# Unlocking the Potential of Generative AI through Neuro-Symbolic Architectures -Benefits and Limitations

Oualid Bougzime, Samir Jabbar, Christophe Cruz, Frédéric Demoly

# Unlocking the Potential of Generative AI through Neuro-Symbolic Architectures – Benefits and Limitations

Oualid Bougzime[1], Samir Jabbar[2], Christophe Cruz[2], and Frédéric Demoly[1,3]

[1]*ICB UMR 6303 CNRS, Université Marie et Louis Pasteur, UTBM, 90010 Belfort Cedex, France*

[2]*ICB UMR 6303 CNRS, Université Bourgogne Europe, 21078 Dijon, France*

[3]*Institut universitaire de France (IUF), Paris, France*

## Abstract

Neuro-symbolic artificial intelligence (NSAI) represents a transformative approach in artificial intelligence (AI) by combining deep learning's ability to handle large-scale and unstructured data with the structured reasoning of symbolic methods. By leveraging their complementary strengths, NSAI enhances generalization, reasoning, and scalability while addressing key challenges such as transparency and data efficiency. This paper systematically studies diverse NSAI architectures, highlighting their unique approaches to integrating neural and symbolic components. It examines the alignment of contemporary AI techniques such as retrieval-augmented generation, graph neural networks, reinforcement learning, and multi-agent systems with NSAI paradigms. This study then evaluates these architectures against comprehensive set of criteria, including generalization, reasoning capabilities, transferability, and interpretability, therefore providing a comparative analysis of their respective strengths and limitations. Notably, the Neuro $\rightarrow$ Symbolic $\leftarrow$ Neuro model consistently outperforms its counterparts across all evaluation metrics. This result aligns with state-of-the-art research that highlight the efficacy of such architectures in harnessing advanced technologies

like multi-agent systems.

Keywords: Neuro-symbolic Artificial Intelligence, Neural Network, Symbolic AI, Generative AI, Retrieval-Augmented Generation (RAG), Reinforcement Learning (RL), Natural Language Processing (NLP), Explainable AI (XAI), Benchmark

# 1 Introduction

Neuro-symbolic artificial intelligence (NSAI) is fundamentally defined as the combination of deep learning and symbolic reasoning [1]. This hybrid approach aims to overcome the limitations of both symbolic and neural artificial intelligence (AI) systems while harnessing their respective strengths. Symbolic AI excels in reasoning and interpretability, whereas neural AI thrives in learning from vast amounts of data. By merging these paradigms, NSAI aspires to embody two fundamental aspects of intelligent cognitive behavior: the ability to learn from experience and the capacity to reason based on acquired knowledge [1, 2].

The importance of NSAI has been increasingly recognized in recent years, especially after the 2019 Montreal AI Debate between Gary Marcus and Yoshua Bengio. This debate highlighted two contrasting perspectives on the future of AI: Marcus argued that "expecting a monolithic architecture to handle abstraction and reasoning is unrealistic," emphasizing the limitations of current AI systems, while Bengio maintained that "sequential reasoning can be performed while staying in a deep learning framework" [3]. This discussion brought attention to the strengths and weaknesses of neural and symbolic approaches, catalyzing a surge of interest in hybrid solutions. Bengio's subsequent remarks at IJCAI 2021 underscored the importance of addressing out-of-distribution (OOD) generalization, stating that "we need a new learning theory" to tackle this critical challenge [4]. This aligns with the broader consensus within the AI community that combining neural and symbolic paradigms is essential to developing more robust and adaptable systems. Drawing on concepts like Daniel Kahne-

man's dual-process theory of reasoning, which compares fast, intuitive thinking (System 1) to deliberate, logical thought (System 2), NSAI seeks to bridge the gap between learning from data and reasoning with structured knowledge [5]. Despite ongoing debates about the optimal architecture for integrating these two paradigms, the 2019 Montreal AI Debate has played a pivotal role in shaping the trajectory of research in this promising field [6, 7, 8, 9, 10, 11].

NSAI offers a promising avenue for addressing limitations of purely symbolic or neural systems. For instance, while neural networks (NNs) often struggle with interpretability, symbolic AI systems are rigid and require extensive domain knowledge. By combining the adaptability of neural models with the explicit reasoning capabilities of symbolic methods, NSAI systems aim to provide enhanced generalization, interpretability, and robustness. These characteristics make NSAI particularly well-suited for solving complex, real-world problems where adaptability and transparency are critical [12]. Several NSAI architectures have been proposed to integrate these paradigms effectively. Examples include Symbolic Neuro Symbolic systems, Symbolic[Neuro], Neuro[Symbolic], Neuro — Symbolic coroutines, $\text{Neuro}_{\text{Symbolic}}$, and others [13]. Each architecture offers unique advantages but also poses specific challenges in terms of scalability, interpretability, and adaptability. A systematic evaluation of these architectures is imperative to understand their potential and limitations, guiding future research in this rapidly evolving field.

Generative AI has witnessed remarkable advancements, encompassing a diverse range of technologies that address various challenges in data processing, reasoning, and decision-making. These advancements can be categorized into several major branches of AI. Natural language processing (NLP) [14] includes technologies such as retrieval-augmented generation (RAG) [15], sequence-to-sequence models [16], semantic parsing [17], named entity recognition (NER) [18], and relation extraction [19], which focus on understanding and generating human language. Reinforcement learning techniques, like reinforcement learning with human feedback (RLHF) [20], enable systems to learn optimal actions through interaction with their

environment. Advanced NNs include innovations such as graph neural networks (GNNs) [21] and generative adversarial networks (GANs) [22], which excel in handling structured data and generating realistic data samples, respectively. Multi-agent systems [23, 24] explore the coordination and decision-making among multiple intelligent agents. Recent advances leverage mixture of experts (MoE) architectures to enhance scalability and specialization in collaborative frameworks. In MoE-based multi-agent systems, each expert operates as an autonomous agent, specializing in distinct sub-tasks or data domains, while a dynamic gating mechanism orchestrates their contributions [25, 26]. Transfer Learning [27], including pre-training [28], fine-tuning [29], and few-shot learning [30], allows AI models to adapt knowledge from one task to another efficiently. Explainable AI (XAI) [31] focuses on making AI systems transparent and interpretable, while efficient learning techniques, such as model distillation [32], aim to optimize resource usage. Reasoning and inference methods like chain-of-thought (CoT) [33] reasoning and link prediction enhance logical decision-making capabilities. Lastly, continuous learning [34] paradigms ensure adaptability over time. Together, these technologies form a comprehensive toolkit for tackling the increasingly complex demands of generative AI applications.

The classification of generative AI technologies within the NSAI framework is crucial for several reasons. Firstly, it provides a structured approach to understanding how these diverse technologies relate to and enhance NSAI capabilities. By mapping these techniques to specific NSAI architectures, researchers and practitioners can better grasp their potential applications and limitations. This classification also facilitates the identification of synergies between different AI approaches, potentially leading to more robust and versatile hybrid systems. Furthermore, it aids in decision-making processes when selecting appropriate technologies for specific tasks, considering factors like interpretability, reasoning capabilities, and generalization. As AI continues to evolve, this systematic categorization becomes increasingly valuable for bridging the gap between cutting-edge research and practical implementation,

ultimately driving the field towards more integrated and powerful AI solutions.

Therefore, this research aims to explores the alignment of generative AI technologies with the core catergories of NASAI and examines the insights this classification provides regarding their strenghts and limitations. The proposed methodology is threefold: (i) to define and analyze existing NSAI architectures, (ii) to classify generative AI technologies within the NSAI framework to provide a unified perspective on their integration, and (iii) to develop a systematic framework for assessing NSAI architectures across various criteria.

# 2 Neuro-Symbolic AI: Combining Learning and Reasoning to Overcome AI's Limitations

NNs have been exemplary in handling unstructured forms of data, e.g., images, sounds, and textual data. The capacity of these networks to acquire sophisticated patterns and representations from voluminous datasets has provided major breakthroughs in a series of disciplines, from computer vision, speech recognition, to NLP [35, 14]. One of the major benefits of NNs is that they learn and become better from raw data without requiring precoded rules or expert knowledge. This makes them highly scalable and efficient to utilize in applications with large raw data. However, despite these benefits, NNs also have some very well-documented disadvantages. One of the major ones of these might be that they are not transparent. Indeed, neural models pose interpretability challenges, making it difficult to understand the process by which they arrive at specific decisions or predictions. Such opacity causes problems for critical applications where explanation is necessary, such as in healthcare, finance, legal frameworks, and engineering. Additionally, NNs have a high requirement for data, requiring substantial amounts of labeled training data in order to operate effectively. This reliance on large data makes them ineffective when applied to data-scarce or data-costly environments. Neural models also struggle with reasoning and generalizing beyond their

training data, which makes their performance less impressive when it comes to tasks in logical inference or commonsense reasoning. Specifically, tasks including understanding causality, sequential problem-solving, and decision-making relying on outside world knowledge.

Symbolic AI is better at handling areas that are weaker for NNs. Symbolic systems function on explicit rules and structured representations, which enables them to achieve reasoning tasks related to complicated issues, such as mathematical proofs, planning, and expert systems. Symbolic AI is most important because it is transparent. Since symbolic methods are grounded in known rules and logical formalisms, decision-making processes are easy to interpret and explain. However, symbolic AI systems have some drawbacks. One of the biggest ones is that they are rigid and difficult to respond to new circumstances. They require rules to be manually defined and require structured input data, leading them difficult to apply to real-world situations where data might contain noise, incompleteness, or unstructured form. They are also susceptible to combinatorial explosions in handling big data or hard reasoning problems, which significantly slows down their performance at scale. Symbolic systems are also not well suited for perception tasks like image or speech recognition since they are unable to draw knowledge from raw data alone.

While traditional NNs are strong at recognizing patterns in collections of data but falter when presented with new situations, symbolic reasoning provides a rational foundation for decision-making but is limited in the manner in which it can learn knowledge from new information and adapt in a dynamic process. The combination of these two approaches in NSAI effectively minimizes these limitations, producing a more flexible, explainable, and effective AI system. Another distinguishing feature of NSAI is that it is able to generalize outside its training set. Traditional AI systems are prone to fail in novel situations; however, NSAI, because of its combination of learning and logical reasoning, works better in such cases. Such a feature is critical for real-world applications such as autonomous transport and medicine, where systems need to perform well in uncontrolled environments. Apart from

that, in an interdisciplinary engineering context such as 4D printing, which brings together materials science, additive manufacturing, and engineering, NSAI holds significant promise for improving both the interpretability and reliability of design decisions on the actuation and mechanical performance, and printability. Although these advantages seem promising, they remain hypotheses requiring more extensive validation and industrial-scale testing. Ongoing research must demonstrate, through empirical studies and real-world implementations, how NSAI can reliably accelerate the discovery of smart materials and structures [36]. The second key benefit point of NSAI is that it has a reduced need for big data sets. Traditional AI systems usually require a tremendous amount of data to operate, which might be very time- and resource-consuming. NSAI, however, is able to do better with a much smaller set of data required, due to its symbolic reasoning ability. This makes it a more sustainable and viable option, especially for small organizations or new research areas with limited resources. Along with the aforementioned data efficiency, NSAI models also have the exceptional transferability, i.e., their capacity for using knowledge learned from one task and applying it in another with less need for retraining. Such a property is highly desirable in situations where there is little data related to a new task.

