EDITOR: Amit Sheth, amit@sc.edu

DEPARTMENT: NEUROSYMBOLIC AI

Causal Neurosymbolic AI: A Synergy Between Causality and Neurosymbolic Methods

Utkarshani Jaimini, *Artificial Intelligence Institute at University of South Carolina, Columbia, SC, 29208, USA*

Cory Henson, *Bosch Center for Artificial Intelligence, Pittsburgh, PA, 15222, USA*

Amit Sheth, *Artificial Intelligence Institute at University of South Carolina, Columbia, SC, 29208, USA*

Causal neurosymbolic AI (NeSyAI) combines the benefits of causality with NeSyAI. More specifically, it 1) enriches NeSyAI systems with explicit representations of causality, 2) integrates causal knowledge with domain knowledge, and 3) enables the use of NeSyAI techniques for causal AI tasks. The explicit causal representation yields insights that predictive models may fail to analyze from observational data. It can also assist people in decision-making scenarios where discerning the cause of an outcome is necessary to choose among various interventions.

Causality is a force by which one process, event, object, or state, known as the *cause*, impacts another process, event, object, or state, known as the *effect*.

Causality has been studied in philosophy, psychology, statistics, economics, medicine, and AI. Since the early ages, these studies have focused on identifying causal relations in the world and deducing laws that govern causality. The features of a system or procedure would be manipulated (through intervention) to see which other features change or not. At the same time, the system is observed without manipulation (through collecting observational data). Famous scientists, like Archimedes, Newton, Galileo, and Pascal, followed the same process of intervention and observation to define the laws and principles of physics (see "Evolution of Causality").[1]

In more recent times, causality has been used 1) in smart manufacturing to detect root causes within the processing pipeline, (2) in medicine for diagnostic care and determining the effect of a treatment plan and treatment intervention, and (3) in autonomous driving for collision understanding. In general, causality can be utilized in any domain requiring an in-depth knowledge of the system's functioning. Performing every possible experiment in sensitive domains, such as autonomous driving and medicine, is not feasible due to cost, time, and safety concerns regarding putting human life in danger. These domains solely rely on observation data and the assumed causal model of the system to infer causal associations. On the other hand, in disciplines like neuroscience and biology, causal relations are often discovered using interventional experiments that are time-consuming and expensive.

Causal neurosymbolic AI (NeSyAI) offers a promising avenue for addressing the inquiry posed by Dr. Judea Pearl: "How can machines (and people) effectively represent causal knowledge, enabling swift access to pertinent information, accurate question answering, and intuitive comprehension akin to a three-year-old child?"[2]

Causal AI is a branch of AI systems that infers causal associations in observation data, as explained in preceding scenarios. Causal discovery methods are used to learn causal relations in the data, and causal inference methods are used to quantify the role of a cause on its effect. The traditional methods for causal discovery encounter issues such as missing data (especially in domains such as health care) and unmeasured confounding bias. Some causal discovery methods do not scale well with an increase in the number of variables due to the combinatorial optimization problem.[4]

NeSyAI, on the other hand, is a hybrid approach that utilizes both symbolic and subsymbolic knowledge,

1541-1672 © 2024 IEEE
Digital Object Identifier 10.1109/MIS.2024.3395936
Date of current version 11 June 2024.

often using neural networks. NeSyAI combines the strength of statistical AI, like machine learning, with symbolic human-like knowledge and reasoning. This enables the development of robust and reliable AI systems that can learn, reason, and interact with humans.

However, current NeSyAI systems do not yet understand and support causality. They do not incorporate causal representations as defined by traditional causal AI. Incorporating causal-AI-based concepts and representations into NeSyAI can enable the ability to perform causal AI tasks, such as causal discovery and inference.

A hybrid causal NeSyAI framework can do the following:

- Integrate heterogeneous domain knowledge from NeSyAI with causal knowledge, which complements the missing data and unmeasured data bias problems often found in causal AI.
- Fuse traditional causal representations into NeSyAI models.
- Map causal AI tasks, such as causal discovery, to NeSyAI tasks, such as knowledge graph (KG) completion.
- Use NeSyAI techniques to solve causal AI tasks, such as causal discovery in large-scale graphs, by integrating external knowledge from KGs.

CAUSAL AI

Causal AI is a branch of AI that deals with identifying and estimating cause-and-effect relations found in the data. It makes predictions based on the causes rather than relying on the correlations present in the data. The techniques used in causal AI help to make models more robust, explainable, and fair.

