

# Neuro-Symbolic Architectures: Formalizing, Extending, Evaluating, and Empowering Generative AI

Journal Title  
XX(X):1–26  
©The Author(s) 2025  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/ToBeAssigned  
www.sagepub.com/

SAGE

Oualid Bougzime<sup>1</sup> and Samir Jabbar<sup>2</sup> and Christophe Cruz<sup>2</sup> and Frédéric Demoly<sup>1,3</sup>

## Abstract

Neuro-symbolic artificial intelligence (NSAI) represents a promising direction in artificial intelligence (AI) by combining deep learning's ability to process large-scale and unstructured data with the structured reasoning capabilities of symbolic methods. By leveraging their complementary strengths, NSAI has the potential to improve generalization, reasoning, and transparency while addressing limitations related to interpretability and data efficiency. This paper clarifies, formalizes, and extends Kautz's NSAI taxonomy, proposes a conditional and illustrative mapping of selected modern generative AI methods to these architectures, and introduces a qualitative, literature-grounded evaluation framework for comparing NSAI paradigms. We systematically examine NSAI architectures and discuss how recent generative AI approaches may relate to them when explicit symbolic components, constraints, or reasoning mechanisms are present. We then compare these architectures across criteria including generalization, reasoning capabilities, transferability, and interpretability, thereby providing a structured analysis of their respective strengths and limitations. Among the architectures considered, the Neuro  $\rightarrow$  Symbolic  $\leftarrow$  Neuro model appears particularly well balanced across our qualitative, literature-grounded criteria; however, this observation should be interpreted as an indicative comparison rather than as a definitive empirical ranking. Finally, we illustrate the practical relevance of the taxonomy in the context of the emerging 4D printing technology by proposing paradigm-specific application scenarios for the design of smart materials and structures. Overall, this work provides a structured reference for discussing architectural choices and for guiding future neuro-symbolic research in generative AI and engineering domains.

## 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, 4D Printing

## 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 Kahneman'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

<sup>1</sup>ICB UMR 6303 CNRS, Université Marie et Louis Pasteur, UTBM, 90010 Belfort Cedex, France

<sup>2</sup>ICB UMR 6303 CNRS, Université Bourgogne Europe, 21078 Dijon, France

<sup>3</sup>Institut universitaire de France (IUF), Paris, France

## Corresponding author:

Oualid Bougzime, ICB UMR 6303 CNRS, Université Marie et Louis Pasteur, UTBM, 90010 Belfort Cedex, France

Email: oualid.bougzime@utbm.fr

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]. Kautz [13] identified several NSAI architectures that effectively integrate these paradigms, 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 examine how selected generative AI technologies may relate to the core categories of NSAI and what insights such a comparison can provide regarding their strengths and limitations. Building on Kautz's taxonomy, our contribution is fourfold: (i) to clarify, formalize, and extend existing NSAI architectures, (ii) to provide an illustrative and conditional mapping of selected recent generative AI methods onto NSAI paradigms, explicitly distinguishing literature-grounded instantiations from prospective integrations, (iii) to develop a qualitative and literature-grounded evaluation framework for comparing NSAI architectures across multiple criteria, and (iv) to propose and illustrate NSAI-based application scenarios for 4D printing, showing how each paradigm can be instantiated in this domain and what opportunities it may support for design workflows.

## 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 pre-coded 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.

## 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—most notably the fibring architecture—and by introducing more granular subclassifications within the compiled architecture to capture the evolving landscape of NSAI systems. **These proposed extensions should be considered as analytical refinements rather than as an established community consensus.** These approaches range from symbolic systems augmented by neural modules for specialized tasks to deeply integrated models in which 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.

### *Sequential: Symbolic → Neuro → Symbolic*

As part of the sequential NSAI, the *Symbolic → Neuro → Symbolic* architecture involves systems where both 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:

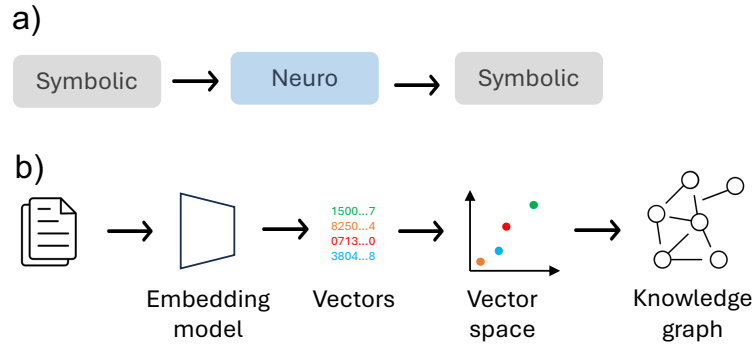
$$y = f_{\text{neural}}(x) \quad (1)$$

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

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.

### *Nested: Symbolic[Neuro] and Neuro[Symbolic]*

The nested NSAI category is composed of two different architectures. The first – *Symbolic[Neuro]* – places a



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

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:

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

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:

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

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.

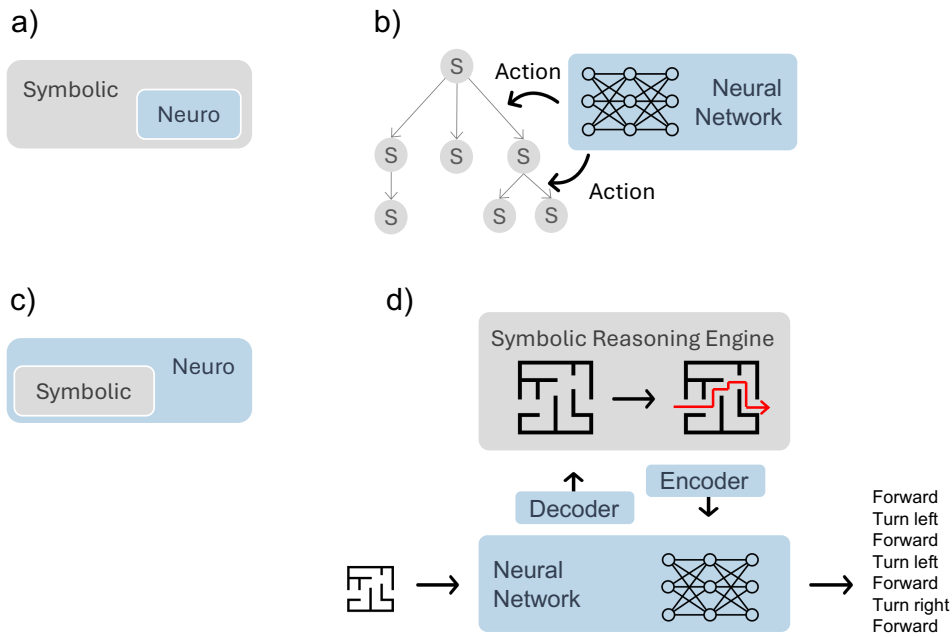
Figure 2d illustrates this framework, a symbolic reasoning engine processes structured data, such as a maze, to generate a solution path. A NN encodes the problem into a latent representation and decodes it into a symbolic sequence of actions (e.g., forward, turn left, turn right).

### Cooperative: *Neuro | Symbolic*

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:

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

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



**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.

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 be 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).

### Compiled: $\text{Neuro}_{\text{SymbolicLoss}}$ , $\text{Neuro}_{\text{SymbolicNeuro}}$ and $\text{Neuro:Symbolic} \rightarrow \text{Neuro}$

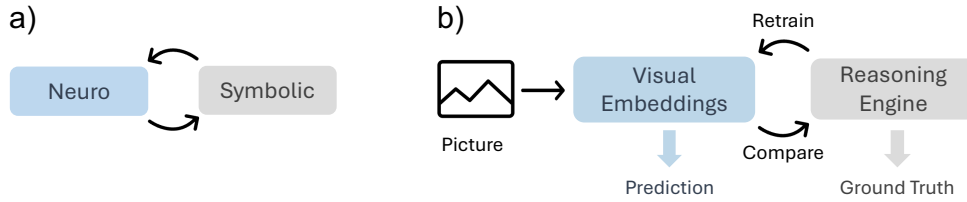
As part of the compiled NSAI,  $\text{Neuro}_{\text{SymbolicLoss}}$  uses symbolic reasoning into the loss function of a NN (Figure 4a). The loss function is typically used to measure the discrepancy between the model’s predictions and the true outputs. By incorporating symbolic rules or constraints, 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.

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

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  $\text{Neuro}_{\text{SymbolicNeuro}}$ , 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



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

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:

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

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:

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

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.

