

Differentiable Modal Logic for Multi-Agent Diagnosis, Orchestration and Communication

Antonin Sulc^{*}

*Lawrence Berkeley National Lab
Berkeley, U.S.A.*

ASULC@LBL.GOV

Abstract

As multi-agent AI systems evolve from simple chatbots to autonomous swarms, debugging semantic failures requires reasoning about knowledge, belief, causality, and obligation, precisely what modal logic was designed to formalize. However, traditional modal logic requires manual specification of relationship structures that are unknown or dynamic in real systems. This tutorial demonstrates differentiable modal logic (DML), implemented via Modal Logical Neural Networks (MLNNs), enabling systems to learn trust networks, causal chains, and regulatory boundaries from behavioral data alone.

We present a unified neurosymbolic debugging framework through four modalities: epistemic (who to trust), temporal (when events cause failures), deontic (what actions are permitted), and doxastic (how to interpret agent confidence). Each modality is demonstrated on concrete multi-agent scenarios, from discovering deceptive alliances in diplomacy games to detecting LLM hallucinations, with complete implementations showing how logical contradictions become learnable optimization objectives. Key contributions for the neurosymbolic community: (1) interpretable learned structures where trust and causality are explicit parameters, not opaque embeddings; (2) knowledge injection via differentiable axioms that guide learning with sparse data (3) compositional multi-modal reasoning that combines epistemic, temporal, and deontic constraints; and (4) practical deployment patterns for monitoring, active control and communication of multi-agent systems. All code provided as executable Jupyter notebooks.

1. Introduction

Imagine monitoring a system where twenty AI agents coordinate to complete complex tasks. Your dashboard shows green lights everywhere: messages are flowing, latencies are low, tasks are completing. Yet somehow, the system’s overall performance is degrading. Users are complaining. Something is wrong, but your metrics cannot tell you *what*.

This scenario often arises in production multi-agent systems. The fundamental problem is that traditional monitoring tracks *behavior* without capturing *semantics*.

It can tell you that Agent A sent a message to Agent B, but not whether B understood it. It can tell you that Agent C completed a task, but not whether C kept the promise it made to Agent D. It can detect that a group of agents is behaving differently, but not whether they have formed an implicit alliance or are experiencing a shared failure mode.

^{*} Code is available at https://github.com/sulcantonin/MLNN_public and on PyPI as torchmodal.

1.1. The Semantic Gap in Multi-Agent Monitoring

Consider three scenarios with identical telemetry: (A) Alice and Bob successfully coordinate a task; (B) Bob defects despite understanding the agreement; (C) both execute incorrectly due to mismatched interpretations of "urgent." All produce identical logs: high message frequency, low latency, 100% completion, until failure reveals the difference. Yet each requires different interventions: nothing, trust recalibration, or protocol clarification. A monitoring system that cannot distinguish these cases cannot guide appropriate responses.

1.2. Modal Logic: The Right Mathematical Vocabulary

The key insight of this tutorial is that the concepts needed to distinguish these scenarios, knowledge, belief, trust, obligation, causality, are precisely what *modal logic* was designed to formalize. Developed by philosophers to reason about necessity and possibility, modal logic extends classical logic with operators that qualify truth relative to *contexts*: what an agent knows, what time it is, what norms apply.

Key Insight

Classical logic asks: "Is ϕ true?" Modal logic asks: "Is ϕ true *according to whom? At what time? Under what obligations?*"

This contextual reasoning is exactly what multi-agent debugging requires. The challenge is that traditional modal logic requires rigid, manual specification of the underlying model structure, who trusts whom, how time flows, which norms apply. This is impractical for complex systems where these relationships are unknown or change dynamically.

Differentiable modal logic (DML) addresses this by making the underlying Kripke structure learnable from data. Modal Logical Neural Networks (MLNNs) (Sulc, 2025) is one concrete framework that can do that: MLNN parameterizes accessibility relations and optimize them by minimizing logical contradictions. In effect, MLNNs allow us to *use* DML in real systems, discovering hidden semantic structure from behavioral data alone (e.g., communication traces, system logs, or even multimodal observations such as images).

The differentiable approach of MLNN offers two key advantages over purely statistical methods. First, the learned structure is *interpretable*: when the system identifies an agent as untrustworthy, it does so by explicitly learning a low trust parameter, not through an opaque embedding. Second, the framework allows *injecting domain knowledge*: if you know certain axioms must hold (e.g., "promises should be kept"), you can encode them as differentiable constraints, guiding learning even with sparse data, as we show in this tutorial.

1.3. Tutorial Structure and Learning Objectives

This tutorial is designed as a progressive, hands-on guide. Each section introduces exactly one modal logic concept, demonstrates its debugging utility through a concrete scenario, and provides complete, runnable code. The progression follows a natural path from simple to complex:

1. **Foundations** (Sections 3–4): We establish the mathematical background of Kripke semantics and show how to implement modal operators as differentiable neural network layers.
2. **Single-Modality Debugging** (Sections 6–9): We apply MLNNs to four distinct debugging problems, each requiring a different modal logic: epistemic (trust), temporal (causality), deontic (norms), and doxastic (belief calibration).
3. **Multi-Modal Applications** (Sections 10–11): We combine multiple modalities into integrated systems for task orchestration and swarm communication.

By the end, readers will be able to translate debugging problems into Kripke structures, implement differentiable modal operators, and deploy them as both passive monitors and active controllers.