Dr. Judea Pearl introduced a taxonomy, the ladder of causation, to distinguish correlation from causation.[2] The ladder consists of three rungs. The first rung on the ladder of causation deals with the associations between observational data measured by conditional probability. The act of association simply means to observe the world around us. It uses conditional probability and conditional expectations to infer associations from the observed data. It predicts actions, events, states, and objects based on past observations. Current machine learning methods are based on this first rung. It can answer questions of this type: What does a symptom tell me about the disease?

The second rung on the ladder of causation is based on the act of "doing" or "intervening." It deals with the interventions in a system and analyzes the effect of the intervention. It can answer questions of this type: What if we change a parameter in the system? What if I take aspirin for my headache?

The third and topmost rung on the ladder of causation deals with counterfactuals. It is based on imagining a world and reasoning about observed phenomena. It can answer questions of this type: Was it the aspirin that cured my headache? The counterfactual considers scenarios that are absent, unthinkable, and inexpressible at the first and second rungs of the ladder. It is important to understand that correlation is not causation, irrespective of how strong the relation between two variables is. The strong correlation is not conclusive evidence for a causal relationship. Climbing up the ladder of causation leads to detailed and better explainability.

The current state of the art in modeling causality within the causal AI community revolves around the use of graphical models, such as causal networks (also known as causal Bayesian networks), structural causal models, and do-calculus.[3]

The structure of a causal network is either learned from the data using structure learning algorithms or otherwise determined by domain experts.[1] Causal networks are often used for causal reasoning, including intervention and counterfactual reasoning. An intervention on an event leads to a new causal model, where the value of the event is set to the intervention value, and the causal effect weights are estimated using the new model. A causal network evaluates the effect of an intervention on a given model using do-calculus.[3] Causal AI tasks can be divided into two major categories: causal discovery and causal inference.[4,5]

Causal Discovery

Causal discovery is the task of inferring new causal relations from observation data. There are three types of causal discovery methods: 1) constraint based, 2) score based, and 3) non-Gaussian and nonlinear based on structural causal models. The discovery of causal relations from observational data may suffer from insufficient information about the underlying causal relations and the vast search space during the discovery phase. In domains like health care, missing data are quite prevalent, and causal discovery may lead to incorrect conclusions.

Causal Inference

Causal inference is the task of measuring the effect a change in the cause has on the outcome variable. Causal networks are the foundations for causal inference. Causal inference quantifies the causal relations between two variables by intervening on one variable and observing its effect on another. On the ladder of

causation, causal inference helps to answer intervention questions at rung two and counterfactual questions at rung three.

NeSyAI

NeSyAI is a hybrid approach that merges symbolic knowledge-based methods with neural-network-based methods, improving the overall performance of AI systems.[6,7,8]

The symbolic knowledge in NeSyAI is often represented using KGs and logic rules. KGs are heterogeneous semantic networks of entities and their relations. They can combine data across multiple domains. Entities in a KG can be objects, states, processes, or events. The KG standardizes the representation of knowledge, facilitating the establishment of inference, integration, and relations among various data sources. The neural part of NeSyAI includes the use of machine learning and neural-network-based approaches for downstream tasks. While they can learn with vast amounts of data, they also remain black box due to a lack of explicit representation of background knowledge. The combination of neural networks with symbolic knowledge has the potential to answer counterfactual questions at the top rung of the ladder of causation.[7]

KGs contain domain knowledge, but they often suffer from incompleteness.[9] The task of KG completion aims to discover new knowledge using the existing knowledge represented in the form of a KG. The new knowledge can be either new entities or new relations. The task of discovering new relations or links in the KG is often implemented using link prediction algorithms.

Link prediction infers new links in a KG using the existing links, thus completing the graph. The knowledge in the KG is encoded as a triple of the form <subject, predicate, object>. Given a specified triple with either the subject or object missing (<?, predicate, object> or <subject, predicate, ?>), a link prediction algorithm aims to predict the missing subject or object in the triple. It has been used in applications such as question answering, relation extraction, and recommender systems.

CAUSAL NeSyAI

Causal NeSyAI (as shown in Figure 1) is a fusion of causality with NeSyAI that does the following:

- Aims to enhance NeSyAI systems with definite representations of causality and its concepts.
- Integrates causal information (from causal networks) with prior domain knowledge to improve causal AI tasks.
- Enables the use of NeSyAI techniques for causal AI tasks.

