

From Global Search to Local Lookup: Scalable Knowledge Base Completion via Differentiable Guarded Logic

YIZHENG ZHAO, State Key Laboratory of Novel Software Technology & School of Artificial Intelligence, Nanjing University, China

State-of-the-art knowledge base completion models, ranging from geometric embeddings to GNNs, achieve impressive predictive performance but share a critical limitation: they act as black boxes that often ignore data consistency, producing predictions that violate rigorous ontological constraints. While neuro-symbolic frameworks attempt to bridge this gap by integrating logical rules, they face a prohibitive scalability barrier. Existing approaches rely on unguarded universal quantification that demands global search, causing combinatorial explosions and execution failures on large-scale knowledge bases. In this paper, we propose GUARDNET, a differentiable reasoning framework designed to break this expressivity-scalability deadlock. Our core contribution is leveraging the Guarded Fragment (GF) of first-order logic to fundamentally restructure the computational graph of reasoning. We demonstrate that the GF's syntactic "guard" acts as a topological constraint that transforms intractable global quantification into efficient neighborhood-restricted lookups that strictly align logical evaluation with the underlying graph structure. For sparse knowledge bases, this paradigm shift eliminates the quadratic grounding overhead that plagues traditional neuro-symbolic systems, reducing complexity to linear in the number of edges without sacrificing the semantic expressiveness required for complex reasoning. This construction is the differentiable counterpart of the guarded chase in database theory: the guard atom acts as an index that confines grounding to existing tuples, inheriting the PTIME data complexity of guarded tuple-generating dependencies. Experiments on benchmarks with up to 377K concepts show that GUARDNET is the first neuro-symbolic framework to scale to such large knowledge bases, succeeding where neuro-symbolic and probabilistic baselines fail to converge within 72 hours, while significantly outperforming state-of-the-art geometric embedding and GNN models in both accuracy & consistency. We release our code at <https://doi.org/10.5281/zenodo.21809118> to support reproducibility and future research.

CCS Concepts: • **Theory of computation** → **Logic and databases**; • **Computing methodologies** → **Knowledge representation and reasoning**; • **Information systems** → **Graph-based database models**.

Additional Key Words and Phrases: Knowledge Graphs, Knowledge Base Completion, Neuro-Symbolic AI, Guarded Fragment, Description Logics, Differentiable Fuzzy Logic, Scalable Reasoning

ACM Reference Format:

Yizheng Zhao. 2026. From Global Search to Local Lookup: Scalable Knowledge Base Completion via Differentiable Guarded Logic. *Proc. ACM Manag. Data* 4, 4 (SIGMOD), Article 275 (September 2026), 27 pages. <https://doi.org/10.1145/3837113>

1 Introduction

Knowledge Base Completion (KBC) is a fundamental task in data management, requiring systems to infer missing information from large-scale, incomplete knowledge bases [2, 11, 20, 22, 29, 30]. In the Description Logic (DL) [6] context, a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is typically defined as the union of a TBox (aka ontology) $\mathcal{T}$ (terminological axioms defining concept hierarchies and relationships) and an ABox $\mathcal{A}$ (assertional facts about individuals). Equivalently, from a knowledge graph (KG)

Author's Contact Information: Yizheng Zhao, State Key Laboratory of Novel Software Technology & School of Artificial Intelligence, Nanjing University, Nanjing, China, zhaoyz@nju.edu.cn.



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2026 Copyright held by the owner/author(s).

ACM 2836-6573/2026/9-ART275

<https://doi.org/10.1145/3837113>

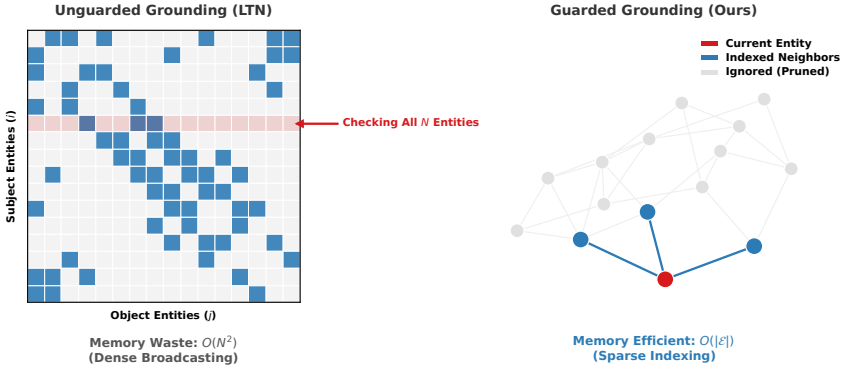


Fig. 1. **The Grounding Bottleneck.** **Left (Unguarded):** Standard neuro-symbolic approaches materialize the full $N \times N$ Cartesian product: for the current entity (highlighted row) the implication is evaluated against all N candidates, even though only the sparse blue cells correspond to actual edges. The remaining cells yield vacuous truths that consume $O(N^2)$ memory without contributing gradient signal. **Right (Guarded):** GUARDNET uses the guard atom as a topological index, evaluating only the indexed neighbors of the current entity (blue) and pruning the rest (gray), so the cost is $O(|E|)$ rather than $O(N^2)$.

perspective, the TBox corresponds to the ontological schema while the ABox corresponds to the instance-level graph of entities and relations. Modern biomedical ontologies such as SNOMED CT [47] (377K concepts) and Gene Ontology [4] (44K concepts; abbreviation: GO) encode rich taxonomic hierarchies in their TBoxes, while protein interaction networks demand reasoning over millions of relational facts in their ABoxes [51]. These applications require *multi-hop logical inference*—deducing that a concept belongs to a category by composing intermediate subsumption and role restrictions—rather than shallow pattern matching [21, 27, 35, 40, 41, 49, 68].

Prominent approaches to KBC fall into two camps, each with fundamental limitations. *Geometric embedding methods* [11, 31, 50, 55, 63, 67] and their ontology-aware extensions [1, 23, 26, 65, 70] scale linearly but sacrifice logical fidelity, learning implicit patterns rather than explicit rules. *Neuro-symbolic frameworks* [7, 14, 43, 57] preserve logical semantics but suffer catastrophic scalability failures on real-world knowledge bases. This creates an **expressivity–scalability dilemma** that has stymied progress toward production-grade logical reasoning systems [15, 36].

The Grounding Bottleneck. We identify the root cause of this dilemma as the *grounding bottleneck* inherent in existing neuro-symbolic architectures, as illustrated in **Figure 1**. Consider evaluating a universally quantified formula $\forall x, y. R(x, y) \rightarrow S(x, y)$ over a domain of N entities. Standard frameworks such as LTN and Neural LP [64] implement this via *dense tensor broadcasting*: they materialize the full Cartesian product $N \times N$, allocating $O(N^2)$ memory to represent all possible variable bindings. For SNOMED CT with $N = 377K$, this requires 1.4×10^{11} tensor elements—far exceeding any GPU’s capacity.

This is analogous to executing a nested-loop join without indices in a database system. The vast majority of computed entries correspond to *vacuous truths*—pairs (x, y) where $R(x, y)$ is false, making the implication trivially satisfied. These computations are semantically meaningless yet consume identical resources to genuine inferences, creating massive waste that triggers out-of-memory (OOM) failures at modest scales. As we demonstrate empirically, LTN and Neural LP crash at merely 5K entities, while production knowledge bases contain two orders of magnitude more.

LTN does offer a mechanism called *guarded quantifier* [7] that applies a Boolean mask to exclude certain variable bindings from aggregation. However, this mask is evaluated at *runtime*: the full

N^2 domain is materialized first, each pair is tested against the mask condition, and only then are non-qualifying pairs discarded. The mask evaluation itself remains $O(N^2)$, so no asymptotic complexity reduction is achieved. This raises a deeper question: can the restriction be imposed at *formula-writing time*, so that the grounding scope is determined by the syntactic structure of the logic before any computation begins?

The Landscape of Existing Solutions. Before presenting our answer, we situate this question within the broader landscape of neuro-symbolic research. Within the neuro-symbolic camp, three major paradigms have emerged to integrate neural learning with symbolic reasoning [10, 16], each representing a different trade-off:

- *Probabilistic Logic Programming.* Systems such as ProbLog [39], DeepProbLog [33], NeurASP [66], DeepStochLog [61], and NeuPSL [38] integrate neural networks with probabilistic inference. While principled for handling uncertainty, their reliance on explicit symbolic grounding limits scalability to knowledge bases with tens of thousands of facts.
- *Logic as Soft Constraint.* Another paradigm treats logical axioms as regularization terms during training. This includes Semantic-Based Regularization [17], Semantic Loss [62], Deep Logic Models [37], and differentiable SAT solvers [60]. These methods achieve efficiency by relaxing logical strictness, but struggle to enforce complex multi-hop dependencies.
- *Differentiable Fuzzy Logic.* Methods including DFL [57], LTN [7], logLTN [8], Neural Theorem Provers [43], TensorLog [14], and semantic entailment systems [52, 53] embed first-order logic (FOL) into continuous vector spaces via fuzzy semantics [24]. While maximally expressive, they inherit FOL's undecidability [13, 56] and the quadratic grounding bottleneck.

Outside the neuro-symbolic camp, graph neural network (GNN)-based approaches [54, 58, 59, 69, 71] achieve scalability through implicit message passing, but lack explicit logical semantics and cannot express negation, disjunction, or universal quantification [28].

Guard-based Tractability. Our answer draws on a line of work that is largely absent from the neuro-symbolic literature: the use of syntactic restrictions to secure tractability. The idea of restricting a logical language to achieve tractable differentiable reasoning has precedent: TensorLog [14] identifies polytree-limited stochastic deductive knowledge graphs (SDKGs) as a tractable fragment with linear-time inference, targeting the #P-hardness of probabilistic inference. However, TensorLog's restriction is orthogonal to ours: it constrains the variable-sharing structure of clauses (polytree) to avoid exponential proof paths, whereas GUARDNET constrains the quantifier scope via a guard atom to reduce *grounding size* from $O(N^2)$ to $O(|\mathcal{E}|)$. More broadly, several neuro-symbolic systems restrict quantification in practice (e.g., LTN's runtime mask, Neural-LP's sparse-matrix multiplication, NTP's backward chaining against observed facts, PSL/NeuPSL's typed domains, pLogicNet's variational selection), but none provides a provable grounding-size bound from a named logical fragment. In database theory, guarded tuple-generating dependencies (GTGDs) [12] use a syntactic guard condition to confine the chase to existing tuples, achieving PTIME data complexity for atomic queries. Vadalog [9] operationalizes a generalization of this idea (Warded Datalog[±], strictly more expressive than Guarded Datalog[±]) for enterprise-scale reasoning. GUARDNET is the first system to embed this guard-based tractability mechanism as a differentiable inductive bias, connecting neuro-symbolic grounding to the database-theoretic tractability of GTGDs.

Our Insight: The Guarded Fragment as Computational Primitive. We answer affirmatively the question posed above—whether the grounding restriction can be imposed at formula-writing time. Building on the GTGD tractability framework discussed above, we identify the guard condition of the **Guarded Fragment (GF)** as the missing inductive bias for scalable neuro-symbolic reasoning. Concretely, we adopt GF [3], the logical foundation underlying GTGDs. In GF, every quantified

variable must appear within a *guard atom*—a relational atom that restricts the variable’s scope to a local neighborhood. Unlike LTN’s runtime mask, the GF guard condition $FV(\psi) \cup \bar{x} \subseteq \text{Vars}(\alpha)$ is a *syntactic* property of the formula itself: it determines the grounding scope at formula-writing time, before any tensor is allocated. Operationally, the guard atom functions as a topological index, the differentiable analogue of a database index that converts a nested-loop join into a direct lookup. Rather than broadcasting over the entire domain, quantification is restricted to entities connected by existing edges in the knowledge base. For a guarded formula $\forall \bar{x}(\alpha(\bar{x}) \rightarrow \psi)$, this transforms the enumeration over all $N^{|\bar{x}|}$ possible tuples into a sparse traversal over only those tuples that actually satisfy the guard α in the knowledge base. In sparse knowledge bases—the predominant structure in real-world applications where the number of edges $|\mathcal{E}| \ll N^2$ —this yields substantial complexity reductions. Crucially, GF is not a toy fragment: it subsumes $\mathcal{ALC}$ [46] and covers nearly all constructs of $\mathcal{EL}^{++}$ [5] (the only exceptions are role composition and transitivity; see §5), while remaining decidable [6, 19]. This means the vast repository of industrial ontologies and knowledge bases can be directly translated into GF without loss of semantics, enabling rigorous logical reasoning at unprecedented scale.

The GUARDNET Framework. We operationalize this insight in GUARDNET, the first end-to-end differentiable framework that leverages the GF as a structured inductive bias for neural reasoning over knowledge bases. Our contributions are:

- **Principled Fuzzy Semantics (§3).** Building on theoretical foundations of fuzzy DLs [48] and differentiable fuzzy operators [57], we present a semantics grounded in the Product t-norm with a Sigmoidal Reichenbach implication. This combination ensures smooth, non-vanishing gradients that prevent dead zones plaguing existing systems.
- **Guarded Grounding with Linear Complexity (§4.1).** We develop a novel symbol grounding mechanism that uses guard atoms as topological indices, transforming quantified formulas into sparse, neighborhood-restricted tensor operations. This reduces memory complexity from $O(N^2)$ to $O(|\mathcal{E}|)$, enabling—for the first time—rigorous FOL-based reasoning on knowledge bases with hundreds of thousands of concepts. This neuralizes the tractability mechanism of GTGDs [12], establishing a formal bridge between database-theoretic query answering and differentiable neural reasoning. Unlike parallel embedding training approaches [25] that distribute existing algorithms, our method fundamentally restructures the computation itself.
- **Hybrid Domain Strategy (§4.2, 4.3, 4.4).** We propose a novel witness construction mechanism that systematically prevents the model from satisfying axioms vacuously. This addresses a fundamental learning pathology—where models learn trivial “all-false” solutions—ensuring that logical satisfaction reflects genuine semantic understanding.