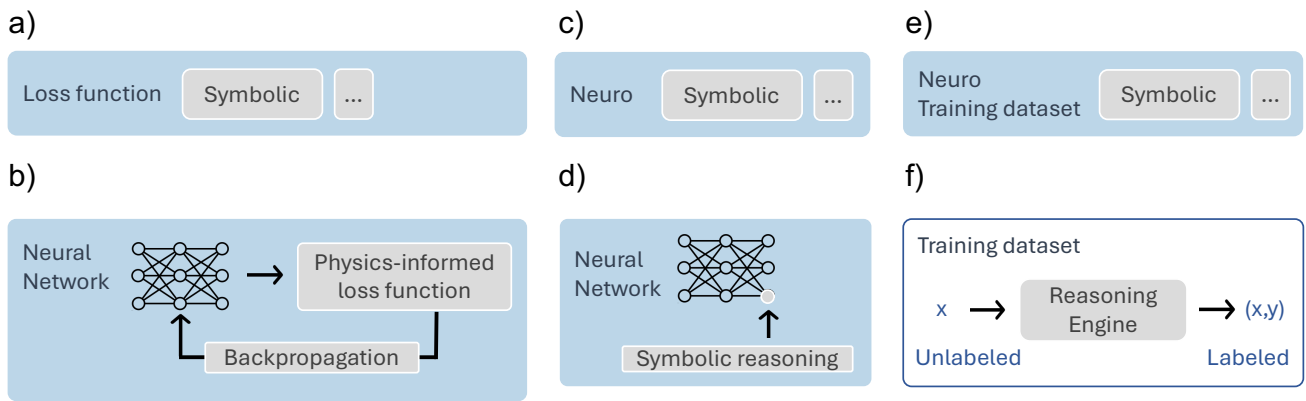
### Fibring: *Neuro* $\rightarrow$ *Symbolic* $\leftarrow$ *Neuro*

Another architecture that we propose to distinguish taxonomically is *Neuro*  $\rightarrow$  *Symbolic*  $\leftarrow$  *Neuro*, in which multiple interconnected NNs are coordinated through a symbolic fibring function that enables them to collaborate and share information while remaining consistent with explicit symbolic constraints (Figure 5a). The symbolic function acts

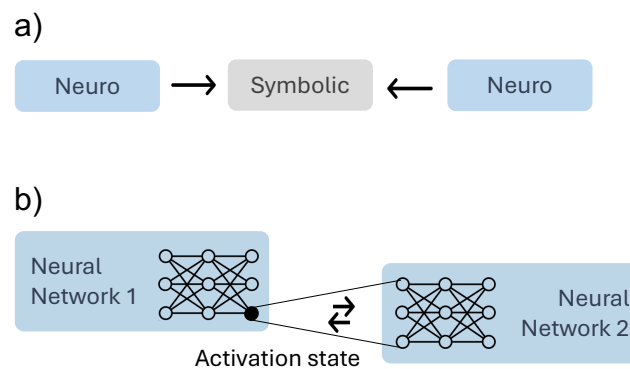
as an intermediary, facilitating communication between the networks by ensuring that their interactions respect predefined symbolic rules or structures. This enables the networks to exchange information in a structured manner, allowing them to jointly solve problems while benefiting from both the statistical learning power of NNs and the logical constraints imposed by the symbolic system [47]. We present fibring here as a proposed extension to Kautz's taxonomy rather than as an established consensus category. Its motivation is to capture an integration principle that is not fully expressed by the existing cooperative or nested paradigms. In cooperative architectures, neural and symbolic components typically interact within a single iterative problem-solving loop. In nested architectures, one component is structurally embedded inside the other and remains subordinate to its host pipeline. By contrast, fibring is intended to describe settings in which multiple specialized neural components must remain mutually consistent through a shared symbolic coordination layer that performs aggregation, conflict resolution, and cross-model constraint enforcement. This distinction becomes especially relevant when no single neural-symbolic pair is sufficient to characterize the system, such as in multi-agent coordination, multi-scale design, or multi-view decision-making, where symbolic reasoning must mediate among several neural subsystems simultaneously. This architecture can formally defined as follows:

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

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



**Figure 5.** Fibring 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.

## 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 examines how selected contemporary AI techniques can be discussed in relation to the NSAI framework. Rather than proposing definitive taxonomic assignments, it provides a structured perspective on where explicit symbolic elements are already present and where connections with NSAI remain indirect, conditional, or emerging.

Accordingly, the discussion distinguishes between (i) approaches that are already instantiated in the literature as genuinely neuro-symbolic systems and (ii) approaches whose relation to NSAI remains partial or prospective. This distinction is especially important for techniques which may support neuro-symbolic integration in some settings while relying on substantially different design principles.

In this way, the section helps clarify how these techniques relate to different forms of neural-symbolic integration and highlights where future neuro-symbolic extensions may plausibly emerge. It also supports a more structured discussion of their roles in addressing challenges such as interpretability, reasoning, and generalization.

## 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.

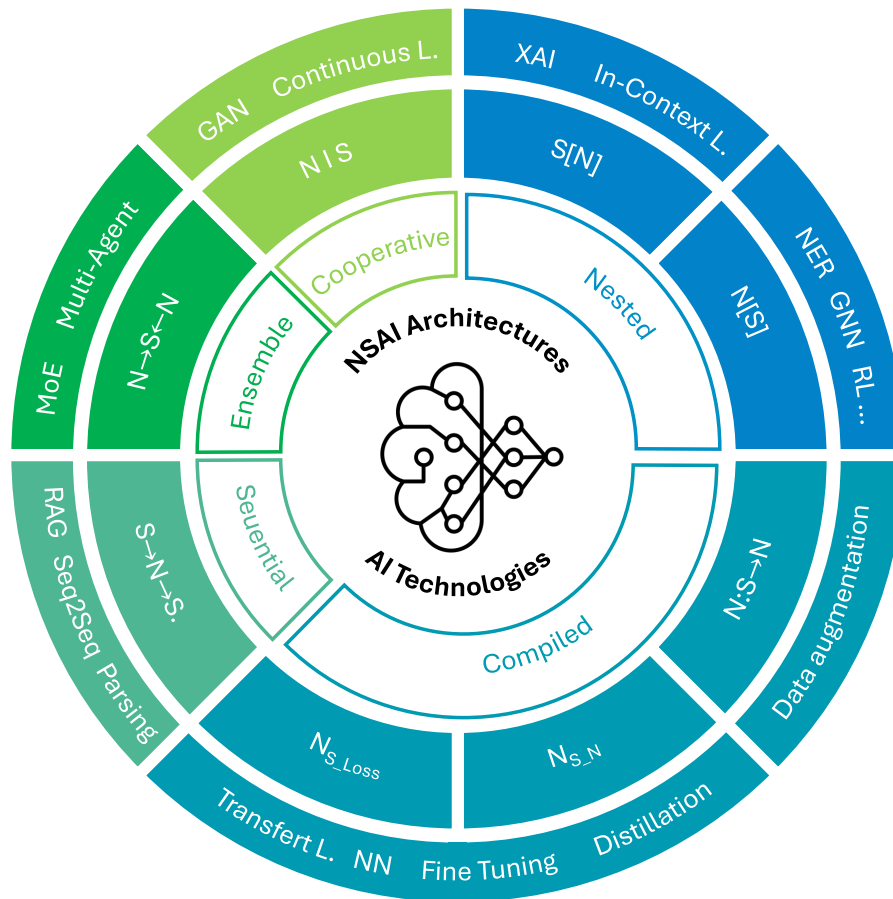
Reinforcement learning and its variant RLHF [68], focus on training agents to make sequential decisions in dynamic environments. RLHF further aligns agent behavior with human preferences, fostering ethical and adaptive AI systems. Finally, continuous learning, or lifelong learning, addresses the challenge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

These techniques represent prominent directions in contemporary generative AI, each contributing to the resolution of complex challenges across diverse applications. The discussion of these methods through an NSAI lens, developed in the following sections, provides a structured perspective on their synergies and practical relevance, without implying that all of them constitute neuro-symbolic systems by default or that their relationship to NSAI is equally established in the literature.

## Classification of AI Technologies within NSAI Architectures

This section discusses how selected generative AI techniques may be interpreted in relation to the eight NSAI architectures, with attention to both their underlying integration principles and their practical use cases. The aim is not to impose definitive taxonomic assignments, but rather to clarify which techniques are already instantiated in the literature as neuro-symbolic systems and which are better understood as partial, approximate, or prospective alignments, as summarized in Figure 6.

*The Sequential Paradigm: From Symbolic to Neural Reasoning* Techniques like RAG, GraphRAG, and Seq2Seq models (including LLMs, e.g., GPT [70]) often align with this paradigm because they encode symbolic inputs (e.g., text or structured information) into neural representations to perform complex transformations, and then produce outputs that are either discrete symbolic sequences (natural language) or, in some pipelines, explicitly structured symbolic forms. 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



**Figure 6.** Classification of AI technologies into NSAI architectures.

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.

*The Nested Paradigm: Embedding Symbolic Logic in Neural Systems* In-context learning mechanisms, such as few-shot learning and CoT reasoning, can instantiate different nested paradigms depending on how symbolic structure and neural generation are combined. In settings where an explicit symbolic framework (e.g., rules, knowledge bases, or logical templates) structures the reasoning task and the neural model operates within this framework, these mechanisms can be viewed through the lens of the *Symbolic[Neuro]* paradigm. XAI techniques can also be associated with this category, since a significant subset of post-hoc methods combine neural feature extraction with symbolic or rule-based structures to generate human-interpretable explanations [72, 73, 74].

Zhang et al. [75] 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 [76] 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. Complementarily, Ma et al.’s Neural Symbolic Reinforcement Learning (NSRL) framework [77] encodes states and policies in first-order logic and uses a differentiable reasoning module to learn symbolic rules, illustrating a nested variant in which a predominantly symbolic RL agent incorporates a neural subcomponent for rule learning and multi-step reasoning.

*The Cooperative Paradigm: Iterative Interaction Between Neural and Symbolic Modules* Rather than generic GANs, cooperative training schemes inspired by the GAN setting can instantiate the cooperative paradigm when at least one of the components is explicitly tied to symbolic rules or logic-based constraints. In such configurations, a neural generator proposes candidate solutions while a symbolic or rule-based critic evaluates them against formal constraints or domain knowledge, providing structured feedback that guides the generator. The resulting iterative feedback loop between neural generation and symbolic evaluation fits the *Neuro | Symbolic* framework, as it combines adaptive learning with constraint-driven refinement.

Continuous learning is inherently facilitated in this cooperative paradigm, 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 [78], 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.