To further explain the causal NeSyAI framework, we map a causal AI task to a NeSyAI task (Figure 2). The tasks of KG completion in NeSyAI and causal discovery in causal AI both deal with discovering new information. Causal discovery is focused on discovering new causal relations from observational data, while KG completion is focused on inferring missing relations between entities in the KG. We can intuitively map the task of causal discovery to KG completion. Hence, if we can represent causal relations within a KG, then we can use the existing NeSyAI techniques for KG completion to perform causal discovery of new causal relations in the KG.

The traditional causal discovery methods in causal AI suffer from the problems of missing data, unmeasured confounding bias, and the inability to use heterogeneous data from different sources. In the case of insufficient information and missing data during causal discovery, the relevant domain knowledge from KG can supplement and boost the causal discovery process.

To overcome the problem of inadequate information and missing data, causal NeSyAI combines the vast knowledge captured in a KG with the concepts in causal AI. In the case of missing data, the causal discovery can utilize the domain knowledge captured in KG to infer new causal links. A KG may have a few samples of causal links, which may have been either learned from domain experts or partially learned from observation data. The task of causal discovery in KG involves predicting new causal links.

KGs in causal discovery enable the ability to integrate heterogeneous domain knowledge and help with the issue of unmeasured confounding bias and missing data.

Causal NeSyAI requires a good representation of causal knowledge, following the paradigm of "representation first, acquisition second."[2,10] The causal ontology serves this purpose by defining and representing causal AI concepts, such as causal relation, causal event roles (i.e., treatment, mediator, and outcome), and causal effects.

The causal ontology is the first known ontology to 1) model concepts from causal AI and 2) integrate symbolic knowledge abstracted from the causal network.[12] The causal ontology provides a minimal and intuitive ontology that captures the essential structure and semantics of causal relations. This allows compatibility with current state-of-the-art models and applications while also adopting best practices and designs from the ontology community. The nodes in a causal