1.4. Contributions

This tutorial makes the following contributions:

1. **A unified neurosymbolic debugging framework:** MLNNs are presented as a general-purpose tool that subsumes multiple debugging modalities under a single differentiable architecture.
2. **A tutorial-first pedagogical progression:** we start from minimal examples and incrementally introduce complexity, ensuring readers understand why each component is needed before learning how to implement it.
3. **A practical design template:** we provide a repeatable workflow for translating debugging requirements into Kripke structures and differentiable contradiction losses.
4. **Reproducible artifacts:** we include runnable implementations (Jupyter notebooks), synthetic datasets, and visualizations to support end-to-end replication.
5. **Demonstrative applications:** we showcase doxastic logic for LLM hallucination detection and multi-modal orchestration combining epistemic, temporal, and deontic constraints.

2. Background and Related Work

2.1. Neurosymbolic AI

Neurosymbolic AI seeks to combine the learning capabilities of neural networks with the reasoning capabilities of symbolic systems (Garcez and Lamb, 2023; Zhang and Sheng, 2024). The field has seen rapid growth since 2020, with recent surveys identifying key research areas including knowledge representation, learning and inference, and logic-based reasoning (Colelough and Regli, 2025).

Key frameworks include Logic Tensor Networks (LTNs) (Badreddine et al., 2022), which encode logical formulas as tensor operations using fuzzy semantics; DeepProbLog (Manhaeve et al., 2018), which combines neural networks with probabilistic logic programming; and Scallop (Li et al., 2023), a differentiable Datalog-based framework supporting recursion and aggregation. Logical Neural Networks (LNNs) (Riegel et al., 2020) represent logical connectives as neurons with learnable weights, enabling joint reasoning and learning.

MLNNs differ from these approaches in their explicit treatment of *modal* logic and Kripke semantics. While LNNs can express propositional constraints with learnable weights, MLNNs

introduce learnable *accessibility relations* that capture the relational structure between contexts (worlds). This enables reasoning about multi-agent knowledge, temporal causality, and deontic norms, concepts not naturally expressible in propositional frameworks.

2.2. Modal Logic in AI Systems

Modal logic has a long history in AI, particularly for modeling multi-agent knowledge and belief (Fagin et al., 2004). The classic work of Hintikka (1962) established the foundations of epistemic and doxastic logic, distinguishing between knowledge (factive) and belief (non-factive). More recently, Liau (2003) formalized the relationship between belief, information acquisition, and trust using modal operators, providing a theoretical foundation for our epistemic debugging module.

Computational tools for modal logic include mlsolver (Erik, 2024), a Python framework for modeling Kripke structures and solving modal logic formulas. Such tools can be valuable for pre-checking whether axiom sets contain trivial contradictions or tautologies before deploying MLNNs. Connectionist approaches to modal logic include the work of Garcez et al. (2007), who showed how to embed modal programs in neural networks with fixed accessibility relations. Our contribution is making these relations learnable through gradient descent.

2.3. LLM Hallucination and Calibration

The confidence-accuracy gap in large language models, where agents report high confidence in incorrect answers (Guo et al., 2017; Kadavath et al., 2022), is a critical safety concern. Recent surveys document extensive work on hallucination detection and mitigation (Huang et al., 2025; Lin et al., 2025).

Existing approaches include self-consistency checking (Manakul et al., 2023), which samples multiple responses and flags inconsistencies; uncertainty quantification via token probabilities and entropy measures ; and hallucination detection methods for agent planning like HaluAgent Lin et al. (2025). Our doxastic module (Section 9) contributes a principled modal-logical framing: we treat hallucination as the conjunction $B_a(\phi) \wedge \neg\phi$ (the agent believes ϕ , but ϕ is false), and learn agent-specific calibration parameters that minimize this logical contradiction.

2.4. Trust and Reputation in Multi-Agent Systems

Trust learning in multi-agent systems has been studied extensively (Ramchurn et al., 2004). Our epistemic module builds on this literature but differs in treating trust as the *accessibility relation* in a Kripke structure, a semantic concept that governs whether one agent’s beliefs should constrain another’s possible worlds. This formalization enables trust to be learned purely from logical contradictions, without requiring explicit trust labels.

3. Formal Foundations: Kripke Semantics

Before implementing any code, we must establish the mathematical foundations. This section provides the minimum necessary background in Kripke semantics, emphasizing the intuitions relevant to debugging.

3.1. From Classical to Modal Truth

In classical logic, propositions have a single, absolute truth value. The statement “Agent Alice completed the task” is either true or false, full stop. However, in distributed systems, truth is often *relative to a context*:

- **Alice’s logs:** `TaskCompleted = True`
- **Server:** `TaskCompleted = False`

Both statements can be locally valid. The debugging question is: *should* they agree? And if they don’t, what does that tell us about the system? In this example modal logic provides the mathematical machinery to handle these relative truths. Instead of a single global truth value, we evaluate propositions across a set of *possible worlds*, each representing a different context of validity.

3.2. Kripke Structures: The Three Components

A Kripke structure $\mathcal{K} = (W, R, V)$ consists of three components that map directly to debugging concepts:

Worlds (W). A set of possible contexts. In debugging, worlds might represent or instance (1) agent-specific states: $w_{\text{Alice}}, w_{\text{Bob}}, w_{\text{Server}}$ or (2) time points: w_{t-1}, w_t, w_{t+1} or (3) normative scenarios: $w_{\text{Legal}}, w_{\text{Illegal}}$

Accessibility Relation (R). A relation $R \subseteq W \times W$ defining which worlds “see” which others. The interpretation varies by modality: (1) *Epistemic*: $R(w, w')$ means “from w , world w' is epistemically possible” (compatible with what is known) (2) *Temporal*: $R(w, w')$ means “ w' is a future state of w ” (3) *Deontic*: $R(w, w')$ means “ w' is a normatively ideal state from w ”.