*The Compiled Paradigm: Embedding Symbolic Reasoning Within Neural Computation* Approaches such as model distillation, fine-tuning, pre-training, and transfer learning are not neuro-symbolic methods by themselves. However, they can align with the *NeuroSymbolic* approach when symbolic constraints or objectives (e.g., logical consistency,

relational structures) are explicitly integrated into the neural training process—either through the loss function or via neuron-level mechanisms (e.g., constrained activations). In such settings, the symbolic component is effectively compiled into the neural model during training, enabling the resulting network to adhere to predefined rules and support structured forms of reasoning. Parametric activation functions that implement fuzzy or logic-like operators provide one example of neuron-level compilation: as surveyed by Pusztaházi et al. [79], certain neuro-fuzzy activations can be interpreted as learnable AND/OR operators, embedding symbolic structure directly into the activation behaviour of neurons while remaining differentiable. Thus, these techniques become part of the compiled paradigm only when they explicitly transfer or enforce symbolic knowledge within the neural architecture. Mendez-Lucero and al. [80] exemplify this perspective by embedding logical constraints directly into the loss function. Their distribution-based method 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 such as the Fisher–Rao distance or Kullback–Leibler divergence, guiding the neural network to satisfy symbolic requirements. A related line of work is the Semantic Loss framework of Arrotta et al. [81], which introduces a logic-based loss term for context-aware human activity recognition; the loss penalizes predictions that violate symbolic constraints expressing activity patterns, thereby steering the network towards solutions that are both accurate and consistent with domain knowledge. Such integrations ensure that the model learns from data while simultaneously inheriting explicit logical structure, thereby strengthening the connection between neural learning and symbolic reasoning and making the approach suitable for use within regimes such as knowledge distillation, fine-tuning, pre-training, and transfer learning.

Rather than generic data augmentation, we focus here on rule-constrained synthetic data generation, which leverages the *Neuro:Symbolic*  $\rightarrow$  *Neuro* approach. In this setting, symbolic reasoning is used to generate logically valid synthetic examples that augment the training corpus of neural models. By producing high-quality labeled data through logical inference, symbolic solvers effectively compile structured knowledge into the data distribution, which is then absorbed by the neural network during training. 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. [82] 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 to increase difficulty). Each transformation results in a new

problem–solution instance with its corresponding solution, forming labeled pairs  $(x', y')$  that enrich the training corpus with controlled complexity levels and logically grounded synthetic data. In a complementary vein, Llugiqi et al. [83] use knowledge graphs as a symbolic backbone to perform semantic data augmentation for tabular prediction tasks: symbolic relations in the graph are first exploited to derive enriched semantic descriptors (e.g., graph-based embeddings and distance features), which are then used to train purely neural predictors.

*The Fibring Paradigm: Connecting Neural Models Through Symbolic Constraints* Multi-agent AI 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. [84] explored how the combination of symbolic reasoning and agents can enable the development of advanced systems that are closer to human-like intelligence. They discuss 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.

MoE models are neural architectures in which multiple specialized subnetworks (“experts”) are trained jointly, and a learned gating/router network dynamically selects a small subset of experts for each input, enabling conditional computation and scalable capacity (i.e., large parameter counts with limited active compute). In modern LLMs such as Mixtral  $8 \times 7B$ , each Transformer layer contains several feed-forward experts, and a purely neural router chooses (e.g., top-2) experts per token and combines their outputs, without any explicit symbolic mediator or rule-based coordination layer [85]. DeepSeek-R1 builds on a large MoE base model and improves reasoning via reinforcement learning, but the interaction among experts remains governed by learned neural routing rather than formal symbolic constraints [86]. Therefore, standard MoE systems are not neuro-symbolic by default; they can only be interpreted as instantiations of a fibring-like neuro-symbolic paradigm in variants where expert routing or aggregation is explicitly controlled or constrained by symbolic rules, logical constraints, or knowledge-based controllers [87].

## Evaluation of NSAI Architectures

Ensuring the reliability and practical applicability of NSAI 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.

### 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* [88]: The ability to maintain performance on data that deviate from the training distribution.
- *Contextual flexibility* [89]: The capacity to adapt seamlessly to changes in context or domain with minimal retraining.
- *Relational accuracy* [90]: 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* [91]: The ability to process and derive insights from massive datasets.
- *Hardware efficiency* [92]: Optimal utilization of available computational resources, enabling operation on both low-resource devices and high-performance infrastructures.
- *Complexity management* [93]: The ability to accommodate increased architectural complexity without compromising speed or deployment feasibility.

**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* [94]: Achieving high performance with a reduced amount of training data.
- *Data optimization* [95]: Maximizing the utility of available data (both labeled and unlabeled), potentially through semi-supervised learning techniques.
- *Incremental adaptability* [96]: 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* [97]: The systematic application of explicit rules to derive precise and consistent inferences.

- *Relational understanding* [98]: The comprehension of complex relationships between entities within the data.
- *Cognitive versatility* [99]: 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* [100]: The ability to sustain stable performance despite the presence of noise or adversarial data.
- *Adaptive resilience* [101]: The maintenance of functionality under changing or unpredictable conditions.
- *Bias resilience* [102]: 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* [103]: The capacity to generalize across diverse domains with minimal modifications.
- *Multi-task learning* [104]: The capability to handle multiple tasks simultaneously through shared knowledge representations.
- *Personalization* [105]: 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* [106]: The clarity with which the internal mechanisms and decision processes of the model are revealed.
- *Explanation* [107]: The ability to provide comprehensible justifications for predictions or decisions.
- *Traceability* [108]: The capability to reconstruct the sequence of operations and factors that contributed to a given outcome.

By systematically 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.

## 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 main criterion, architectures are rated on a four-point scale according to how many of its three sub-criteria using binary (*yes/no*) judgements grounded in the above methodology. If all three sub-criteria are met, the rating is *High*, reflecting consistently exceptional performance; if two are met, the rating is *Medium*, indicating generally satisfactory results with some limitations; if only one sub-criterion is met, the rating is *Low–Medium*, denoting limited strengths; and if none are met, the rating is *Low*, signifying significant weaknesses or inconsistent outcomes. In this way, each architecture receives a quantitative score ranging from 0 to 3, ensuring a balanced and evidence-based evaluation.

This evaluation framework should nevertheless be interpreted with caution. The proposed ratings do not constitute standardized benchmark scores obtained under a shared experimental protocol, but rather a qualitative synthesis of published evidence, architectural design principles, and representative empirical findings. As a result, the “yes/no” judgements and the resulting “High/Medium/Low” summaries should be read as structured comparative indicators rather than as definitive rankings. They are intended to make trade-offs explicit and to support discussion across paradigms, especially in a field where architectures are often evaluated on heterogeneous tasks, datasets, and objectives.

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.

## Results and Discussion

Tables 2–3–4 provide the resulting qualitative evaluation of NSAI paradigms. Each table reports, for every main criterion, the satisfaction of its three sub-criteria (*yes/no*) for the corresponding architectures, together with a summary level (*High/Medium/Low*) derived directly from the number of satisfied sub-criteria. This organization allows Sequential, Fibring, Cooperative, Nested, and Compiled paradigms to be compared in a consistent and transparent manner with respect to their strengths and limitations. Taken together, the

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

Architecture	References
<i>Symbolic</i> $\rightarrow$ <i>Neuro</i> $\rightarrow$ <i>Symbolic</i>	[109], [110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129], [130], [131], [132], [133]
<i>Neuro</i> [ <i>Symbolic</i> ]	[43], [44]
<i>Symbolic</i> [ <i>Neuro</i> ]	[41], [134], [135], [136], [137], [138]
<i>Neuro</i>   <i>Symbolic</i>	[45], [139], [140], [141], [142], [143], [144]
<i>Neuro</i> $\rightarrow$ <i>Symbolic</i> $\leftarrow$ <i>Neuro</i>	[145], [47], [84], [86], [85], [23], [24], [25], [26]
<i>Neuro:Symbolic</i> $\rightarrow$ <i>Neuro</i>	[37], [146], [147], [148], [149], [150], [151], [152], [153], [154], [155], [156], [157], [158], [159], [160]
<i>Neuro</i> <sub><i>Symbolic</i></sub> <sub><i>Loss</i></sub>	[161], [162], [163], [164], [165], [166]
<i>Neuro</i> <sub><i>Symbolic</i></sub> <sub><i>Neuro</i></sub>	[167], [168]

results reveal complementary profiles rather than a single universally optimal design. Sequential architectures perform strongly on reasoning, interpretability, and transferability, but typically remain more limited in scalability and robustness. Cooperative architectures provide rich iterative reasoning and good interpretability, yet their scalability and robustness profiles remain more constrained in large-scale settings. Within the Nested family, *Symbolic*[*Neuro*] models excel in data efficiency, reasoning, and interpretability but struggle with large-scale adaptation, whereas *Neuro*[*Symbolic*] variants often improve generalization and robustness by embedding symbolic modules into neural pipelines, while sacrificing some versatility and transferability. Compiled architectures show that embedding symbolic knowledge directly into losses or neuron-level mechanisms can strengthen reasoning and interpretability, although their generalization and multi-domain transfer capabilities remain more task-dependent.

Within this qualitative and literature-grounded comparison, the *Neuro*  $\rightarrow$  *Symbolic*  $\leftarrow$  *Neuro* architecture receives the broadest support across the selected criteria and appears particularly well balanced for scenarios requiring coordination, explicit constraints, robustness, and interpretability. However, this observation should not be interpreted as a definitive empirical ranking. Rather, it indicates that this architecture offers a comparatively strong profile under the present evaluative lens, while alternative paradigms may remain preferable depending on the task, the integration mechanism, the data regime, and the deployment constraints. This interpretation is consistent with recent work showing that neuro-symbolic multi-agent settings can improve interpretability and probabilistic decision-making under uncertainty [169, 170]. Accordingly, the main contribution of these tables is not to identify a universally superior architecture, but to make explicit the trade-offs that characterize different NSAI integration strategies.

