EDITOR: Amit Sheth, amit@sc.edu

DEPARTMENT: KNOWLEDGE GRAPH

Knowledge-Enhanced Neurosymbolic Artificial Intelligence for Cybersecurity and Privacy

Aritran Piplai, *University of Texas at El Paso, El Paso, TX, 79968, USA*

Anantaa Kotal, Seyedreza Mohseni, and Manas Gaur, *Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Catonsville, MD, 21250, USA*

Sudip Mittal, *Department of Computer Science, Mississippi State University, Starkville, MS, 39762, USA*

Anupam Joshi, *Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Catonsville, MD, 21250, USA*

Neurosymbolic artificial intelligence (AI) is an emerging and quickly advancing field that combines the subsymbolic strengths of (deep) neural networks and the explicit, symbolic knowledge contained in knowledge graphs (KGs) to enhance explainability and safety in AI systems. This approach addresses a key criticism of current generation systems, namely, their inability to generate human-understandable explanations for their outcomes and ensure safe behaviors, especially in scenarios with unknown unknowns (e.g., cybersecurity, privacy). The integration of neural networks, which excel at exploring complex data spaces, and symbolic KGs, which represent domain knowledge, allows AI systems to reason, learn, and generalize in a manner understandable to experts. This article describes how applications in cybersecurity and privacy, two of the most demanding domains in terms of the need for AI to be explainable while being highly accurate in complex environments, can benefit from neurosymbolic AI.

Neurosymbolic artificial intelligence (*AI*) refers to the integration of neural network-based methods with symbolic knowledge-based approaches. It combines the strengths of both approaches to leverage the representational power of neural networks and the logical reasoning capabilities of symbolic approaches. Neural networks are excellent at large-scale data processing and the extraction of intricate patterns and characteristics from unprocessed input. However, they often struggle to provide explicit explanations for their decisions.[1] This is one of the reasons why, after a promising decade of work from the mid-1980s, neural networks never made it beyond academic and industrial research labs. On the other hand, symbolic knowledge-based approaches, such as rule-based systems or expert systems, utilize explicit knowledge representations and logical reasoning mechanisms. They can capture domain-specific knowledge and provide transparent explanations for their conclusions.[1,12] However, these approaches may struggle with handling uncertain or incomplete information and have limited capacity to learn from large-scale data.[1]

A combination of these two paradigms, called *neurosymbolic AI*, has begun to see popularity among the AI community over the last five years. The idea for this combination is not new, with the term *neural symbolic* being used at least as far back as the early 2000s.[2] In the 1990s, for instance, there were several efforts to marry connectionist approaches with fuzzy rules.[11] Indeed, it can (and has) been argued that the kernel of this idea can be found in the McColloch and Pitts' paper "A Logical Calculus of the Ideas Immanent in

1089-7801 © 2023 IEEE
Digital Object Identifier 10.1109/MIC.2023.3299435
Date of current version 14 September 2023.

Nervous Activity." There are multiple reasons for this renewed popularity. We describe these reasons in the context of cybersecurity.

First, a combination of symbolic reasoning and data-driven methods can be used to extract the sequence of steps or events that triggered the conclusion that the model reached. This is a great motivation for neurosymbolic approaches to be used in cybersecurity and privacy, especially in solving problems such as threat detection and analysis, which require not just patterns to be detected, but for such detected patterns from disparate systems across time to be put into a common context.[5] Neurosymbolic approaches can do this while preserving privacy (e.g., incorporating privacy policies, regulations, and compliance). For example, a neurosymbolic model can reason about sensitive network flow data usage by the neural network detector based on explicit privacy policies, and ensure compliance by incorporating privacy-preserving techniques such as differential privacy or secure multiparty computation.

Second is keeping AI systems *secure and safe.* The rise in data-driven models for automating vulnerability assessment of systems limits safety as the system learns only the vulnerabilities on which it is trained. In a neurosymbolic approach, the AI-based software systems would be trained with experts acting as adversaries and the AI model would learn to infer policies/rules dynamically.[1] Further, with knowledge in security specification documents explicitly captured using symbolic approaches and used as constraints on behavior, the AI system would be more robust and safe. This is of immediate interest to regulators and legislators in many countries as advanced AI systems have a high probability for generating risky/harmful information, without the control of human knowledge/expertise.