Empirical Validation. We conduct extensive experiments on four challenging benchmarks. GUARDNET consistently outperforms 15 state-of-the-art baselines spanning five paradigms. Notably, while state-of-the-art neuro-symbolic methods fail with OOM at 5K/10K entities, GUARDNET scales linearly to 377K concepts—a **75× improvement in scalability** with simultaneous gains in accuracy. Our ablation studies further validate the critical role of the **Guarded Grounding** and **Hybrid Domain** strategies in achieving these results.

2 Preliminaries: Guarded Fragment

In this section, we introduce the **Guarded Fragment (GF)** [3], a decidable fragment of first-order logic (FOL) that strikes a balance between expressivity and decidability & complexity.

Let $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{V}$ be countably infinite and pairwise disjoint sets of **predicate**, **constant**, and **variable** symbols, respectively. A **signature** $\Sigma = (\mathcal{P}, \mathcal{C})$ specifies the non-logical vocabulary used

to construct well-formed GF formulas. Following the standard definition of GF, **terms** are restricted to constants and variables only, excluding function symbols present in full FOL.

An **atom** has the form $P(t_1, \dots, t_n)$, where $P \in \mathcal{P}$ is an n -ary predicate and $t_1, \dots, t_n$ are terms. For any atom $\alpha = P(t_1, \dots, t_n)$, we define $\text{Vars}(\alpha) = \{t_i \mid t_i \in \mathcal{V}\}$ as the set of variables occurring in α . In a quantified formula $\forall x \phi / \exists x \phi$, we say the quantifier **binds** the variable x , and ϕ is the **scope** of the quantifier. An **occurrence** of a variable x in a formula ϕ is **bound** if it lies within the scope of a quantifier that binds x ; otherwise, the occurrence is **free**. A variable x is a **free variable** of ϕ if x has at least one free occurrence in ϕ . We denote by $\text{FV}(\phi)$ the set of all free variables of ϕ . A GF formula with no free variables is called a **sentence**.

DEFINITION 1 (SYNTAX OF GF). *GF formulas are defined inductively as follows:*

- Every atom is a GF formula;
- If ϕ and ψ are GF formulas, then so are $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, and $\phi \rightarrow \psi$;
- If ψ is a GF formula, $\bar{x}$ is a tuple of variables, and α is an atom satisfying the **guard condition**:

$$\text{FV}(\psi) \cup \bar{x} \subseteq \text{Vars}(\alpha),$$

then $\exists \bar{x} (\alpha \wedge \psi)$ and $\forall \bar{x} (\alpha \rightarrow \psi)$ are GF formulas.

In guarded quantification, the atom α serves as a **guard** for the **body** ψ . The guard condition requires that all variables involved in the quantification—both the free variables of the body and the quantified variables—appear in the guard. This fundamentally relativizes quantification: instead of ranging over the entire domain, quantified variables are restricted to the local neighborhood defined by the guard. For instance, $\exists x (R(x, y) \wedge \psi(x, y))$ restricts x to objects that are R -related to y , effectively performing a one-hop neighbor lookup on a KB. We write $\text{ext}(\alpha)$ for the **guard extension** of an atom α , i.e., the set of tuples that satisfy α in $\mathcal{K}$; §4.1 analyzes its role in the complexity of grounding. We illustrate with two axioms from SNOMED CT, which serve as running examples throughout the paper.

► Example (Running)

Consider the SNOMED CT axiom:

$$\text{Pneumonia} \sqsubseteq \text{LungDisease}$$

stating that every pneumonia is a lung disease. Its GF translation is:

$$\forall x (\text{Pneumonia}(x) \rightarrow \text{LungDisease}(x))$$

Here the unary atom $\text{Pneumonia}(x)$ is the guard. The guard extension $\text{ext}(\text{Pneumonia})$ contains only the entities asserted or inferred to be pneumonia instances. In SNOMED CT ($N=377K$ concepts), this might be a few hundred entities. An unguarded system evaluates the implication for all 377K entities; GUARDNET evaluates it only over $\text{ext}(\text{Pneumonia})$.

Now consider a richer axiom:

$$\text{Pneumonia} \sqsubseteq \exists \text{hasSite.Lung}$$

stating that every pneumonia is located in the lung. Its GF translation is:

$$\forall x (\text{Pneumonia}(x) \rightarrow \exists y (\text{hasSite}(x, y) \wedge \text{Lung}(y)))$$

The guard $\text{Pneumonia}(x)$ restricts the outer quantification, and $\text{hasSite}(x, y)$ guards the inner existential. Together, they confine grounding to the two-hop neighborhood of pneumonia entities in the KG, yielding $O(|\mathcal{E}|)$ instead of $O(N^2)$.

Table 1. Fuzzy operators selected for GUARDNET (**bold**) and the alternatives evaluated. The **Gradient note** column summarizes the optimization behavior that motivates each choice; operators with dead zones or plateaus yield vanishing gradients and are therefore avoided.

Category	Operator	Gradient note
T-norm ($\wedge$)	Product (ours)	Adaptive: $\partial T / \partial a = b$
	Gödel	Winner-take-all
	Łukasiewicz	Dead zone when $a+b \leq 1$
	Yager ($p=2$)	Tunable interpolation
	Hamacher	Coupled sensitivity
T-conorm ($\vee$)	Probabilistic Sum (ours)	Product dual; balanced
Implication ($\rightarrow$)	Sigmoidal Reichenbach	Boundary-focused; no plateaus
	Standard Reichenbach	Linear; loss = $a(1-c)$
	Goguen	Zero plateau ($a \leq c$)
Quantifier ($\forall/\exists$)	Generalized Mean	Tunable; $[0, 1]$ -bounded

Based on this syntax, we define the central data structure used in our framework. A **knowledge base** (KB) $\mathcal{K}$ is a finite set of GF sentences constructed over the signature Σ . Analogous to the terminology in DLs, we partition $\mathcal{K}$ into an ABox $\mathcal{A}$, consisting of ground facts (i.e., variable-free atoms), and a TBox $\mathcal{T}$, consisting of quantified formulas that describe general terminological knowledge. The **size** of the KB, denoted $|\mathcal{K}|$, is defined as the total number of symbols required to write down all formulas in $\mathcal{K}$. Throughout, N denotes the number of entities (domain elements) and $|\mathcal{E}|$ the number of ground facts, i.e., edges of the knowledge graph.

GF inherits its semantics from standard FOL: a structure $\mathfrak{A} = \langle \Delta, \cdot^{\mathfrak{A}} \rangle$ assigns a domain Δ , interprets constants as domain elements and predicates as relations over Δ , and evaluates satisfaction inductively in the usual way. The only difference from unrestricted FOL is that quantifiers in GF are syntactically constrained by guards; the semantic evaluation rules for $\exists \bar{x} (\alpha \wedge \psi)$ and $\forall \bar{x} (\alpha \rightarrow \psi)$ are identical to their FOL counterparts. GF is decidable [3], and enjoys the finite model property [19], i.e., every satisfiable GF formula has a model of a finite domain. These properties make GF a fragment where expressive logical reasoning remains computationally feasible. From a database-theoretic perspective, the guard condition in Definition 1 is precisely the syntactic constraint underlying GTGDs [12]. GUARDNET operationalizes this classical mechanism in a differentiable neural architecture, as detailed in §4.1.

3 A Differentiable Semantics for Learning

Classical GF provides decidable reasoning, yet real-world applications demand the ability to handle the uncertainty and vagueness inherent in noisy data. To address this, we extend GF with fuzzy semantics, mapping atomic formulas like $\text{Lung}(x)$ to continuous truth values in $[0, 1]$.

However, the choice of fuzzy operators is a critical hyperparameter defining the **geometry of the optimization landscape** [57]. A poor choice can lead to vanishing gradients or stagnation. Table 1 summarizes the operator configuration used by GUARDNET.

3.1 Fuzzy Operator Selection Rationale

We select the **Product** t-norm because it provides adaptive gradient scaling: $\partial T / \partial a = b$, so each input receives signal proportional to its partner’s confidence, balancing bottleneck correction

with holistic learning. Alternatives suffer from dead zones (Łukasiewicz), winner-take-all sparsity (Gödel), or premature fixation (Hamacher). The **Probabilistic Sum**, as the Product dual, inherits these balanced properties for disjunction. For implication, the **Sigmoidal Reichenbach** [57] concentrates gradient mass on decision boundaries, addressing the zero-plateau problem of R-implications (Goguen) and the linear treatment of all violations by the standard Reichenbach S-implication.

Concretely, the operators instantiating T , S , and I in Definition 2 are

$$T(a, b) = a b, \quad S(a, b) = a + b - a b, \quad I(a, c) = \text{scale}(\sigma(s(I_R(a, c) + b_0))), \quad (1)$$

where $I_R(a, c) = 1 - a + a c$ is the standard Reichenbach implication, σ is the logistic sigmoid with steepness s and bias b_0 , and $\text{scale}(\cdot)$ is the affine map that renormalizes the attainable sigmoid range onto $[0, 1]$; we give it explicitly in §4.4, where it enters the training loss.

3.2 Guarded Quantifiers as Parameterized Local Pooling

In GF, quantification is always restricted to a local neighborhood defined by the guard atom. To support our framework’s logic-agnostic design—where operators may range from strict (Gödel) to smooth (Product)—the aggregation mechanism must be equally adaptable. We adopt the **Generalized Mean (p -mean)** as a universal, parameterizable pooling operator. By adjusting the exponent $p \in [1, \infty)$, it provides a continuous spectrum: reducing to the arithmetic mean at $p = 1$ (matching Łukasiewicz linearity) and converging to the hard extremum as $p \rightarrow \infty$ (matching Gödel). Let $\mathbf{z} = (z_1, \dots, z_k)$ be the truth values evaluated over k guard-satisfying neighbors.

Existential quantifier ($\exists$). We employ the standard Generalized Mean to aggregate evidence for satisfaction:

$$\mathcal{M}^\exists(\mathbf{z}; p) = \left(\frac{1}{k} \sum_{i=1}^k z_i^p \right)^{\frac{1}{p}} \quad (2)$$

Higher p focuses on the best evidence (approaching max); lower p aggregates uniformly.

Universal quantifier ($\forall$). We employ the **Generalized Mean Error** to aggregate deviation from perfect truth [7]:

$$\mathcal{M}^\forall(\mathbf{z}; p) = 1 - \left(\frac{1}{k} \sum_{i=1}^k (1 - z_i)^p \right)^{\frac{1}{p}} \quad (3)$$

This formulation prevents gradient vanishing when the majority of neighbors satisfy the condition. As $p \rightarrow \infty$, the result converges to the minimum (the “weakest link”). It is convenient to name its dual, the **error aggregator** $\mathcal{D}_p^\forall(\mathbf{l}) = \left(\frac{1}{k} \sum_{i=1}^k l_i^p \right)^{1/p}$, which aggregates per-instance dissatisfaction $l_i = 1 - z_i$ and satisfies $\mathcal{M}^\forall(\mathbf{z}) = 1 - \mathcal{D}_p^\forall(\mathbf{1} - \mathbf{z})$. We use $\mathcal{D}_p^\forall$ when stating loss-side results. Since each $z_i \in [0, 1]$, both $\mathcal{M}^\exists$ and $\mathcal{M}^\forall$ are guaranteed to remain in $[0, 1]$ for all $p \geq 1$, unlike unbounded approximations such as LogSumExp.

Numerical stability. As noted by Badreddine et al. [7], the gradient of the p -mean can explode when inputs approach boundaries (0 or 1) for large p . To ensure stability across arbitrary operator combinations, we implement the *Stable Product* strategy, applying a projection $\pi(x) = (1 - \epsilon)x + \epsilon$ to inputs before aggregation, where $\epsilon > 0$ is a small constant that keeps truth values away from the boundaries at which the p -mean gradient diverges.

3.3 Generalized Fuzzy Satisfaction

Finally, we integrate the operator choices discussed above into a unified semantic framework. GUARDNET treats the logical semantics as a hyperparameter, allowing the system to be configured based on domain constraints (e.g., noise levels, graph sparsity). We define the generalized fuzzy satisfaction degree $\llbracket \phi \rrbracket_{\mathfrak{A}, \rho} \in [0, 1]$ as follows.

DEFINITION 2 (GENERALIZED FUZZY SATISFACTION). *Let $\mathfrak{A}$ be a fuzzy structure, ρ a variable assignment, and $O = \{T, S, I, \mathcal{M}_p^\exists, \mathcal{M}_p^\forall\}$ be a selected set of differentiable fuzzy operators parameterized by p . The satisfaction degree is defined recursively:*

$$\begin{aligned} \llbracket P(\bar{t}) \rrbracket_{\mathfrak{A}, \rho} &= P^{\mathfrak{A}}(\bar{t}^{\mathfrak{A}, \rho}) & \llbracket \neg \phi \rrbracket_{\mathfrak{A}, \rho} &= 1 - \llbracket \phi \rrbracket_{\mathfrak{A}, \rho} \\ \llbracket \phi \wedge \psi \rrbracket_{\mathfrak{A}, \rho} &= T(\llbracket \phi \rrbracket_{\mathfrak{A}, \rho}, \llbracket \psi \rrbracket_{\mathfrak{A}, \rho}) \\ \llbracket \phi \vee \psi \rrbracket_{\mathfrak{A}, \rho} &= S(\llbracket \phi \rrbracket_{\mathfrak{A}, \rho}, \llbracket \psi \rrbracket_{\mathfrak{A}, \rho}) \\ \llbracket \exists \bar{x} (\alpha \wedge \psi) \rrbracket_{\mathfrak{A}, \rho} &= \mathcal{M}_p^\exists \left(\{ \llbracket \alpha \wedge \psi \rrbracket_{\mathfrak{A}, \rho'} \mid \bar{a} \in \Delta^{|\bar{x}|} \} \right) \\ \llbracket \forall \bar{x} (\alpha \rightarrow \psi) \rrbracket_{\mathfrak{A}, \rho} &= \mathcal{M}_p^\forall \left(\{ \llbracket \alpha \rightarrow \psi \rrbracket_{\mathfrak{A}, \rho'} \mid \bar{a} \in \Delta^{|\bar{x}|} \} \right) \end{aligned}$$

where $\rho' = \rho[\bar{x} \mapsto \bar{a}]$ denotes the assignment update.