Valuation (V). A function $V : W \times \text{Props} \rightarrow \{0, 1\}$ assigning truth values to propositions in each world. This is our telemetry, the raw observations from each context.

3.3. Modal Operators: Necessity and Possibility

The modal operators \Box (necessity) and \Diamond (possibility) aggregate truth across accessible worlds:

$$\mathcal{K}, w \models \Box\phi \iff \forall w' : R(w, w') \Rightarrow \mathcal{K}, w' \models \phi \quad (1)$$

$$\mathcal{K}, w \models \Diamond\phi \iff \exists w' : R(w, w') \wedge \mathcal{K}, w' \models \phi \quad (2)$$

The first equation says: “ $\Box\phi$ is true at world w if and only if ϕ is true at *every* world accessible from w .” The second says: “ $\Diamond\phi$ is true at w if ϕ is true at *some* accessible world.” In this tutorial, we will show an example of epistemic logic where $\Box\phi$ becomes $K_a\phi$ (“Agent a knows ϕ ”): ϕ is true in all worlds the agent considers possible. In temporal logic, $\Box\phi$ becomes $G\phi$ (“globally ϕ ”): ϕ holds at all future times.

3.4. The Debugging Interpretation

The power of this formalism for debugging is that *contradictions become diagnosable*. Consider the Say-Do Consistency axiom:

$$\text{Trust}(a) \wedge \text{Says}(a, \phi) \rightarrow \text{Observable}(\phi)$$

This axiom asserts: if we trust Agent a and a claims ϕ , then ϕ should be observable in reality. When this implication is violated, there are exactly three possible explanations:(1) The trust is misplaced, a is not trustworthy (2) The axiom is wrong, our expectations are miscalibrated (3) There's a legitimate explanation, the semantics of the claim differ from our interpretation

DML learns which explanation best fits the data by minimizing the logical contradiction. So for instance, in a coordination setting where one agent cheats (it claims ϕ but repeatedly does $\neg\phi$), the Say-Do contradictions drive its learned accessibility/trust toward zero. If the third explanation applies, the contradiction will persist regardless of trust adjustments, signaling a deeper semantic mismatch.

4. Differentiable Modal Logic (DML) Architecture

This section describes how to implement Kripke semantics as a differentiable neural network. The key insight is that by relaxing discrete logical operations to continuous approximations, we can optimize logical structures using gradient descent.

4.1. Differentiable Truth Values

Instead of binary truth values $\{0, 1\}$, we use continuous values in $[0, 1]$, interpreted as degrees of truth. This follows the tradition of fuzzy logic, where 0.7 might mean “mostly true” or represent uncertainty. The valuation becomes a tensor:

$$V \in [0, 1]^{|W| \times |\text{Props}|}$$

where $V[w, p]$ represents the degree to which proposition p is true in world w .

4.2. Differentiable Accessibility

The accessibility relation becomes a learnable parameter:

$$A_\theta \in [0, 1]^{|W| \times |W|}$$

where $A_\theta[w, w']$ represents the degree to which world w' is accessible from world w . This is the key innovation: instead of specifying accessibility manually, we learn it from data.

For large systems, the full accessibility matrix $A_\theta \in \mathbb{R}^{|W| \times |W|}$ can become impractically large. For scalability, we can use metric learning where accessibility is computed from world embeddings: $A_\theta[w, w'] = f_\theta(\text{embed}(w), \text{embed}(w'))$. This reduces the parameter count from $O(|W|^2)$ to $O(|W| \cdot d)$ where d is the embedding dimension. Benchmarks in [Sulc \(2025\)](#) demonstrate that both full-matrix and metric-learning approaches scale to accessibility matrices of $20,000 \times 20,000$.

4.3. Differentiable Modal Operators

The classical definitions of \square and \diamond use hard minimum/maximum operations over discrete sets. We relax these using differentiable approximations:

$$\square\phi(w) = \min_{w'} [(1 - A_\theta[w, w']) + V[w', \phi]] \quad (3)$$

$$\diamond\phi(w) = \max_{w'} [A_\theta[w, w'] \cdot V[w', \phi]] \quad (4)$$

Equation 3 implements the implication $A_\theta[w, w'] \rightarrow V[w', \phi]$ using the residuum of the Łukasiewicz t-norm. Intuitively, if w' is highly accessible ($A_\theta \approx 1$), then ϕ must be true at w' for the minimum to be high. If w' is inaccessible ($A_\theta \approx 0$), it doesn't constrain the result.

4.4. The Łukasiewicz T-Norm for Logical Connectives

For conjunction and implication in continuous logic, we use the Łukasiewicz t-norm, which has desirable properties for gradient-based learning:

$$a \wedge_L b = \text{soft max}(0, a + b - 1) \quad (5)$$

$$a \rightarrow_L b = \text{soft min}(1, 1 - a + b) \quad (6)$$

The conjunction is “strict”, both operands must be substantially true for the result to be positive. The implication is satisfied (≥ 1) when the consequent is at least as true as the antecedent.