Another reason why a combination of rules and data-driven methods may be useful is the lack of high-quality data to make reliable inferences. This problem can be found in domains where the data for experiments are either hard to obtain or difficult to share because they may be sensitive. However, alternate sources such as text descriptors of the sensitive data may be available. Rules can be derived from these alternate sources that are shareable. If the available data alone are insufficient to infer reliable conclusions, these rules can be used to augment the conclusions derived from the data.[12] They can also be provided as an input to the data-driven model during the learning process.

Some domains may also be very dynamic, and so the data may be representative for only a brief period of time. The conclusions derived from the data may also be valid for a brief period of time. This is a problem in domains such as cybersecurity and fraud detection. The patterns that we derive from our existing dataset may be useful for some cyberattacks that are happening right now, but they may not be useful in the future. Combining deep network-based detectors with explicit rules that capture data drift or the temporal limits on the usefulness of a model can help in such situations.

NEUROSYMBOLIC AI IN CYBERSECURITY

In the context of cybersecurity, neurosymbolic AI can be applied to enhance various aspects of security systems, such as intrusion detection, malware analysis, vulnerability assessment, and threat intelligence. Very broadly, it can assist in creating the next-generation Security Operations Center (SoC), which combines AI approaches with a human, either in or on the loop.

Let us consider a scenario of security analysts who work in an SoC and play a major role in ensuring the security of an organization. The amount of background knowledge they have about evolving and new attacks makes a significant difference in their ability to detect attacks from the output of deep neural networks or machine learning (ML)-based systems that today analyze the sensed data stream. We can assist an analyst by capturing information available in open source threat intelligence sources, such as text descriptions of cyberattacks or threat feeds, and store them in a structured fashion in a cybersecurity knowledge graph (CKG). We describe two methods in which the structured cyber information present in CKGs can be used for downstream tasks, with a focus on explainability (reasoning and inference). The first method involves creating sophisticated rules based on real data, and an existing knowledge engine (*rule-based framework*). The second method involves using existing rules in downstream data-driven AI models and creating new policies for cybersecurity (*knowledge-guided models*).

In the *rule-based framework*, the ultimate goal is to create the strongest and closest rules for target machines to protect them from any type of threats and adversary behaviors. The rules can be simple to complex and will be consumed by any system or subsystem that needs protection. In the *knowledge-guided models*, we aim to tackle novel cyberthreats or mutated versions of older cyberthreats that do not exist in existing datasets for data-driven experimentation. Exploratory modeling techniques, such as reinforcement learning (RL), are needed to discover new adversaries that can further lead to new defenses. In our experiments, we see that CKGs can guide these exploratory learning strategies to be faster, more effective, and explainable.

Modeling Cyber Events

A plethora of information is available in unstructured text for cybersecurity. This information can come from various sources such as social media posts, user-written blogs, or published reports from large organizations. Through our research, we were able to extract this unstructured information and convert it into structured knowledge.[3] To achieve this, we utilize semantic triples, which consist of a subject-predicate-object relationship. In other words, when encountering two entities in a text, we aimed to deduce the relationship between them. We use Bidirectional Encoder Representations from Transformers (BERT) embeddings, as well as neural models, to generate these semantic triples. By analyzing 474 technical reports and numerous smaller technical posts, we construct a KG using these semantic triples.

In addition, software companies release information concerning cybersecurity threats to help consumers identify vulnerabilities in their products. These data can be accessed through *Trusted Automated eXchange of Intelligence Information* servers and integrated into a centralized KG.

We model our CKG ontology based on the concepts used in the *Structured Threat-Intelligence Exchange*, an industry standard for exchanging cyberthreat information. We further enrich this ontology with system-attribute concepts that describe the malware's behavior. The *rule-based framework* and the *knowledge-guided models* for cybersecurity can use this ontology for generating new rules and policies and employ it in downstream models for explainable intrusion and malware detection.[7]

Reasoning and Inference Examples

The semantic triples extracted from open source text are asserted with a CKG. This CKG can be leveraged, not just by human security analysts but also by other data-driven models. We can see some examples of how security analysts can use this CKG to uncover important insights about malware using the following SPARQL queries:

```
SELECT ? x where {
  ? x a FusedCKG:Malware;
    FusedCKG:uses
      FusedCKG:588f41bbc[ ... ].}
```