Definition 2 states the semantics in its unrestricted form, ranging over all of $\Delta^{|\bar{x}|}$. §4.1 shows that the guard condition lets us evaluate the quantifier cases over $\text{ext}(\alpha)$ alone without changing the resulting truth value, which is what yields the $O(|\mathcal{E}|)$ bound.

This formulation provides GUARDNET with continuous adaptability: it can interpolate between **strict, rule-bound logic** (using Gödel operators and large p) for precise reasoning in clean domains, and **probabilistic, soft reasoning** (using Product operators and small p) for robust learning in noisy, uncertain KBs.

4 A Differentiable Fuzzy GF Framework

To operationalize the theoretical advantages of GF for learning, we present GUARDNET, a neuro-symbolic framework built upon three core components: (1) a scalable neural grounding mechanism; (2) a unified fuzzy loss function; and (3) a semantically-aware domain construction strategy. The overall architecture is illustrated in **Figure 2**.

4.1 Neural Grounding Mechanism

The bridge between logic and neural networks is the **grounding** function $\mathcal{G}_\theta$. Before detailing our implementation, we formally analyze the bottleneck that necessitates our architectural departure from standard neuro-symbolic approaches.

4.1.1 The Bottleneck: Unguarded Grounding and Gradient Dilution. Standard approaches (e.g., LTN) typically ground FOL rules by instantiating variables over the entire domain Δ . For a rule like $\forall x, y (R(x, y) \rightarrow C(y))$, standard semantics treat quantifiers as **unguarded**, aggregating over the full Cartesian product:

$$\llbracket \forall x, y \phi(x, y) \rrbracket = \bigotimes_{u, v \in \Delta} \llbracket \phi(u, v) \rrbracket \quad (4)$$

This formulation introduces two key failures in large-scale learning:

- **System Implementation Bottleneck:** Physically evaluating Equation 4 in deep learning frameworks requires **tensor broadcasting** to materialize all variable combinations. For a domain size of $|\Delta| \approx 3.7 \times 10^5$ (e.g., SNOMED CT), constructing a dense pairwise tensor of shape $[|\Delta|, |\Delta|, d]$ attempts to allocate hundreds of gigabytes of GPU memory, triggering immediate OOM errors.

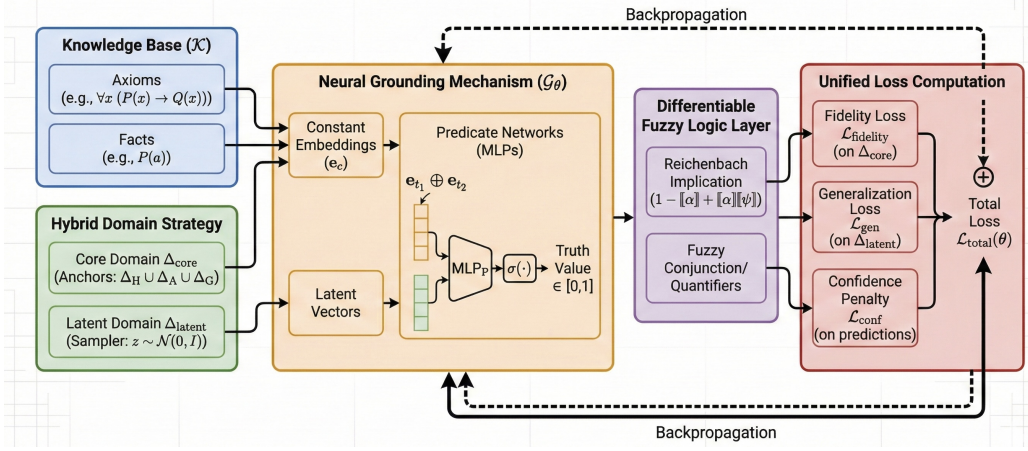


Fig. 2. **The GUARDNET framework.** A GF axiom is grounded by the **Neural Grounding** mechanism, which treats the guard atom as a topological index and enumerates only the tuples in $\text{ext}(\alpha)$ (§4.1). Terms are mapped to learnable embeddings and predicates to neural networks, yielding fuzzy truth values. Groundings are drawn from the **Hybrid Domain**: the Core Domain Δ_{core} supplies semantic anchors for fidelity, while the Latent Domain Δ_{latent} supplies sampled probes for generalization (§4.3). Both streams feed the unified objective $\mathcal{L}_{\text{total}}$, whose gradients update the embeddings and predicate networks jointly (§4.4).

- **Optimization Bottleneck (Gradient Dilution):** Even if memory permits, existing frameworks remain agnostic to the sparsity of R . They evaluate the implication $R(u, v) \rightarrow C(v)$ for *every* pair. In sparse KBs, $R(u, v)$ is false for the vast majority of pairs. Since the logical implication $\text{False} \rightarrow \dots$ trivially evaluates to True (1), the model wastes $O(|\Delta|^2)$ computation verifying these tautologies. Crucially, this creates a gradient dilution problem: the learning signal from the few actual edges (where R is true) is mathematically drowned out by millions of vacuously satisfied instances, stalling convergence.

4.1.2 The Solution: Guarded Sparse Indexing. GUARDNET resolves this by strictly enforcing the **Guarded Grounding** paradigm. Instead of broadcasting over the domain, we leverage the GF syntax to treat the guard atom α as a topological index. Contrasting with Equation 4, guarded aggregation ranges only over the tuples that satisfy the guard:

$$\llbracket \forall \bar{x} (\alpha(\bar{x}) \rightarrow \psi(\bar{x})) \rrbracket = \bigotimes_{\bar{a} \in \text{ext}(\alpha)} \llbracket \alpha(\bar{a}) \rightarrow \psi(\bar{a}) \rrbracket, \quad (5)$$

which is violation-equivalent to the full-domain form by Proposition 1.

Complexity Analysis. Let $\phi = \forall \bar{x} (\alpha(\bar{x}) \rightarrow \psi(\bar{x}))$ be a guarded axiom where α is the guard atom and $|\bar{x}| = k$ is the number of quantified variables. We define the *guard extension* as:

$$\text{ext}(\alpha) = \{(a_1, \dots, a_k) \mid \alpha(a_1, \dots, a_k) \text{ holds in } \mathcal{K}\}$$

- **Unguarded Grounding:** Standard approaches enumerate all N^k variable assignments (where $N = |\Delta|$ is the domain size), yielding $O(N^k)$ complexity per axiom. For binary predicates ($k = 2$), this is $O(N^2)$.
- **Guarded Grounding:** Our approach enumerates only tuples in $\text{ext}(\alpha)$, yielding $O(|\text{ext}(\alpha)|)$ complexity per axiom.

For a KB $\mathcal{K}$ with m axioms, the total grounding complexity per epoch is:

$$O\left(\sum_{i=1}^m |\text{ext}(\alpha_i)|\right) \leq O(m \cdot |\mathcal{E}|_{\max})$$

where $|\mathcal{E}|_{\max} = \max_i |\text{ext}(\alpha_i)|$ is the size of the largest guard extension in $\mathcal{K}$.

Sparsity Assumption. This complexity advantage is predicated on the *sparsity* of real-world KBs. When guard predicates have bounded extension size (i.e., $|\mathcal{E}|_{\max} = O(N)$), guarded grounding achieves complexity linear in the domain size. We note three caveats:

- In dense KBs where guard extensions approach $\Theta(N^2)$, guarded and unguarded grounding have equivalent asymptotic complexity.
- Our experimental datasets (SNOMED CT, GO, PPI networks) exhibit sparse structure with average degree $\bar{d} \ll N$, empirically validating this assumption.
- Even in scale-free networks with high-degree hub nodes, the worst-case guard extension $|\text{ext}(\alpha)|_{\max}$ remains far below N^2 . In our densest experimental dataset ($N=6,438$, $\bar{d}=34.3$; see §5), power-law degree analysis estimates $d_{\max} \approx 300$, so even the largest hub processes $d_{\max}^2/N^2 < 0.3\%$ of the pairs an unguarded approach would evaluate. Hub load imbalance is a scheduling concern, not an asymptotic one; it is handled by our mini-batch strategy which distributes guard extensions across batches.

► Example (Guarded Grounding Trace)

Consider a 3-entity subgraph of SNOMED CT with entities $\{\text{case}_1, \text{case}_2, \text{lung}_1\}$ and a single edge:

$\text{hasSite}(\text{case}_1, \text{lung}_1)$

For the axiom:

$$\forall x (\text{Pneumonia}(x) \rightarrow \exists y (\text{hasSite}(x, y) \wedge \text{Lung}(y)))$$

Unguarded (LTN): Enumerate all $3^2 = 9$ pairs for hasSite , evaluate the implication for each. 8 out of 9 produce vacuous truths ($\text{hasSite} = \text{false} \Rightarrow \text{implication} = 1$), wasting computation and diluting gradients.

Guarded (GUARDNET): Look up:

$$\text{ext}(\text{hasSite}) = \{(\text{case}_1, \text{lung}_1)\}$$

Evaluate the implication only for this single pair. Total cost: $O(|\mathcal{E}|) = O(1)$ instead of $O(N^2) = O(9)$.

Computational vs. Semantic Properties. The $O(|\mathcal{E}|)$ bound is a *computational* property derived from the syntactic structure of the guard atom and the KG topology. It does not depend on the choice of fuzzy operators: whether conjunction is interpreted via Product, Gödel, or any other t-norm, the set of tuples to be evaluated remains $\text{ext}(\alpha)$. Classical semantic properties of GF (decidability, finite model property) do not automatically transfer to the fuzzy relaxation; the complexity bound, however, does, because it is determined before any truth-value computation begins.

For sparse grounding, GUARDNET uses a *closed-world guard index*: for each atomic guard relation α_i , the observed guard extension $\text{ext}(\alpha_i)$ is constructed from the KB, and tuples outside this extension are assigned zero guard truth for the purpose of grounding. The following proposition makes explicit that, under this masked grounding semantics, only tuples inside the guard extension can contribute nonzero violation.

PROPOSITION 1 (SOUNDNESS OF MASKED GUARDED EVALUATION). *Consider a guarded axiom $\phi_i = \forall \bar{x} (\alpha_i(\bar{x}) \rightarrow \psi_i(\bar{x}))$ where α_i is the atomic guard relation. Under the masked guard-index semantics, $\llbracket \alpha_i(c) \rrbracket_{\text{mask}} = 0$ for all $c \notin \text{ext}(\alpha_i)$. If the fuzzy implication I satisfies $I(0, t) = 1$ for all $t \in [0, 1]$, then every tuple outside the guard extension is vacuously satisfied:*

$$c \notin \text{ext}(\alpha_i) \Rightarrow I(\llbracket \alpha_i(c) \rrbracket_{\text{mask}}, \llbracket \psi_i(c) \rrbracket) = 1.$$

Hence, sparse evaluation over $\text{ext}(\alpha_i)$ is violation-equivalent to full-domain evaluation under masked guard-index semantics.

The proof is immediate from $I(0, t) = 1$. This condition is satisfied by the normalized Reichenbach-style implication used in our fuzzy semantics.

REMARK. *The closed-world guard index is a computational convention for sparse grounding. It specifies which observed guard tuples participate in the differentiable objective and should not be interpreted as a claim that all unobserved ontology facts are false under classical open-world semantics.*

Below we detail the specific components of this efficient grounding mechanism:

- **Constants as Learnable Embeddings:** Each constant symbol $c \in C$ is grounded as a dense vector embedding $\mathbf{e}_c \in \mathbb{R}^d$. Unlike traditional symbolic reasoning where constants are rigid identifiers, here they act as malleable parameters. This optimization process allows the embeddings to capture the semantic properties of entities specifically as constrained by the logical axioms (e.g., learning that two constants appearing in similar logical contexts have similar vector representations). This design choice parallels KG embedding methods such as TransE [11] and RotatE [50], but with a key difference: in GUARDNET, embeddings are jointly optimized under both relational and logical constraints, so the learned geometry must simultaneously satisfy entity-level facts and universal axioms. We ablate this design in §5.3, comparing learnable embeddings against fixed pretrained alternatives.
- **Predicates as Differentiable Functions:** To ground our n -ary predicates in a way that is both expressive and scalable, we employ Multi-Layer Perceptrons (MLPs) as universal function approximators [45]. This choice provides the flexibility to learn arbitrary non-linear relationships without imposing strong prior assumptions on their geometric structure, which is a crucial feature for a general-purpose GF framework.

For an n -ary predicate, the input strategy to this MLP is straightforward: it is formed by concatenating the n individual term embeddings [18]. Concatenation is chosen as it is a parameter-free, information-preserving operation that makes no prior assumptions about the relationships between predicate arguments, leaving this task entirely to the learnable layers of the predicate network (see **Figure 2**). The truth value for an atom $P(t_1, \dots, t_n)$ is then explicitly given by:

$$\llbracket P(t_1, \dots, t_n) \rrbracket_{\mathcal{G}_\theta} = \sigma(\text{MLP}_P(\mathbf{e}_{t_1} \oplus \dots \oplus \mathbf{e}_{t_n})) \quad (6)$$

where $\mathbf{e}_{t_i} \in \mathbb{R}^d$ is the embedding of term t_i and $\oplus$ denotes concatenation. Our predicate MLPs utilize the ReLU activation function in hidden layers for its robustness against vanishing gradients. The final output layer employs a Sigmoid function $\sigma(\cdot)$ to map the network's logit to a fuzzy truth value in the $[0, 1]$ interval, aligning with a probabilistic interpretation of satisfaction.