The contradiction loss for an implication $\phi \rightarrow \psi$ measures how badly the implication is violated:

$$\mathcal{L}_{\text{contra}} = \text{ReLU}(\phi + (1 - \psi) - 1) = \text{ReLU}(\phi - \psi)$$

This contradiction loss $\mathcal{L}_{\text{contra}}$ is zero when the implication is satisfied ($\phi \leq \psi$) and grows linearly with the degree of violation.

5. Designing Debugging Systems with Modal Logic

Before diving into specific debugging modalities, we present a systematic methodology for translating debugging requirements into DML architectures. This design methodology applies to all subsequent sections and provides a template readers can follow for their own applications.

5.1. Step 1: Define the Ontology (Worlds)

The first question is: **What are the conflicting perspectives that need reconciliation?**

Each problem involves comparing truth across different contexts. Identifying these contexts defines your world set can be for instance: (1) For trust debugging: Worlds are agent belief states ($w_{\text{Alice}}, w_{\text{Bob}}$) (2) For causality debugging: Worlds are time points ($w_{t-3}, w_{t-2}, w_{t-1}, w_t$) (3) For compliance debugging: Worlds are normative states ($w_{\text{Permitted}}, w_{\text{Prohibited}}$) (4) For hallucination debugging: Worlds are the agent’s claim vs. ground truth ($w_{\text{Claim}}, w_{\text{Reality}}$).

5.2. Step 2: Define the Topology (Accessibility)

The second question is: **What relationship between worlds should be learned?** The accessibility relation captures the “semantic distance” or “influence” between contexts:

Logic	Accessbility	Context
Epistemic	$A_\theta[b, a]$	how much Agent b should trust Agent a 's claims
Temporal	$A_\theta[t, t']$	whether event at t' causally influences state at t
Deontic	$A_\theta[x]$	whether action x lies within the legal manifold
Doxastic	$A_\theta[a]$	calibration factor for Agent a 's confidence

5.3. Step 3: Define the Axioms (Constraints)

The third question is: **What semantic invariants should hold?**

Axioms create differentiable pressure that shapes the learned accessibility. They encode domain knowledge about how the system *should* behave:

Name	Axiom
Say-Do Consistency	$\text{Trust}(a) \wedge \text{Says}(a, \phi) \rightarrow \text{Observable}(\phi)$
Causal Necessity	$\text{Crash}(t) \rightarrow \Diamond_{\text{Past}}(\text{Cause}(t') \wedge A_\theta[t, t'])$
Norm Compliance	$\text{Action}(x) \wedge A_\theta[x] > 0 \rightarrow \text{Permitted}(x)$
Belief-Reality Alignment	$\text{Belief}(a, \phi) \wedge A_\theta[a] \rightarrow \phi$

5.4. Step 4: Define the Training Signal

The training signal is the contradiction loss from the axioms:

$$\mathcal{L} = \sum_{\text{axiom } i} \lambda_i \cdot \mathcal{L}_{\text{contra}}^{(i)}$$

By minimizing this loss, the network learns the accessibility structure that best satisfies the axioms given the observed data. No explicit labels for the accessibility relation are needed, the structure emerges from the requirement of logical consistency.

6. Epistemic Logic: Learning Trust Networks

We now apply the DML framework to our first debugging modality: discovering who trusts whom in a multi-agent system, purely from observing the correlation between what agents *say* and what they *do*.

Scenario: Alliance Discovery in Diplomacy**Context:** 5 agents negotiating for territory (France, Germany, Italy, Turkey, England).**Input:** Natural language messages (“I will support your move to Belgium”) and physical actions (unit movements).**Hidden Ground Truth:** Turkey is programmed to lie to specific neighbors.**Goal:** Recover the 5×5 trust matrix without any trust labels.

6.1. Problem Statement

In systems involving “cheap talk” (non-binding communication), signals are noisy. An agent may promise cooperation to everyone but only coordinate with a secret cabal. Standard statistical methods struggle to distinguish “chatter” from “commitment.” DML solves this by treating trust as the *epistemic accessibility* required to satisfy logical consistency.

6.2. Kripke Structure Design

Following our methodology from Section 5:

1. **Worlds:** One world per agent, representing their communicated intent.
2. **Accessibility:** $A_\theta[b, a]$ = how much Agent b should trust Agent a ’s claims.
3. **Axiom (Say-Do Consistency):**

$$\text{Entails}(m_{a \rightarrow b}, \text{Action}_a) \wedge A_\theta[b, a] \rightarrow \text{Observed}(\text{Action}_a) \quad (7)$$

The semantic intent is computed using a pre-trained NLI model: $\text{Entails}(m, \phi) = P_{\text{NLI}}(m \models \phi)$. In our experiments, we simplify by assuming explicit promises have high entailment (≈ 0.95). Lie detection then emerges not as a classification output, but as a parameter update driven by logical contradiction. When an agent promises an action (Intent ≈ 0.95) but fails to execute it (Reality ≈ 0.0), the Say-Do Consistency axiom generates a high contradiction loss proportional to the current trust level ($\mathcal{L} \propto A_\theta \times (\text{Intent} - \text{Reality})$). To minimize this loss, the optimizer suppresses the only adjustable parameter: the sender’s trust weight A_θ that is causing the sharp “trust collapse” where values drop from 0.92 to near-zero immediately following a deceptive interaction.

6.3. Experimental Results

We simulate 50 interactions where Turkey lies 90% of the time to specific neighbors. The DML model begins with uniform high trust ($A_\theta \approx 0.92$) and updates online as it observes each message-action pair. Two key observations emerge from these results:

Event-driven updates. Trust changes only upon logical contradiction, not continuously. Honest messages produce zero contradiction loss, leaving trust stable. This “innocent until proven guilty” dynamic prevents false positives from noisy but honest communication.