This query asks what malware(s) match a particular hash value.

```
SELECT DISTINCT ?x ?y ?z WHERE {
  ?x a FusedKG:Malware.
    FusedKG:hasHash [b9d ..... ]
    FusedKG:uses ?y.
  ?y a FusedKG:Attack-Pattern.
  ?z FusedKG:parameterchange
    FusedKG:increases_meanchange.}
```

The query asks which malware has a particular hash and the associated attack pattern for that malware. It also asks which system parameters show an increase when the malware is active. If a framework employs a KG where such dynamic information, in the form of observations, are recorded, then an AI model can leverage such structured knowledge in creating rules when defending against attacks.

Rule-Based Framework

The rule-based framework is one such architecture that employs existing AI models and transforms them into rule generators. AI models focus on the knowledge of events, their interlinking with other events or malware in the CKG, and a method for generating hypotheses or rules for inference (see Figure 1). The three broad sections of this framework are as follows:

1) *Event parsing engine*: The network block grants access to uncontrolled and unregulated networks. The packets are classified and distinguished to organize and handle data according to their specific types. Administrators, typically network admins, establish administrative policies that contribute to the control and dependability of network traffic. These policies assist in the creation of effective rules for the system.
2) *Symbolic engine*: The KG constructor serves as a graph builder, while the CKG stores the contextual knowledge used to generate hypotheses. The CKG constructor ensures the provision of valid and accurate context knowledge. The observation constructor produces clear and consistent observations derived from network packets and admin policies for integration into the KG. Also, the knowledge test and verification process will be applied to test and verify the alignment (e.g., simply using similarity measures) of both knowledge datasets and observations prior to their incorporation into the CKG.
3) *Reasoning engine*: In a neurosymbolic framework, the knowledge extractor component retrieves hypotheses from a KG and combines them with observations to form entailments. These entailments, organized with start and separator tokens, are then processed by a reasoning engine composed of transformer-based AI models [e.g., Generative Pre-Trained Transformer (GPT)]. This engine generates rankings and selects the most suitable hypothesis.[4] The neurosymbolic rule-based framework collects and categorizes network packets for use in symbolic engines, generating hypotheses

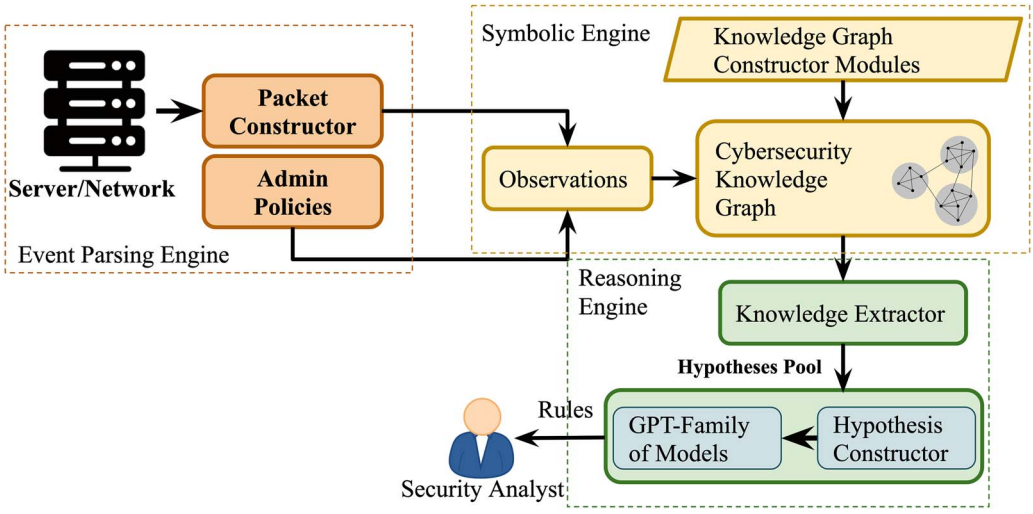

FIGURE 1. Neurosymbolic rule-based framework for dynamic rule inference and generation in cybersecurity. admin: administrative; GPT: Generative Pre-Trained Transformer.

through observation sequences and correlation scores. These hypotheses are combined with the GPT-Family models to create cybersecurity rules, providing improved accuracy, dependability, and explainable behavior in AI malware detection, surpassing traditional rule-based systems like SNORT.