4.2 The Challenge: Representational Collapse

A critical vulnerability in neuro-symbolic learning is the propensity for models to discover *reasoning shortcuts* [34]—devious paths to minimizing the loss function that bypass meaningful learning. We trace the origin of many such shortcuts to a foundational flaw: an improperly constructed

interpretation domain. This flaw permits a catastrophic failure mode we term **representational collapse**. Consider the axiom from Example 1:

$$\forall x (\text{Pneumonia}(x) \rightarrow \exists y (\text{hasSite}(x, y) \wedge \text{Lung}(y)))$$

Representational collapse manifests in two primary forms, as depicted in **Figure 3**:

- **Collapse to Emptiness:** The model achieves zero loss by simply learning to predict $\llbracket \text{Pneumonia}(c) \rrbracket \approx 0$ for all constants c . In this degenerate solution, the axiom is *vacuously satisfied* (since $\text{False} \rightarrow \text{Anything}$ is True), and the model learns nothing about the actual relationship between pneumonia and its anatomical site.
- **Collapse to Ambiguity:** The neural networks learn to output a constant value of 0.5 for all inputs, representing maximum uncertainty. This strategy exploits the properties of certain fuzzy loss landscapes to minimize penalties without the risk of making incorrect decisive predictions.

Figure: Illustration of Representational Collapse Modes

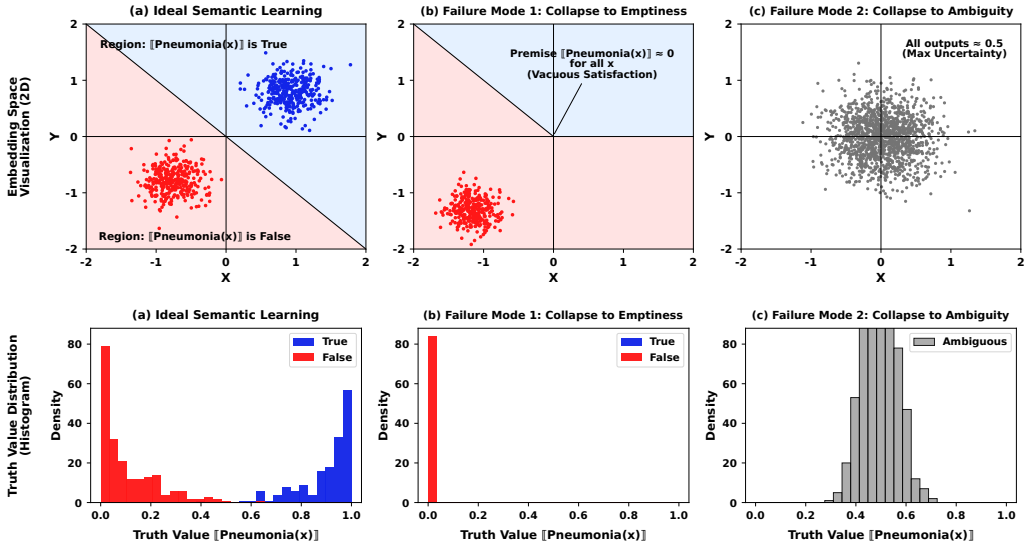


Fig. 3. Illustration of Representational Collapse Modes: Collapse to Emptiness (vacuous satisfaction) and Collapse to Ambiguity (maximum uncertainty).

While prior frameworks such as LTN can be effective given dense factual data, real-world ontologies often consist primarily of universal rules (axioms) with sparse factual assertions. This lack of explicit counter-examples makes them highly vulnerable to the aforementioned collapse modes.

Two natural and standard approaches initially seem plausible to address this data scarcity, but they ultimately fail:

- **Failure of Naive Sampling:** First, one might propose relying solely on randomly sampled vectors (“latent constants”) from a prior distribution to drive learning. However, when faced with random inputs producing ambiguous outputs (e.g., $\llbracket \text{Pneumonia}(\text{latent}_1) \rrbracket \approx 0.5$), the model quickly discovers a trivial optimal strategy: setting $\llbracket \text{Pneumonia}(x) \rrbracket \approx 0$ for all inputs. This allows the model to achieve zero loss by making the axiom vacuously satisfied, without learning any semantic structure.

- **Failure of Standard Regularization:** Second, one might employ standard regularization techniques [18] to mitigate collapse. We do actually incorporate such methods: a truth confidence penalty $\mathcal{L}_{\text{conf}} = \sum_i p_i (1 - p_i)$ that softly encourages predicate outputs p_i to be closer to 0 or 1, along with standard L2 weight decay.

However, these techniques prove fundamentally insufficient. The confidence penalty is “direction-agnostic”—it encourages confident predictions but provides no guidance on what those predictions should be. Even with $\mathcal{L}_{\text{conf}}$, the model can achieve perfect satisfaction by confidently predicting $\llbracket \text{Pneumonia}(x) \rrbracket \approx 0$ for all inputs. This simultaneously satisfies the confidence penalty (since 0 is a decisive prediction) and makes the axiom vacuously satisfied (since the premise is never met), while the model learns nothing meaningful about the relationship between pneumonia and its anatomical site.

4.3 The Solution: Hybrid Domain Strategy

These failures reveal that robust learning requires explicit semantic anchors that force non-trivial engagement with logical structure. To address this, GUARDNET implements a **Hybrid Domain Strategy** that synergistically combines a Core Domain with a Latent Domain, as shown in **Figure 4**.

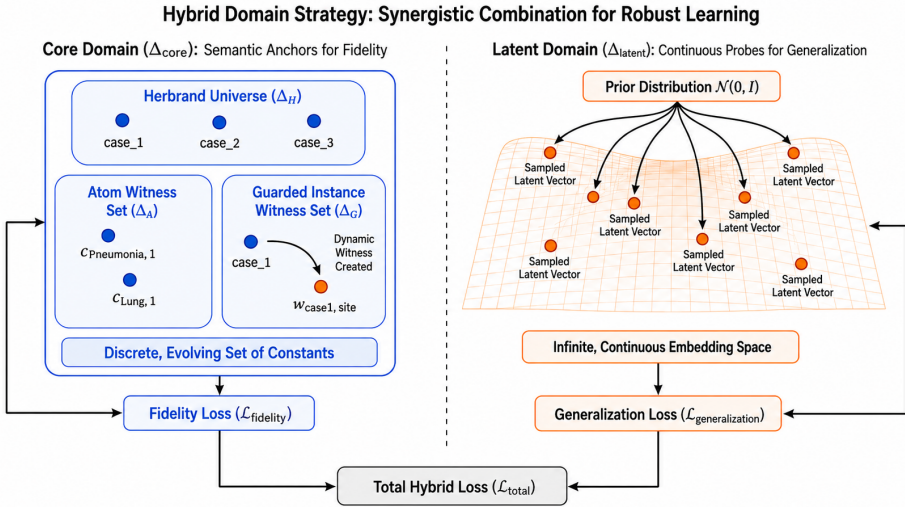


Fig. 4. Hybrid Domain Strategy: The Core Domain (Δ_{core}) provides semantic anchors to ensure fidelity; the Latent Domain (Δ_{latent}) drives generalization via continuous sampling.

4.3.1 Core Domain: Ensuring Fidelity via Semantically-Aware Construction. The Core Domain, Δ_{core} , provides a set of evolving “semantic anchors” that prevent representational collapse through a principled, multi-layered construction process addressing distinct failure modes.

- **Herbrand Domain (Δ_H):** A natural and canonical starting point rooted in classical logic is the **Herbrand Universe** [42]. In the context of a function-free language like GF, this domain consists of all constant symbols that explicitly appear as arguments to predicates within the KB $\mathcal{K}$:

$$\Delta_H = \{c \mid c \in \mathcal{C} \text{ and } c \text{ appears as an argument in } \mathcal{K}\} \quad (7)$$

This domain serves to ground the model’s reasoning in the concrete entities for which prior factual knowledge exists. However, relying solely on Δ_H is theoretically insufficient for robust

learning. Real-world KBs often consist primarily of universal rules without specific instances, rendering Δ_H sparse or even empty. Consequently, a domain constrained only to these explicit constants remains highly vulnerable to the representational collapse modes described previously.

- **Atom Witness Set (Δ_A):** This component acts as our primary defense against global collapse to emptiness, addressing the foundational problem of predicate existence in the absence of explicit facts. For every n -ary predicate $P \in \mathcal{P}$ that is not currently grounded by factual assertions in $\mathcal{K}$, we introduce a unique tuple of n fresh witness constants, $(c_{P,1}, \dots, c_{P,n})$. Through dedicated loss terms, we actively enforce a high truth value for these grounded atoms, i.e., $\llbracket P(c_{P,1}, \dots, c_{P,n}) \rrbracket \approx 1$. This mechanism guarantees that every concept in the theory—such as $\text{Pneumonia}(c_{\text{Pneumonia},1})$ or $\text{Lung}(c_{\text{Lung},1})$ —is interpreted as a non-empty set from the outset, thereby rendering a total collapse to emptiness impossible.
- **Guarded Instance Witness Set (Δ_G):** While atom witnesses break the initial stalemate of emptiness, they are insufficient on their own. A model might satisfy atom witnesses while still learning only vague statistical associations—maintaining moderate truth values that achieve acceptable loss through quantifier aggregation without learning precise relationships. To combat this, we introduce a dynamic mechanism designed to make such ambiguity computationally untenable for complex axioms.

This mechanism creates specific evidence “on-demand” when guarded rules are triggered. Its dynamic nature is crucial: the set of witnesses grows as the model learns. Consider the sample axiom $\forall x(\text{Pneumonia}(x) \rightarrow \exists y(\text{hasSite}(x, y) \wedge \text{Lung}(y)))$.

- **Early Training:** At epoch 1, model predictions are poor. The premise $\llbracket \text{Pneumonia}(c) \rrbracket$ is low for all constants, so no witnesses are created.
- **Triggering Event:** As training progresses (e.g., epoch 50), the model learns to recognize specific entities, such as ‘case₁’, as instances of Pneumonia with high confidence:

$$\llbracket \text{Pneumonia}(\text{case}_1) \rrbracket \approx 0.95.$$

- **Dynamic Creation:** This high-confidence guard triggers the creation of a *fresh* witness constant, $w_{\text{case}_1, \text{site}}$, which is dynamically added to the domain.

Each such event immediately augments the learning objective with a strong enforcement loss specific to that instance:

$$\begin{aligned} \mathcal{L}_{\text{witness}}(\text{axiom}, \text{case}_1) = \\ 1 - \llbracket \text{hasSite}(\text{case}_1, w_{\text{case}_1, \text{site}}) \wedge \text{Lung}(w_{\text{case}_1, \text{site}}) \rrbracket \end{aligned} \quad (8)$$

Critically, this mechanism renders the **collapse to ambiguity** a prohibitively costly strategy. Since our framework interprets conjunction ($\wedge$) with the Product t-norm, an ambiguous output where both atoms in the conclusion have a truth value of 0.5 would yield a conjunction value of 0.25. With a high-confidence guard ($\llbracket \text{Pneumonia}(\text{case}_1) \rrbracket \approx 1$), the witness loss becomes approximately $1 - 0.25 = 0.75$, a substantial penalty that the optimizer must eliminate. To minimize this loss, the model is mathematically compelled to push the truth values of *both* atoms towards 1, enforcing high-confidence predictions. These dynamically generated witnesses are fully integrated into the learning process. Their vector embeddings are initialized randomly (via Xavier) and become trainable parameters. Once added to Δ_{core} , they are available for sampling across all axioms, ensuring their semantic properties are constrained by the entire KB.

The entire **Core Domain** is thus the dynamically evolving set $\Delta_{\text{core}} = \Delta_H \cup \Delta_A \cup \Delta_G$.

4.3.2 Latent Domain Δ_{latent} : Driving Generalization. While the Core Domain Δ_{core} ensures logical fidelity by grounding the model in specific entities, relying on it alone risks overfitting. The model might memorize the properties of these particular constants without capturing the universal nature

of the logical rules. To address this and ensure robust generalization, we introduce the **Latent Domain** Δ_{latent} , which represents the infinite, continuous embedding space itself, rather than a fixed set of discrete points.

Our strategy for utilizing the latent domain involves a continuous sampling mechanism. In each training step, we dynamically generate a mini-batch of “latent constants” by sampling vectors from a standard prior distribution (e.g., $\mathcal{N}(0, I)$ in $\mathbb{R}^d$). These temporary vectors act as random probes of the learned geometric space, used exclusively to evaluate universal axioms. This process defines our **generalization loss**, which forces the predicate networks to learn decision boundaries that are logically coherent everywhere in the embedding space, not just at the locations of known constants in Δ_{core} . Effectively, this ensures the model learns the general concept of what it means to be a pneumonia case, rather than merely memorizing that a specific entity such as case_1 is one.

4.4 The Hybrid Training Objective

The central learning objective in GUARDNET is to find the optimal parameters θ that maximize the satisfiability of the KB $\mathcal{K}$. We derive our training objective directly from our principled fuzzy semantics defined in §3, constructing a loss function that minimizes the dissatisfaction of each axiom using our selected differentiable operators.

Loss for a Single Axiom Instance. For any universally quantified axiom of the form $\forall \bar{x}(\alpha \rightarrow \psi)$, the loss for a grounded instance is derived using the **Sigmoidal Reichenbach implication** recommended in §3.1. Let $I_R = 1 - \llbracket \alpha \rrbracket + \llbracket \alpha \rrbracket \llbracket \psi \rrbracket$ be the standard Reichenbach implication score. The instance loss is defined as:

$$\begin{aligned} \mathcal{L}_{\text{instance}}(\forall \bar{x}(\alpha \rightarrow \psi); \bar{a}) &= 1 - \llbracket \alpha \rightarrow \psi \rrbracket_{\rho[\bar{x} \mapsto \bar{a}]} \\ &= 1 - \text{scale}(\sigma(s \cdot (I_R(\bar{a}) + b_0))) \end{aligned} \quad (9)$$

where $\sigma(\cdot)$ is the Sigmoid function with steepness s and bias b_0 . Since $I_R \in [0, 1]$, the sigmoid attains only the sub-interval $[\sigma(sb_0), \sigma(s(1+b_0))]$; the map

$$\text{scale}(y) = \frac{y - \sigma(sb_0)}{\sigma(s(1+b_0)) - \sigma(sb_0)}$$

rescales it linearly onto $[0, 1]$, so that the implication remains a well-defined $[0, 1]$ -valued truth function. While this formulation introduces non-linearity to sharpen gradients around the decision boundary, it fundamentally preserves the deductive structure of the underlying Reichenbach logic (I_R). The core term I_R dictates that the penalty is driven by $\llbracket \alpha \rrbracket(1 - \llbracket \psi \rrbracket)$, aligning with the classical inference rules of Modus Ponens and Modus Tollens. By wrapping this logic in the Sigmoidal transformation, we ensure robust, non-vanishing gradients even for satisfied or violated constraints.