## Neuro-Symbolic AI Architectures for 4D Printing: Proposed Applications

4D printing is an advanced manufacturing paradigm that integrates additive manufacturing (AM) with smart materials (SMs) capable of changing their shape, properties, or functionality over time when exposed to external stimuli such as heat, light, pH, magnetic or electric fields, or humidity [175, 176]. At the core of this technology lie stimuli-responsive materials such as shape memory

polymers (SMPs), hydrogels, liquid crystal elastomers (LCEs), electroactive polymers (EAPs), and shape memory alloys (SMAs), which enable the autonomous transformation of structures without mechanical intervention [177, 178, 179, 180]. These materials can be processed using AM techniques such as fused filament fabrication (FFF), direct ink writing (DIW), digital light processing (DLP), and powder bed fusion (PBF), enabling the fabrication of functional and adaptive geometries for applications in biomedicine, soft robotics, aerospace, textiles, and structural monitoring [181, 182, 183, 184, 185].

In recent years, artificial NNs have been explored in the 4D printing domain to tackle challenges such as predicting material response under stimuli, learning complex shape evolution dynamics, and supporting the inverse design of shape-changing structures [186, 187, 188, 189, 162, 190, 191, 192]. However, despite their effectiveness in pattern recognition and data-driven modeling, NN approaches face key limitations when applied to 4D printing. First, they often require large, high-quality datasets to achieve meaningful generalization, which is problematic in emerging fields where experimental data are sparse or costly to obtain [6]. Second, the black-box nature of NNs hinders the interpretability of results, limiting their adoption in safety-critical design contexts such as biomedical or aerospace applications. Third, NNs typically struggle with tasks that demand reasoning over structured knowledge, including the symbolic relationships between materials, geometries, and functional outcomes. These shortcomings impede the development of explainable and trustworthy design workflows, especially when targeting multifunctional or multi-material 4D-printed systems.

The integration of NSAI with 4D printing technology presents unprecedented opportunities to address the complex challenges inherent in designing, optimizing, and manufacturing stimuli-responsive structures. By combining the learning capabilities of NNs with the logical reasoning and interpretability of symbolic AI, researchers can develop robust frameworks that address the multidisciplinary demands of 4D printing [36]. This section proposes specific applications of various NSAI architectures throughout the 4D printing workflow, from material discovery to transformation control and structural optimization illustrated in Figure 7.

**Table 2.** Comparison of NSAI architectures (Part 1): Sequential, Fibring, and Cooperative

Main Criterion	Sub-Criterion	Symbolic Neuro	Symbolic ← Neuro	Neuro → Symbolic	Neuro Symbolic
<b>Generalization</b>	Out-of-dist.	yes [131]	yes [84]	yes [84]	yes [139]
	Continuous flex.	yes [109]	yes [86]	yes [86]	yes [143]
	Relative prec.	yes [125]	yes [47]	yes [47]	no
	Summary	High	High	High	Medium
<b>Scalability</b>	Large-scale adapt.	yes [114]	yes [26]	yes [26]	yes [141]
	Hardware efficiency	no	yes [85]	yes [85]	no
	Complexity	yes [122]	yes [24]	yes [24]	no
	Summary	Medium	High	High	Low–Medium
<b>Data Efficiency</b>	Reduction	no	yes [25]	yes [25]	yes [45]
	Optimization	yes [130]	yes [145]	yes [145]	yes [141]
	Incremental adapt.	yes [123]	yes [23]	yes [23]	no
	Summary	Medium	High	High	Medium
<b>Reasoning</b>	Logical reason.	yes [127]	yes [86]	yes [86]	yes [140]
	Comprehension	yes [128]	yes [24]	yes [24]	yes [144]
	Versatility	yes [118]	yes [47]	yes [47]	yes [142]
	Summary	High	High	High	High
<b>Robustness</b>	Perturbations	no	yes [86]	yes [86]	yes [141]
	Adaptability	yes [123]	yes [25]	yes [25]	yes [45]
	Bias handling	no	yes [23]	yes [23]	no
	Summary	Low–Medium	High	High	Medium
<b>Transferability</b>	Multi-domain	yes [110]	yes [24]	yes [24]	yes [143]
	Multi-task	yes [124]	yes [85]	yes [85]	no
	Personalization	yes [116]	yes [23]	yes [23]	no
	Summary	High	High	High	Low–Medium
<b>Interpretability</b>	Transparency	yes [117]	yes [84]	yes [84]	yes [143]
	Explanation	yes [121]	yes [47]	yes [47]	yes [142]
	Traceability	yes [133]	yes [145]	yes [145]	yes [140]
	Summary	High	High	High	High

### Sequential Architecture Applications

Sequential architecture (Symbolic → Neuro → Symbolic) offer promising applications in material property mapping for 4D printing. By first encoding material compositions and structural parameters into symbolic representations, then processing these through NNs, and finally decoding outputs back into interpretable symbolic form, researchers can establish explicit relationships between formulations and transformation behaviors. A sequential NSAI system

for the design of shape memory polymers (SMPs), liquid crystal elastomers (LCEs), and hydrogels would begin with symbolic representations of polymer composition, crosslinking density, and thermal properties, encode these into continuous vectors for neural processing, utilize neural networks to predict complex transformation patterns, and conclude by decoding predictions into symbolic rules describing temperature-dependent folding angles and recovery rates. This approach maintains scientific traceability while leveraging NNs' capacity to model

**Table 3.** Comparison of NSAI architectures (Part 2): Nested

Main Criterion	Sub-Criterion	Symbolic[Neuro]	Neuro[Symbolic]
<b>Generalization</b>	Out-of-dist.	no	yes [44]
	Continuous flex.	yes [134]	no
	Relative prec.	yes [138]	no
	Summary	Medium	Low–Medium
<b>Scalability</b>	Large-scale adapt.	no	yes [43]
	Hardware efficiency	no	no
	Complexity	no	no
	Summary	Low	Low–Medium
<b>Data Efficiency</b>	Reduction	yes [135]	yes [44]
	Optimization	yes [136]	yes [44]
	Incremental adapt.	yes [41]	no
	Summary	High	Medium
<b>Reasoning</b>	Logical reason.	yes [134]	yes [43]
	Comprehension	yes [137]	no
	Versatility	yes [138]	no
	Summary	High	Low–Medium
<b>Robustness</b>	Perturbations	yes [41]	yes [44]
	Adaptability	yes [137]	no
	Bias handling	no	no
	Summary	Medium	Low–Medium
<b>Transferability</b>	Multi-domain	yes [137]	no
	Multi-task	no	no
	Personalization	no	no
	Summary	Low–Medium	Low
<b>Interpretability</b>	Transparency	yes [134]	yes [44]
	Explanation	yes [135]	yes [43]
	Traceability	yes [138]	yes [44]
	Summary	High	High

complex structure-property relationships that are difficult to formulate using purely symbolic methods. The resulting framework would facilitate rapid design iteration with

interpretable outputs that can guide experimental validation and material refinement.

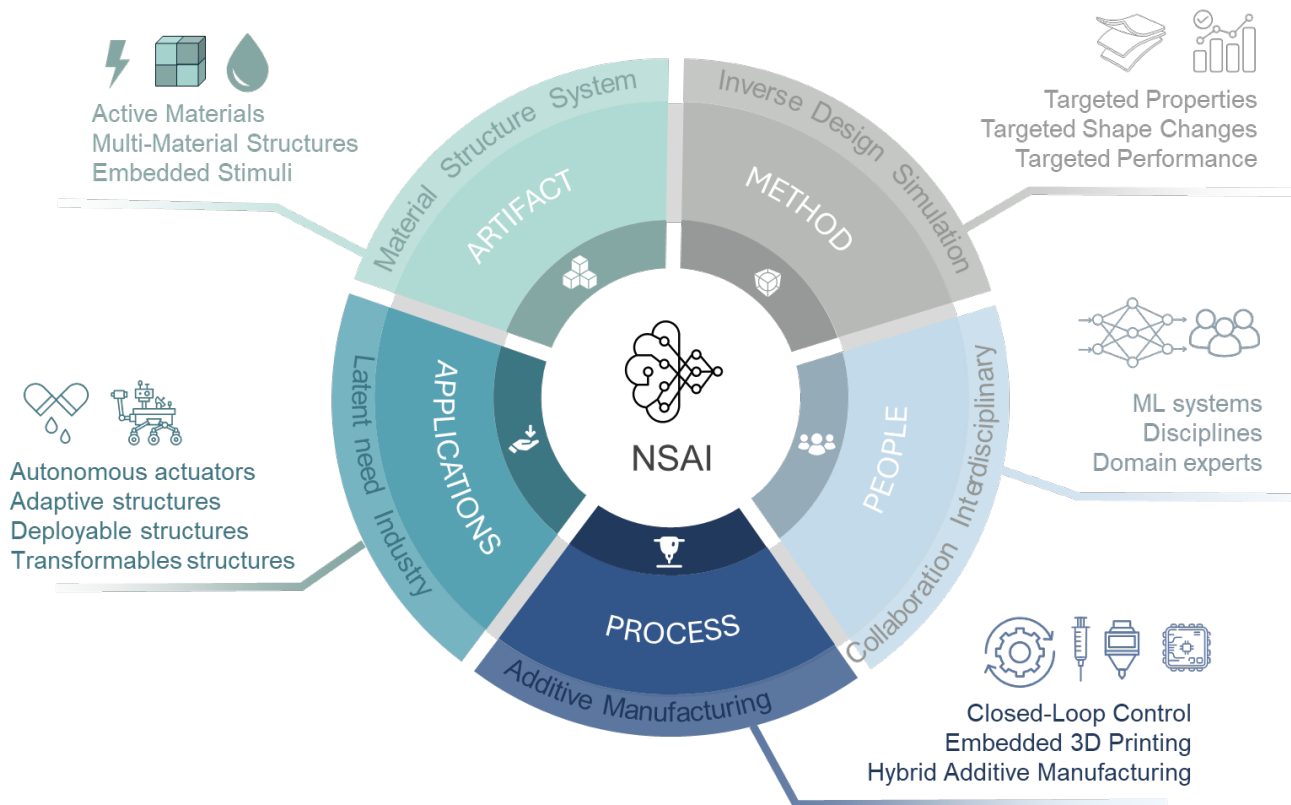
**Table 4.** Comparison of NSAI architectures (Part 3): Compiled