# 3 Neuro-Symbolic AI Architectures

This section provides an overview of various NSAI architectures, offering insights into their design principles, integration strategies, and unique capabilities. While Kautz's classification [13] serves as a foundational framework, we extend it by incorporating additional architectural perspectives to capture the evolving landscape of NSAI systems. These approaches range from symbolic systems augmented by neural modules for specialized tasks to deeply integrated models where explicit reasoning engines operate within neural frameworks. This expanded categorization highlights the diversity of design strategies and the broad applicability of NSAI techniques, emphasizing their potential for more interpretable, robust, and data-efficient AI

solutions.

## 3.1 Sequential

As part of the sequential NSAI, the *Symbolic* $\rightarrow$ *Neuro* $\rightarrow$ *Symbolic* architecture involves systems where both the input and output are symbolic, with a NN acting as a mediator for processing (Figure 1a). Symbolic input, such as logical expressions or structured data, is first mapped into a continuous vector space through an encoding process. The NN operates on this encoded representation, enabling it to learn complex transformations or patterns that are difficult to model symbolically. Once the processing is complete, the resulting vector is decoded back into symbolic form, ensuring that the final output aligns with the structure and semantics of the input domain. This framework is especially useful for tasks that require leveraging the generalization capabilities of NNs while preserving symbolic interpretability [37, 38]. A formulation of this architecture is presented below:

$$y = f_{\text{neural}}(x) \tag{1}$$

where $x$ is the symbolic input, $f_{\text{neural}}(x)$ represents the NN that processes the input, and $y$ is the symbolic output.

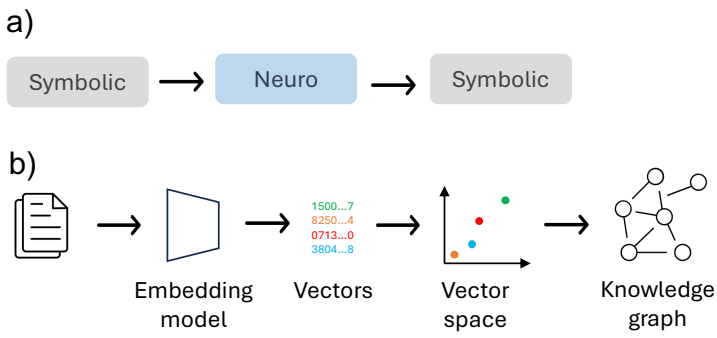


Figure 1: Sequential architecture: (a) Principle and (b) application to knowledge graph construction.

This architecture can be used in a semantic parsing task, where the input is a sequence of symbolic tokens (e.g., words). Here, each token is mapped to a continuous embedding via

word2vec, GloVe, or a similar method [39, 40]. The NN then processes these embeddings to learn compositional patterns or transformations. From this, the network's output layer decodes the processed information back into a structured logical form (such as knowledge-graph triples), as illustrated in Figure 1b.

## 3.2 Nested

The nested NSAI category is composed of two different architectures. The first – *Symbolic[Neuro]* – places a NN as a subcomponent within a predominantly symbolic system (Figure 2a). Here, the NN is used to perform tasks that require statistical pattern recognition, such as extracting features from raw data or making probabilistic inferences, which are then utilized by the symbolic system. The symbolic framework orchestrates the overall reasoning process, incorporating the neural outputs as intermediate results [41]. This architecture can formally defined as follows:

$$y = g_{\text{symbolic}}(x, f_{\text{neural}}(z)) \tag{2}$$

where $x$ represents the symbolic context, $z$ is the input passed from the symbolic reasoner to the NN, $f_{\text{neural}}(z)$ expresses the neural model processing the input, and $g_{\text{symbolic}}$ the symbolic reasoning engine that integrates neural outputs. A well-known instance of this architecture is AlphaGo [41], where a symbolic Monte-Carlo tree search orchestrates high-level decision-making, while a NN evaluates board states, providing a data-driven heuristic to guide the symbolic search process [42] (Figure 2b). Similarly, in a medical diagnosis scenario, a rule-based engine oversees the core diagnostic process by applying expert guidelines to patient history, symptoms, and lab results. At the same time, a NN interprets unstructured radiological images, delivering key indicators such as tumor likelihood. The symbolic system then integrates these indicators into its final decision, combining transparent and rule-driven logic with robust pattern recognition.

The second architecture – *Neuro[Symbolic]* – integrates a symbolic reasoning engine as

a component within a neural system, allowing the network to incorporate explicit symbolic rules or relationships during its operation (Figure 2c). The symbolic engine provides structured reasoning capabilities, such as rule-based inference or logic, which complement the NN's ability to generalize from data. By embedding symbolic reasoning within the neural framework, the system gains interpretability and structured decision-making while retaining the flexibility and scalability of neural computation. This integration is particularly effective for tasks that require reasoning under constraints or adherence to predefined logical frameworks [43, 44]. This configuration can be described as follows:

$$y = f_{\text{neural}}(x, g_{\text{symbolic}}(z)) \tag{3}$$

where $x$ represents the input data to the neural system, $z$ is the input passed from the NN to the symbolic reasoner, $g_{\text{symbolic}}$ is the symbolic reasoning function, and $f_{\text{neural}}$ denotes the NN processing the combined inputs.

This architecture is currently applied in automated warehouse, where a robot navigates dynamically changing aisles. During normal operation, it relies on a neural policy to select routes based on learned patterns. When it encounters an unexpected obstacle, it offloads route computation to a symbolic solver (e.g., a pathfinding or constraint-satisfaction algorithm), which returns an alternative path. The solver's output is then integrated back into the neural policy, and the robot resumes its usual pattern-based navigation. Over time, the robot also learns to identify which challenges call for the symbolic solver, effectively blending fast pattern recognition with precise combinatorial planning.

Figure 2d illustrates this framework, a symbolic reasoning engine processes structured data, such as a maze, to generate a solution path. A NN encodes the problem into a latent representation and decodes it into a symbolic sequence of actions (e.g., forward, turn left, turn right).

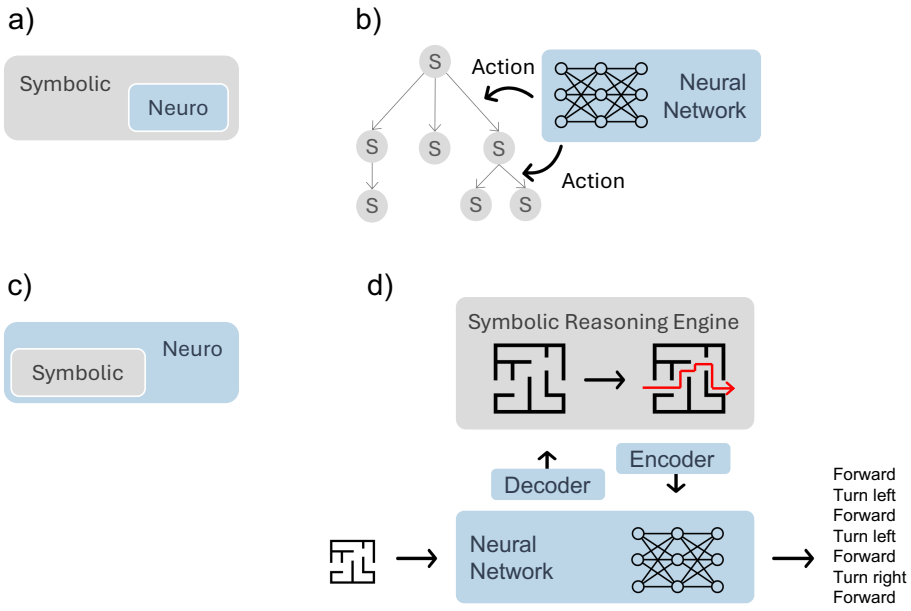


Figure 2: Nested architectures: (a) *Symbolic[Neuro]* principle and (b) its application to tree Search, (c) *Neuro[Symbolic]* principle and (d) its application to maze-solving.

### 3.3 Cooperative

As a cooperative framework, *Neuro* | *Symbolic* uses neural and symbolic components as interconnected coroutines, collaborating iteratively to solve a task (Figure 3a). NNs process unstructured data, such as images or text, and convert it into symbolic representations that are easier to reason about. The symbolic reasoning component then evaluates and refines these representations, providing structured feedback to guide the NN's updates. This feedback loop continues over multiple iterations until the system converges on a solution that meets predefined symbolic constraints or criteria. By combining the strengths of NNs for generalization and symbolic reasoning for interpretability, this approach achieves robust and adaptive problem-solving [45]. This architecture can be described as follows:

$$z^{(t+1)} = f_{\text{neural}}(x, y^{(t)}), \quad y^{(t+1)} = g_{\text{symbolic}}(z^{(t+1)}), \quad \forall t \in \{0, 1, \ldots, n\} \tag{4}$$

where $x$ represents non-symbolic data input, $z^{(t)}$ is the intermediate symbolic representation at iteration $t$, $y^{(t)}$ is the symbolic reasoning output at iteration $t$, $f_{\text{neural}}(x, y^{(t)})$ expresses the

NN that processes the input $x$ and feedback from the symbolic output $y^{(t)}$, $g_{\text{symbolic}}(z^{(t+1)})$ is the symbolic reasoning engine that updates $y^{(t+1)}$ based on the neural output $z^{(t+1)}$, and $n$ is the maximum number of iterations or a convergence threshold. The hybrid reasoning halts when the outputs $y^{(t)}$ converge (e.g., $|y^{(t+1)} - y^{(t)}| < \epsilon$)), where $\epsilon$ is a small threshold denoting minimal change between successive outputs, or when the maximum iterations $n$ is reached.

For instance, this architecture can applied in autonomous driving systems, where a NN processes real-time images from vehicle cameras to detect and classify traffic signs. It identifies shapes, colors, and patterns to suggest potential signs, such as speed limits or stop signs. A symbolic reasoning engine then evaluates these detections based on contextual rules—like verifying if a detected speed limit sign matches the road type or if a stop sign appears in a logical position (e.g., near intersections). If inconsistencies are detected, such as a stop sign identified in the middle of a highway, the symbolic system flags the issue and prompts the neural network to re-evaluate the scene. This iterative feedback loop continues until the system reaches consistent, high-confidence decisions, ensuring robust and reliable traffic sign recognition, even in challenging conditions like poor lighting or partial occlusions (Figure 3b).

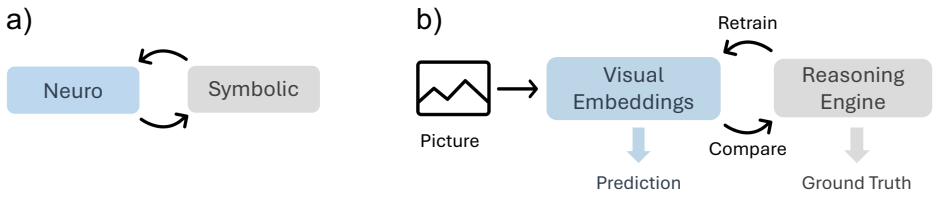


Figure 3: Cooperative architecture: (a) principle and (b) application to visual reasoning.