Theoretical Total Loss. The total loss aggregates the dissatisfaction over all axioms. Consistent with our choice of the **Generalized Mean** for quantifier aggregation (§3.2), we define the theoretical objective not as a simple linear expectation, but using the p -mean error aggregator $\mathcal{D}_p^\vee$:

$$\mathcal{L}_{\text{theoretical}}(\theta) = \sum_{\phi_i \in \mathcal{K}} \mathcal{D}_p^\vee \left(\left\{ \mathcal{L}_{\text{instance}}(\phi_i; \bar{a}) \mid \bar{a} \in \Delta^{|\text{FV}(\phi_i)|} \right\} \right) \quad (10)$$

This formulation ensures that the model focuses on the most violated instances (behaving like a soft-max) rather than being diluted by the majority of satisfied cases, a critical feature for learning from sparse signals in KBs.

The Practical Hybrid Loss Function. To align with our **Hybrid Domain Strategy**, we realize this objective via stratified sampling. The total loss is a weighted combination:

$$\mathcal{L}_{\text{total}}(\theta) = \lambda \cdot \mathcal{L}_{\text{fidelity}} + (1 - \lambda) \cdot \mathcal{L}_{\text{gen}} + \gamma \cdot \mathcal{L}_{\text{conf}} \quad (11)$$

where $\lambda \in (0, 1]$ balances fidelity and generalization, and γ controls the confidence regularization $\mathcal{L}_{\text{conf}} = \sum_i p_i(1 - p_i)$. The two main loss terms are explicitly defined as follows:

- **Fidelity Loss ($\mathcal{L}_{\text{fidelity}}$):** Computed on the Core Domain Δ_{core} , this term enforces consistency with known facts and dynamic witnesses:

$$\mathcal{L}_{\text{fidelity}} = \mathcal{D}_p^{\forall} (\{\mathcal{L}_{\text{instance}}(\phi; \bar{c}) \mid \bar{c} \sim \Delta_{\text{core}}\}) + \sum \mathcal{L}_{\text{witness}} \quad (12)$$

It acts as “textbook examples”, grounding the model in the given theory and explicitly preventing collapse via the witness term.

- **Generalization Loss ($\mathcal{L}_{\text{gen}}$):** Computed on the Latent Domain Δ_{latent} , this term enforces universal consistency:

$$\mathcal{L}_{\text{gen}} = \mathcal{D}_p^{\forall} (\{\mathcal{L}_{\text{instance}}(\phi; \bar{z}) \mid \bar{z} \sim \mathcal{N}(0, I)\}) \quad (13)$$

It acts as the “final exam”, testing the model with countless unseen inputs to ensure it has learned the universal principles rather than memorizing specific constants.

By jointly optimizing these complementary loss terms, GUARDNET learns a model that is both faithful to the provided data and robustly generalizable.

THEOREM 1 (ZERO-LOSS SEMANTIC CONSISTENCY). *Let $\mathcal{K}$ be a KB consisting of a finite set of GF formulas. If a GUARDNET model trained on $\mathcal{K}$ achieves $\mathcal{L}_{\text{fidelity}}(\theta) = 0$, then the learned fuzzy interpretation $\mathcal{G}_{\theta}$ satisfies all axioms on the Core Domain Δ_{core} . That is, for every axiom $\phi \in \mathcal{K}$ and every grounding $\bar{c} \in \Delta_{\text{core}}^{|\text{FV}(\phi)|}$, we have $\llbracket \phi \rrbracket_{\mathcal{G}_{\theta}, \bar{c}} = 1$.*

Theorem 1 establishes that the loss function and the logical semantics are formally aligned: zero loss implies full satisfaction. This is a consistency guarantee for the training objective, not a claim about the practically relevant nonzero-loss regime. The following corollary addresses the nonzero-loss setting, while the main scalability guarantee is provided by the complexity result in Theorem 2.

PROOF SKETCH OF THEOREM 1. $\mathcal{L}_{\text{fidelity}}$ decomposes into a p -mean error over instance losses plus the witness terms, both non-negative. The generalized mean satisfies a strict zero-property: $\mathcal{D}_p^{\forall}(\mathbf{1}) = 0$ iff $l_i = 0$ for every i . Hence zero fidelity loss forces $\mathcal{L}_{\text{instance}}(\phi; \bar{c}) = 0$ for every axiom ϕ and every grounding $\bar{c} \in \Delta_{\text{core}}$, which by Eq. 9 means $\llbracket \phi \rrbracket_{\bar{c}} = 1$. Since the Sigmoidal Reichenbach implication attains 1 only when the premise is false or the conclusion is true, this coincides with classical satisfaction. The full argument is given in the supplementary material. $\square$

COROLLARY 1 (APPROXIMATE SATISFACTION). *For axiom $\phi_i = \forall \bar{x} (\alpha_i \rightarrow \psi_i)$, define the instance dissatisfaction $z_{i,c} = 1 - \llbracket \alpha_i \rightarrow \psi_i \rrbracket_c$ for $c \in \text{ext}(\alpha_i)$, and let $k_i = |\text{ext}(\alpha_i)| > 0$. If $\mathcal{D}_p^{\forall}(\mathbf{z}_i) \leq \varepsilon_i$ for $p \geq 1$, then:*

$$(i) \quad \frac{1}{k_i} \sum_{c \in \text{ext}(\alpha_i)} z_{i,c} \leq \varepsilon_i \quad (ii) \quad \max_{c \in \text{ext}(\alpha_i)} z_{i,c} \leq \min\{1, k_i^{1/p} \varepsilon_i\}$$

PROOF OF COROLLARY 1. (i) By Jensen’s inequality applied to the convex function $f(x) = x^p$ for $p \geq 1$,

$$\left(\frac{1}{k_i} \sum_c z_{i,c} \right)^p \leq \frac{1}{k_i} \sum_c z_{i,c}^p \leq \varepsilon_i^p.$$

Taking p -th roots yields $\frac{1}{k_i} \sum_c z_{i,c} \leq \varepsilon_i$. The Jensen step is tight when all instance violations are identical.

(ii) Since

$$\max_c z_{i,c}^p \leq \sum_c z_{i,c}^p \leq k_i \varepsilon_i^p,$$

taking p -th roots gives $\max_c z_{i,c} \leq k_i^{1/p} \varepsilon_i$. Combining with $z_{i,c} \leq 1$ yields $\max_c z_{i,c} \leq \min\{1, k_i^{1/p} \varepsilon_i\}$. The uncapped L_p -to- L_∞ conversion is attained when all dissatisfaction is concentrated on a single grounding. $\square$

Axioms with $k_i = 0$ are treated as vacuously satisfied under the masked grounding semantics and contribute zero universal grounding loss. Note that the global fidelity loss $\frac{1}{m} \sum_i \varepsilon_i \leq \varepsilon$ controls the average over axioms, not each ε_i uniformly; uniform per-axiom guarantees require controlling each ε_i separately.

THEOREM 2 (TRAINING COMPLEXITY). *Let $\mathcal{K}$ be a KB with m guarded axioms $\{\forall \bar{x}_i(\alpha_i \rightarrow \psi_i)\}_{i=1}^m$ over a domain of N entities. Let d denote the embedding dimension, L the number of layers in predicate MLPs, B the batch size, and $|\mathcal{P}|$ the number of predicates. Recall that $\text{ext}(\alpha)$ denotes the guard extension (defined in §2) and Δ_{core} the Core Domain (§4.3). A single training iteration of GUARDNET has:*

Time Complexity:

$$O\left(\sum_{i=1}^m |\text{ext}(\alpha_i)| \cdot L \cdot d^2\right) = O(m \cdot |\mathcal{E}|_{\max} \cdot L \cdot d^2) \quad (14)$$

where $|\mathcal{E}|_{\max} = \max_i |\text{ext}(\alpha_i)|$ is the largest guard extension. The d^2 factor arises from matrix operations in the MLP forward pass.

Space Complexity:

$$O(|\Delta_{\text{core}}| \cdot d + |\mathcal{P}| \cdot L \cdot d^2 + B \cdot |\mathcal{E}|_{\max}) \quad (15)$$

where the three terms account for entity embeddings, predicate MLP parameters, and the batch of grounded instances, respectively.

In sparse KBs where $|\mathcal{E}|_{\max} = O(N)$, both complexities are **linear in the domain size N** , circumventing the $O(N^2)$ bottleneck of unguarded approaches.

PROOF SKETCH OF THEOREM 2. Per axiom ϕ_i , guarded grounding performs an index lookup returning $|\text{ext}(\alpha_i)|$ tuples, so no tuple outside the guard extension is ever materialized (Proposition 1). Each retrieved tuple triggers one forward pass through the predicate MLP, costing $O(Ld^2)$, and the truth values are combined by a single p -mean pass, costing $O(|\text{ext}(\alpha_i)|)$. Summing over the m axioms gives the stated time bound. For space, the three resident structures are the embedding table $O(|\Delta_{\text{core}}|d)$, the predicate networks $O(|\mathcal{P}|Ld^2)$, and one mini-batch of groundings $O(B|\mathcal{E}|_{\max})$; none depends on N^2 . The detailed accounting appears in the supplementary material. $\square$

5 Empirical Evaluation

In the previous sections, we theoretically argued that GUARDNET resolves the fundamental conflict between expressivity and scalability. This is achieved through our novel **Guarded Inductive Bias**, which restricts computation to valid topological neighborhoods, and our **Hybrid Domain Strategy**, which ensures robust generalization. This section provides the **empirical substantiation** of these claims. We structure our evaluation to progress from system-level feasibility to task-level performance, and finally to architectural component analysis. Specifically, we design our experiments to answer three core Research Questions (RQs):

- **RQ1 (Scalability & System Efficiency):** Can GUARDNET physically overcome the “tensor broadcasting” bottleneck analyzed in §4.1 and scale to large KBs (e.g., SNOMED CT) where existing neuro-symbolic baselines fail?
- **RQ2 (Comparative Performance):** Once feasibility is established, does GUARDNET outperform state-of-the-art Geometric Embeddings, GNNs, and neuro-symbolic models across diverse reasoning tasks, demonstrating robust generalization capabilities?
- **RQ3 (Ablation & Mechanism Analysis):** What are the sources of these performance gains? We conduct a comprehensive ablation study to verify the impact of the **Hybrid Domain Strategy** in preventing representational collapse, and analyze the sensitivity of learning dynamics to different **Fuzzy Operators**.

Data Translation Strategy. Since no native benchmark datasets exist specifically for the GF, we construct our experimental testbed by converting standard ontologies into GF formulas. This approach is theoretically well-grounded due to the deep syntactic connections among these formalisms. GF was originally introduced in [3] as a natural generalization of modal logic to the polyadic (multi-variable) setting, preserving key properties such as decidability and the finite model property. DLs, in turn, are well-known to be notational variants of modal logics [6]. As a result, standard DL axioms translate directly into GF sentences. This translation ensures that our experiments faithfully evaluate reasoning capabilities on semantically equivalent logical structures.

Evaluation Tasks and Datasets. We evaluate GUARDNET on two reasoning tasks. **TBox Completion** tests concept subsumption: given a test axiom $C \sqsubseteq D$, we rank the true axiom against corruptions generated by replacing C (or D) with all candidate concepts. **ABox Completion** tests link prediction [44]: given a test triple $r(h, t)$, we rank candidates for head prediction $(?, r, t)$ and tail prediction $(h, r, ?)$. Both tasks employ the standard **filtered setting** [11], removing known true facts from the candidate list.

For TBox completion, we use **SNOMED CT** (377K concepts) [47] as a scalability benchmark in medical terminology and **Gene Ontology (GO)** (44K concepts) [4] for its hierarchically rich biological taxonomy. These ontologies are expressed in the DL $\mathcal{EL}^{++}$ [5], whose core axiom patterns (concept inclusion, existential restriction, conjunction, role hierarchy) translate directly into GF sentences. For ABox completion, we use two protein-protein interaction (PPI) datasets (**Yeast PPI**: 6.4K entities, 110K edges, **Human PPI**: 17.6K entities, 75K edges), combining factual interactions from the STRING database [51] with TBox constraints from GO. Following standard conventions, all datasets are randomly split into 80% training, 10% validation, and 10% test sets.

GF Coverage and Non-Guarded Axioms. The only $\mathcal{EL}^{++}$ constructs that fall outside GF are role chains ($R_1 \circ R_2 \sqsubseteq R_3$) and transitivity declarations, both of which require three variables with no single binary guard atom. In practice, these constitute a negligible fraction of our benchmarks: SNOMED CT contains only 4 such axioms out of 391K total TBox axioms, and GO similarly few. All remaining axioms receive full $O(|\mathcal{E}|)$ guarded grounding. For the excluded axioms, GUARDNET applies bounded-depth neighborhood sampling as a fallback, incurring the same $O(N^3)$ worst-case cost that any neuro-symbolic system pays for three-variable formulas. The overall complexity remains dominated by the guarded majority.

Baselines. We benchmark GUARDNET against 15 baselines spanning five paradigms in KBC tasks:

- **Geometric Ontology Embeddings:** EL Embeddings (ELEM) [26], Box²EL [23], and TransBox [65]. As state-of-the-art approaches that are specifically designed to capture the semantics of the DL $\mathcal{EL}^{++}$, these constitute the strongest and most direct baselines for ontology-based reasoning tasks.