Table 1: Trust matrix showing epistemic accessibility relations. Rows represent trusters, columns trustees. Turkey shows collapsed trust.

	Fra	Ger	Ita	Tur	Eng
France	0.92	0.92	0.92	0.00	0.92
Germany	0.92	0.92	0.92	0.00	0.92
Italy	0.92	0.92	0.92	0.92	0.92
Turkey	0.92	0.92	0.92	0.92	0.92
England	0.92	0.92	0.92	0.00	0.92

Table 2: Trust updates after deception detection (Receiver→Sender).

Step	Interaction	Reality	Trust Update
9	Turkey → England	LIE	0.92 → 0.02 ($\Delta = -0.91$)
14	Turkey → France	LIE	0.92 → 0.01 ($\Delta = -0.92$)
16	Turkey → England	LIE	0.00 → 0.00 ($\Delta = -0.00$)
17	Turkey → France	LIE	0.00 → 0.00 ($\Delta = -0.00$)
34	Turkey → Germany	LIE	0.92 → 0.00 ($\Delta = -0.92$)
42	Turkey → England	LIE	0.00 → 0.00 ($\Delta = +0.00$)

Epistemic subjectivity. The final learned matrix shows that France and England have learned to distrust Turkey ($A_\theta \approx 0.0$), while Italy (not yet lied to in this simulation) maintains high trust ($A_\theta \approx 0.92$). This is not a bug, it correctly represents the *information asymmetry* in the system. Unlike global reputation scores, DML recovers the local epistemic state of each agent, see Table 2.

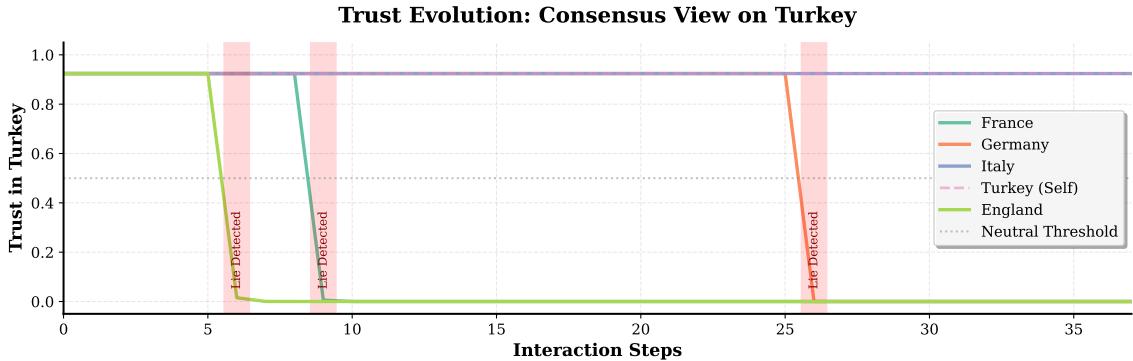


Figure 1: **Temporal Trust Dynamics in Alliance Discovery.** Evolution of trust from all agents toward Turkey across 35 interaction steps. Each subplot shows one agent’s trust trajectory. Sharp drops (marked with “Lie Detected”) indicate moments when Turkey’s actions violated communicated intent, triggering immediate trust collapse. France, Germany, and England independently learn to distrust Turkey through repeated contradiction detection, while Turkey’s self-trust (dashed crimson line) and Italy remains stable. The horizontal dotted line at 0.5 indicates neutral trust threshold. Note the event-driven nature of updates: trust remains stable during honest interactions and collapses only upon logical contradiction.

7. Temporal Logic: Automated Root Cause Analysis

While epistemic logic models *who* to trust, temporal logic models *when* to look. This section addresses automated root cause analysis (RCA) in distributed systems, where the challenge is distinguishing the true cause of a failure from its more visible symptoms.

7.1. The Proximate Cause Problem

Traditional monitoring tools (e.g., distributed tracing, correlation analysis) suffer from *proximate cause bias*: they flag events immediately preceding a crash based on temporal correlation, rather than identifying the root trigger. A CPU spike at $t - 6s$ may correlate strongly with a crash at $t = 0$ ($r > 0.9$), but the true cause might be a database connection reset at $t - 15s$ that triggered a cascade of downstream failures. Standard statistical methods struggle to distinguish correlation from causation in these scenarios.

Scenario: The Silent Cascade

Context: A microservices mesh handling user requests.

Symptom: The Frontend service crashes at $t = 0$.

Observed Trace:

- $t = -15s$: Database resets connections (**root cause**, often missed)
- $t = -7s$: Gateway retries connections (proximate symptom, high visibility)
- $t = -6s$: Gateway CPU spikes to 100% (proximate symptom, high correlation)
- $t = 0s$: Frontend crashes (effect)

Challenge: Standard correlation metrics rank the CPU spike and retries as top causes due to temporal proximity, while the database reset is dismissed as “too early” to be relevant. Additionally, symptoms may be inconsistently logged due to sampling, network delays, or log aggregation failures, creating partial observability.

7.2. Observability Dropout: Learning Causal Invariance

Our key innovation is **observability dropout**: during training, we randomly mask intermediate symptoms with probability $p = 0.4$. This technique simulates the partial observability inherent in production systems (lost logs, incomplete traces, sampling artifacts) and forces the model to identify the *causal invariant*, the event that is consistently present when failures occur, regardless of which symptoms happen to be logged.