## 3.4 Compiled

As part of the compiled NSAI, $Neuro_{Symbolic_{Loss}}$ uses symbolic reasoning into the loss function of a NN (Figure 4a). The loss function is typically used to measure the discrepancy between the model's predictions and the true outputs. By incorporating symbolic rules or constraints, the network's training process not only minimizes prediction error but also ensures that the output aligns with symbolic logic or predefined relational structures. This allows the model

to learn not just from data but also from symbolic reasoning, helping to guide its learning process toward solutions that are both accurate and consistent with symbolic principles.

$$\mathcal{L} = \mathcal{L}_{\text{task}}(y, y_{\text{target}}) + \lambda \cdot \mathcal{L}_{\text{symbolic}}(y) \tag{5}$$

where $y$ is the model prediction,$y_{\text{target}}$ represents the ground truth labels, $\mathcal{L}_{\text{task}}$ is the task-specific loss (e.g., cross-entropy), $\mathcal{L}_{\text{symbolic}}$ is the penalization for violating symbolic rules, $\lambda$ the Weight balancing the two loss components, and $\mathcal{L}$ the final loss, combining both the task-specific loss and the symbolic constraint penalty to guide model optimization. This architecture is typically useful in the field of 4D printing, where structures need to be optimized at the material level to achieve a target shape. In such a case, a NN predicts the material distribution and geometric configuration that allows the structure to adapt under external stimuli. The training process incorporates a physics-informed loss function, where, in addition to minimizing the difference between predicted and desired mechanical behavior, the model is penalized whenever the predicted deformation violates symbolic mechanical constraints, such as equilibrium equations or the stress-strain relationship (Figure 4b). By embedding these symbolic equations directly into the loss function, the NN learns to generate designs that are not only data-driven but also physically consistent, ensuring that the final 4D-printed structure maintains the desired shape across different operational conditions.

A second compiled NSAI architecture, called $Neuro_{Symbolic_{Neuro}}$, uses symbolic reasoning at the neuron level by replacing traditional activation functions with mechanisms that incorporate symbolic reasoning (Figure 4c). Rather than using standard mathematical operations like ReLU or sigmoid, the neuron activation is governed by symbolic rules or logic. This allows the NN to reason symbolically at a more granular level, integrating explicit reasoning steps into the learning process. This fusion of symbolic and neural operations enables more interpretable and constrained decision-making within the network, enhancing its ability to reason in a structured and rule-based manner while retaining the flexibility of neural computations. This architecture can be described as follows:

$$y = g_{\text{symbolic}}(x) \tag{6}$$

where: $x$ represents the pre-activation input, $g_{\text{symbolic}}(x)$ is the symbolic reasoning-based activation function, and $y$ the final neuron. This architecture can find application in lean approval systems, where neural activations are driven by symbolic financial rules rather than traditional functions. One example is the collateral-based constraint neuron, which dynamically adjusts the risk score based on the value of the pledged collateral. When the collateral's value falls below a predefined threshold relative to the loan amount, the neuron applies a strict penalty that substantially increases the risk score, effectively preventing the system from underestimating the associated financial risk. This symbolic constraint ensures that, regardless of favorable patterns identified in other data, the model consistently accounts for the critical impact of insufficient collateral, leading to more reliable and regulation-compliant credit decisions (Figure 4d).

Finally, the last compiled architecture, *Neuro:Symbolic* $\rightarrow$ *Neuro*, uses a symbolic reasoner to generate labeled data pairs $(x, y)$, where $y$ is produced by applying symbolic rules or reasoning to the input $x$ (Figure 4e). These pairs are then used to train a NN, which learns to map from the symbolic input $x$ to the corresponding output $y$. The symbolic reasoner acts as a supervisor, providing high-quality, structured labels that guide the NN's learning process [46]. This architecture can be governed as follows:

$$\mathcal{D}_{\text{train}} = \{(x, g_{\text{symbolic}}(x)) \mid x \in \mathcal{X}\} \tag{7}$$

where $\mathcal{D}_{\text{train}}$ is the training dataset, $x$ denotes the unlabeled data, $g_{\text{symbolic}}(x)$ represents symbolic rules generating labeled data, and $\mathcal{X}$ the set of all input data (Figure 4b).

Figure 4f illustrates this architecture, where a reasoning engine is used to label unlabeled training data, transforming raw inputs into structured $(x, y)$ pairs, where symbolic rules enhance the data quality.

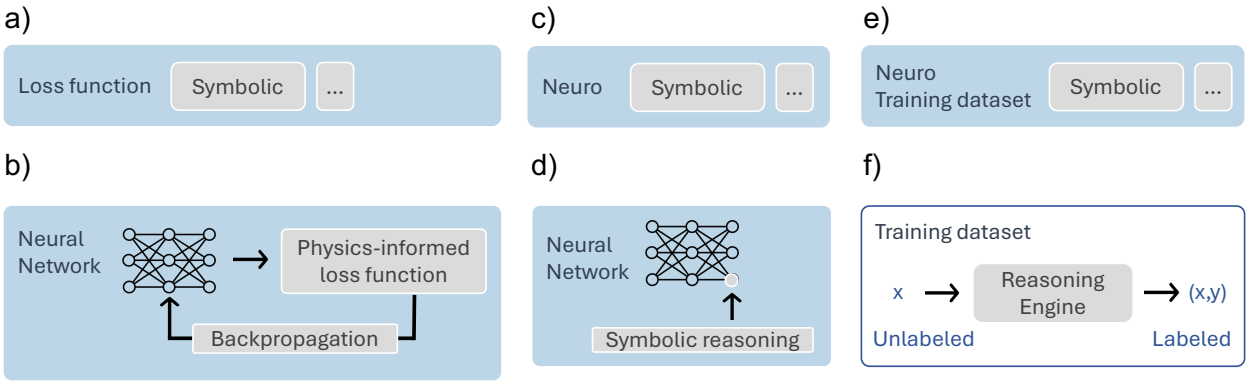


Figure 4: Compiled architectures: (a) $Neuro_{Symbolic_{Loss}}$ principle and (b) application to physics-informed learning; (c) $Neuro_{Symbolic_{Neuro}}$ principle and (d) application of symbolic reasoning in NNs; (e) *Neuro:Symbolic* $\rightarrow$ *Neuro* principle and (f) application to data Llabeling.

## 3.5 Ensemble

Another promising architecture, called *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* uses multiple interconnected NNs via a symbolic fibring function, which enables them to collaborate and share information while adhering to symbolic constraints (Figure 5a). The symbolic function acts as an intermediary, facilitating communication between the networks by ensuring that their interactions respect predefined symbolic rules or structures. This enables the networks to exchange information in a structured manner, allowing them to jointly solve problems while benefiting from both the statistical learning power of NNs and the logical constraints imposed by the symbolic system [47]. This architecture can formally defined as follows:

$$y = g_{\text{fibring}}(\{f_i\}_{i=1}^{n}) \tag{8}$$

where $f_i$ represents the individual NN, $g_{\text{fibring}}$ is the logic-aware aggregator that enforces symbolic constraints while unifying the outputs of multiple NNs, $n$ the umber of NNs, and $y$ is the combined output of interconnected NNs, produced through the symbolic fibring function $g_{\text{fibring}}$. For instance in smart cities and urban planning, multiple NNs can be employed, each handle a different urban data stream—such as real-time traffic flow, energy consumption, and air quality measurements. A symbolic fibring function then harmonizes these outputs,

enforcing city-level constraints (e.g., ensuring pollution alerts match local environmental regulations, verifying that traffic predictions align with current road network rules). If one network forecasts a surge in vehicle congestion that would push pollution levels beyond acceptable thresholds, the symbolic aggregator identifies the conflict and directs all networks to converge on a coordinated strategy—such as adjusting traffic signals or advising public transport usage. By leveraging each network's specialized insight within logical urban-planning constraints, the system delivers efficient, consistent decisions across the city's complex infrastructure.

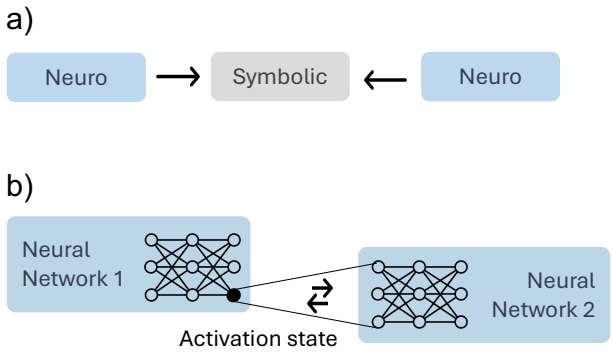


Figure 5: Ensemble architecture: (a) principle and (b) application to NN collaboration.

Figure 5b illustrates this architecture, where two NNs (Neural Network 1 and Neural Network 2) communicate through activation states, which enables dynamic exchange of learned representations.

# 4 Leveraging NSAI in AI Technologies

Generative AI is advancing at a remarkable pace, addressing increasingly complex challenges through the integration of diverse methodologies. A key development is the combination of NNs with symbolic reasoning, resulting in hybrid systems that leverage both strengths. Recent studies have demonstrated the effectiveness of this approach in various applications, including design generation and enhancing instructability in generative models [48, 49]. This section aims to classify contemporary AI techniques such as RAG, GNNs, agent-based AI,

and transfer learning within the NSAI framework. This classification clarifies how generative AI aligns with neuro-symbolic approaches, bridging cutting-edge research with established paradigms. It also reveals how generative AI increasingly embodies both neural and symbolic characteristics, moving beyond siloed methods.

Additionally, this classification enhances our understanding of these techniques' roles in AI's broader landscape, particularly in addressing challenges like interpretability, reasoning, and generalization. It identifies synergies between methods, fostering robust hybrid models that combine neural learning's adaptability with symbolic reasoning's precision. Lastly, it supports informed decision-making, guiding researchers and practitioners in selecting the most suitable AI techniques for specific tasks.

## 4.1 Overview of Key AI Technologies

One of the most significant advancements is RAG, which integrates information retrieval with generative models to perform knowledge-intensive tasks. By combining a retrieval mechanism to extract relevant external data with Seq2Seq models for generation [50], RAG excels in applications such as question answering and knowledge-driven conversational AI [51]. Seq2Seq models themselves, built as encoder-decoder architectures, have been pivotal in machine translation, text summarization, and conversational modeling, providing the foundation for many generative AI systems. An extension of RAG is the GraphRAG approach [52], which incorporates graph-based reasoning into the retrieval and generation process. By leveraging knowledge graph (KGq) and ontologies structures to represent relationships between information elements, GraphRAG enhances query-focused summarization and reasoning tasks [53, 54]. This method has demonstrated success in producing coherent and contextually rich summaries by integrating local and global reasoning.

GNNs [55] represent a breakthrough in extending neural architectures to graph-structured data, enabling advanced reasoning over interconnected entities. Their ability to model relationships between entities makes them indispensable for a range of tasks, including link

prediction, node classification, and recommendation systems, with notable success in KG reasoning. GNNs have also proven highly effective in named entity recognition (NER) [56], where they can leverage graph representations to capture contextual dependencies and relationships between entities in text. This capability extends to relation extraction [57], where GNNs identify and classify semantic relationships between entities, crucial for building and enhancing KG.