- **Graph Neural Networks (GNNs):** CompGCN [58], GRAIL [54], and NBFNet [71]. These represent the leading GNN-based approaches for the KBC task, whose reliance on implicit message propagation provides a crucial contrast to GUARDNET’s explicit, logic-defined neighborhoods.
- **Neuro-Symbolic (FOL-based):** LTN [7], logLTN [8], and Neural LP [64]. Rooted in more general, unguarded FOL, these frameworks highlight the computational complexity that GUARDNET’s principled restriction to GF is designed to overcome.¹
- **Probabilistic Frameworks:** NeurASP [66], DeepProbLog [33], and DeepStochLog [61]. These are prominent probabilistic logic programming systems that integrate neural networks with probabilistic inference, representing an orthogonal approach to neuro-symbolic integration.
- **Standard KG Embedding Models:** TransE [11], RotatE [50], and ComplEx [55]. These widely adopted embedding methods learn logical patterns *implicitly* through geometric formulations, providing a clear baseline to quantify the performance gains from explicit logical reasoning.

Evaluation Metrics. For both completion tasks, we report **Hits@K** ($K \in \{1, 10, 100\}$) and **MRR**. GUARDNET ranks candidates by fuzzy satisfaction degree $\llbracket \alpha \rrbracket \in [0, 1]$ in descending order; geometric ontology embedding baselines rank by ascending distance, and KG embedding models use their respective scoring functions.

Implementation Details. All experiments were run on NVIDIA RTX 4090 GPUs using PyTorch 2.0 with CUDA 11.8. For all models, the AdamW optimizer [32] was used with an early stopping strategy (patience of 15 epochs) based on validation MRR.

For training, we employ **negative sampling** with $\omega = 64$ corrupted samples per positive instance. Within our fuzzy semantics, observed facts $P(a, b)$ are assigned target truth value 1, while corrupted facts $P(a', b)$ or $P(a, b')$ are assigned target truth value 0. This discriminative signal complements the axiom-based losses in Equation 11: while $\mathcal{L}_{\text{fidelity}}$ and $\mathcal{L}_{\text{gen}}$ enforce logical consistency across the embedding space, the negative sampling loss provides explicit supervision for individual facts. We adopt a self-adversarial margin-based loss (margin $\delta = 1.0$) to sharpen the discrimination boundary. L2 regularization ($\lambda_{\text{reg}} = 10^{-5}$) is applied to all embeddings.

All hyperparameters were selected via grid search with sensitivity analysis on validation sets. For baselines, search spaces included learning rates $\in \{10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}\}$, embedding dimensions $d \in \{100, 200, 400\}$, and batch sizes $\in \{256, 512, 1024\}$. For GUARDNET, the optimal configuration is: learning rate 2×10^{-4} , embedding dimension $d = 200$, batch size 512 (SNOMED CT) or 1024 (others), predicate MLPs with two 256-unit ReLU layers. For fuzzy operators, we use steepness $s = 10$ and bias $b_0 = -0.5$ in the Sigmoidal implication, and aggregation exponent $p = 2$ for quantifiers. The hybrid loss weights are $\lambda = 0.7$ (fidelity-generalization balance) and $\gamma = 0.01$ (confidence penalty). These values were determined through systematic sensitivity analysis.

To guarantee reproducibility, all reported metrics are the **mean $\pm$ standard deviation over 5 independent runs** with different random seeds. Paired t-tests between GUARDNET and the strongest baseline confirm statistical significance ($p < 0.01$) across all benchmarks and metrics, except for H@1 on SNOMED CT ($p < 0.05$).

5.1 Scalability and System Efficiency (RQ1)

We begin by addressing the most critical bottleneck in neuro-symbolic reasoning: physical scalability. To answer **RQ1**, we conducted a stress test comparing GUARDNET against representative neuro-symbolic baselines (LTN, logLTN, and Neural LP) on increasingly large subsets of SNOMED CT.

¹Note that to enable these baselines to run on large datasets for Table 2, we applied stochastic mini-batching. As discussed in §5.1, this engineering compromise avoids OOM errors but severely degrades logical reasoning performance.

The results in **Figure 5** offer definitive validation of our linear complexity advantages. Crucially, we must distinguish between computational feasibility and logical validity. While baselines can technically run on smaller-scale ABox tasks (as reported in Table 2) by employing aggressive stochastic mini-batching, this strategy fundamentally breaks the global dependency chains required for logical inference. As shown in Table 2, even with such engineering compromises, neuro-symbolic baselines completely fail (DNF) on large-scale TBox reasoning tasks such as SNOMED CT and GO, while achieving only moderate performance on the smaller PPI datasets (e.g., LTN attains MRR of 0.285 on Yeast PPI, far below GUARDNET’s 0.405). **Figure 5** reveals the root cause: when forced to perform *rigorous* full grounding to preserve logic, these models hit a quadratic memory wall at just 5K entities. In sharp contrast, GUARDNET maintains linear scalability without sacrificing logical fidelity, achieving the unique combination of state-of-the-art accuracy and linear efficiency.

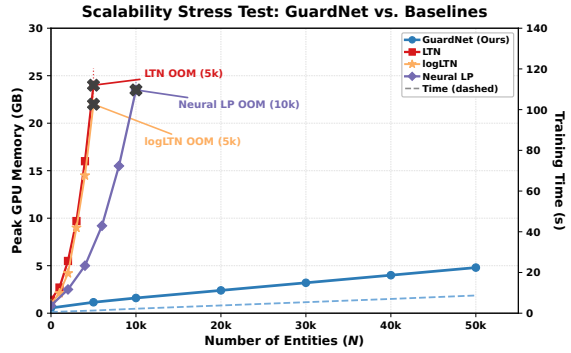


Fig. 5. **Scalability stress test on SNOMED CT subsets.** Peak GPU memory (solid) and training time (dashed) as the number of entities N grows. LTN and logLTN hit the 24 GB limit at 5K entities, NEURAL LP at 10K, while GUARDNET stays linear in the number of edges $|E|$ throughout.

5.2 Comparative Performance (RQ2)

Having established physical feasibility, we evaluate the predictive performance of GUARDNET on full-scale datasets. Table 2 presents the “Grand Slam” benchmark results, comparing our model against 15 state-of-the-art baselines across TBox and ABox tasks.

GUARDNET demonstrates consistent superiority across all datasets and reasoning tasks. On the massive **SNOMED CT** ontology, GUARDNET attains an MRR of 0.125, surpassing the strongest geometric baseline TRANSBox (0.119) and Box²EL (0.114), while the leading GNN baseline NBFNet lags significantly behind (0.055) due to its inability to capture strict logical entailment in deep hierarchies. This advantage extends to the **GO**, where GUARDNET achieves an MRR of 0.133, outperforming TransBox (0.126) by a notable margin, while neuro-symbolic baselines like LTN fail to complete (DNF) on this scale.

The primary reason for this success lies in our framework’s **high logical fidelity**. Unlike geometric models that approximate logical constraints through spatial relationships (e.g., box containment or vector rotation), GUARDNET’s loss function is derived directly from the **GF semantics**. This forces the learned embeddings to maintain strict consistency with the ontological structure, pruning the search space to valid neighborhoods. This is particularly evident in precision-oriented metrics, where GUARDNET excels by avoiding the false positives that plague purely statistical methods.

On ABox reasoning tasks (Yeast and Human PPI), the dramatic performance differential with standard KG embedding models confirms the fundamental importance of incorporating structured logical reasoning. GUARDNET achieves an MRR of 0.405 on Yeast PPI, significantly outperforming the leading GNN baseline NBFNet (0.351) by over 15%. This confirms that the *Guarded Inductive Bias* is not limited to taxonomic reasoning but also generalizes robustly to inductive link prediction, capturing multi-hop logical dependencies that standard message-passing schemes miss.

Table 2. **Overall KBC Performance Comparison.** We evaluate GUARDNET against state-of-the-art baselines across four benchmarks spanning TBox reasoning (concept subsumption) and ABox reasoning (link prediction). GUARDNET (Product t-norm + Sigmoidal Reichenbach implication) achieves the best performance across all metrics. **DNF:** Did Not Finish (OOM/Timeout after 72h).

Model / Configuration	SNOMED CT (377K)				GO (44K)				Yeast PPI (110K edges)				Human PPI (75K edges)			
	TBox Reasoning (Concept Subsumption)				ABox Reasoning (Link Prediction)				ABox Reasoning (Link Prediction)				ABox Reasoning (Link Prediction)			
	H@1	H@10	H@100	MRR	H@1	H@10	H@100	MRR	H@1	H@10	H@100	MRR	H@1	H@10	H@100	MRR
GUARDNET (Prod+Sig)	5.1±1	28.3±4	70.5±3	.125±0	5.5±2	29.8±3	73.4±2	.133±0	30.5±4	60.2±5	91.1±3	.405±0	29.2±5	57.9±6	88.9±4	.388±0
Geometric Ontology Embedding Models																
ELEM	2.1±1	20.1±6	38.9±5	.078±0	2.4±1	23.8±5	43.4±4	.089±0	20.5±6	45.0±8	74.8±6	.301±0	18.2±8	40.1±11	69.7±8	.268±0
Box ² EL	3.4±2	25.5±6	68.1±5	.114±0	3.9±2	26.5±5	70.9±4	.120±0	27.1±5	55.1±6	86.9±5	.368±0	25.5±6	52.3±8	82.7±6	.346±0
TransBox	3.6±2	26.8±5	69.2±4	.119±0	4.1±2	27.8±4	71.8±3	.126±0	28.5±5	57.5±6	88.5±4	.385±0	26.8±6	54.8±7	85.2±5	.365±0
Graph Neural Networks																
CompGCN	0.8±1	8.1±5	20.3±4	.041±0	1.2±1	9.9±4	24.1±3	.052±0	23.5±6	50.5±7	78.3±5	.328±0	21.5±7	45.2±8	72.1±6	.302±0
GRAIL	1.0±1	9.8±5	24.0±5	.050±0	1.5±1	11.9±5	28.1±4	.063±0	25.8±6	52.6±7	80.2±5	.345±0	23.8±7	48.0±8	75.5±6	.321±0
NBFNet	1.2±1	10.5±4	25.8±4	.055±0	1.8±1	13.0±4	30.2±3	.071±0	26.5±5	53.5±6	81.1±5	.351±0	24.5±6	49.1±7	76.9±6	.332±0
Expressive Neuro-Symbolic Models (Unguarded, FOL-based)																
LTN (Prod+Sig)		DNF				DNF			17.5±5	42.5±9	70.1±7	.285±0	16.0±6	39.8±11	65.4±8	.265±0
logLTN (Prod+Sig)		DNF				DNF			19.5±6	45.1±11	72.7±9	.302±0	18.5±7	42.0±13	68.2±9	.284±0
Neural LP		DNF				DNF			18.0±5	43.3±9	71.1±8	.291±0	16.5±6	40.5±12	66.8±9	.272±0
Probabilistic Frameworks																
NeurASP		DNF				DNF			14.1±3	35.5±6	60.4±5	.225±0	12.8±4	32.8±7	58.5±6	.208±0
DeepProbLog		DNF				DNF			12.8±2	32.2±5	58.1±4	.208±0	11.5±3	29.5±6	55.8±5	.192±0
DeepStochLog		DNF				DNF			16.5±4	38.8±7	65.2±6	.255±0	15.1±5	36.2±8	62.9±7	.235±0
Standard KG Embedding Models																
TransE	0.2±1	2.1±3	8.7±2	.018±0	0.3±1	2.8±3	12.3±2	.023±0	15.5±4	38.0±7	64.9±5	.238±0	14.0±5	35.0±8	60.7±6	.219±0
RotatE	0.3±1	3.2±4	11.8±3	.025±0	0.4±1	4.1±4	16.7±3	.032±0	18.5±5	42.6±8	68.1±6	.268±0	16.0±6	38.3±9	64.3±7	.245±0
CompLex	0.3±1	2.6±3	10.1±3	.022±0	0.4±1	3.5±3	14.9±3	.028±0	17.0±5	40.5±8	66.4±6	.252±0	15.5±6	36.1±9	62.8±7	.232±0

5.3 Ablation and Mechanism Analysis (RQ3)

To isolate the source of the performance gains, we perform a comprehensive component analysis and investigate the model’s sensitivity to operator choices.

Hybrid Domain Strategy. Table 3 reveals that the **Hybrid Domain Strategy** is critical for preventing representational collapse.

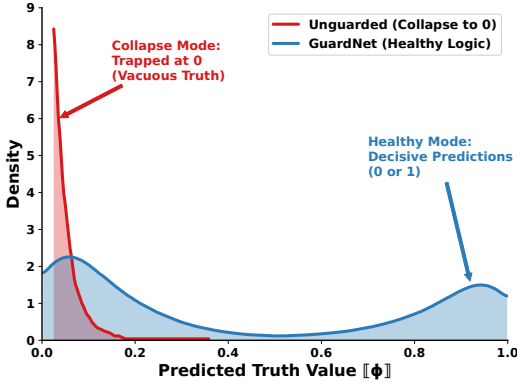
- **Preventing Emptiness:** The base model (Row 2) relying solely on the Herbrand domain (Δ_H) suffers from an 82% collapse rate on SNOMED CT, as the model learns to predict logical falsity to vacuously satisfy axioms. Introducing **Atom Witnesses** (Δ_A , Row 3) provides the necessary static anchors, drastically reducing collapse to 12% and nearly doubling the MRR.
- **Enforcing Decisiveness:** The addition of **Guarded Witnesses** (Δ_G , Row 4) further refines the decision boundaries, handling dynamic triggering of rules. Finally, the full model with **Latent Generalization** (Δ_{latent} , Row 5) eliminates collapse almost entirely (<1%), proving that continuous sampling is essential for learning universal logical principles.
- **Role of Learnable Embeddings:** Row 6 isolates the contribution of joint embedding optimization by freezing TransE-pretrained embeddings. The Hybrid Domain Strategy remains fully active, so collapse rates stay below 1%, confirming that collapse prevention is independent of embedding trainability. However, MRR drops by $\approx 20\%$ on TBox tasks, where pretrained relational embeddings lack alignment with concept-hierarchy axioms and the frozen geometry cannot adapt. On PPI tasks, where the pretrained structure already encodes relational patterns, the drop is only $\approx 8\%$. This asymmetry demonstrates that logic-driven embedding fine-tuning is essential for concept-level reasoning but less critical for relation-level link prediction.

