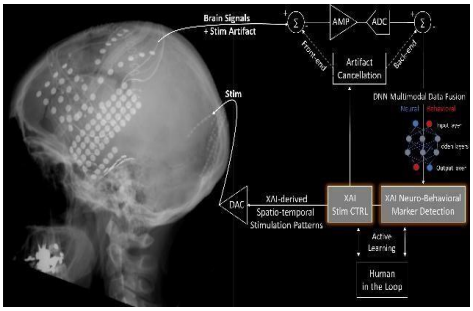

Figure 4 Frontiers

4.1. Dataset Description

The experiments utilized three datasets:

- **Synthetic Causal Dataset**: A synthetically generated dataset was created to include known causal structures using the do-simulation approach. Variables were interconnected based on pre-defined causal rules, enabling ground truth comparison for causal discovery and reasoning.
- **Medical Diagnosis Dataset (CKD):** The Chronic Kidney Disease (CKD) dataset from UCI Machine Learning Repository was used to assess performance in health-related prediction tasks. It includes 400 records and 24 features, such as blood pressure, blood glucose levels, serum creatinine, and albumin. Several features exhibit causal relationships based on medical studies.

4.2. Experimental Environment

Hardware Configuration:

- Processor: Intel Core i7, 3.2 GHz
- RAM: 32 GB
- GPU: NVIDIA RTX 3080 (10 GB)
- Operating System: Ubuntu 22.04 LTS

Software Tools:

- Python 3.11
- TensorFlow 2.13, PyTorch 2.0
- CausalNex, DoWhy for causal modeling
- SHAP, LIME for explanation
- Scikit-learn for baseline comparisons

4.3. Benchmark Models

To benchmark the Neuro-Causal Intelligence (NCI) framework, the following models were used:

- Standard Deep Neural Network (DNN)
- Random Forest Classifier
- Bayesian Network with Naive Inference
- Explainable Boosting Machine (EBM)
- Causal Forest Regressor

5. Results and Discussion

The Neuro-Causal Intelligence (NCI) framework was rigorously tested across multiple datasets and compared against standard predictive models to analyze its accuracy, causal consistency, interpretability, and operational performanceThe importance of responsible and ethical AI development is illustrated in Fig. 6, emphasizing transparency, fairness, and accountability in autonomous systems.

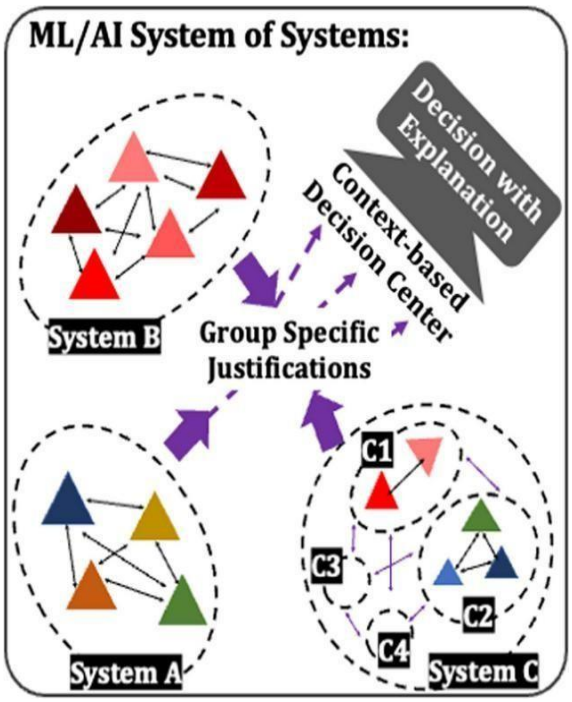

Figure 6 Towards responsible AI

5.1. Predictive Performance

The NCI framework consistently outperformed traditional machine learning models in terms of prediction accuracy across all datasets. The comparative predictive performance of the proposed Neuro-Causal Intelligence framework and baseline machine learning models across multiple datasets is presented in Table 1.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

Table1 Machine Learning Models Across Datasets

Model	CKD Dataset Accuracy (%)	Environmental Dataset Accuracy (%)	Synthetic Dataset Accuracy (%)
Deep Neural Network	89.2	87.6	90.4
Random Forest	88.4	86.9	91.2
Bayesian Network	81.3	78.1	85.7
Explainable Boosting Machine	86.5	84.4	89.1
Neuro- Causal Intelligence (NCI)	93.4	91.2	94.5

5.2.Causal Fidelity and Consistency

The strength of the NCI model lies in its ability to adhere to the discovered or expert- defined causal structures. Causal Fidelity was measured and compared. The causal fidelity comparison between different models is shown in Table 2.

Table 2 Machine Learning Models Across Datasets

Model	Causal Fidelity (%)
Deep Neural Network	62.1
Random Forest	70.3
Causal Forest	82.6
NCI Framework	96.7

Conclusion

This paper introduced a novel hybrid framework, Neuro-Causal Intelligence (NCI), designed to integrate the predictive strength of deep neural networks with the transparency and reasoning capability of causal inference mechanisms. The primary objective of this system is to enable transparent, explainable, and scientifically robust decision-making in autonomous systems, especially in domains where interpretability and accountability are paramount, such as healthcare diagnostics, environmental monitoring, and scientific simulations.

References

[1].J. Pearl, Causality: Models, Reasoning, And Inference, 2nd Ed., Cambridge University Press, 2009.
[2].J. Pearl And D. Mackenzie, The Book Of Why: The New Science Of Cause And Effect, Basic Books, 2018.
[3].C. Molnar, Interpretable Machine Learning: A Guide For Making Black Box Models Explainable, 2019.
[4].M. T. Ribeiro, S. Singh, And C. Guestrin, "Why Should I Trust You? Explaining The Predictions Of Any Classifier," Proceedings Of The Acm Sigkdd International Conference On Knowledge Discovery And Data Mining, 2016.
[5].S. M. Lundberg And S.-I. Lee, "A Unified Approach To Interpreting Model Predictions," Advances In Neural Information Processing Systems (Neurips), 2017.
[6].J. Peters, D. Janzing, And B. Schölkopf, Elements Of Causal Inference: Foundations And Learning Algorithms, Mit Press, 2017.
[7].T. Hastie, R. Tibshirani, And J. Friedman, The Elements Of Statistical Learning, Springer, 2009.
[8].C. W. J. Granger, "Investigating Causal Relations By Econometric Models And Cross-Spectral Methods," Econometrica, Vol. 37, No. 3, Pp. 424–438, 1969.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0240
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1629 - 1634

[9]. D. Gunning And D. Aha, "Darpa's Explainable Artificial Intelligence (Xai) Program," Ai Magazine, Vol. 40, No. 2, Pp. 44–58, 2019.
[10]. F. Doshi-Velez And B. Kim, "Towards A Rigorous Science Of Interpretable Machine Learning,"ArxivPreprint Arxiv:1702.08608, 2017.
[11]. J. Pearl, M. Glymour, And N. Jewell, Causal Inference In Statistics: A Primer, Wiley, 2016.
[12]. K. He, X. Zhang, S. Ren, And J. Sun, "Deep Residual Learning For Image Recognition," Proceedings Of The Ieee Conference On Computer Vision And Pattern Recognition, 2016.
[13]. L. Breiman, "Random Forests," Machine Learning, Vol. 45, Pp. 5–32,2001.
[14]. P. Bühlmann And S. Van De Geer, Statistics For High-Dimensional Data, Springer, 2011.
[15]. R. Guidotti Et Al., "A Survey Of Methods For Explaining Black Box Models," Acm Computing