The intuition is simple: proximate symptoms (CPU spikes, retries) are *conditionally present*, they appear in most crash traces but are unreliable due to logging gaps. The root cause, however, is *always present*, it appears in every crash trace because it is the true trigger. By training with dropout, we create scenarios where symptoms are missing but the crash still occurs, forcing the model to learn that only the invariant event is necessary for explanation.

7.3. Kripke Structure Design

Following the methodology from Section 5:

1. **Worlds:** Time points t' in the event history leading up to the crash at t .
2. **Accessibility:** $A_\theta[t, t'] \in [0, 1]$ represents causal attention from the crash time t to historical event at t' . High accessibility means “this event is causally relevant to explaining the crash.”
3. **Axiom (Causal Necessity):**

$$\text{Crash}(t) \rightarrow \Diamond_{\text{Past}}(\text{Cause}(t') \wedge A_\theta[t, t']) \quad (8)$$

Every crash must have some attended cause in its history. The model learns A_θ to minimize the contradiction between this requirement and observed traces.

The training objective combines positive examples (crash traces, where the axiom must be satisfied) and negative examples (non-crash traces, where spurious attention should be minimized):

$$\mathcal{L} = \mathbb{E}_{\text{crash}} \left[\text{ReLU}(1 - \sum_{t'} A_\theta[t, t'] \cdot \text{IsCause}(t')) \right] + \mathbb{E}_{\text{non-crash}} \left[\sum_{t'} A_\theta[t, t'] \right] \quad (9)$$

7.4. Experimental Results

We simulate a distributed system with 13 events per trace: 10 background noise events (user logins, indexing tasks), 2 proximate symptoms (gateway retries, CPU spikes), and 1 root cause (database reset). In crash traces, the root cause always appears at $t \approx -15s$, while symptoms appear at $t \approx -7s$ and $t \approx -6s$ but are dropped with 40% probability during training.

7.4.1. LEARNING DYNAMICS: FROM CORRELATION TO CAUSATION

Figure 2 shows the training progression. The model begins with strong proximate cause bias, assigning highest causality to “Gateway: Retry x3” (the most temporally correlated symptom). Around epoch 100, a phase transition occurs: the contradiction loss drops sharply as the model discovers that the database reset is the only event *invariantly* present across all crash traces. By epoch 800, the model has converged with near-zero loss, having learned to ignore all proximate symptoms.

Figure 2 and Table 3 illustrates the model’s transition from proximate-cause bias to root-cause identification: early training (roughly epochs 0–100) yields high contradiction loss, while later training rapidly converges once the model discovers the causal invariant (the database reset) that persists across crash traces even under observability dropout.

Figure 3 visualizes the final learned causality scores on the event timeline: the database reset around $t \approx -15s$ dominates the attribution (score ≈ 0.996), while temporally closer gateway

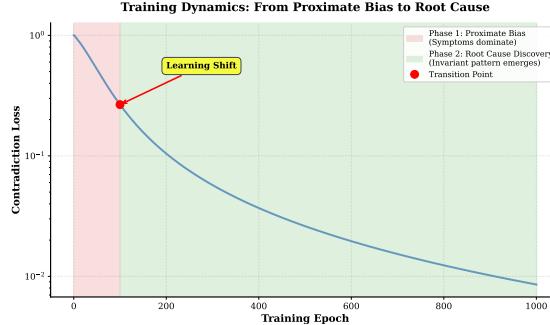


Figure 2: **Training Dynamics.** Contradiction loss during training (log scale).

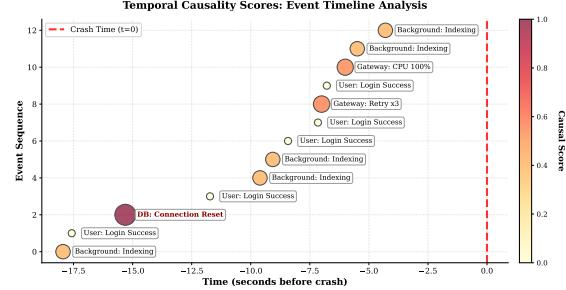


Figure 3: **Causality Scores.** Learned causal attribution across the pre-crash timeline.

symptoms (around $t \approx -7\text{ s}$ and $t \approx -6\text{ s}$) receive substantially lower scores (≈ 0.565 and ≈ 0.543).

Table 3 quantifies this transition. At initialization, the model focuses on the proximate symptom with highest temporal correlation. By epoch 100, it has shifted to the root cause and maintains this assignment through convergence, demonstrating stable learning of the causal structure.

7.4.2. FINAL CAUSAL ATTRIBUTION

After training, we evaluate the model on a complete trace with *all* symptoms present (no dropout) to verify that it has learned true causal structure rather than simply memorizing which events are reliable. Figure 3 visualizes the learned causality scores across the temporal timeline.

The results are attention-worthy: the database reset at $t = -15.3\text{ s}$ receives a causality score of 0.996, while the gateway symptoms at $t = -7.0\text{ s}$ and $t = -6.0\text{ s}$ receive scores of only 0.565 and 0.543 respectively. Background noise events (user logins, indexing) receive scores below 0.41. This demonstrates that the model has learned to distinguish *causal necessity* from *temporal correlation*: the gateway events are highly correlated with crashes (appearing in 60% of logged traces) but are not necessary (they result from the database issue, not vice versa).