Mechanism Analysis: Visualizing Collapse. To visually confirm the elimination of representational collapse, we plotted the density distribution of predicted truth values on the test set in **Figure 6a**.

Table 3. **Comprehensive Component Analysis (RQ3)**. We systematically dissect GUARDNET’s performance gains. **Col.** (Collapse Rate): % of predictions stuck in trivial states. Low is good. **RSR** (Rule Satisfaction Rate): Average truth value of axioms. High is good. **FPE** (Row 6): Fixed Pretrained Embedding (TransE-initialized, frozen during training). **Note**: On PPI (Graph tasks), **Collapse Rate** is consistently low (<1%) for all GUARDNET variants due to explicit positive edges, but **RSR** steadily improves, showing that the model learns to satisfy logical rules more strictly. Percentage improvements are computed relative to Row 2 (GUARDNET Base) for TBox tasks and Row 1 (LTN Optimized) for ABox tasks.

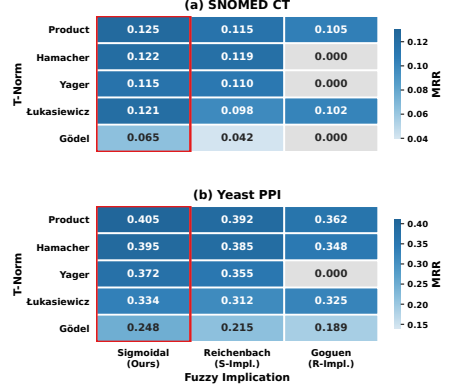
Model Configuration	Domain	SNOMED CT						GO						Yeast PPI					Human PPI				
		H@1	H@10	H@100	MRR	Col.	RSR	H@1	H@10	H@100	MRR	Col.	RSR	H@1	H@10	H@100	MRR	RSR	H@1	H@10	H@100	MRR	RSR
1. LTN (Optimized) (Product + Sigmoidal)	Unguarded (Δ_H)	DNF (OOM)						DNF (OOM)						17.5	42.5	70.1	.285	.82	16.0	39.8	65.4	.265	.80
2. GUARDNET (Base) (Product + Sigmoidal)	Guarded (Δ_H)	1.2	11.5	24.8	.055	82%	.65	1.4	12.8	26.5	.062	79%	.68	22.1	49.8	80.5	.335	.91	20.5	46.5	75.2	.312	.89
		(High Collapse due to Emptiness)						(High Collapse due to Emptiness)						+26%	+17%	+14%	+17%	+11%	+28%	+16%	+15%	+17%	+11%
3. + Atom Witness (+ Static Anchors)	Hybrid ($\cup \Delta_A$)	3.8	23.5	61.5	.105	12%	.92	4.2	25.8	65.4	.115	10%	.93	27.5	55.2	86.8	.375	.96	26.2	53.5	82.5	.352	.94
		+216%	+104%	+148%	+91%	↓70pt	+27	+200%	+101%	+146%	+85%	↓69pt	+25	+24%	+10%	+7.8%	+12%	+05	+27%	+15%	+9.7%	+13%	+05
4. + Guarded Witness (+ Dynamic Anchors)	Hybrid ($\cup \Delta_G$)	4.8	26.8	68.2	.118	5%	.96	5.1	28.2	71.5	.126	4%	.97	29.8	59.1	90.5	.396	.98	28.5	57.0	87.5	.378	.96
		+26%	+14%	+11%	+12%	↓7pt	+04	+21%	+9.3%	+9.3%	+9.6%	↓6pt	+04	+8.3%	+7.1%	+4.2%	+5.6%	+02	+8.7%	+6.5%	+6.0%	+7.3%	+02
5. Full GUARDNET (+ Latent Generalization)	Hybrid ($\cup \Delta_{\text{latent}}$)	5.1	28.3	70.5	.125	<1%	.99	5.5	29.8	73.4	.133	<1%	.99	30.5	60.2	91.1	.405	.99	29.2	57.9	88.9	.388	.98
		+6.2%	+5.6%	+3.4%	+5.9%	↓4pt	+03	+7.8%	+5.7%	+2.7%	+5.6%	↓3pt	+02	+2.3%	+1.9%	+0.7%	+2.3%	+01	+2.4%	+1.6%	+1.6%	+2.6%	+02
6. FPE (TransE, frozen)	Hybrid ($\Delta_{\text{core}} \cup \Delta_{\text{latent}}$)	4.0	22.5	57.8	.100	<1%	.95	4.4	24.2	60.8	.109	<1%	.95	27.8	55.8	87.2	.372	.97	26.8	53.8	84.5	.357	.96
		-22%	-20%	-18%	-20%	·	-.04	-20%	-19%	-17%	-18%	·	-.04	-8.9%	-7.3%	-4.3%	-8.1%	-.02	-8.2%	-7.1%	-5.0%	-8.0%	-.02

Mechanism Analysis: Avoiding Representational Collapse



(a) Truth-value distribution.

Sensitivity Analysis: T-Norm × Implication Operator



(b) Operator heatmap.

Fig. 6. **Mechanism analysis on Yeast PPI.** (a) The ungarded model (red) collapses to vacuous truths (trapped at 0), whereas GUARDNET (blue) maintains a healthy bimodal distribution with decisive predictions near 0 or 1. (b) The heatmap isolates the impact of fuzzy operators on MRR: the choice of T-norm (rows) has marginal variance, while the choice of **Implication** (columns) is decisive.

The **Unguarded** baseline (red curve) exhibits a pathological distribution heavily skewed towards 0. This empirically validates our theoretical concern that without proper grounding, the model minimizes loss by learning “vacuous truths” (i.e., predicting all atoms as False to trivially satisfy implications like $\text{False} \rightarrow \dots$). In contrast, GUARDNET (blue curve) displays a healthy **bimodal distribution** with distinct peaks near 0 (False) and 1 (True). This indicates that our Hybrid Domain

Strategy successfully forces the model to make decisive, semantically meaningful predictions, avoiding the collapse modes that plague standard neuro-symbolic approaches.

Sensitivity to Fuzzy Operators. Finally, **Figure 6b** highlights the pivotal role of the **Sigmoidal Implication**. While the choice of T-norm (rows) shows marginal variance, the implication operator (columns) is decisive. Standard R-implications (Goguen) frequently lead to vanishing gradients (“Dead Zones”), causing models to fail (DNF) or underperform on large datasets. The Sigmoidal Implication (Red Box) consistently resolves this, providing the robust gradient signal required for deep reasoning architectures to converge.

Sensitivity of Sigmoidal Parameters. Figure 7 analyzes the sensitivity to steepness s and bias b_0 of the Sigmoidal Reichenbach implication. Panel (a) shows that performance peaks at $(s=10, b_0=-0.5)$ with a broad plateau across $s \in [5, 20]$ and $b_0 \in [-0.75, -0.25]$, confirming robustness. Panel (b) reveals the interplay between s and training dynamics at fixed $b_0=-0.5$: as s increases from 1 to 10, MRR improves and collapse rate drops below 1%, while $|\Delta_G|$ decreases from $\approx 22.8k$ to $\approx 12.4k$ as sharper gradients enable more precise witness creation. Beyond $s=20$, the sigmoid approaches a step function and gradients vanish away from the boundary, degrading both MRR and collapse rate.

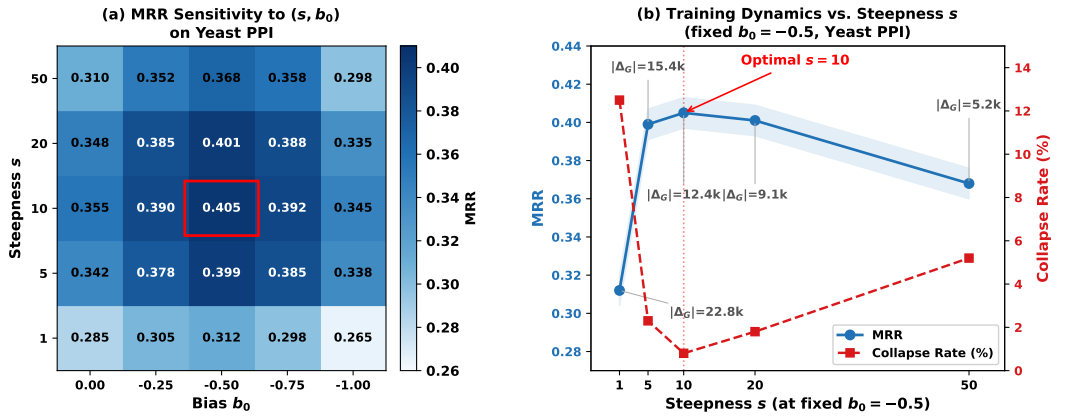


Fig. 7. **Sensitivity of Sigmoidal Implication Parameters on Yeast PPI.** (a) MRR heatmap across the grid of steepness s and bias b_0 ; the red box marks the optimum ($s=10, b_0=-0.5$). (b) MRR and Collapse Rate vs. steepness s at fixed $b_0=-0.5$, annotated with the size of the guarded instance witness set $|\Delta_G|$.

Table 4. Core domain composition at training convergence. $|\Delta_H|$ is the Herbrand Universe (constants in the KB), $|\mathcal{P}|$ the number of predicates, $|\Delta_A|$ the atom witnesses (bounded by $2|\mathcal{P}|$), and $|\Delta_G|$ the dynamically created guarded instance witnesses. The last column reports the growth of Δ_G relative to Δ_H .

Dataset	$ \Delta_H $	$ \mathcal{P} $	$ \Delta_A $	$ \Delta_G $	$ \Delta_G / \Delta_H $
Yeast PPI	6,438	8	16	$\approx 12,400$	193%
Human PPI	17,620	12	24	$\approx 22,800$	129%

Core Domain Growth. Table 4 reports the final size of the Guarded Instance Witness Set $|\Delta_G|$ at training convergence. While Δ_H (Herbrand constants) and Δ_A (atom witnesses, bounded by $2|\mathcal{P}|$) are fixed before training, Δ_G grows dynamically as the model’s confidence in guard predicates

increases. The growth is convergence-coupled: new witnesses require a guard premise above the confidence threshold, so creation decelerates as the fidelity loss converges.

The ratio $|\Delta_G|/|\Delta_H|$ indicates how many additional witness constants the model created relative to the original Herbrand domain. The higher ratio for Yeast PPI reflects its denser topology (average degree 34.3 vs. 8.5 for Human PPI), which triggers more guard predicates above the confidence threshold. Despite this growth, the resulting embedding table remains modest: $|\Delta_{\text{core}}| \cdot d = (|\Delta_H| + |\Delta_A| + |\Delta_G|) \cdot 200$ parameters, i.e., 15.1 MB for Yeast PPI and 32.4 MB for Human PPI in float32.

6 Conclusion and Future Work

In this work, we address the longstanding dichotomy between logical expressivity and computational scalability in neuro-symbolic learning. By reinterpreting the **Guarded Fragment** of FOL not as a restriction, but as a powerful topological inductive bias, we fundamentally restructured the computational graph of reasoning: transforming intractable global search into efficient, neighborhood-constrained lookups. This approach neuralizes the tractability mechanism of GTGDs from database theory, establishing a formal bridge between classical query answering and differentiable neural reasoning. GUARDNET demonstrates that rigorous logical constraints need not be computational bottlenecks; on the contrary, the “guard” acts as a topological index that guides neural optimization, enabling linear scalability ($O(|E|)$) on massive ontologies like SNOMED CT where prior methods fail. Furthermore, our Hybrid Domain Strategy successfully reconciles data fidelity with generalization, preventing representational collapse.

For future work, we envision two strategic directions:

- **In-Database Reasoning:** Integrating GUARDNET’s sparse grounding kernels directly into graph database engines (e.g., logic-aware stored procedures) to support real-time, zero-shot logical query answering without external model export.
- **Beyond Static Graphs:** Extending the guarded mechanism to *Temporal Guarded Logic* to handle dynamic KGs where validity intervals act as temporal guards.

Acknowledgments

We thank the reviewers for their useful comments. This work was supported by the National Natural Science Foundation of China (No. 62476125), the Noncommunicable Chronic Diseases–National Science and Technology Major Project (No. 2025ZD0550704), the Fundamental Research Funds for the Central Universities (No. 0221-14380025), the “111 Center” (No. B26023), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM118), and the Nanjing University International Collaboration Initiative.

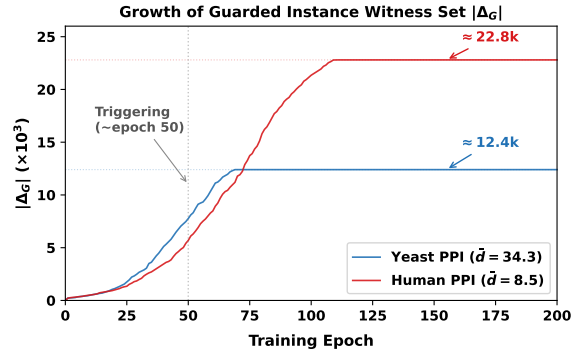


Fig. 8. Growth of the guarded instance witness set Δ_G during training. Both curves exhibit logistic growth consistent with the convergence-coupled creation mechanism: witnesses accumulate rapidly once guard confidence rises (~epoch 50), then plateau as fidelity loss converges. Yeast PPI (denser, average degree $\bar{d}=34.3$) triggers witnesses earlier but saturates at a lower count; Human PPI (sparser, $\bar{d}=8.5$) grows later but reaches a higher plateau due to its larger entity set.