Advances in agentic AI systems, which leverage Large Language Models (LLMs), have shown significant potential in enabling autonomous decision-making and task execution. These systems are designed to function independently, interacting with environments, coordinating with other agents, and adapting to dynamic situations without human intervention. A notable example is AutoGen [58], a framework that enables the creation of autonomous agents that can interact with each other to solve tasks and improve through continual interactions. Recent work has further enhanced these systems through MoE architectures, which integrate specialized sub-models ("experts") into multi-agent frameworks to optimize task-specific performance and computational efficiency. For instance, MoE-based coordination allows agents to dynamically activate subsets of experts based on context, enabling scalable specialization in complex environments [59, 60]. Xie et al. [61] explored the role of LLMs in these agentic systems, discussing their ability to facilitate autonomous cooperation and communication between agents in complex environments, and marking an important step toward scalable and self-sufficient AI. By combining MoE principles with multi-agent collaboration, systems can achieve hierarchical decision-making: LLMs act as meta-controllers, routing tasks to specialized agents (e.g., vision, planning, or language experts) while maintaining global coherence.

However, the growing autonomy of such systems underscores the importance of XAI [62] to ensure transparency and trust. XAI has gained prominence as a means to enhance accountability and support ethical AI adoption. By providing insights into model behavior, XAI ensures that even highly autonomous systems remain interpretable and accountable, addressing concerns about their decisions and actions in sensitive and dynamic environments.

Recent advancements in AI have demonstrated the potential of integrating fine-tuning, distillation, and in-context learning to enhance model performance. Huang et al. [63] introduced in-context learning distillation, a novel method that transfers few-shot learning capabilities from large pre-trained LLMs to smaller models. By combining in-context learning objectives with traditional language modeling, this approach allows smaller models to perform effectively with limited data while maintaining computational efficiency.

Transfer learning [64] has similarly emerged as a foundational technique, enabling pre-trained models to adapt their extensive knowledge to new domains using minimal data. This capability is particularly advantageous in resource-constrained scenarios. Techniques such as feature extraction, where pre-trained model layers are repurposed for specific tasks, and fine-tuning, which involves adjusting the weights of the pre-trained model for new tasks, further illustrate its adaptability.

Complementing these methods, prompt engineering empowers LLMs to perform task-specific functions through carefully designed prompts. Techniques such as CoT prompting [33], zero-shot [65], and few-shot prompting enhance the ability of LLMs to reason and generalize across diverse tasks without extensive retraining [66]. Additionally, knowledge distillation plays a crucial role in optimizing AI models by transferring knowledge from larger, more complex models to smaller, efficient ones [67]. Variants of distillation, such as task-specific distillation, feature distillation, and response-based distillation, further streamline the process for edge computing and resource-limited environments.

Reinforcement learning and its variant RLHF [68], focus on training agents to make sequential decisions in dynamic environments. RLHF further aligns agent behavior with human preferences, fostering ethical and adaptive AI systems. Finally, continuous learning, or lifelong learning, addresses the challenge of adapting AI systems to new data while retaining previously learned knowledge, ensuring AI remains effective in changing environments [69].

These techniques represent the cutting edge of generative AI, each contributing to solving complex challenges across diverse applications. The classification of these methods within

NSAI paradigm, explored in the following sections, offers a structured perspective on their synergies and practical relevance.

## 4.2 Classification of AI Technologies within NSAI Architectures

This section categorizes generative AI techniques within the eight distinct NSAI architectures, highlighting their underlying principles and practical applications. By classifying these approaches, we gain a clearer understanding of how neural and symbolic methods synergize to address diverse challenges in AI, as summarized in Figure 6.

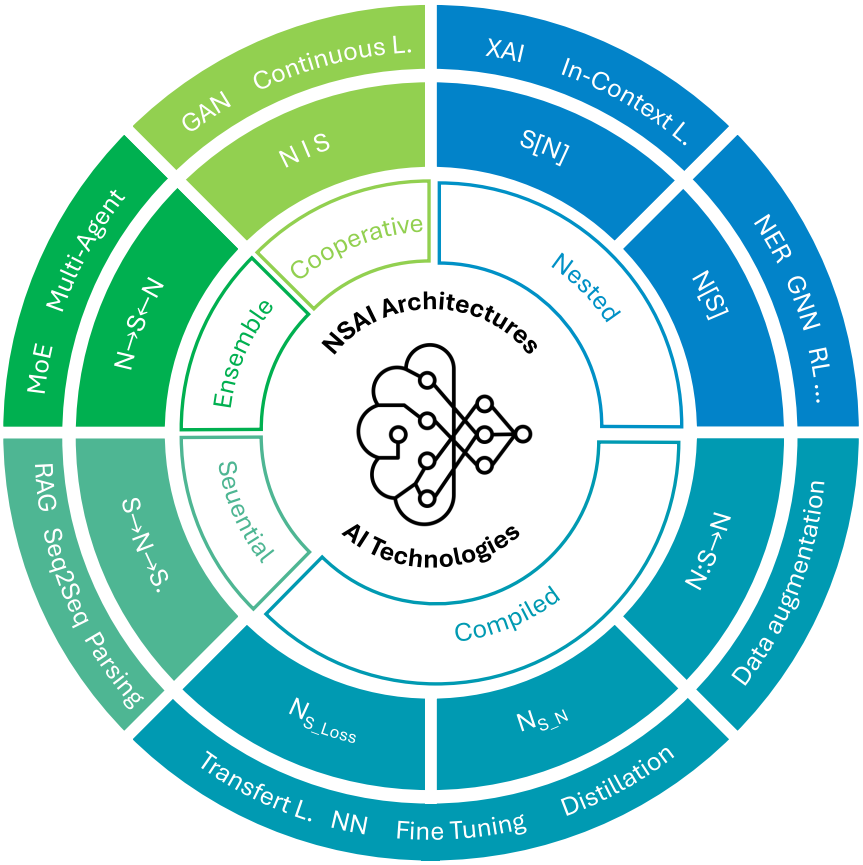


Figure 6: Classification of AI technologies into NSAI architectures.

### 4.2.1 The Sequential Paradigm: From Symbolic to Neural Reasoning

Techniques like RAG, GraphRAG, and Seq2Seq models (including LLMs, e.g., GPT [70]) align with this method due to their reliance on neural encodings of symbolic data (e.g., text or structured information) to perform complex transformations before outputting results in symbolic form. Similarly, semantic parsing benefits from this framework by leveraging NNs to uncover latent patterns in symbolic inputs and generating interpretable symbolic conclusions. For instance, RAG-Logic proposes a dynamic example-based framework using RAG to enhance logical reasoning capabilities by integrating relevant, contextually appropriate examples [71]. It first encodes symbolic input (e.g., logical premises) into neural representations using the RAG knowledge base search module. Neural processing occurs through the translation module, which transforms the input into formal logical formulas. Finally, the fix module ensures syntactic correctness, and the solver module evaluates the logical consistency of the formulas, decoding the results back into symbolic output. This process maintains the interpretability of symbolic reasoning while leveraging the power of NNs to improve flexibility and performance.

### 4.2.2 The Nested Paradigm: Embedding Symbolic Logic in Neural Systems

In-context learning, such as few-shot learning and CoT reasoning, aligns with the *Symbolic[Neuro]* approach by leveraging NNs for context-aware predictions, while symbolic systems facilitate higher-order reasoning. Similarly, XAI falls into this category, as it often combines neural models for extracting features with symbolic frameworks to produce explanations that are easily understood by humans.

Zhang et al. [72] presented a framework in which symbolic reasoning is enhanced by NNs. CoT is used as a method to generate prompts that combine symbolic rules with neural reasoning. For example, the task of reasoning about relationships between entities, such as "Joseph's sister is Katherine" is approached by generating a reasoning path through CoT. The reasoning path is structured using symbolic rules, such as $Sister(A, C) \leftarrow Brother(A, B) \wedge$

$Sister(B, C)$, which define the relationships between entities. These rules are then used to form CoT prompts that guide the model through the reasoning steps. The NN processes these prompts, performing feature extraction and probabilistic inference, while the symbolic system (including the knowledge base and logic rules) orchestrates the overall reasoning process. In this approach, the symbolic framework is the primary system for structuring the reasoning task, and the NN acts as a subcomponent that processes raw data and interprets the symbolic rules in the context of the query.

Methods like GNNs, NER, link prediction, and relation extraction fit into the *Neuro[Symbolic]* category. These methods often leverage symbolic relationships, such as ontologies or graphs, as integral components to enhance neural processing. In addition, they integrate symbolic reasoning subroutines to perform higher-order logical operations, enforce consistency, or derive insights from structured representations. RL and RLHF exemplify this approach, where symbolic reasoning is integrated into the reward shaping and policy optimization stages to enforce logical constraints, ensure decision-making consistency, and align neural outputs with human-like decision-making criteria. For instance, NeuSTIP [73] exemplifies this approach by combining GNN-based neural processing with symbolic reasoning to tackle link prediction and time interval prediction in temporal knowledge graphs (TKGs). NeuSTIP employs temporal logic rules, extracted via "all-walks" on TKGs, to enforce consistency and strengthen reasoning. By embedding symbolic reasoning subroutines into the neural framework, NeuSTIP demonstrates how such models can effectively derive structured insights and perform reasoning under constraints.

### 4.2.3 The Cooperative Paradigm: Iterative Interaction Between Neural and Symbolic Modules

GANs align with this paradigm as their iterative interplay mirrors a cooperative dynamic between two distinct components: the generator creates outputs, while the discriminator evaluates them against predefined criteria, providing structured feedback to improve the generator's

performance. This iterative feedback loop exemplifies the *Neuro* | *Symbolic* framework, where neural networks and symbolic reasoning components collaborate to achieve robust and adaptive problem-solving while adhering to symbolic constraints or logical consistency. Moreover, this cooperative dynamic inherently facilitates continuous learning, a process in which both neural and symbolic modules undergo iterative refinement to enhance their performance over time. In this paradigm, NN continuously updates its internal representations and model parameters in response to feedback derived from the symbolic module's logical inferences and constraint evaluations. This adaptive process enables the NN to generalize more effectively across diverse and evolving data distributions. Simultaneously, the symbolic module is not static; it dynamically revises its rule-based reasoning mechanisms and knowledge structures by integrating new information extracted from the NN's learned representations. An example of this approach in reinforcement learning is the detect-understand-act (DUA) framework [74], where neural and symbolic components collaborate iteratively to solve tasks in a structured manner. In DUA, the detect module uses a traditional computer vision object detector and tracker to process unstructured environmental data into symbolic representations. The understand component, which integrates symbolic reasoning, processes this data using answer set programming (ASP) and inductive logic programming (ILP), ensuring that decisions align with symbolic rules and constraints. The act component, composed of pre-trained reinforcement learning policies, acts as a feedback loop to refine the symbolic representations, allowing the system to converge on solutions that meet predefined criteria.