Rules For Knowledge-Guided RL in Cybersecurity

The semantic knowledge embodied in KGs has the potential to direct algorithms like RL through rewards. Although KGs encompass well-established facts regarding cyber incidents, they might not possess details about emerging threats in unpredictable environments. Because of the exploratory nature of RL algorithms, new rules can be inferred based on the static knowledge in KGs, systems' data, and rules extracted from the rule-based framework.[5] The reward-based learning in RL allows the intrinsic deep learning AI algorithm to be expressive over the rules and explainable in prediction.

In the scenario of malware detection, we show that knowledge-guided RL sees even faster convergence, better efficiency, and better response time.[5] The acquired knowledge is incorporated into the exploration phase of the RL algorithm, and higher rewards are assigned to states that align with our knowledge sources. Specifically when we integrate prior knowledge, we observe an 8% reduction in average episode time.

Additionally, we conduct experiments with offline RL algorithms to examine the influence of prior knowledge. We discover that incorporating prior knowledge leads to a 4% rise in detection in three out of four malware families.[6]

In another study by Piplai et al.,[7] we show that employing knowledge-guided RL with rules on extensive Packet Capture files, which ranged from 3 to 4 GB in size, leads to more accurate decision making in countering attackers. In a two-player zero-sum game setup, both the attacker and defender were simulated using a knowledge-guided RL algorithm. By incorporating informed actions based on this approach, we demonstrate a remarkable preservation of 78% network availability, in comparison to a mere 25% when knowledge guidance was not utilized.[7] In both of the studies, the knowledge-guided RL strategy would preserve trace to knowledge sources for providing explanations to analysts.

CYBERSECURITY AND PRIVACY WITH KNOWLEDGE-GUIDED AI AND A RULE-BASED FRAMEWORK

Ensuring privacy in AI models is vital, similar to the importance of cybersecurity. AI models that unintentionally reveal personally identifiable information (PII) during defense against attacks can leave systems vulnerable to future unpredictable attacks. The challenge lies in training AI models to handle various cyberattacks while optimizing response time and preserving privacy. Reliable datasets for privacy-preserving training are scarce as organizations are hesitant to share data that contain PII or sensitive insights.

To overcome this, generative modeling methods like generative adversarial networks (GANs) can produce surrogate datasets that protect privacy while remaining useful for learning tasks. We utilized conditional GANs (CGANs) along with the t-closeness principle to preserve privacy in tabular data containing continuous and discrete variables.[8] However, training standard CGANs on sensitive data has limitations as they struggle to model conditionally continuous variables and can only repeat the discrete variable values seen in the original dataset.

To address these limitations, we propose a dual approach using privacy-preserving deep learning models, combining generative modeling and symbolic KGs that express domain knowledge. Domain-specific KGs like the Unified Cyber Ontology can guide the generative model by providing standardized information.[9] By querying the KG, the CGAN can be trained using a mix of original dataset and discrete values, ensuring that the generated dataset contains observed values and other alternatives from the KG. This approach enhances privacy preservation in generated datasets for downstream ML tasks.

CONCLUSION AND FUTURE WORK

This article discussed two main approaches of neurosymbolic AI in cybersecurity and privacy: 1) the rule-based framework and 2) knowledge-guided AI using RL. Both approaches intrinsically involve partitioning the CKG for solving tasks in our concerned domains and ensuring that the AI system is explainable. Our perspective is based on the noticeable improvements over traditional AI systems. However, additional research endeavors are required to develop reasoning engines in cybersecurity and privacy for optimal and explainable decision making that focuses on the users, such as security analysts. The combination of rules and ML/RL models can further be improved by focusing on which information is more beneficial to the model. The application of transformers, in this case by selecting specific paths in the graph based on real data, is a promising idea. We can also improve this with the help of graph embeddings of state spaces appended to data representation during training time.

Neurosymbolic AI has the potential to address critical challenges in the future, especially in domains like privacy in health care, where there is a growing demand for explainability. The techniques we explored can be expanded to biomedicine, where medical treatments and procedures are confidential. When limited to specific domains, GANs can only replicate treatment procedures based on the information upon which they were trained, however, incorporating a biomedical KG can enhance the GAN model by providing a broader range of knowledge.[10] By delving into such methods of knowledge infusion, not only can we contribute to the cause we can effectively tackle the limitations associated with purely data-driven approaches.