Main Criterion	Sub-Criterion	Neuro:Symbolic $\rightarrow$ Neuro	Neuro <sub>SymbolicLoss</sub>	Neuro <sub>SymbolicNeuro</sub>
<b>Generalization</b>	Out-of-dist.	yes [160]	no	no
	Continuous flex.	no	no	no
	Relative prec.	no	no	no
	Summary	Low–Medium	Low	Low
<b>Scalability</b>	Large-scale adapt.	yes [155]	yes [166]	yes [168]
	Hardware efficiency	yes [37]	yes [161]	yes [171]
	Complexity	no	no	no
	Summary	Medium	Medium	Medium
<b>Data Efficiency</b>	Reduction	no	yes [162]	yes [167]
	Optimization	yes [149]	yes [164]	yes [172]
	Incremental adapt.	no	no	no
	Summary	Low–Medium	Medium	Medium
<b>Reasoning</b>	Logical reason.	yes [152]	yes [161]	yes [167]
	Comprehension	yes [157]	yes [163]	yes [168]
	Versatility	yes [151]	no	no
	Summary	High	Medium	Medium
<b>Robustness</b>	Perturbations	yes [147]	no	yes [173]
	Adaptability	yes [154]	no	no
	Bias handling	yes [159]	no	yes [174]
	Summary	High	Low	Medium
<b>Transferability</b>	Multi-domain	no	no	no
	Multi-task	no	yes [166]	yes [167]
	Personalization	no	no	no
	Summary	Low	Low–Medium	Low–Medium
<b>Interpretability</b>	Transparency	yes [148]	yes [165]	yes [168]
	Explanation	yes [158]	yes [166]	yes [167]
	Traceability	yes [146]	yes [163]	yes [168]
	Summary	High	High	High

### *Nested Architectures for Multi-scale Optimization*

The nested Symbolic[Neuro] architecture presents significant advantages for rule-based design systems that incorporate neural material prediction. In this configuration, a

symbolic reasoning framework establishes design constraints based on application requirements, while embedded NNs predict material distributions that achieve desired transformation properties. For instance, an expert system for



**Figure 7.** Overview of NSAI in accelerated design for 4D printing with regard to AM, people involved from various fields, artifacts, and applications [36] (Adapted with permission from Elsevier).

designing 4D-printed medical stents could utilize symbolic reasoning to establish design constraints based on clinical requirements, employ NN subcomponents to predict material distribution patterns that achieve desired transformation properties, and evaluate neural predictions against clinical constraints to iteratively refine the design. This architecture maintains interpretable decision logic while leveraging NNs' pattern recognition capabilities for complex material behavior prediction.

Conversely, the Neuro[Symbolic] architecture enables generative design capabilities by embedding physical constraint solvers within neural frameworks. A design system for 4D-printed adaptive structures could leverage a generative NN to propose innovative material distributions, activate embedded symbolic solvers to verify physical feasibility when faced with complex transformation sequences, and integrate results back into the neural framework to guide subsequent design iterations. This approach would enable the exploration of novel design spaces while ensuring physical feasibility, effectively reducing the gap between computational design and manufacturability that currently hampers innovation in 4D printing applications.

### *Cooperative Architecture for Transformation Control*

Cooperative NSAI architectures (Neuro | Symbolic) present compelling solutions for interactive transformation planning systems in 4D printing. Through iterative refinement of predictions based on physical models, these systems enable

precise control of complex transformation sequences that are difficult to model using either purely data-driven or analytical approaches. A cooperative system for designing multi-stage transformation sequences would utilize NNs to process visual information of printed structures during initial transformation, employ symbolic reasoning to evaluate transformation against desired geometric goals and physical constraints, identify discrepancies between predicted and observed behavior, and continue this feedback loop until achieving the desired transformation sequence. This architecture is particularly suitable for applications requiring precise spatiotemporal control of transformation, such as soft robotics and adaptive medical devices, where transformation pathways are as critical as final configurations.

### *Compiled Architectures for Physics-Informed Design*

The incorporation of physical laws directly into neural network architectures through compiled approaches offers significant advantages for 4D printing applications. A physics-informed material design system utilizing  $\text{Neuro}_{\text{Symbolic, Loss}}$  functions would enable more physically accurate prediction of material behavior by penalizing predictions that violate fundamental principles. For instance, a design system for 4D-printed hydrogel structures could employ NNs to predict swelling behavior based on material composition while incorporating physics-based loss terms that enforce conservation of mass, diffusion dynamics, and mechanical equilibrium constraints. This approach ensures that model

predictions remain physically plausible even for novel material combinations or environmental conditions, substantially reducing the need for extensive experimental validation that currently limits exploration of new material formulations.

Similarly, constraint-aware material distribution optimization can be achieved through  $\text{Neuro}_{\text{Symbolic}}^{\text{Neuro}}$  architectures, where traditional activation functions are replaced with logic-based elements that enforce manufacturing constraints. A voxel-based material distribution optimizer with symbolic neurons would enforce printability rules like minimum feature size, support requirements, and material interface limitations directly within the optimization process, eliminating post-processing steps and streamlining the transition from computational design to fabrication. This integration of manufacturing constraints at the neuron level represents a significant advancement over current approaches that typically separate design optimization from manufacturability verification.

The  $\text{Neuro:Symbolic} \rightarrow \text{Neuro}$  approach addresses the critical challenge of limited experimental data in 4D printing through supervised training with synthetic data generation. A predictive model for LCEs behavior could utilize symbolic reasoning based on physical models to generate a comprehensive dataset of material compositions and corresponding transformation behaviors. The resulting neural model would combine the physical accuracy of analytical models with the computational efficiency of neural inference, enabling rapid exploration of material design spaces that would be prohibitively expensive to characterize experimentally.

### Fibering Architecture for Multi-domain Optimization

The fibering architecture offers powerful capabilities for multi-scale material and structure optimization in 4D printing by connecting specialized neural networks through symbolic relationships. A comprehensive design system utilizing this approach would deploy specialized NNs for different aspects of the problem (molecular-scale material properties, meso-scale structural behavior, and macro-scale system performance), connect these networks through symbolic fibering functions that ensure consistency across scales, and resolve contradictions between predictions at different scales through the symbolic component. This architecture addresses the fundamental multi-scale challenge in 4D printing, where molecular-level interactions ultimately determine system-level performance, but direct modeling across all scales remains computationally intractable. By decomposing the problem into specialized networks while maintaining consistency through symbolic relationships, researchers can achieve simultaneous optimization of material composition and structural configuration.

The proposed applications of NSAI architectures in 4D printing demonstrate the potential for significant advancements throughout the material-design-manufacturing workflow. Each architecture addresses specific challenges inherent in 4D printing: sequential architectures provide interpretable material-property relationships; nested architectures balance creativity with physical constraints; cooperative

architectures enable precise transformation control; compiled architectures integrate physical laws and manufacturing constraints; and fibering architectures connect predictions across multiple scales and domains. By systematically applying these approaches, researchers can accelerate innovation while maintaining scientific interpretability, potentially reducing the current gap between fundamental research and practical applications in 4D printing technology.

## Conclusion

This study examined several NSAI architectures through a common conceptual framework and compared them across criteria including generalization, scalability, data efficiency, reasoning, robustness, transferability, and interpretability. The resulting analysis highlights the distinct strengths and limitations of each architecture and supports a more structured discussion of their potential roles in real-world applications. Among the architectures considered,  $\text{Neuro} \rightarrow \text{Symbolic} \leftarrow \text{Neuro}$  appears especially promising within the scope of our qualitative comparison, particularly for scenarios that require coordination across multiple components, explicit reasoning under constraints, and robust decision-making. However, this observation remains interpretive and literature-grounded rather than empirically definitive, and it should not be read as establishing universal superiority over other NSAI paradigms.