#### 4.2.4 The Compiled Paradigm: Embedding Symbolic Reasoning Within Neural Computation

Approaches such as model distillation, fine-tuning, pre-training, and transfer learning align with the $Neuro_{Symbolic}$ approach by integrating symbolic constraints or objectives (e.g., logical consistency, relational structures) directly into the learning process of NNs, either through the loss function or at the neuron level via activation functions. This ensures that outputs adhere

to predefined symbolic rules, enabling structured reasoning within the network. Consequently, all NN models can be modeled by this paradigm, by embedding symbolic logic into neural architectures, bridging data-driven learning with symbolic reasoning. Mendez-Lucero et al. [75] complemented this perspective by embedding logical constraints within the loss function. The authors propose a distribution-based method that incorporates symbolic logic, such as propositional formulas and first-order logic, into the learning process. These constraints are encoded as a distribution and incorporated into the optimization procedure using measures like the Fisher-Rao distance or Kullback-Leibler divergence, effectively guiding the NN to adhere to symbolic constraints. This integration of symbolic knowledge into the loss function ensures that the neural model not only learns from data but also incorporates predefined logical rules, reinforcing the connection between neural learning and symbolic reasoning in the context of model distillation, fine-tuning, pre-training, and transfer learning.

Data augmentation leverages the *Neuro:Symbolic* $\rightarrow$ *Neuro* approach, which uses symbolic reasoning to generate synthetic examples, enabling effective data augmentation. By producing high-quality labeled data through logical inference, it enhances the training process of NNs. This method seamlessly integrates the precision and structure of symbolic logic with the scalability and adaptability of NNs, resulting in more robust and efficient learning. Li et al. [76] proposed a methodological framework that exemplifies this approach. Their framework systematically generates labeled data pairs $(x, y)$, where $y$ is derived from $x$ through symbolic transformations based on formal logical rules. The process begins with the formalization of mathematical problems in a symbolic space using mathematical solvers, ensuring the logical validity of the generated instances. Subsequently, mutation mechanisms are applied to diversify the examples, including simplification strategies (reducing the complexity of expressions) and complication strategies (adding constraints or variables). Each transformation results in a new problem instance with its corresponding solution, forming labeled pairs $(x', y')$ that enrich the training corpus with controlled complexity levels.

### 4.2.5 The Fibring Paradigm: Connecting Neural Models Through Symbolic Constraints

Techniques such as multi-agent AI and MoE systems align with this paradigm by leveraging symbolic functions to facilitate communication and coordination between agents (i.e., neural models). Symbolic reasoning mediates interactions, enforces constraints, and ensures alignment with predefined rules, while neural components adapt and learn from collective behaviors. This interplay enables robust and scalable problem-solving in complex, multi-agent environments. Belle et al. [77] explored how the combination of symbolic reasoning and agents can enable the development of advanced systems that are closer to human-like intelligence. They discusses how symbolic reasoning can mediate communication between agents, ensuring that they adhere to predefined rules while allowing the neural components to learn and adapt from collective behaviors. This directly aligns with the fibring paradigm, where multiple NNs are interconnected via a symbolic fibring function, enabling them to collaborate and share information in a structured manner.

Similarly, the recent DeepSeek-R1 [78] framework employs a MoE architecture to enhance reasoning capabilities in large-scale AI systems. DeepSeek's MoE approach activates only a subset of its parameters for each task, mimicking a team of specialized experts. These experts coordinate effectively using reinforcement learning rewards and symbolic constraints, enabling efficient resource utilization while ensuring adherence to reasoning rules. The symbolic constraints act as an intermediary layer, guiding the interactions between experts in a structured manner, aligning their individual outputs to form a cohesive solution.

Likewise, Mixtral 8x7B [79] employs a sparse mixture-of-experts (SMoE) framework, where each layer selects specific expert groups to process input tokens. This architecture not only reduces computational costs but also ensures that the model specializes in handling different tasks through expert routing. Mixtral's ability to adaptively select experts for tasks requiring mathematical reasoning or multilingual understanding exemplifies how MoE-based systems achieve scalability and specialization while maintaining efficiency. The

symbolic mediator within Mixtral ensures that expert selection follows a structured process governed by logical rules, promoting an orderly exchange of information between the experts while adhering to predefined symbolic constraints.

# 5 Evaluation of NSAI Architectures

Ensuring the reliability and practical applicability of NASAI architectures requires a systematic evaluation across multiple well-defined criteria. Such an evaluation not only identifies the strengths and limitations of the architectures but also fosters trust among stakeholders by emphasizing interpretability, transparency, and robustness—qualities essential in domains such as healthcare, finance, and autonomous systems. Moreover, a rigorous assessment provides benchmarks that can stimulate the development of next-generation models. The following sections delineate the key criteria for evaluating NSAI architectures.

## 5.1 Core Criteria

The evaluation framework for NSAI architectures is built upon several fundamental criteria: generalization, scalability, data efficiency, reasoning, robustness, transferability, and interpretability. Each criterion is elaborated below.

**Generalization:** Generalization is defined as the capability of a model to extend its learned representations beyond the training dataset to perform effectively in novel or unforeseen situations. This criterion is evaluated based on:

- *Out-of-distribution (OOD) performance*: The ability to maintain performance on data that deviate from the training distribution.
- *Contextual flexibility*: The capacity to adapt seamlessly to changes in context or domain with minimal retraining.

– *Relational accuracy*: The capacity to identify and exploit relevant relationships in data while mitigating the influence of spurious correlations.

**Scalability:** Scalability assesses the performance of NSAI architecture under increasing data volumes or computational demands. A scalable system should remain efficient and effective as it scales. Key aspects include:

– *Large-scale adaptation*: The ability to process and derive insights from massive datasets.

– *Hardware efficiency*: Optimal utilization of available computational resources, enabling operation on both low-resource devices and high-performance infrastructures.

– *Complexity management*: The ability to accommodate increased architectural complexity without compromising speed or deployment feasability.

**Data Efficiency:** Data efficiency measures how effectively an NSAI model learns from limited data, an important consideration in scenarios where labeled data are scarce or expensive to obtain. This criterion encompasses:

– *Data reduction*: Achieving high performance with a reduced amount of training data.

– *Data optimization*: Maximizing the utility of available data (both labeled and unlabeled), potentially through semi-supervised learning techniques.

– *Incremental adaptability*: The capacity to incorporate new data progressively without undergoing complete retraining.

**Reasoning:** Reasoning reflects the model's ability to analyze data, extract insights, and draw logical conclusions. This criterion underscores the unique advantage of NSAI architectures, which combine neural learning with symbolic reasoning. This criterion evaluates:

– *Logical reasoning*: The systematic application of explicit rules to derive precise and consistent inferences.

- *Relational understanding*: The comprehension of complex relationships between entities within the data.
- *Cognitive versatility*: The integration of various reasoning paradigms (e.g., deductive, inductive, and abductive reasoning) to tackle diverse challenges.

**Robustness:** Robustness measures the system's reliability and resilience to disruptions, including noisy data, adversarial inputs, or dynamic environments. The evaluation considers:

- *Resilience to perturbations/anomalies*: The ability to sustain stable performance despite the presence of noise or adversarial data.
- *Adaptive resilience*: The maintenance of functionality under changing or unpredictable conditions.
- *Bias resilience*: The effectiveness in detecting and correcting biases to ensure fairness and accuracy in predictions.

**Transferability:** Transferability assesses the model's ability in applying learned knowledge to new contexts, domains, or tasks. This is essential for reducing the effort and time required for model adaptation. Its evaluation involves:

- *Multi-domain adaptation*: The capacity to generalize across diverse domains with minimal modifications.
- *Multi-task learning*: The capability to handle multiple tasks simultaneously through shared knowledge representations.
- *Personalization*: The adaptability of the model to meet specific user or application requirements with limited additional effort.

Interpretability: Interpretability evaluates the model's ability to explain its decisions, ensuring transparency and trust in NSAI systems. This criterion assesses:

– *Transparency*: The clarity with which the internal mechanisms and decision processes of the model are revealed.

– *Explanation*: The ability to provide comprehensible justifications for predictions or decisions.

– *Traceability*: The capability to reconstruct the sequence of operations and factors that contributed to a given outcome.

By systematicaly addressing these criteria, researchers and practitioners can ensure that NSAI architectures are not only scientifically rigorous but also practical, adaptable, and ready for real-world applications. This evaluation framework not only facilitates continuous improvement and innovation but also supports the broad adoption of NSAI systems across various industries and application domains.

## 5.2 Evaluation Methodology

The evaluation of NSAI architectures was conducted using a systematic approach to ensure a robust and transparent assessment of their performance across multiple criteria. This process relied on three key sources: scientific literature, empirical findings, and an analysis of the design principles underlying each architecture. **Table 1** summarizes the relevant research works associated with the identified NSAI architectures in Section 3. The scientific literature served as the primary source of qualitative insights, offering detailed analyses of the strengths and limitations of various architectures. Foundational research and state-of-the-art studies provided evidence of performance in areas such as scalability, reasoning, and interpretability, helping to guide the evaluation. Additionally, empirical results from experimental studies and benchmarks offered quantitative data, enabling objective comparisons across architectures. Metrics such as accuracy, adaptability, and efficiency were particularly valuable in validating the claims made in research papers. The design principles of each technology were also considered to understand how neural and symbolic components were integrated. This analysis

provided insights into the inherent capabilities and constraints of each architecture, such as its suitability for handling complex reasoning tasks, scalability to large datasets, or adaptability to dynamic environments.

For each criterion, the ratings were assigned as follows:

- *High:* Awarded to architectures that consistently demonstrated exceptional performance across multiple studies and benchmarks, showcasing clear advantages in the specific criterion.
- *Medium:* Assigned to architectures with satisfactory performance, excelling in certain aspects but with notable limitations in others.
- *Low:* Given to architectures with significant weaknesses, such as inconsistent results or an inability to effectively address the criterion.

By combining insights from literature, empirical findings, and design analysis, this methodology ensures a balanced and evidence-based evaluation. It provides a clear understanding of the strengths and weaknesses of each architecture, enabling meaningful comparisons and guiding future advancements in NSAI research and applications.

Table 1: Set of relevant published NSAI architectures considered in the proposed study.

| **Architecture** | **References** |
|---|---|
| *Symbolic* $\rightarrow$ *Neuro* $\rightarrow$ *Symbolic* | [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104] |
| *Neuro[Symbolic]* | [43], [44] |
| *Symbolic[Neuro]* | [41], [105], [106], [107], [108], [109] |
| *Neuro* \| *Symbolic* | [45], [110], [111], [112], [113], [114], [115] |
| *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* | [116], [47], [77], [78], [79], [23], [24], [25], [26] |
| *Neuro:Symbolic* $\rightarrow$ *Neuro* | [37], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129], [130], [131] |
| $Neuro_{Symbolic_{Loss}}$ | [132], [133], [134], [135], [136], [137] |
| $Neuro_{Symbolic_{Neuro}}$ | [138] [139] |

## 5.3 Results and Discussion

**Figure 7** provides a comparative analysis of various NSAI architectures across seven main evaluation criteria and their respective sub-criteria. This comprehensive evaluation highlights the strengths and weaknesses of each architecture, offering insights into their performance, adaptability, and interpretability.