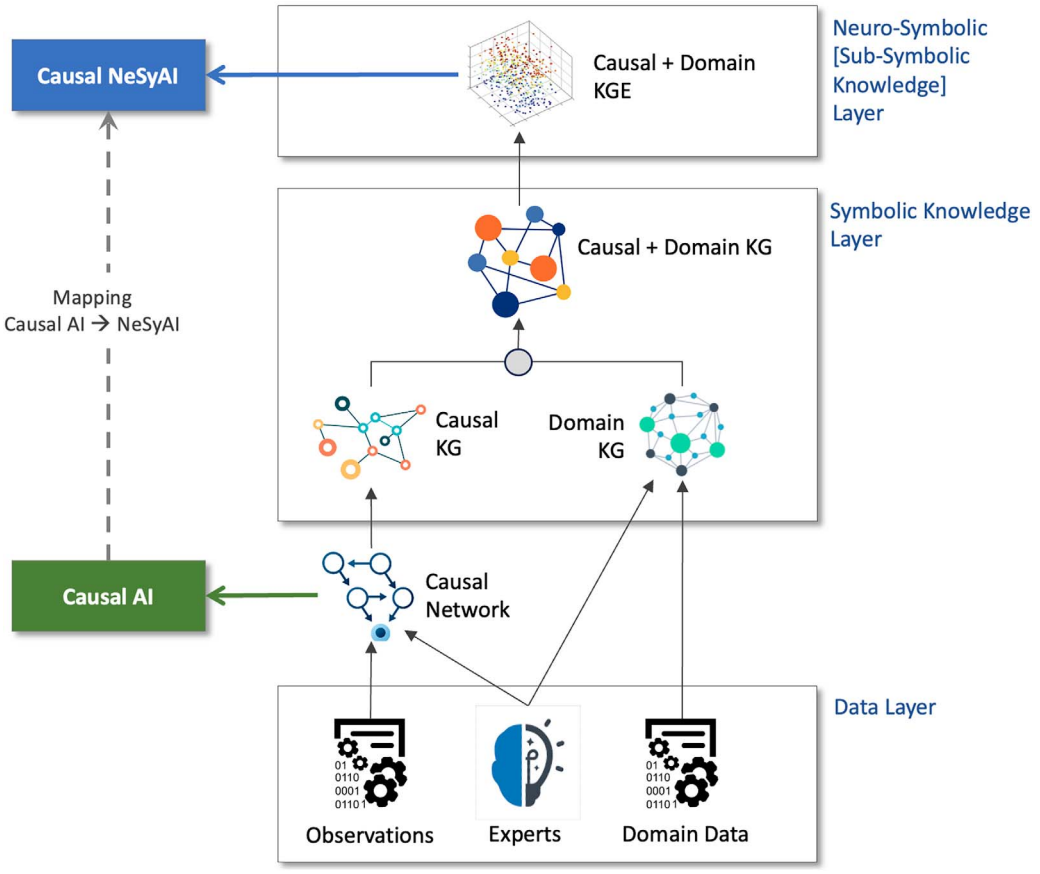

FIGURE 1. Causal NeSyAI architecture. It consists of three layers: 1) the data layer, which includes observations, domain data, and expert knowledge; 2) the symbolic knowledge layer, which merges the causal KG with the domain KG to create an integrated causal + domain KG; 3) the neurosymbolic (subsymbolic knowledge) layer, which converts the integrated causal + domain KG into a KG embedding, which can be used to perform downstream causal AI tasks using NeSyAI methods. KG: knowledge graph; KGE: knowledge graph embedding; NeSyAI: neurosymbolic AI.

network are considered to be events. The role played by an event in a causal relation is represented using the causal event role (i.e., the treatment, mediator, and outcome), and the strength of the causal relation between events is represented using a causal effect weight literal value. Since the causal ontology contains all of the information from a causal network, we can query a conformant KG for this information. In other words, the structure of the causal ontology is aligned as closely as possible with causal networks.

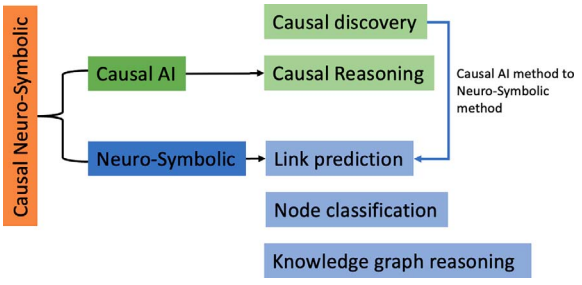

FIGURE 2. Types of causal AI methods and their mapping to NeSyAI methods.

Figure 1 describes the causal NeSyAI architecture. It comprises three layers: the data layer, symbolic knowledge layer, and neurosymbolic (subsymbolic knowledge) layer. The data layer includes observations, domain data, and expert domain knowledge. A causal network is learned from the data using traditional causal discovery methods[1] and/or expert knowledge. Within the symbolic knowledge layer, the learned causal network is then converted into a causal KG that is conformant with the causal ontology. The nodes and causal edges in the causal network are the entities and

causal relations in the KG. The transformation of the causal network to KG-based representation is crucial for using KG completion methods for causal discovery. The KG link prediction methods can be used to discover the missing causal relations. The causal KG is then merged with the domain KG to create an integrated causal + domain KG. The neurosymbolic (subsymbolic knowledge) layer converts the causal + domain KG into a KG embedding space (KGE) where NeSyAI methods are used to discover missing causal links.

CAUSAL NeSyAI IN SMART MANUFACTURING

Causal NeSyAI can be used effectively for smart manufacturing.[11] Given a manufacturing production line to assemble toy rocket (Figure 3), the goal is to detect the root cause of an anomaly found during the assembly process from the observation data. An anomaly in the assembly process is an event that led to the rocket's incomplete assembly. The rocket assembly line shown in Figure 3 consists of four robots (R01, R02, R03, and R04), four conveyor belts, a material handling station, four stoppers, two safety doors, and a toy rocket with four parts.

The assembly process records measurements of 1) the gripper potentiometer, gripper load, and angle (L, U, S, R, B, and T angles) for each robot; 2) the temperature and speed for four conveyors; 3) the status of four stoppers (as Boolean values); 4) the status of material handling (as a Boolean value); 5) the status of two safety doors (as Boolean values); and 6) the status of the HMI stop button (as a Boolean value). The robot arm is involved in two events: 1) picking up an object and 2) placing an object. The toy rocket has four parts: a base, two body parts, and a nose. An anomaly in this process can occur due to the stopper, conveyor belt, safety door failure, the robot arm missing picking up or placing the object, gripper potentiometer or load sensor failure, etc.

A causal network is learned from the data, described earlier, with the help of domain experts and traditional causal discovery methods. The causal network is then transformed into a causal KG that is conformant with the causal ontology. Next, the causal KG is integrated with the existing smart manufacturing KG for this assembly process. The smart manufacturing KG includes information about the robots, sensors and their ranges, event abstractions from the sensor measurements, etc. The integrated causal and smart manufacturing KG is used to train a KGE model. This KGE model can support discovering causal relations among the events, including possible reasons for the anomaly. Existing NeSyAI KG link prediction methods are used for causal discovery.