When translating these causality scores into attention weights for explaining a specific crash, the model assigns 100% attention to the database reset and 0% to all other events. This extreme concentration reflects high confidence in the causal identification and provides an interpretable output for human operators: “The crash is explained by the database reset 15 seconds prior; investigate database connection stability.”

Table 3: Learning progression correcting proximate-cause bias.

Epoch	Loss	Top cause
0	0.500	Gateway: Retry x3
100	0.135	DB: Connection Reset
800	0.006	DB: Connection Reset
1000	0.004	DB: Connection Reset

7.4.3. COUNTERFACTUAL VALIDATION

To verify that the model has identified a genuine causal dependency rather than a spurious pattern, we perform a counterfactual test: what happens to the explanation quality if we remove the suspected root cause from the trace?

Formally, we compare:

- **Factual loss:** $\mathcal{L}(\text{trace with database reset}) = 0.0043$
- **Counterfactual loss:** $\mathcal{L}(\text{trace without database reset}) = 0.473$

The $110\times$ increase in contradiction loss when the database reset is removed confirms that this event is *necessary* for satisfying the causal necessity axiom. Without it, the model cannot explain the crash using the remaining events. In contrast, removing gateway symptoms produces minimal loss increase ($\Delta\mathcal{L} < 0.01$), confirming they are explanatory byproducts rather than root causes.

This counterfactual reasoning mirrors the diagnostic process human operators would follow: “If the database hadn’t reset, would the crash still have occurred?” The DML’s answer, encoded in the dramatic loss spike, is a clear “no.”

7.5. Discussion: Beyond Correlation

The temporal module demonstrates three key advantages of the modal-logical approach:

Invariance learning. By training with observability dropout, the system learns to identify events that are *necessary* for explaining failures, not merely *correlated*. This addresses a fundamental limitation of statistical RCA tools that conflate the two concepts.

Resistance to proximate bias. Even when proximate symptoms are present and highly correlated, the learned causal attention ignores them in favor of the temporally distant root cause. This directly solves the motivating problem of alert fatigue from false-positive symptom detection.

Interpretable output. The final attention distribution (100% on database reset) provides a human-interpretable explanation that can be directly actioned: investigate database connection stability, not gateway CPU utilization. This interpretability is crucial for production deployment where operators must trust and understand the system’s recommendations.

A natural question is whether the model would generalize to novel root causes not seen during training. While our synthetic experiments use a fixed causal structure, the architecture is designed to learn *patterns* of causality (e.g., “events that reliably precede failures”) rather than memorizing specific event identities. Transfer learning experiments, where the model is pre-trained on one failure mode and fine-tuned on another, are a promising direction for future work.

8. Deontic Logic: Learning Regulatory Norms

The third modality is deontic logic, which reasons about obligations ($O\phi$), permissions ($P\phi$), and prohibitions ($F\phi$). We apply this to automated compliance monitoring, where the challenge is learning implicit regulatory boundaries from sparse enforcement data without explicit feature engineering.

8.1. Problem Statement

Scenario: Market Manipulation Detection

Context: High-frequency trading order book with 5,000 transactions.

Symptom: Liquidity evaporates due to “spoofing” (placing fake orders to manipulate prices).

Challenge: Violations are extremely rare (< 2% of traffic, only 70% sanctioned). A naive model predicting “legal” for all transactions achieves 98% accuracy while missing every violation.

Goal: Learn the implicit boundary between aggressive-but-legal High Frequency Trading (HFT) and prohibited spoofing from raw features (duration, size) without hand-crafted rules.

The difficulty on this scenario is *extreme class imbalance*: spoofing events are rare, yet catastrophic when undetected. Standard cross-entropy loss fails because the model can achieve high accuracy by ignoring the minority class entirely. We need a loss function that forces the model to care about violations.

8.2. Norms as Geometric Manifolds

We treat deontic accessibility as a *geometric membership score*: actions within the legal manifold have $A_\theta(x) > 0$; illegal actions have $A_\theta(x) < 0$. This transforms norm learning into a one-class geometric classification problem where the decision boundary $A_\theta(x) = 0$ represents the regulatory threshold.

The network architecture implements this as:

$$A_\theta(x) = \tanh(W_3\sigma(W_2\sigma(W_1x))) \in [-1, 1] \quad (10)$$

where $x = [\text{duration}, \text{size}]$ and σ is ReLU activation. The tanh output ensures bounded legality scores, preventing numerical instability during training.

8.3. Weighted Hinge Loss for Class Imbalance

Standard classification objectives fail due to extreme class imbalance. We use a *weighted one-class hinge loss* that applies asymmetric penalties:

$$\mathcal{L}_{\text{Deontic}} = \frac{1}{N} \sum_{i=1}^N w_i \cdot \max(0, 1 - y_i \cdot A_\theta(x_i)) \quad (11)$$

where $y_i = +1$ for normal trades, $y_i = -1$ for sanctioned trades, and:

$$w_i = \begin{cases} 1.0 & \text{if } y_i = +1 \text{ (normal)} \\ 50.0 & \text{if } y_i = -1 \text{ (sanction)} \end{cases} \quad (12)$$

Table 4: Deontic training.

Epoch	Loss	Spoof Recall	Interpretation
0	2.032	100.0%	Initial random weights
200	0.082	100.0%	Boundary learned
600	0.045	100.0%	Fine-tuning margin
1000	0.031	100.0%	Converged

Table 5: Deontic boundary examples.