For example, under the "generalization" criterion, *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* and *Neuro* | *Symbolic* perform well in generalization scenarios, demonstrating strong generalization capabilities, particularly in handling relational accuracy, making it suitable for complex, real-world applications. However, $Neuro_{Symbolic_{Loss}}$ and $Neuro_{Symbolic_{Neuro}}$ demonstrates notable shortcomings in continuous flexibility and OOD generalization, highlighting its difficulty in adapting to dynamic and evolving contexts without the need for extensive retraining. As for the "scalability" criterion, *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* and $Neuro_{Symbolic_{Neuro}}$ excel across all sub-criteria, including large-scale adaptation and hardware efficiency, demonstrating their capacity to handle industrial-scale applications. Conversely, *Symbolic[Neuro]* achieves only medium performance in scalability, reflecting challenges in balancing its rule-based reasoning with the demands of large-scale or resource-intensive tasks. In particular, *Neuro* | *Symbolic*, rated low, struggles to maintain efficiency and adaptability when scaling to more complex systems, highlighting a need for improved coordination between its neural and symbolic components.

In terms of "data efficiency", architectures such as *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro*, *Symbolic Neuro Symbolic*, and $Neuro_{Symbolic_{Neuro}}$ consistently achieve high ratings, excelling in both data reduction and optimization. This indicates their ability to learn effectively with limited data. However, *Symbolic[Neuro]* demonstrates only medium adaptability when incorporating incremental data updates. When evaluating the "Reasoning" criterion, architectures such as *Symbolic[Neuro]*, *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro*, and $Neuro_{Symbolic_{Neuro}}$ show strong capabilities in logical reasoning and relational understanding. However, *Neuro:Symbolic* $\rightarrow$ *Neuro* displays lower versatility in combining diverse reasoning methods, reflecting limitations in solving

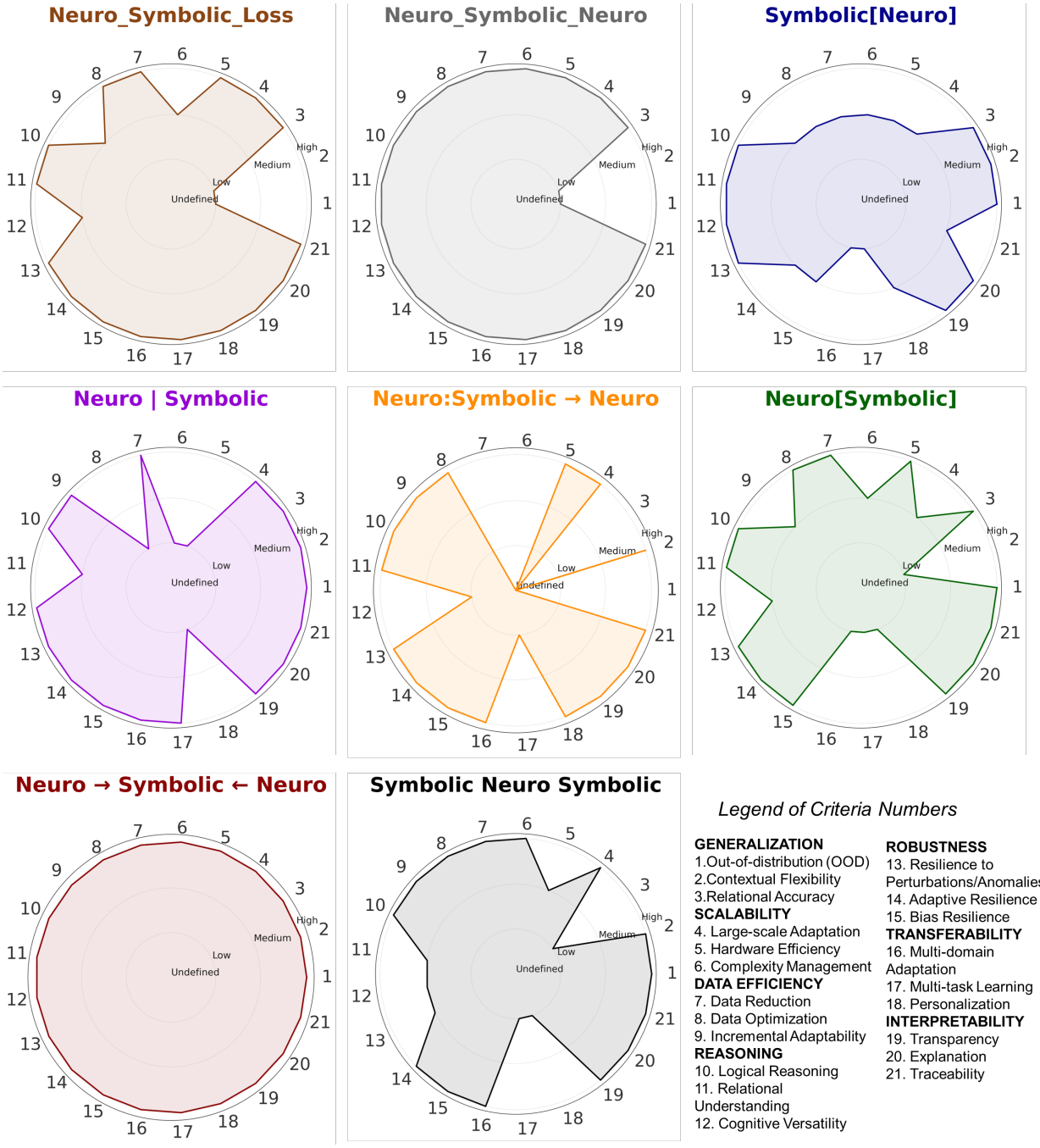


Figure 7: Comparison of NSAI architectures based on various criteria and sub-criteria.

complex problems. For "Robustness", most architectures perform well, demonstrating high resilience to perturbations and effective bias handling. However, *Symbolic[Neuro]* and *Symbolic Neuro Symbolic* architectures exhibit weaknesses in adapting to dynamic environments and mitigating biases effectively.

Regarding "Transferability", architectures like *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro*, $Neuro_{Symbolic_{Loss}}$, and $Neuro_{Symbolic_{Neuro}}$ excel in multi-task learning and multi-domain adaptation, enabling effective reuse of knowledge across domains. In contrast, *Symbolic Neuro Symbolic*, *Neuro:Symbolic* $\rightarrow$ *Neuro*, and nested architectures demonstrate lower adaptability to personalized applications. Lastly, in "Interpretability", most architectures perform well, achieving high marks in transparency and traceability. *Symbolic[Neuro]* also achieves commendable results in this criterion, demonstrating its ability to explain decisions effectively, which is essential for sensitive applications like healthcare and finance.

Overall, the *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* architecture emerges as the best-performing model, consistently achieving high ratings across all criteria. Its exceptional performance in generalization, scalability, and interpretability makes it highly suitable for real-world applications that demand reliability, adaptability, and transparency. While other architectures also perform well in specific areas, the versatility and robustness of *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* set it apart as the most balanced and capable solution. This conclusion aligns with findings in the state of the art, which highlight the effectiveness of *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* architectures in leveraging advanced AI technologies, such as multi-agent systems. Multi-agent systems are well-documented for their robustness, particularly in dynamic and distributed environments, where their ability to coordinate, adapt, and reason collectively enables superior performance. For instance, Subramanian et al. [140] demonstrated that incorporating neuro-symbolic approaches into multi-agent RL enhances both interpretability and probabilistic decision-making. This makes such systems highly robust in environments with partial observability or uncertainties. Similarly, Keren et al. [141] highlighted that collaboration

among agents in multi-agent frameworks promotes group resilience, enabling these systems to adapt effectively to dynamic or adversarial conditions. These attributes are particularly valuable in *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* architectures, as they address the critical need for transparency and robustness in complex real-world applications.

# 6 Conclusion

This study evaluates several NSAI architectures against a comprehensive set of criteria, including generalization, scalability, data efficiency, reasoning, robustness, transferability, and interpretability. The results highlight the strengths and weaknesses of each architecture, offering valuable insights into their capabilities for real-world applications. Among the architectures investigated, *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* emerges as the most balanced and robust solution. It consistently demonstrates superior performance across multiple criteria, excelling in generalization, scalability, and interpretability. These results align with recent advancements in the field, which emphasize the role of multi-agent systems in enhancing robustness and adaptability. As shown in recent studies, multi-agent frameworks, when integrated with neuro-symbolic methods, provide significant advantages in handling uncertainty, fostering collaboration, and maintaining resilience in dynamic environments. This integration not only enables better decision-making but also ensures transparency and traceability, which are critical for sensitive applications. Moreover, its ability to leverage advanced AI technologies, such as multi-agent systems, positions *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* as a leading candidate for addressing the demands of generative AI applications.

Future work will be focused on exploring the scalability of this architecture in even larger and more diverse environments. Additionally, advancing the integration of symbolic reasoning within multi-agent systems may further enhance their robustness and cognitive versatility. As the field evolves, *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro* architectures are likely to remain at the forefront of innovation, offering practical and scientifically grounded solutions to the most

pressing challenges in AI.

## CRediT authorship contribution statement

**Oualid Bougzime:** Writing – original draft, Methodology, Investigation. **Samir Jabbar:** Writing – original draft, Methodology, Investigation. **Christophe Cruz:** Writing – review & editing, Methodology, Supervision. **Frédéric Demoly:** Writing – review & editing, Methodology, Supervision, Funding acquisition, Project administration.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was funded by the IUF, Innovation Chair on 4D Printing, the French National Research Agency under the "France 2030 Initiative" and the "DIADEM Program", grant number 22-PEXD-0016 ("ARTEMIS").

## References

[1] Artur d'Avila Garcez and Luis C Lamb. Neurosymbolic ai: The 3 rd wave. *Artificial Intelligence Review*, 56(11):12387–12406, 2023.

[2] Leslie G Valiant. Three problems in computer science. *Journal of the ACM (JACM)*, 50(1):96–99, 2003.

[3] Yoshua Bengio, Gary Marcus, and Vincent Boucher. AI DEBATE! Yoshua Bengio vs Gary Marcus, 2019.

[4] Y Bengio. System 2 deep learning: Higher-level cognition, agency, out-of-distribution generalization and causality. In *30th International Joint Conference on Artificial Intelligence. https://ijcai-21. org/invited-talks*, 2022.

[5] Gary Marcus and Ernest Davis. *Rebooting AI: Building artificial intelligence we can trust*. Vintage, 2019.

[6] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[7] Zhixuan Liu, Zihao Wang, Yuan Lin, and Hang Li. A neural-symbolic approach to natural language understanding. *arXiv preprint arXiv:2203.10557*, 2022.

[8] Jing Zhang, Bo Chen, Lingxi Zhang, Xirui Ke, and Haipeng Ding. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open*, 2:14–35, 2021.

[9] Luís C Lamb, Artur Garcez, Marco Gori, Marcelo Prates, Pedro Avelar, and Moshe Vardi. Graph neural networks meet neural-symbolic computing: A survey and perspective. *arXiv preprint arXiv:2003.00330*, 2020.

[10] Laura Von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, et al. Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):614–633, 2021.

[11] Vaishak Belle. Symbolic logic meets machine learning: A brief survey in infinite domains. In *International conference on scalable uncertainty management*, pages 3–16. Springer, 2020.