References

- [1] Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. 2020. BoxE: A Box Embedding Model for Knowledge Base Completion. In *Advances in Neural Information Processing Systems 33 (NeurIPS'20)*.
- [2] Farahnaz Akrami, Mohammed Samiul Saeef, Qingheng Zhang, Wei Hu, and Chengkai Li. 2020. Realistic Re-evaluation of Knowledge Graph Completion Methods: An Experimental Study. In *Proc. SIGMOD'20*. ACM, 1995–2010.
- [3] Hajnal Andr  ka, Istv  n N  meti, and Johan van Benthem. 1998. Modal Languages and Bounded Fragments of Predicate Logic. *J. Philos. Log.* 27, 3 (1998), 217–274.
- [4] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, Midori A Harris, David P Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C Matese, Joel E Richardson, Martin Ringwald, Gerald M Rubin, and Gavin Sherlock. 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25, 1 (2000), 25–29.
- [5] Franz Baader, Sebastian Brandt, and Carsten Lutz. 2005. Pushing the $\mathcal{EL}$ Envelope. In *Proc. IJCAI'05*. Professional Book Center, 364–369.
- [6] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. 2017. *An Introduction to Description Logic*. Cambridge University Press.
- [7] Samy Badreddine, Artur S. d'Avila Garcez, Luciano Serafini, and Michael Spranger. 2022. Logic Tensor Networks. *Artif. Intell.* 303 (2022), 103649.
- [8] Samy Badreddine, Luciano Serafini, and Michael Spranger. 2023. logLTN: Differentiable Fuzzy Logic in the Logarithm Space. *CoRR* abs/2306.14546 (2023). <https://doi.org/10.48550/arXiv.2306.14546>
- [9] Luigi Bellomarini, Emanuel Sallinger, and Georg Gottlob. 2018. The Vadalog System: Datalog-based Reasoning for Knowledge Graphs. *Proc. VLDB Endow.* 11, 9 (2018), 975–987.
- [10] Tarek R. Besold, Artur S. d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro M. Domingos, Pascal Hitzler, Kai-Uwe K  hnberger, Luis C. Lamb, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. 2021. Neural-Symbolic Learning and Reasoning: A Survey and Interpretation. In *Neuro-Symbolic Artificial Intelligence: The State of the Art*. Frontiers in Artificial Intelligence and Applications, Vol. 342. IOS Press, 1–51.
- [11] Antoine Bordes, Nicolas Usunier, Alberto Garc  a-Dur  n, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26 (NeurIPS'13)*. 2787–2795.
- [12] Andrea Cali, Georg Gottlob, and Michael Kifer. 2013. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. *J. Artif. Intell. Res.* 48 (2013), 115–174.
- [13] Alonzo Church. 1936. An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics* 58 (1936), 345.
- [14] William W. Cohen, Fan Yang, and Kathryn Mazaitis. 2020. TensorLog: A Probabilistic Database Implemented Using Deep-Learning Infrastructure. *J. Artif. Intell. Res.* 67 (2020), 285–325.
- [15] Artur d'Avila Garcez and Luis C. Lamb. 2023. Neurosymbolic AI: the 3rd wave. *Artif. Intell. Rev.* 56, 11 (2023), 12387–12406.
- [16] Artur S. d'Avila Garcez, Krysia Broda, and Dov M. Gabbay. 2002. *Neural-symbolic learning systems - foundations and applications*. Springer.
- [17] Michelangelo Diligenti, Marco Gori, and Claudio Sacc  . 2017. Semantic-based regularization for learning and inference. *Artif. Intell.* 244 (2017), 143–165.
- [18] Ian J. Goodfellow, Yoshua Bengio, and Aaron C. Courville. 2016. *Deep Learning*. MIT Press.
- [19] Erich Gr  del. 1999. On The Restraining Power of Guards. *J. Symb. Log.* 64, 4 (1999), 1719–1742.
- [20] V  ctor Guti  rrez-Basulto and Steven Schockaert. 2018. From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. In *Proc. KR'18*. AAAI Press, 379–388.
- [21] William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. 2018. Embedding Logical Queries on Knowledge Graphs. In *Advances in Neural Information Processing Systems 31 (NeurIPS'18)*. 2030–2041.
- [22] Ihab F. Ilyas, Theodoros Rekatsinas, Vishnu Konda, Jeffrey Pound, Xiaoguang Qi, and Mohamed A. Soliman. 2022. Saga: A Platform for Continuous Construction and Serving of Knowledge at Scale. In *Proc. SIGMOD'22*. ACM, 2259–2272.
- [23] Mathias Jackermeier, Jiaoyan Chen, and Ian Horrocks. 2024. Dual Box Embeddings for the Description Logic $\mathcal{EL}^+ + \perp$. In *Proc. WWW'24*. ACM, 2250–2258.
- [24] Erich-Peter Klement, Radko Mesiar, and Endre Pap. 2000. *Triangular Norms*. Trends in Logic, Vol. 8. Springer.
- [25] Adrian Kochsiek and Rainer Gemulla. 2021. Parallel Training of Knowledge Graph Embedding Models: A Comparison of Techniques. *Proc. VLDB Endow.* 15, 3 (2021), 633–645.
- [26] Maxat Kulmanov, Wang Liu-Wei, Yuan Yan, and Robert Hoehndorf. 2019. EL Embeddings: Geometric Construction of Models for the Description Logic $\mathcal{EL}^+ + \perp$. In *Proc. IJCAI'19*. ijcai.org, 6103–6109.
- [27] Brenden M. Lake and Marco Baroni. 2018. Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks. In *Proc. ICML'18 (Proceedings of Machine Learning Research, Vol. 80)*. PMLR,

- 2879–2888.
- [28] Na Li, Thomas Bailleux, Zied Bouraoui, and Steven Schockaert. 2024. Ontology Completion with Natural Language Inference and Concept Embeddings: An Analysis. *CoRR* abs/2403.17216 (2024).
 - [29] Na Li, Zied Bouraoui, and Steven Schockaert. 2019. Ontology Completion Using Graph Convolutional Networks. In *Proc. ISWC'19 (LNCS, Vol. 11778)*. Springer, 435–452.
 - [30] Xueling Lin, Haoyang Li, Hao Xin, Zijian Li, and Lei Chen. 2020. KBPearl: A Knowledge Base Population System Supported by Joint Entity and Relation Linking. *Proc. VLDB Endow.* 13, 7 (2020), 1035–1049.
 - [31] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proc. AAAI'15*. AAAI Press, 2181–2187.
 - [32] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *Proc. ICLR'19*. OpenReview.net.
 - [33] Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. 2018. DeepProbLog: Neural Probabilistic Logic Programming. In *Advances in Neural Information Processing Systems 31 (NeurIPS'18)*. 3753–3763.
 - [34] Emanuele Marconato, Stefano Teso, Antonio Vergari, and Andrea Passerini. 2023. Not All Neuro-Symbolic Concepts Are Created Equal: Analysis and Mitigation of Reasoning Shortcuts. In *Advances in Neural Information Processing Systems 36 (NeurIPS'23)*.
 - [35] Gary Marcus. 2018. Deep Learning: A Critical Appraisal. *CoRR* abs/1801.00631 (2018).
 - [36] Gary Marcus. 2020. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. *CoRR* abs/2002.06177 (2020).
 - [37] Giuseppe Marra, Francesco Giannini, Michelangelo Diligenti, and Marco Gori. 2019. Integrating Learning and Reasoning with Deep Logic Models. In *Proc. ECML-PKDD'19 (LNCS, Vol. 11907)*. Springer, 517–532.
 - [38] Connor Pryor, Charles Dickens, Eriq Augustine, Alon Albalak, William Yang Wang, and Lise Getoor. 2023. NeuPSL: Neural Probabilistic Soft Logic. In *Proc. IJCAI'23*. ijcai.org, 4145–4153.
 - [39] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. 2007. ProbLog: A Probabilistic Prolog and Its Application in Link Discovery. In *Proc. IJCAI'07*. 2462–2467.
 - [40] Hongyu Ren, Hanjun Dai, Bo Dai, Xinyun Chen, Michihiro Yasunaga, Haitian Sun, Dale Schuurmans, Jure Leskovec, and Denny Zhou. 2021. LEGO: Latent Execution-Guided Reasoning for Multi-Hop Question Answering on Knowledge Graphs. In *Proc. ICML'21 (Proceedings of Machine Learning Research, Vol. 139)*. PMLR, 8959–8970.
 - [41] Hongyu Ren, Hanjun Dai, Bo Dai, Xinyun Chen, Denny Zhou, Jure Leskovec, and Dale Schuurmans. 2022. SMORE: Knowledge Graph Completion and Multi-hop Reasoning in Massive Knowledge Graphs. In *Proc. KDD'22*. ACM, 1472–1482.
 - [42] John Alan Robinson and Andrei Voronkov (Eds.). 2001. *Handbook of Automated Reasoning (in 2 volumes)*. Elsevier and MIT Press.
 - [43] Tim Rocktäschel and Sebastian Riedel. 2017. End-to-end Differentiable Proving. In *Advances in Neural Information Processing Systems 30 (NeurIPS'17)*. 3788–3800.
 - [44] Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. 2022. Explaining Link Prediction Systems based on Knowledge Graph Embeddings. In *Proc. SIGMOD'22*. ACM, 2062–2075.
 - [45] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning representations by back-propagating errors. *Nature* 323, 6088 (1986), 533–536.
 - [46] Manfred Schmidt-Schauß and Gert Smolka. 1991. Attributive Concept Descriptions with Complements. *Artif. Intell.* 48, 1 (1991), 1–26.
 - [47] Kent A. Spackman, Keith E. Campbell, and Roger A. Côté. 1997. SNOMED RT: a reference terminology for health care. In *Proc. AMIA'97*. AMIA.
 - [48] Umberto Straccia. 2001. Reasoning within Fuzzy Description Logics. *J. Artif. Intell. Res.* 14 (2001), 137–166.
 - [49] Ye Sun, Lei Shi, and Yongxin Tong. 2025. eXpath: Explaining Knowledge Graph Link Prediction with Ontological Closed Path Rules. *Proc. VLDB Endow.* 18, 9 (2025), 2818–2830.
 - [50] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *Proc. ICLR'19*. OpenReview.net.
 - [51] Damian Szklarczyk, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T. Doncheva, John H. Morris, Peer Bork, Lars Juhl Jensen, and Christian von Mering. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, Database-Issue (2019), D607–D613.
 - [52] Zhenwei Tang, Tilman Hinnerichs, Xi Peng, Xiangliang Zhang, and Robert Hoehndorf. 2022. F $\mathcal{AL}$ CON: Sound and Complete Neural Semantic Entailment over $\mathcal{ALC}$ Ontologies. *CoRR* abs/2208.07628 (2022).
 - [53] Zhenwei Tang, Shichao Pei, Xi Peng, Fuzhen Zhuang, Xiangliang Zhang, and Robert Hoehndorf. 2022. Joint Abductive and Inductive Neural Logical Reasoning. *CoRR* abs/2205.14591 (2022).

- [54] Komal K. Teru, Etienne G. Denis, and William L. Hamilton. 2020. Inductive Relation Prediction by Subgraph Reasoning. In *Proc. ICML'20 (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 9448–9457.
- [55] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proc. ICML'16 (JMLR Workshop and Conference Proceedings, Vol. 48)*. JMLR.org, 2071–2080.
- [56] Alan M. Turing. 1937. On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* s2-42, 1 (1937), 230–265.
- [57] Emile van Krieken, Erman Acar, and Frank van Harmelen. 2022. Analyzing Differentiable Fuzzy Logic Operators. *Artif. Intell.* 302 (2022), 103602.
- [58] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. 2020. Composition-based Multi-Relational Graph Convolutional Networks. In *Proc. ICLR'20*. OpenReview.net.
- [59] Kai Wang, Yuwei Xu, and Siqiang Luo. 2024. TIGER: Training Inductive Graph Neural Network for Large-scale Knowledge Graph Reasoning. *Proc. VLDB Endow.* 17, 10 (2024), 2459–2472.
- [60] Po-Wei Wang, Priya L. Donti, Bryan Wilder, and J. Zico Kolter. 2019. SATNet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *Proc. ICML'19 (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 6545–6554.
- [61] Thomas Winters, Giuseppe Marra, Robin Manhaeve, and Luc De Raedt. 2022. DeepStochLog: Neural Stochastic Logic Programming. In *Proc. AAAI'22*. AAAI Press, 10090–10100.
- [62] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Van den Broeck. 2018. A Semantic Loss Function for Deep Learning with Symbolic Knowledge. In *Proc. ICML'18 (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, 5498–5507.
- [63] Zezhong Xu, Yincen Qu, Wen Zhang, Lei Liang, and Huajun Chen. 2024. InBox: Recommendation with Knowledge Graph using Interest Box Embedding. *Proc. VLDB Endow.* 17, 13 (2024), 4641–4654.
- [64] Fan Yang, Zhilin Yang, and William W. Cohen. 2017. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. In *Advances in Neural Information Processing Systems 30 (NeurIPS'17)*. 2319–2328.
- [65] Hui Yang, Jiaoyan Chen, and Uli Sattler. 2025. TransBox: $\mathcal{EL}^{++}$ -closed Ontology Embedding. In *Proc. WWW'25*. ACM, 22–34.
- [66] Zhun Yang, Adam Ishay, and Joohyung Lee. 2020. NeurASP: Embracing Neural Networks into Answer Set Programming. In *Proc. IJCAI'20*. ijcai.org, 1755–1762.
- [67] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. In *Proc. AAAI'20*. AAAI Press, 3065–3072.
- [68] Zhanqiu Zhang, Jie Wang, Jiajun Chen, Shuiwang Ji, and Feng Wu. 2021. ConE: Cone Embeddings for Multi-Hop Reasoning over Knowledge Graphs. In *Advances in Neural Information Processing Systems 34 (NeurIPS'21)*. 19172–19183.
- [69] Zhanqiu Zhang, Jie Wang, Jieping Ye, and Feng Wu. 2022. Rethinking Graph Convolutional Networks in Knowledge Graph Completion. In *Proc. WWW'22*. ACM, 798–807.
- [70] Fernando Zhapa-Camacho and Robert Hoehndorf. 2024. Lattice-Preserving $\mathcal{ALC}$ Ontology Embeddings. In *Proc. NeSy'24 (LNCS, Vol. 14979)*. Springer, 355–369.
- [71] Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. 2021. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. In *Advances in Neural Information Processing Systems 34 (NeurIPS'21)*. 29476–29490.

Received 17 January 2026; revised 19 May 2026; accepted 12 June 2026