More broadly, this work argues that NSAI architectures are best understood as complementary integration strategies rather than as competitors in a single ranking. Their relevance depends on the form of neural-symbolic interaction required by the task, the level of interpretability sought, and the operational constraints of the target domain. Future work should therefore focus on comparative benchmarks under shared protocols, clearer reporting standards for architectural claims, and more explicit distinctions between established neuro-symbolic instantiations and prospective integrations, particularly in the rapidly evolving context of generative 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] Garcez Ad and Lamb LC. Neurosymbolic ai: The 3rd wave. *Artificial Intelligence Review* 2023; 56(11): 12387–12406.
- [2] Valiant LG. Three problems in computer science. *Journal of the ACM (JACM)* 2003; 50(1): 96–99.
- [3] Bengio Y, Marcus G and Boucher V. AI DEBATE! Yoshua Bengio vs Gary Marcus, 2019. URL <https://montrealartificialintelligence.com/aidebate/>.
- [4] Bengio Y. 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>.
- [5] Marcus G and Davis E. *Rebooting AI: Building artificial intelligence we can trust*. Vintage, 2019.
- [6] Marcus G. Deep learning: A critical appraisal. *arXiv preprint arXiv:180100631* 2018; .
- [7] Liu Z, Wang Z, Lin Y et al. A neural-symbolic approach to natural language understanding. *arXiv preprint arXiv:220310557* 2022; .
- [8] Zhang J, Chen B, Zhang L et al. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open* 2021; 2: 14–35.
- [9] Lamb LC, d’Avila Garcez A, Gori M et al. Graph neural networks meet neural-symbolic computing: a survey and perspective. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. IJCAI’20. ISBN 9780999241165.
- [10] Von Rueden L, Mayer S, Beckh K et al. Informed machine learning—a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering* 2021; 35(1): 614–633.
- [11] Belle V. Symbolic logic meets machine learning: A brief survey in infinite domains. In *International conference on scalable uncertainty management*. Springer, pp. 3–16.
- [12] Hamilton K, Nayak A, Božić B et al. Is neuro-symbolic ai meeting its promises in natural language processing? a structured review. *Semantic Web* 2024; 15(4): 1265–1306.
- [13] Kautz H. The third ai summer: Aaai robert s. engelmore memorial lecture. *Ai magazine* 2022; 43(1): 105–125.
- [14] Vaswani A. Attention is all you need. *Advances in Neural Information Processing Systems* 2017; .
- [15] Lewis P, Perez E, Piktus A et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 2020; 33: 9459–9474.
- [16] Sutskever I, Vinyals O and Le QV. Sequence to sequence learning with neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. NIPS’14, MIT Press, p. 3104–3112.
- [17] Jiang P and Cai X. A survey of semantic parsing techniques. *Symmetry* 2024; 16(9): 1201.
- [18] Marrero M, Urbano J, Sánchez-Cuadrado S et al. Named entity recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces* 2013; 35(5): 482–489.
- [19] Zhao X, Deng Y, Yang M et al. A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys* 2024; 56(11): 1–39.
- [20] Christiano PF, Leike J, Brown T et al. Deep reinforcement learning from human preferences. *Advances in neural information processing systems* 2017; 30.
- [21] Zhou J, Cui G, Hu S et al. Graph neural networks: A review of methods and applications. *AI open* 2020; 1: 57–81.
- [22] Goodfellow I, Pouget-Abadie J, Mirza M et al. Generative adversarial nets. *Advances in neural information processing systems* 2014; 27.
- [23] Guo T, Chen X, Wang Y et al. Large language model based multi-agents: A survey of progress and challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI 2024)*. pp. 8048–8057. DOI:10.24963/ijcai.2024/890. URL <https://doi.org/10.24963/ijcai.2024/890>.
- [24] Maldonado D, Cruz E, Torres JA et al. Multi-agent systems: A survey about its components, framework and workflow. *IEEE Access* 2024; .
- [25] He XO. Mixture of a million experts, 2024. URL <https://arxiv.org/abs/2407.04153>. 2407.04153.
- [26] Lo KM, Huang Z, Qiu Z et al. A closer look into mixture-of-experts in large language models, 2024. URL <https://arxiv.org/abs/2406.18219>. 2406.18219.
- [27] Alyafeai Z, AlShaibani MS and Ahmad I. A survey on transfer learning in natural language processing. *arXiv preprint arXiv:200704239* 2020; .
- [28] Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:181004805* 2018; .
- [29] Howard J and Ruder S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:180106146* 2018; .

- [30] Parnami A and Lee M. Learning from few examples: A summary of approaches to few-shot learning. *arXiv preprint arXiv:220304291* 2022; .
- [31] Arrieta AB, Díaz-Rodríguez N, Del Ser J et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion* 2020; 58: 82–115.
- [32] Hinton G. Distilling the knowledge in a neural network. *arXiv preprint arXiv:150302531* 2015; .
- [33] Wei J, Wang X, Schuurmans D et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 2022; 35: 24824–24837.
- [34] Chen Z and Liu B. *Lifelong machine learning*. Morgan & Claypool Publishers, 2018.
- [35] Kenton JMW and Toutanova LK. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1. Minneapolis, Minnesota, p. 2.
- [36] Bougzime O, Cruz C, André JC et al. Neuro-symbolic artificial intelligence in accelerated design for 4d printing: Status, challenges, and perspectives. *Materials & Design* 2025; Under review.
- [37] Lample G and Charton F. Deep learning for symbolic mathematics ; .
- [38] Dimassi S, Demoly F, Cruz C et al. An ontology-based framework to formalize and represent 4d printing knowledge in design. *Computers in Industry* 2021; 126: 103374.
- [39] Mikolov T. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:13013781* 2013; 3781.
- [40] Pennington J, Socher R and Manning CD. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543.
- [41] Silver D, Huang A, Maddison CJ et al. Mastering the game of go with deep neural networks and tree search. *nature* 2016; 529(7587): 484–489.
- [42] Coulom R. Efficient selectivity and backup operators in monte-carlo tree search. In *International conference on computers and games*. Springer, pp. 72–83.
- [43] Heule MJ, Kullmann O and Marek VW. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *International Conference on Theory and Applications of Satisfiability Testing*. Springer, pp. 228–245.
- [44] Madan K, Ke NR, Goyal A et al. Fast and slow learning of recurrent independent mechanisms. *arXiv preprint arXiv:210508710* 2021; .
- [45] Mao J, Gan C, Kohli P et al. The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. In *International Conference on Learning Representations*. URL <https://openreview.net/forum?id=rJgMlhRctm>.
- [46] Riegel R, Gray A, Luus F et al. Logical neural networks. *arXiv preprint arXiv:200613155* 2020; .
- [47] Garcez ASd and Gabbay DM. Fibring neural networks. In *AAAI*. pp. 342–347.
- [48] Sheth A, Pallagani V and Roy K. Neurosymbolic ai for enhancing instructability in generative ai. *IEEE Intelligent Systems* 2024; 39(5): 5–11.
- [49] Jacobson MJ and Xue Y. Integrating symbolic reasoning into neural generative models for design generation. *Artificial Intelligence* 2025; 339: 104257.
- [50] Yin X and Wan X. 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)*. pp. 7701–7710.
- [51] Yang D, Rao J, Chen K et al. 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*. pp. 730–740.
- [52] Edge D, Trinh H, Cheng N et al. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:240416130* 2024; .
- [53] Chen X, Jia S and Xiang Y. A review: Knowledge reasoning over knowledge graph. *Expert systems with applications* 2020; 141: 112948.
- [54] Antoniou G and Harmelen Fv. Web ontology language: Owl. *Handbook on ontologies* 2009; : 91–110.
- [55] Mavromatis C and Karypis G. GNN-RAG: Graph neural retrieval for large language model reasoning, 2025. URL <https://openreview.net/forum?id=EVuANndP1X>.
- [56] Roy A. Recent trends in named entity recognition (ner). *arXiv preprint arXiv:210111420* 2021; .
- [57] Wu T, You X, Xian X et al. Towards deep understanding of graph convolutional networks for relation extraction. *Data & Knowledge Engineering* 2024; 149: 102265.
- [58] Wu Q, Bansal G, Zhang J et al. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*. URL <https://openreview.net/forum?id=BAakY1hNKS>.