[12] Kyle Hamilton, Aparna Nayak, Bojan Božić, and Luca Longo. Is neuro-symbolic ai meeting its promises in natural language processing? a structured review. *Semantic Web*, 15(4):1265–1306, 2024.

[13] Henry Kautz. The third ai summer: Aaai robert s. engelmore memorial lecture. *Ai magazine*, 43(1):105–125, 2022.

[14] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[15] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[16] I Sutskever. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014.

[17] Peng Jiang and Xiaodong Cai. A survey of semantic parsing techniques. *Symmetry*, 16(9):1201, 2024.

[18] Mónica Marrero, Julián Urbano, Sonia Sánchez-Cuadrado, Jorge Morato, and Juan Miguel Gómez-Berbís. Named entity recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces*, 35(5):482–489, 2013.

[19] Xiaoyan Zhao, Yang Deng, Min Yang, Lingzhi Wang, Rui Zhang, Hong Cheng, Wai Lam, Ying Shen, and Ruifeng Xu. A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys*, 56(11):1–39, 2024.

[20] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

[21] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.

[22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[23] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024.

[24] Diego Maldonado, Edison Cruz, Jackeline Abad Torres, Patricio J Cruz, and Silvana Gamboa. Multi-agent systems: A survey about its components, framework and workflow. *IEEE Access*, 2024.

[25] Xu Owen He. Mixture of a million experts, 2024.

[26] Ka Man Lo, Zeyu Huang, Zihan Qiu, Zili Wang, and Jie Fu. A closer look into mixture-of-experts in large language models, 2024.

[27] Zaid Alyafeai, Maged Saeed AlShaibani, and Irfan Ahmad. A survey on transfer learning in natural language processing. *arXiv preprint arXiv:2007.04239*, 2020.

[28] Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[29] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.

[30] Archit Parnami and Minwoo Lee. Learning from few examples: A summary of approaches to few-shot learning. *arXiv preprint arXiv:2203.04291*, 2022.

[31] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, tax-

onomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[32] Geoffrey Hinton. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[33] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[34] Zhiyuan Chen and Bing Liu. *Lifelong machine learning*. Morgan & Claypool Publishers, 2018.

[35] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2. Minneapolis, Minnesota, 2019.

[36] Oualid Bougzime, Christophe Cruz, Jean-Claude André, Kun Zhou, H. Jerry Qi, and Frédéric Demoly. Neuro-symbolic artificial intelligence in accelerated design for 4d printing: Status, challenges, and perspectives. *Materials & Design*, 2025. Under review.

[37] Guillaume Lample and François Charton. Deep learning for symbolic mathematics. *arXiv preprint arXiv:1912.01412*, 2019.

[38] Saoussen Dimassi, Frédéric Demoly, Christophe Cruz, H Jerry Qi, Kyoung-Yun Kim, Jean-Claude André, and Samuel Gomes. An ontology-based framework to formalize and represent 4d printing knowledge in design. *Computers in Industry*, 126:103374, 2021.

[39] Tomas Mikolov. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 3781, 2013.

[40] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[41] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[42] Rémi Coulom. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*, pages 72–83. Springer, 2006.

[43] Marijn JH Heule, Oliver Kullmann, and Victor W Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 228–245. Springer, 2016.

[44] Kanika Madan, Nan Rosemary Ke, Anirudh Goyal, Bernhard Schölkopf, and Yoshua Bengio. Fast and slow learning of recurrent independent mechanisms. *arXiv preprint arXiv:2105.08710*, 2021.

[45] Jiayuan Mao, Chuang Gan, Pushmeet Kohli, Joshua B Tenenbaum, and Jiajun Wu. The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. *arXiv preprint arXiv:1904.12584*, 2019.

[46] Ryan Riegel, Alexander Gray, Francois Luus, Naweed Khan, Ndivhuwo Makondo, Ismail Yunus Akhalwaya, Haifeng Qian, Ronald Fagin, Francisco Barahona, Udit Sharma, et al. Logical neural networks. *arXiv preprint arXiv:2006.13155*, 2020.

[47] Artur S d'Avila Garcez and Dov M Gabbay. Fibring neural networks. In *AAAI*, pages 342–347, 2004.

[48] Amit Sheth, Vishal Pallagani, and Kaushik Roy. Neurosymbolic ai for enhancing instructability in generative ai. *IEEE Intelligent Systems*, 39(5):5–11, 2024.

[49] Maxwell J Jacobson and Yexiang Xue. Integrating symbolic reasoning into neural generative models for design generation. *Artificial Intelligence*, 339:104257, 2025.

[50] Xunjian Yin and Xiaojun Wan. How do seq2seq models perform on end-to-end data-to-text generation? In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7701–7710, 2022.

[51] Diji Yang, Jinmeng Rao, Kezhen Chen, Xiaoyuan Guo, Yawen Zhang, Jie Yang, and Yi Zhang. Im-rag: Multi-round retrieval-augmented generation through learning inner monologues. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 730–740, 2024.

[52] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

[53] Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. *Expert systems with applications*, 141:112948, 2020.

[54] Grigoris Antoniou and Frank van Harmelen. Web ontology language: Owl. *Handbook on ontologies*, pages 91–110, 2009.

[55] Costas Mavromatis and George Karypis. Gnn-rag: Graph neural retrieval for large language model reasoning. *arXiv preprint arXiv:2405.20139*, 2024.

[56] Arya Roy. Recent trends in named entity recognition (ner). *arXiv preprint arXiv:2101.11420*, 2021.

[57] Tao Wu, Xiaolin You, Xingping Xian, Xiao Pu, Shaojie Qiao, and Chao Wang. Towards deep understanding of graph convolutional networks for relation extraction. *Data & Knowledge Engineering*, 149:102265, 2024.

[58] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 2023.

[59] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, 2017.

[60] Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. Gshard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*, 2020.

[61] Junlin Xie, Zhihong Chen, Ruifei Zhang, Xiang Wan, and Guanbin Li. Large multimodal agents: A survey. *arXiv preprint arXiv:2402.15116*, 2024.

[62] Weiping Ding, Mohamed Abdel-Basset, Hossam Hawash, and Ahmed M Ali. Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. *Information Sciences*, 615:238–292, 2022.

[63] Yukun Huang, Yanda Chen, Zhou Yu, and Kathleen McKeown. In-context learning distillation: Transferring few-shot learning ability of pre-trained language models. *arXiv preprint arXiv:2212.10670*, 2022.

[64] Mohammadreza Iman, Hamid Reza Arabnia, and Khaled Rasheed. A review of deep transfer learning and recent advancements. *Technologies*, 11(2):40, 2023.

[65] Farhad Pourpanah, Moloud Abdar, Yuxuan Luo, Xinlei Zhou, Ran Wang, Chee Peng Lim, Xi-Zhao Wang, and QM Jonathan Wu. A review of generalized zero-shot learning methods. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4051–4070, 2022.

[66] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, pages 1–7, 2021.

[67] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.

[68] Josef Dai, Xuehai Pan, Ruiyang Sun, Jiaming Ji, Xinbo Xu, Mickel Liu, Yizhou Wang, and Yaodong Yang. Safe rlhf: Safe reinforcement learning from human feedback. *arXiv preprint arXiv:2310.12773*, 2023.

[69] Venkata Rama Padmaja Chinimilli and Lakshminarayana Sadasivuni. The rise of ai: a comprehensive research review. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 13:2226, 06 2024.

[70] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene,

Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama,

Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. Gpt-4 technical report, 2024.

[71] Anonymous. RAG-logic: Enhance neuro-symbolic approaches for logical reasoning with retrieval-augmented generation. In *Submitted to ACL Rolling Review - June 2024*, 2024. under review.

[72] Hanlin Zhang, YiFan Zhang, Li Erran Li, and Eric Xing. The impact of symbolic representations on in-context learning for few-shot reasoning. In *NeurIPS 2022 Workshop on Neuro Causal and Symbolic AI (nCSI)*, 2022.

[73] Ishaan Singh, Navdeep Kaur, Garima Gaur, et al. Neustip: A novel neuro-symbolic model for link and time prediction in temporal knowledge graphs. *arXiv preprint arXiv:2305.11301*, 2023.

[74] Ludovico Mitchener, David Tuckey, Matthew Crosby, and Alessandra Russo. Detect, understand, act: A neuro-symbolic hierarchical reinforcement learning framework. *Machine Learning*, 111(4):1523–1549, 2022.

[75] Miguel Angel Mendez-Lucero, Enrique Bojorquez Gallardo, and Vaishak Belle. Semantic objective functions: A distribution-aware method for adding logical constraints in deep learning. *arXiv preprint arXiv:2405.15789*, 2024.

[76] Zenan Li, Zhi Zhou, Yuan Yao, Yu-Feng Li, Chun Cao, Fan Yang, Xian Zhang, and Xiaoxing Ma. Neuro-symbolic data generation for math reasoning. *arXiv preprint arXiv:2412.04857*, 2024.

[77] Vaishak Belle, Michael Fisher, Alessandra Russo, Ekaterina Komendantskaya, and Alistair Nottle. Neuro-symbolic ai+ agent systems: A first reflection on trends, opportunities and challenges. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 180–200. Springer, 2023.

[78] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[79] Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. Mixtral of experts. *arXiv preprint arXiv:2401.04088*, 2024.

[80] Panagiotis Kouris, Georgios Alexandridis, and Andreas Stafylopatis. Abstractive text summarization: Enhancing sequence-to-sequence models using word sense disambiguation and semantic content generalization. *Computational Linguistics*, 47(4):813–859, 2021.

[81] Alexander Sutherland, Sven Magg, and Stefan Wermter. Leveraging recursive processing for neural-symbolic affect-target associations. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2019.

[82] Yu Gu, Jeff Z Pan, Gong Cheng, Heiko Paulheim, and Giorgos Stoilos. Local abox consistency prediction with transparent tboxes using gated graph neural networks. In *NeSy@ IJCAI*, 2019.

[83] Qingyao Cui, Yanquan Zhou, and Mingming Zheng. Sememes-based framework for knowledge graph embedding with comprehensive-information. In *Knowledge Science,*

*Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14*, pages 419–426. Springer, 2021.

[84] Canran Xu and Ruijiang Li. Relation embedding with dihedral group in knowledge graph. *arXiv preprint arXiv:1906.00687*, 2019.

[85] Alexander I Cowen-Rivers, Pasquale Minervini, Tim Rocktaschel, Matko Bosnjak, Sebastian Riedel, and Jun Wang. Neural variational inference for estimating uncertainty in knowledge graph embeddings. *arXiv preprint arXiv:1906.04985*, 2019.

[86] Mariem Bounabi, Karim Elmoutaouakil, and Khalid Satori. A new neutrosophic tf-idf term weighting for text mining tasks: text classification use case. *International Journal of Web Information Systems*, 17(3):229–249, 2021.

[87] Fatima Es-Sabery, Abdellatif Hair, Junaid Qadir, Beatriz Sainz-De-Abajo, Begoña García-Zapirain, and Isabel De La Torre-Díez. Sentence-level classification using parallel fuzzy deep learning classifier. *IEEE Access*, 9:17943–17985, 2021.

[88] Rinaldo Lima, Bernard Espinasse, and Frederico Freitas. The impact of semantic linguistic features in relation extraction: A logical relational learning approach. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 648–654, 2019.