An integration of causality with NeSyAI leads to more robust and explainable AI systems. While we have shown a use case of causal discovery using KG link prediction, there are other synergies to find and problems to solve using this framework.

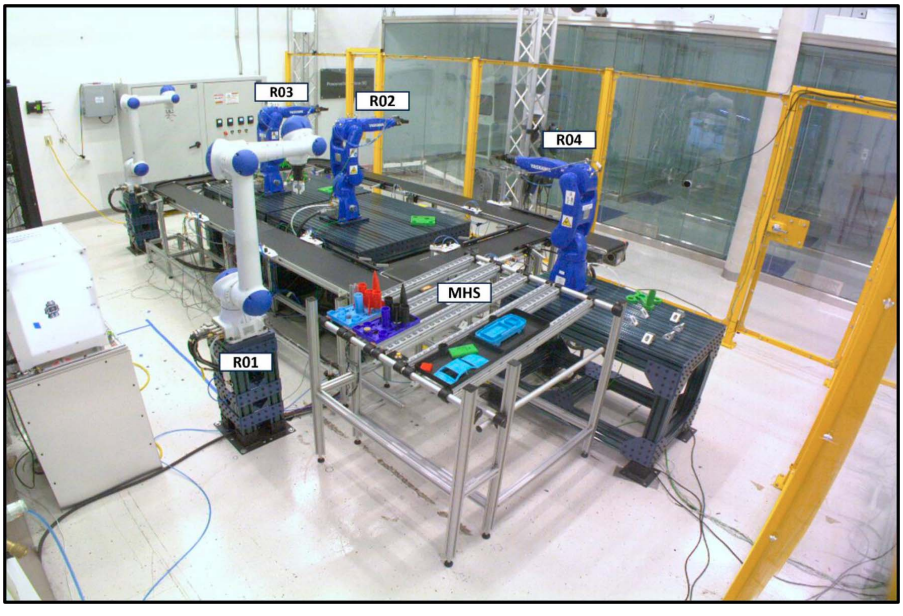

FIGURE 3. Smart manufacturing with a production line to assemble toy rockets.

EVOLUTION OF CAUSALITY

Plato (4427–348 BCE), driven by a profound intellectual curiosity, sought to understand the cause of things around us—why each thing comes into existence, why it goes out of existence, and why it exists. He believed that the question of why is crucial to our understanding of the world.

Aristotle (384–322 BCE) said his predecessors needed a complete understanding of possible causes and their systematic interrelations. Their use of causality needed to be supported by an adequate theory of causality. He believed we have proper knowledge of things only when we have understood their cause. The final cause is the most important, as the others would not have happened without it.

Sir Francis Bacon (1617–1621) was the father of empiricism; he said that knowledge comes from sensory experiences. He was also one of the founders of modem sciences and helped develop scientific methods in modem science. According to Bacon, if one wants to investigate the cause of a phenomenon, then one must list all things in which the phenomenon we are trying to explain occurs and things in which it does not. Then, rank the list according to the degree to which the phenomenon occurs in each one. Using this method, we can deduce the factors that cause the phenomenon's occurrence in one list and do not occur in others. He gave an example: An army is victorious when commanded by Essex and unsuccessful when not commanded by Essex. The success also depends on Essex's degree of involvement. Being commanded by Essex is the cause of the army's success.

Galileo (1564–1642), a true innovator, approached causality through controlled experiments involving numerous control factors. In an experiment, isolate one factor, manipulate its value while keeping others constant, observe the effect on the outcome, and understand what causes the process. Today, this approach is known as an *intervention*.

Udeny Yule (1871–1951) investigated causal questions using techniques like regression. He explored causal questions in various fields. In England, he investigated whether putting people on welfare led to their becoming dependent on the government or self-reliant and getting back on their feet. This interdisciplinary approach to studying causality is a testament to its broad applicability and relevance.

Charles Spearman (1863–1945) was the first to take statistical evidence seriously to determine the hidden variables we cannot observe directly. He came up with general intelligence. He argued that if multiple tests were given to people on math and reading, and we looked at the correlation among those tests, we could observe a pattern of constraints in those correlations, which would confirm this general intelligence. The constraint is called a tetrad equation, which says if one has four variables, then the product of two correlations permuted all three ways are equal.