REFERENCES

1. 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.
2. S. Bader and P. Hitzler, "Dimensions of neural-symbolic integration - A structured survey," 2005, *arXiv:cs/0511042v1*.
3. A. Piplai, S. Mittal, A. Joshi, T. Finin, J. Holt, and R. Zak, "Creating cybersecurity knowledge graphs from malware after action reports," *IEEE Access*, vol. 8, pp. 211,691–211,703, 2020, doi: 10.1109/ACCESS.2020.3039234.
4. D. Paul and A. Frank, "Social commonsense reasoning with multi-head knowledge attention," 2020, *arXiv:2010.05587v1*.
5. A. Piplai, P. Ranade, A. Kotal, S. Mittal, S. N. Narayanan, and A. Joshi, "Using knowledge graphs and reinforcement learning for malware analysis," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2020, pp. 2626–2633, doi: 10.1109/BigData50022.2020.9378491.
6. A. Piplai, A. Joshi, and T. Finin, "Offline RL+ CKG: A hybrid AI model for cybersecurity tasks," in *Proc. AAAI Make*, 2023.
7. A. Piplai, M. Anoruo, K. Fasaye, A. Joshi, T. Finin, and A. Ridley, "Knowledge guided two-player reinforcement learning for cyber attacks and defenses," in *Proc. Int. Conf. Mach. Learn. Appl.*, 2022, pp. 1342–1349, doi: 10.1109/ICMLA55696.2022.00213.
8. A. Kotal, A. Piplai, S. S. Laya Chukkapalli, and A. Joshi, "PriveTAB: Secure and privacy-preserving sharing of tabular data," in *Proc. ACM Int. Workshop Secur. Privacy Analytics*, 2022, pp. 35–45, doi: 10.1145/3510548.3519377.
9. Z. Syed, A. Padia, M. L. Mathews, T. Finin, and A. Joshi, "UCO: A unified cybersecurity ontology," in *Proc. AAAI Workshop Artif. Intell. Cyber Secur.*, 2016.
10. L. Elluri, A. Piplai, A. Kotal, A. Joshi, and K. P. Joshi, "A policy-driven approach to secure extraction of COVID-19 data from research papers," *Frontiers Big Data*, vol. 4, Aug. 2021, Art. no. 701966, doi: 10.3389/fdata.2021.701966.
11. A. Joshi, N. Ramakrishman, E. N. Houstis, and J. R. Rice, "On neurobiological, neuro-fuzzy, machine learning, and statistical pattern recognition techniques," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 18–31, Jan. 1997, doi: 10.1109/72.554188.

12. S. Kambhampati, "Polanyi's revenge and AI's new romance with tacit knowledge," *Commun. ACM*, vol. 64, no. 2, pp. 31–32, Feb. 2021, doi: 10.1145/3446369.

ARITRAN PIPLAI is an assistant professor at the University of Texas at El Paso, El Paso, TX, 79968, USA. Contact him at apiplai1@umbc.edu.

ANANTAA KOTAL is a Ph.D. candidate in the Department of Computer Science and Electrical Engineering, the University of Maryland, Baltimore County, MD, 21250, USA. Contact her at akkotal1@umbc.edu.

SEYEDREZA MOHSENI is a Ph.D. candidate in the Department of Computer Science and Electrical Engineering, the University of Maryland, Baltimore County, MD, 21250, USA. Contact him at mohseni1@umbc.edu.

MANAS GAUR is an assistant professor with the Department of Computer Science and Electrical Engineering, the University of Maryland, Baltimore County, MD, USA. Contact him at manas@umbc.edu.

SUDIP MITTAL is an assistant professor with the Department of Computer Science at Mississippi State University, Starkville, MS, 39762, USA. Contact him at mittal@cse.msstate.edu.

ANUPAM JOSHI is the Oros Family Professor with the Department of Computer Science and Electrical Engineering, the University of Maryland, Baltimore County (UMBC), MD, 21250, USA, and acting dean of the College of Engineering and Information Technology at UMBC. He is also the director of UMBC's Center for Cybersecurity. Contact him at joshi@umbc.edu.