[89] Mengjia Zhou, Donghong Ji, and Fei Li. Relation extraction in dialogues: A deep learning model based on the generality and specialty of dialogue text. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2015–2026, 2021.

[90] Jibing Gong, Zhiyong Teng, Qi Teng, Hekai Zhang, Linfeng Du, Shuai Chen, Md Zakirul Alam Bhuiyan, Jianhua Li, Mingsheng Liu, and Hongyuan Ma. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification. *IEEE Access*, 8:30885–30896, 2020.

[91] Ange Adrienne Nyamen Tato, Roger Nkambou, and Aude Dufresne. Hybrid deep neural networks to predict socio-moral reasoning skills. In *EDM*, 2019.

[92] John Langton and Krishna Srihasam. Applied medical code mapping with character-based deep learning models and word-based logic. In *Proceedings of the 1st and 2nd Workshops on Natural Logic Meets Machine Learning (NALOMA)*, pages 7–11, 2021.

[93] Adrian MP Braşoveanu and Răzvan Andonie. Semantic fake news detection: a machine learning perspective. In *Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part I 15*, pages 656–667. Springer, 2019.

[94] Claudio Pinhanez, Paulo Cavalin, Victor Henrique Alves Ribeiro, Ana Appel, Heloisa Candello, Julio Nogima, Mauro Pichiliani, Melina Guerra, Maira de Bayser, Gabriel Malfatti, et al. Using meta-knowledge mined from identifiers to improve intent recognition in conversational systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7014–7027, 2021.

[95] Chen Dehua, Zhong Keting, and He Jianrong. Bdcn: Semantic embedding self-explanatory breast diagnostic capsules network. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1178–1189, 2021.

[96] Lejla Begic Fazlic, Ahmed Hallawa, Anke Schmeink, Arne Peine, Lukas Martin, and Guido Dartmann. A novel nlp-fuzzy system prototype for information extraction from medical guidelines. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1025–1030. IEEE, 2019.

[97] Jennifer D'Souza, Isaiah Onando Mulang, and Sören Auer. Team svmrank: Leveraging feature-rich support vector machines for ranking explanations to elementary science

questions. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*, pages 90–100, 2019.

[98] Raja Ayyanar, George Koomullil, and Hariharan Ramasangu. Causal relation classification using convolutional neural networks and grammar tags. In *2019 IEEE 16th India Council International Conference (INDICON)*, pages 1–3. IEEE, 2019.

[99] Dou Hu, Lingwei Wei, and Xiaoyong Huai. Dialoguecrn: Contextual reasoning networks for emotion recognition in conversations. *arXiv preprint arXiv:2106.01978*, 2021.

[100] Kunlong Chen, Weidi Xu, Xingyi Cheng, Zou Xiaochuan, Yuyu Zhang, Le Song, Taifeng Wang, Yuan Qi, and Wei Chu. Question directed graph attention network for numerical reasoning over text. *arXiv preprint arXiv:2009.07448*, 2020.

[101] Prashanti Manda, Saed SayedAhmed, and Somya D Mohanty. Automated ontology-based annotation of scientific literature using deep learning. In *Proceedings of the international workshop on semantic Big Data*, pages 1–6, 2020.

[102] Hiroshi Honda and Masafumi Hagiwara. Question answering systems with deep learning-based symbolic processing. *IEEE Access*, 7:152368–152378, 2019.

[103] Claudia Schon, Sophie Siebert, and Frieder Stolzenburg. The corg project: cognitive reasoning. *KI-Künstliche Intelligenz*, 33:293–299, 2019.

[104] Kareem Amin. Cases without borders: automating knowledge acquisition approach using deep autoencoders and siamese networks in case-based reasoning. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 133–140. IEEE, 2019.

[105] Qiaochu Chen, Aaron Lamoreaux, Xinyu Wang, Greg Durrett, Osbert Bastani, and Isil Dillig. Web question answering with neurosymbolic program synthesis. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pages 328–343, 2021.

[106] Zeming Chen, Qiyue Gao, and Lawrence S Moss. Neurallog: Natural language inference with joint neural and logical reasoning. *arXiv preprint arXiv:2105.14167*, 2021.

[107] Maria Leonor Pacheco and Dan Goldwasser. Modeling content and context with deep relational learning. *Transactions of the Association for Computational Linguistics*, 9:100–119, 2021.

[108] Iti Chaturvedi, Ranjan Satapathy, Sandro Cavallari, and Erik Cambria. Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recognition Letters*, 125:264–270, 2019.

[109] Jinghui Qin, Xiaodan Liang, Yining Hong, Jianheng Tang, and Liang Lin. Neural-symbolic solver for math word problems with auxiliary tasks. *arXiv preprint arXiv:2107.01431*, 2021.

[110] Yiqun Yao, Jiaming Xu, Jing Shi, and Bo Xu. Learning to activate logic rules for textual reasoning. *Neural Networks*, 106:42–49, 2018.

[111] Jihao Shi, Xiao Ding, Li Du, Ting Liu, and Bing Qin. Neural natural logic inference for interpretable question answering. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3673–3684, 2021.

[112] Blaž Škrlj, Matej Martinc, Nada Lavrač, and Senja Pollak. autobot: evolving neuro-symbolic representations for explainable low resource text classification. *Machine Learning*, 110(5):989–1028, 2021.

[113] Wenya Wang and Sinno Jialin Pan. Variational deep logic network for joint inference of entities and relations. *Computational Linguistics*, 47(4):775–812, 2021.

[114] Henrique Lemos, Pedro Avelar, Marcelo Prates, Artur Garcez, and Luís Lamb. Neural-symbolic relational reasoning on graph models: Effective link inference and computation from knowledge bases. In *International Conference on Artificial Neural Networks*, pages 647–659. Springer, 2020.

[115] Qiuyuan Huang, Li Deng, Dapeng Wu, Chang Liu, and Xiaodong He. Attentive tensor product learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1344–1351, 2019.

[116] Rajarshi Das, Manzil Zaheer, Dung Thai, Ameya Godbole, Ethan Perez, Jay-Yoon Lee, Lizhen Tan, Lazaros Polymenakos, and Andrew McCallum. Case-based reasoning for natural language queries over knowledge bases. *arXiv preprint arXiv:2104.08762*, 2021.

[117] Len Yabloko. Ethan at semeval-2020 task 5: Modelling causal reasoning in language using neuro-symbolic cloud computing. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 645–652, 2020.

[118] Ben Zhou, Kyle Richardson, Qiang Ning, Tushar Khot, Ashish Sabharwal, and Dan Roth. Temporal reasoning on implicit events from distant supervision. *arXiv preprint arXiv:2010.12753*, 2020.

[119] Ekaterina Saveleva, Volha Petukhova, Marius Mosbach, and Dietrich Klakow. Graph-based argument quality assessment. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1268–1280, 2021.

[120] Komal Gupta, Tirthankar Ghosal, and Asif Ekbal. A neuro-symbolic approach for question answering on research articles. In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pages 40–49, 2021.

[121] David Demeter and Doug Downey. Just add functions: A neural-symbolic language model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7634–7642, 2020.

[122] Hang Jiang, Sairam Gurajada, Qiuhao Lu, Sumit Neelam, Lucian Popa, Prithviraj Sen, Yunyao Li, and Alexander Gray. Lnn-el: A neuro-symbolic approach to short-text entity linking. *arXiv preprint arXiv:2106.09795*, 2021.

[123] Konstantinos Kogkalidis, Michael Moortgat, and Richard Moot. Neural proof nets. *arXiv preprint arXiv:2009.12702*, 2020.

[124] Qiyuan Zhang, Lei Wang, Sicheng Yu, Shuohang Wang, Yang Wang, Jing Jiang, and Ee-Peng Lim. Noahqa: Numerical reasoning with interpretable graph question answering dataset. *arXiv preprint arXiv:2109.10604*, 2021.

[125] Prithviraj Sen, Marina Danilevsky, Yunyao Li, Siddhartha Brahma, Matthias Boehm, Laura Chiticariu, and Rajasekar Krishnamurthy. Learning explainable linguistic expressions with neural inductive logic programming for sentence classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4211–4221, 2020.

[126] Siyu Huo, Tengfei Ma, Jie Chen, Maria Chang, Lingfei Wu, and Michael J Witbrock. Graph enhanced cross-domain text-to-sql generation. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*, pages 159–163, 2019.

[127] Jingchi Jiang, Huanzheng Wang, Jing Xie, Xitong Guo, Yi Guan, and Qiubin Yu. Medical knowledge embedding based on recursive neural network for multi-disease diagnosis. *Artificial Intelligence in Medicine*, 103:101772, 2020.

[128] Wenge Liu, Jianheng Tang, Xiaodan Liang, and Qingling Cai. Heterogeneous graph reasoning for knowledge-grounded medical dialogue system. *Neurocomputing*, 442:260–268, 2021.

[129] Subhajit Chaudhury, Prithviraj Sen, Masaki Ono, Daiki Kimura, Michiaki Tatsubori, and Asim Munawar. Neuro-symbolic approaches for text-based policy learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3073–3078, 2021.

[130] Pat Verga, Haitian Sun, Livio Baldini Soares, and William W Cohen. Facts as experts: Adaptable and interpretable neural memory over symbolic knowledge. *arXiv preprint arXiv:2007.00849*, 2020.

[131] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. *Advances in neural information processing systems*, 26, 2013.

[132] Luciano Serafini and Artur d'Avila Garcez. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422*, 2016.

[133] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[134] Kezhen Chen, Qiuyuan Huang, Hamid Palangi, Paul Smolensky, Ken Forbus, and Jianfeng Gao. Mapping natural-language problems to formal-language solutions using structured neural representations. In *International Conference on Machine Learning*, pages 1566–1575. PMLR, 2020.

[135] Lisa Graziani, Stefano Melacci, and Marco Gori. Jointly learning to detect emotions and predict facebook reactions. In *Artificial Neural Networks and Machine Learning–ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part IV 28*, pages 185–197. Springer, 2019.

[136] Edgar Jaim Altszyler Lemcovich, Pablo Brusco, Nikoletta Basiou, John Byrnes, and Dimitra Vergyri. Zero-shot multi-domain dialog state tracking using descriptive rules. 2020.

[137] Amir Hussain and Erik Cambria. Semi-supervised learning for big social data analysis. *Neurocomputing*, 275:1662–1673, 2018.

[138] Paul Smolensky, Moontae Lee, Xiaodong He, Wen tau Yih, Jianfeng Gao, and Li Deng. Basic reasoning with tensor product representations, 2016.

[139] Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1-2):159–216, 1990.

[140] Chitra Subramanian, Miao Liu, Naweed Khan, Jonathan Lenchner, Aporva Amarnath, Sarathkrishna Swaminathan, Ryan Riegel, and Alexander Gray. A neuro-symbolic approach to multi-agent rl for interpretability and probabilistic decision making. *arXiv preprint arXiv:2402.13440*, 2024.

[141] Sarah Keren, Matthias Gerstgrasser, Ofir Abu, and Jeffrey Rosenschein. Collaboration promotes group resilience in multi-agent ai. *arXiv preprint arXiv:2111.06614*, 2021.