Sewall Wright (1889–1988), a prominent biologist and geneticist, was the first to represent causality mathematically. He developed a path diagram, a graphical representation of causal relationships, which became a precursor to modem causal graphs. His innovative approach allowed for a more visual and intuitive understanding of causality.

Ronald Fisher (1890–1962) took Galileo's approach further; instead of controlling every factor in the experiment, randomly assign one of the factors, and let all of the other factors distribute (as they will then, at least, probabilistically and statistically). This can accomplish the same results as Galileo's technique and make a statistical inference about whether or not we are effective with this random assignment.

Jerzy Neyman (1894–1981) formulated the potential outcome framework. Suppose we were to give a treatment to one individual and a control to another. We are interested in finding what would have happened had we given a different treatment than we did (the unmeasured, missing, and unobserved data problem). This idea was later extensively used by **Don Rubin.** This framework is also very widely used in epidemiology and biology. Potential outcomes are evident and powerful if one is confident about the underlying model and the parameters to estimate.

The work of **Jamie Robins**, **Sewall Wright**, and **Judea Pearl**, leading figures in the field of artificial intelligence, led to the development of the formal structure of graphical causal models. These models, called Bayesian networks, are graphical representations of causal relationships. They provide a visual and intuitive way to understand complex causal systems, making them a valuable tool in causal analysis.

ACKNOWLEDGMENTS

This research is supported in part by the first author's summer internship at Bosch, National Science Foundation (NSF) Awards 2335967, "EAGER: Knowledge-Guided Neurosymbolic AI With Guardrails for Safe Virtual Health Assistants," and 2119654, "RII Track 2 FEC: Enabling Factory to Factory (F2F) Networking for Future Manufacturing." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

1. C. Glymour and K. Zhang, "Review of causal discovery methods based on graphical models," *Frontiers Genetics*, vol. 10, Jun. 2019, Art. no. 524, doi: 10.3389/fgene.2019.00524.
2. J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. New York, NY, USA: Basic Books, 2018.
3. J. Pearl, *Causality*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
4. P. Spirtes and K. Zhang, "Causal discovery and inference: Concepts and recent methodological advances," *Appl. Inf.*, vol. 3, pp. 1–28, Feb. 2016, doi: 10.1186/s40535-016-0018-x.
5. A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," *Wiley Interdisciplinary Rev., Data Mining Knowl. Discovery*, vol. 12, no. 2, 2022, Art. no. e1449.
6. A. Sheth, K. Roy, and M. Gaur, "Neurosymbolic artificial intelligence (why, what, and how)," *IEEE Intell. Syst.*, vol. 38, no. 3, pp. 56–62, May/Jun. 2023, doi: 10.1109/MIS.2023.3268724.
7. A. d. Garcez and L. C. Lamb, "Neurosymbolic AI: The 3rd wave," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12,387–12,406, 2023, doi: 10.1007/s10462-023-10448-w.
8. P. Hitzler, A. Eberhart, M. Ebrahimi, M. K. Sarker, and L. Zhou, "Neuro-symbolic approaches in artificial intelligence," *Nat. Sci. Rev.*, vol. 9, no. 6, 2022, Art. no. nwac035, doi: 10.1093/nsr/nwac035.
9. Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao, and Z. Duan, "Knowledge graph completion: A review," *IEEE Access*, vol. 8, pp. 192,435–192,456, 2020, doi: 10.1109/ACCESS.2020.3030076.
10. V. Belle, "On the relevance of logic for AI: Misunderstandings in social media, and the promise of neuro-symbolic learning," *Neurosymbolic AI J.*
11. R. Harik et al., "Analog and multi-modal manufacturing datasets acquired on the future factories platform," 2024, *arXiv:2401.15544*.
12. U. Jaimini, C. Henson, and A. Sheth, "An ontology design pattern for representing causality," in *14th Workshop Ontol. Design Pattern (WOP) 22nd Int. Semantic Web Conf. (ISWC)*, 2023.

UTKARSHANI JAIMINI is a Ph.D. student at the Artificial Intelligence Institute, University of South Carolina, Columbia, SC, 29208, USA. Contact her at ujaimini@email.sc.edu.

CORY HENSON is a lead research scientist at the Bosch Center for Artificial Intelligence in Pittsburgh, PA, 15222, USA. Contact him at cory.henson@us.bosch.com.

AMIT SHETH is the founding director of the Artificial Intelligence Institute at the University of South Carolina, Columbia, SC, 29208, USA. Contact him at amit@sc.edu.