- [59] Shazeer N, Mirhoseini A, Maziarz K et al. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*. URL <https://openreview.net/forum?id=BlckMDq1lg>.
- [60] Lepikhin D, Lee H, Xu Y et al. {GS}hard: Scaling giant models with conditional computation and automatic sharding. In *International Conference on Learning Representations*. URL <https://openreview.net/forum?id=qrwe7XHTmYb>.
- [61] Xie J, Chen Z, Zhang R et al. Large multimodal agents: A survey. *arXiv preprint arXiv:240215116* 2024; .
- [62] Ding W, Abdel-Basset M, Hawash H et al. Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. *Information Sciences* 2022; 615: 238–292.
- [63] Huang Y, Chen Y, Yu Z et al. In-context learning distillation: Transferring few-shot learning ability of pre-trained language models. *arXiv preprint arXiv:221210670* 2022; .
- [64] Iman M, Arabnia HR and Rasheed K. A review of deep transfer learning and recent advancements. *Technologies* 2023; 11(2): 40.
- [65] Pourpanah F, Abdar M, Luo Y et al. A review of generalized zero-shot learning methods. *IEEE transactions on pattern analysis and machine intelligence* 2022; 45(4): 4051–4070.
- [66] Reynolds L and McDonell K. 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*. pp. 1–7.
- [67] Gou J, Yu B, Maybank SJ et al. Knowledge distillation: A survey. *International Journal of Computer Vision* 2021; 129(6): 1789–1819.
- [68] Dai J, Pan X, Sun R et al. Safe RLHF: Safe reinforcement learning from human feedback. In *The Twelfth International Conference on Learning Representations*. URL <https://openreview.net/forum?id=TyFrPOKYXw>.
- [69] Chinimilli VRP and Sadasivuni L. The rise of ai: a comprehensive research review. *IAES International Journal of Artificial Intelligence (IJ-AI)* 2024; 13: 2226. DOI:10.11591/ijai.v13.i2.pp2226-2235.
- [70] OpenAI, Achiam J, Adler S et al. Gpt-4 technical report, 2024. URL <https://arxiv.org/abs/2303.08774>. 2303.08774.
- [71] Anonymous. RAG-logic: Enhance neuro-symbolic approaches for logical reasoning with retrieval-augmented generation. In *Submitted to ACL Rolling Review - June 2024*. URL <https://openreview.net/forum?id=FAUDgLi6RO>. Under review.
- [72] Guidotti R et al. A survey of methods for explaining black box models. *ACM Computing Surveys* 2018; .
- [73] Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 2019; .
- [74] Samek W et al. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, 2021.
- [75] Zhang H, Zhang Y, Li LE et al. The impact of symbolic representations on in-context learning for few-shot reasoning. In *NeurIPS 2022 Workshop on Neuro Causal and Symbolic AI (nCSI)*.
- [76] Singh I, Kaur N, Gaur G et al. NeuSTIP: A neuro-symbolic model for link and time prediction in temporal knowledge graphs. In *The 2023 Conference on Empirical Methods in Natural Language Processing*. URL <https://openreview.net/forum?id=j61Sx05QRj>.
- [77] Ma Z, Zhuang Y, Weng P et al. Learning symbolic rules for interpretable deep reinforcement learning. *CoRR* 2021; abs/2103.08228. URL <https://arxiv.org/abs/2103.08228>. 2103.08228.
- [78] Mitchener L, Tuckey D, Crosby M et al. Detect, understand, act: A neuro-symbolic hierarchical reinforcement learning framework. *Machine Learning* 2022; 111(4): 1523–1549.
- [79] Pusztaházi LS, Eigner G and Csiszár O. Parametric activation functions for neural networks: A tutorial survey. *IEEE Access* 2024; 12: 168626–168644. DOI: 10.1109/ACCESS.2024.3474574.
- [80] Mendez-Lucero MA, Gallardo EB and Belle V. Semantic objective functions: A distribution-aware method for adding logical constraints in deep learning. *arXiv preprint arXiv:240515789* 2024; .
- [81] Arrotta L, Civitarese G and Bettini C. Semantic loss: A new neuro-symbolic approach for context-aware human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2023; 7(4): 147:1–147:29. DOI:10.1145/3631407.
- [82] Li Z, Zhou Z, Yao Y et al. Neuro-symbolic data generation for math reasoning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. URL <https://openreview.net/forum?id=C1cMzGLyZW>.
- [83] Llugiqi M, Ekaputra FJ and Sabou M. Semantic-based data augmentation for machine learning prediction enhancement. *Neurosymbolic Artificial Intelligence* 2025; 1(6). DOI:10.1177/29498732251340160.
- [84] Belle V, Fisher M, Russo A et al. Neuro-symbolic ai+ agent systems: A first reflection on trends, opportunities and challenges. In *International Conference on Autonomous Agents and Multiagent Systems*. Springer, pp. 180–200.

- [85] Jiang AQ, Sablayrolles A, Roux A et al. Mixtral of experts. *arXiv preprint arXiv:240104088* 2024; .
- [86] Guo D, Yang D, Zhang H et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:250112948* 2025; .
- [87] Anonymous. Symbolic mixture-of-experts: Adaptive skill-based routing for heterogeneous reasoning. In *Submitted to The Fourteenth International Conference on Learning Representations*. URL <https://openreview.net/forum?id=RYrFUKraWM>. Under review.
- [88] Yang J, Zhou K, Li Y et al. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision* 2024; 132(12): 5635–5662.
- [89] Patel VM, Gopalan R, Li R et al. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine* 2015; 32(3): 53–69.
- [90] Ye W, Zheng G, Cao X et al. Spurious correlations in machine learning: A survey. *arXiv preprint arXiv:240212715* 2024; .
- [91] Dean J, Corrado G, Monga R et al. Large scale distributed deep networks. *Advances in neural information processing systems* 2012; 25.
- [92] Silvano C, Ielmini D, Ferrandi F et al. A survey on deep learning hardware accelerators for heterogeneous hpc platforms. *ACM Computing Surveys* 2023; .
- [93] Tan M and Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, pp. 6105–6114.
- [94] Song Y, Wang T, Cai P et al. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys* 2023; 55(13s): 1–40.
- [95] Zhu XJ. Semi-supervised learning literature survey 2005; .
- [96] Gunasekara N, Pfahringer B, Gomes HM et al. Survey on online streaming continual learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. pp. 6628–6637.
- [97] Hou Z. Neural-symbolic reasoning: Towards the integration of logical reasoning with large language models. *Authorea Preprints* 2025; .
- [98] Li Z, Li Y, Luo Y et al. Graph neural networks for databases: a survey. *IJCAI '25*. ISBN 978-1-956792-06-5. DOI:10.24963/ijcai.2025/1172. URL <https://doi.org/10.24963/ijcai.2025/1172>.
- [99] Ke Z, Jiao F, Ming Y et al. A survey of frontiers in llm reasoning: Inference scaling, learning to reason, and agentic systems. *arXiv preprint arXiv:250409037* 2025; .
- [100] Meng MH, Bai G, Teo SG et al. Adversarial robustness of deep neural networks: A survey from a formal verification perspective. *IEEE Transactions on Dependable and Secure Computing* 2022; .
- [101] Liu C, Xu F, Gao C et al. Deep learning resilience inference for complex networked systems. *Nature Communications* 2024; 15(1): 9203.
- [102] Hort M, Chen Z, Zhang JM et al. Bias mitigation for machine learning classifiers: A comprehensive survey. *ACM Journal on Responsible Computing* 2024; 1(2): 1–52.
- [103] Zhou K, Liu Z, Qiao Y et al. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence* 2022; 45(4): 4396–4415.
- [104] Zhang Y and Yang Q. A survey on multi-task learning. *IEEE transactions on knowledge and data engineering* 2021; 34(12): 5586–5609.
- [105] Zhang Z, Rossi RA, Kveton B et al. Personalization of large language models: A survey. *Transactions on Machine Learning Research* 2025; URL <https://openreview.net/forum?id=tf6A9EYMo6>. Survey Certification.
- [106] Lipton ZC. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 2018; 16(3): 31–57.
- [107] Ribeiro MT, Singh S and Guestrin C. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:160605386* 2016; .
- [108] Spoczynski M, Melara MS and Szyller S. Atlas: A Framework for ML Lifecycle Provenance Transparency . In *2025 IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*. Los Alamitos, CA, USA: IEEE Computer Society, pp. 448–461. DOI:10.1109/EuroSPW67616.2025.00058. URL <https://doi.ieeecomputersociety.org/10.1109/EuroSPW67616.2025.00058>.
- [109] Kouris P, Alexandridis G and Stafylopatis A. Abstractive text summarization: Enhancing sequence-to-sequence models using word sense disambiguation and semantic content generalization. *Computational Linguistics* 2021; 47(4): 813–859.
- [110] Sutherland A, Magg S and Wermtner S. Leveraging recursive processing for neural-symbolic affect-target associations. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–6.
- [111] Gu Y, Pan JZ, Cheng G et al. Local abox consistency prediction with transparent tboxes using gated graph neural networks. In *NeSy@ IJCAI*.
- [112] Cui Q, Zhou Y and Zheng M. 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*. Springer, pp. 419–426.
- [113] Xu C and Li R. Relation embedding with dihedral group in knowledge graph. *arXiv preprint arXiv:190600687* 2019; .
- [114] Cowen-Rivers AI, Minervini P, Rocktäschel T et al. Neural variational inference for estimating uncertainty in knowledge graph embeddings. In *Proceedings of the International Workshop on Neural-Symbolic Learning and Reasoning (NeSy) / ICLR 2019 Workshop*. URL <https://openreview.net/forum?id=HJM4rsRqFX>. Under review at ICLR 2019.
- [115] Bounabi M, Elmoutaouakil K and Satori K. A new neutrosophic tf-idf term weighting for text mining tasks: text classification use case. *International Journal of Web Information Systems* 2021; 17(3): 229–249.
- [116] Es-Sabery F, Hair A, Qadir J et al. Sentence-level classification using parallel fuzzy deep learning classifier. *IEEE Access* 2021; 9: 17943–17985.
- [117] Lima R, Espinasse B and Freitas F. 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)*. pp. 648–654.
- [118] Zhou M, Ji D and Li F. 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* 2021; 29: 2015–2026.
- [119] Gong J, Teng Z, Teng Q et al. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification. *IEEE Access* 2020; 8: 30885–30896.
- [120] Tato AAN, Nkambou R and Dufresne A. Hybrid deep neural networks to predict socio-moral reasoning skills. In *EDM*.
- [121] Langton J and Srihasam K. 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)*. pp. 7–11.
- [122] Braşoveanu AM and Andonie R. 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*. Springer, pp. 656–667.
- [123] Pinhanez C, Cavalin P, Ribeiro VHA 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)*. pp. 7014–7027.
- [124] Dehua C, Keting Z and Jianrong H. Bdcn: Semantic embedding self-explanatory breast diagnostic capsules network. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*. pp. 1178–1189.
- [125] Fazlic LB, Hallawa A, Schmeink A et al. 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)*. IEEE, pp. 1025–1030.
- [126] D’Souza J, Mulang IO and Auer S. 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)*. pp. 90–100.
- [127] Ayyanar R, Koomullil G and Ramasangu H. Causal relation classification using convolutional neural networks and grammar tags. In *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE, pp. 1–3.
- [128] Hu D, Wei L and Huai X. Dialoguecrn: Contextual reasoning networks for emotion recognition in conversations. *arXiv preprint arXiv:210601978* 2021; .
- [129] Chen K, Xu W, Cheng X et al. Question directed graph attention network for numerical reasoning over text. *arXiv preprint arXiv:200907448* 2020; .
- [130] Manda P, SayedAhmed S and Mohanty SD. Automated ontology-based annotation of scientific literature using deep learning. In *Proceedings of the international workshop on semantic Big Data*. pp. 1–6.
- [131] Honda H and Hagiwara M. Question answering systems with deep learning-based symbolic processing. *IEEE Access* 2019; 7: 152368–152378.
- [132] Schon C, Siebert S and Stolzenburg F. The corg project: cognitive reasoning. *KI-Künstliche Intelligenz* 2019; 33: 293–299.
- [133] Amin K. 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)*. IEEE, pp. 133–140.
- [134] Chen Q, Lamoreaux A, Wang X et al. Web question answering with neurosymbolic program synthesis. In

- Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*. pp. 328–343.
- [135] Chen Z, Gao Q and Moss LS. Neurallog: Natural language inference with joint neural and logical reasoning. *arXiv preprint arXiv:210514167* 2021; .
- [136] Pacheco ML and Goldwasser D. Modeling content and context with deep relational learning. *Transactions of the Association for Computational Linguistics* 2021; 9: 100–119.
- [137] Chaturvedi I, Satapathy R, Cavallari S et al. Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recognition Letters* 2019; 125: 264–270.
- [138] Qin J, Liang X, Hong Y et al. Neural-symbolic solver for math word problems with auxiliary tasks. *arXiv preprint arXiv:210701431* 2021; .
- [139] Yao Y, Xu J, Shi J et al. Learning to activate logic rules for textual reasoning. *Neural Networks* 2018; 106: 42–49.
- [140] Shi J, Ding X, Du L et al. Neural natural logic inference for interpretable question answering. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. pp. 3673–3684.
- [141] Škrlić B, Martinc M, Lavrač N et al. autobot: evolving neuro-symbolic representations for explainable low resource text classification. *Machine Learning* 2021; 110(5): 989–1028.
- [142] Wang W and Pan SJ. Variational deep logic network for joint inference of entities and relations. *Computational Linguistics* 2021; 47(4): 775–812.
- [143] Lemos H, Avelar P, Prates M et al. Neural-symbolic relational reasoning on graph models: Effective link inference and computation from knowledge bases. In *International Conference on Artificial Neural Networks*. Springer, pp. 647–659.
- [144] Huang Q, Deng L, Wu D et al. Attentive tensor product learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33. pp. 1344–1351.
- [145] Das R, Zaheer M, Thai D et al. Case-based reasoning for natural language queries over knowledge bases. *arXiv preprint arXiv:210408762* 2021; .
- [146] Yabloko L. 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*. pp. 645–652.
- [147] Zhou B, Richardson K, Ning Q et al. Temporal reasoning on implicit events from distant supervision. *arXiv preprint arXiv:201012753* 2020; .
- [148] Saveleva E, Petukhova V, Mosbach M et al. Graph-based argument quality assessment. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*. pp. 1268–1280.
- [149] Gupta K, Ghosal T and Ekbal A. A neuro-symbolic approach for question answering on research articles. In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*. pp. 40–49.
- [150] Demeter D and Downey D. Just add functions: A neural-symbolic language model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34. pp. 7634–7642.
- [151] Jiang H, Gurajada S, Lu Q et al. Lnn-el: A neuro-symbolic approach to short-text entity linking. *arXiv preprint arXiv:210609795* 2021; .
- [152] Kogkalidis K, Moortgat M and Moot R. Neural proof nets. *arXiv preprint arXiv:200912702* 2020; .
- [153] Zhang Q, Wang L, Yu S et al. Noahqa: Numerical reasoning with interpretable graph question answering dataset. *arXiv preprint arXiv:210910604* 2021; .
- [154] Sen P, Danilevsky M, Li Y et al. 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)*. pp. 4211–4221.
- [155] Huo S, Ma T, Chen J et al. Graph enhanced cross-domain text-to-sql generation. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*. pp. 159–163.
- [156] Jiang J, Wang H, Xie J et al. Medical knowledge embedding based on recursive neural network for multi-disease diagnosis. *Artificial Intelligence in Medicine* 2020; 103: 101772.
- [157] Liu W, Tang J, Liang X et al. Heterogeneous graph reasoning for knowledge-grounded medical dialogue system. *Neurocomputing* 2021; 442: 260–268.
- [158] Chaudhury S, Sen P, Ono M et al. Neuro-symbolic approaches for text-based policy learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. pp. 3073–3078.
- [159] Verga P, Sun H, Soares LB et al. Facts as experts: Adaptable and interpretable neural memory over symbolic knowledge. *arXiv preprint arXiv:200700849* 2020; .
- [160] Socher R, Chen D, Manning CD et al. Reasoning with neural tensor networks for knowledge base completion. *Advances in neural information processing systems* 2013; 26.

- [161] Serafini L and d'Avila Garcez AS. Logic tensor networks: Deep learning and logical reasoning from data and knowledge 2016; 1768: 1–12.
- [162] Raissi M, Perdikaris P and Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 2019; 378: 686–707.
- [163] Chen K, Huang Q, Palangi H et al. Mapping natural-language problems to formal-language solutions using structured neural representations. In *International Conference on Machine Learning*. PMLR, pp. 1566–1575.
- [164] Graziani L, Melacci S and Gori M. 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. Springer, pp. 185–197.
- [165] Altszyler Lemcovich EJ, Brusco P, Basiou N et al. Zero-shot multi-domain dialog state tracking using descriptive rules 2020; .
- [166] Hussain A and Cambria E. Semi-supervised learning for big social data analysis. *Neurocomputing* 2018; 275: 1662–1673.
- [167] Smolensky P, Lee M, He X et al. Basic reasoning with tensor product representations, 2016. URL <https://arxiv.org/abs/1601.02745>. 1601.02745.
- [168] Smolensky P. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence* 1990; 46(1-2): 159–216.
- [169] Subramanian C, Liu M, Khan N et al. A neuro-symbolic approach to multi-agent rl for interpretability and probabilistic decision making. *arXiv preprint arXiv:240213440* 2024; .
- [170] Shraga I, Azran G, Gerstgrasser M et al. Collaboration promotes group resilience in multi-agent rl 2021; .
- [171] Xu W, Wang J, Xie L et al. LogicMP: A neuro-symbolic approach for encoding first-order logic constraints. In *The Twelfth International Conference on Learning Representations*. URL <https://openreview.net/forum?id=BLGQ3oqldb>.
- [172] Marra G, Giannini F, Diligenti M et al. Integrating learning and reasoning with deep logic models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 517–532.
- [173] Shi S, Chen H, Ma W et al. Neural logic reasoning. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM '20, New York, NY, USA: Association for Computing Machinery. ISBN 9781450368599, p. 1365–1374. DOI:10.1145/3340531.3411949.
- [174] Dragone P, Teso S and Passerini A. Neuro-symbolic constraint programming for structured prediction 2021; .
- [175] Tibbits S. 4d printing: multi-material shape change. *Architectural design* 2014; 84(1): 116–121.
- [176] Ge Q, Dunn CK, Qi HJ et al. Active origami by 4d printing. *Smart materials and structures* 2014; 23(9): 094007.
- [177] Lendlein A and Kelch S. Shape-memory polymers. *Angewandte Chemie International Edition* 2002; 41(12): 2034–2057.
- [178] Gazzaniga A, Foppoli A, Cerea M et al. Towards 4d printing in pharmaceuticals. *International Journal of Pharmaceutics: X* 2023; 5: 100171.
- [179] Bodaghi M, Wang L, Zhang F et al. 4d printing roadmap. *Smart Materials and Structures* 2024; 33(11): 113501.
- [180] Demoly F, Dunn ML, Wood KL et al. The status, barriers, challenges, and future in design for 4d printing. *Materials & Design* 2021; 212: 110193.
- [181] Demoly F and André JC. *4D Printing, Volume 1: Between Disruptive Research and Industrial Applications*. John Wiley & Sons, 2022.
- [182] Demoly F and Andre JC. *4D Printing, Volume 2: Between Science and Technology*. John Wiley & Sons, 2022.
- [183] Simińska-Stanny J, Nizioł M, Szymczyk-Ziółkowska P et al. 4d printing of patterned multimaterial magnetic hydrogel actuators. *Additive Manufacturing* 2022; 49: 102506.
- [184] Zhao C, Lv Q and Wu W. Application and prospects of hydrogel additive manufacturing. *Gels* 2022; 8(5): 297.
- [185] Dong Y, Wang S, Ke Y et al. 4d printed hydrogels: fabrication, materials, and applications. *Advanced Materials Technologies* 2020; 5(6): 2000034.
- [186] Cecen A, Dai H, Yabansu YC et al. Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Materialia* 2018; 146: 76–84.
- [187] Ma C, Chang Y, Wu S et al. Deep learning-accelerated designs of tunable magneto-mechanical metamaterials. *ACS Applied Materials & Interfaces* 2022; 14(29): 33892–33902.
- [188] Sun X, Yue L, Yu L et al. Machine learning-enabled forward prediction and inverse design of 4d-printed active plates. *Nature Communications* 2024; 15(1): 5509.

- [189] Yang Z, Yu CH and Buehler MJ. Deep learning model to predict complex stress and strain fields in hierarchical composites. *Science Advances* 2021; 7(15): eabd7416.
- [190] Liu Z, Zhu C, Shen X et al. Current and future advances in inverse design approaches for materials. *Journal of Materials Science & Technology* 2018; 34(3): 482–491. DOI:10.1016/j.jmst.2017.12.019.
- [191] Sun X, Yue L, Yu L et al. Machine learning-evolutionary algorithm enabled design for 4d-printed active composite structures. *Advanced Functional Materials* 2022; 32(10): 2109805.
- [192] Abdullah T and Okay O. 4d printing of body temperature-responsive hydrogels based on poly (acrylic acid) with shape-memory and self-healing abilities. *ACS Applied Bio Materials* 2023; 6(2): 703–711.