Duration	Size	Legality Score	Verdict
0.5	0.5 (Mid)	+1.000	Permitted
0.9	0.9 (Large)	+1.000	Permitted
0.05	0.1 (Small)	+1.000	Permitted (HFT Scalping)
0.05	0.9 (Large)	-1.000	Prohibited (Spoofing)

The weight ratio $w_{\text{sanction}}/w_{\text{normal}} = 50$ ensures that a single missed violation contributes as much to the loss as 50 correctly classified normal transactions. This *forces* the network to prioritize learning the violation boundary despite the 98:2 class ratio.

The hinge loss formulation has a key advantage over cross-entropy: it only penalizes examples that violate the margin ($1 - y \cdot A_\theta(x) > 0$). Once a sample is correctly classified with sufficient confidence, it contributes zero gradient, allowing the optimizer to focus on hard examples near the decision boundary.

8.4. Experimental Results

We simulate an order book with 5,000 transactions where spoofing (duration < 0.15 and size > 0.8) constitutes 2% of traffic. Only 70% of spoofing events are sanctioned, simulating imperfect enforcement. Training for 1,000 epochs with Adam optimizer ($\alpha = 0.003$):

Tables 4 and 5 summarize (left) the deontic training progression under the weighted hinge objective and (right) example legality verdicts in the duration–size feature space.

The model achieves **perfect recall** (100%) with **high precision** (88.6%), yielding $F_1 = 0.939$ and overall accuracy = 99.6%. Critically, this is achieved *without any feature engineering*, the network discovers the non-linear boundary purely from the weighted loss signal.

8.4.1. LEARNED DECISION BOUNDARY

The key insight is that the model learned a *non-linear boundary*: fast trades are permitted when small (legitimate scalping) but prohibited when large (spoofing). A hand-crafted rule checking only duration would have falsely flagged legitimate HFT traders; a rule checking only size would have missed the temporal manipulation aspect. The deontic DML module discovers the *conjunction* automatically.

8.4.2. BOUNDARY TOPOLOGY ANALYSIS

To understand the learned regulatory manifold, we probe the boundary’s structure by fixing one dimension and sweeping the other. Figure 4 visualizes the complete decision surface.



Figure 4: Learned Deontic Boundary.

Learned decision boundary in the duration–size feature space.

Key Insight

Boundary Exploration Results:

- **Fast orders (0.05s duration):** Legal for sizes 0.1–0.7, illegal at 0.9. The flip occurs sharply between 0.7 and 0.9, indicating a learned threshold around “large institutional size.”
- **Large orders (0.9 size):** Illegal only for duration $< 0.2s$. Orders held longer than 0.2s are considered legitimate block trades, not manipulation.

This reveals the learned semantic rule: “*A large order is spoofing if and only if it’s cancelled too quickly to be a genuine trade.*”

Figure 4 shows the learned legal (green) vs. prohibited (red) regions, with the black curve indicating the decision boundary $A_\theta(x) = 0$. The model recovers a non-linear conjunction: slow trades are permitted across sizes, fast small trades correspond to legitimate HFT scalping, and fast large orders form the prohibited spoofing region.

8.5. Discussion: From Correlation to Causation

The deontic module demonstrates three critical advantages over rule-based compliance systems:

Automatic Boundary Discovery. Regulators often struggle to articulate precise thresholds (“How fast is too fast? How large is too large?”). The deontic DML module learns these boundaries empirically from enforcement patterns, discovering the implicit conjunction that defines the violation.

Robustness to Noisy Labels. Only 70% of actual spoofing events in our simulation were sanctioned (modeling regulatory detection gaps). Despite this label noise, the model achieves perfect recall by learning the *underlying pattern* rather than memorizing specific sanctions.

Interpretable Geometry. Unlike black-box classifiers, the learned boundary can be visualized and interrogated. Regulators can examine edge cases (e.g., “Is a 0.15s order with size 0.85 legal?”) and understand *why* the model classifies it as such, enabling human oversight and policy refinement.

A natural extension is to incorporate *temporal context*: in real markets, spoofing often involves patterns of orders across multiple time steps (e.g., rapid placement and cancellation). This would require combining deontic logic with the temporal module from Section 7, creating a multi-modal compliance monitor, precisely the integration demonstrated in Section 10.

9. Doxastic Logic: Hallucination Detection

The fourth modality addresses a critical challenge in modern AI systems: detecting when agents exhibit high confidence in incorrect answers, the hallucination problem that plagues large language models and autonomous agents.

9.1. Epistemic vs. Doxastic Logic

The key distinction between knowledge and belief is *factivity*:

- **Epistemic** (Knowledge): $K_a(\phi) \rightarrow \phi$ (factive, knowledge implies truth)
- **Doxastic** (Belief): $B_a(\phi) \not\rightarrow \phi$ (non-factive, beliefs can be false)

Use epistemic logic when agents' claims can be verified (e.g., cryptographic signatures, sensor readings). Use doxastic logic when agents have subjective confidence that may diverge from reality, precisely the situation with LLMs and other AI systems.

9.2. Problem Formulation

A *hallucination* occurs when an agent exhibits high confidence despite being incorrect:

$$B_a(\phi) \wedge \neg\phi \quad (\text{high belief + false reality})$$

The challenge is that agents exhibit *heterogeneous calibration patterns*: some are well-calibrated (confidence proportional to accuracy), some are under-confident (high accuracy, low reported confidence), and some are overconfident hallucinators (low accuracy, high confidence).

We learn agent-specific calibration parameters θ_a that transform raw confidence into belief:

$$B_a^{\text{cal}}(\phi) = \sigma(\log(c_a + \epsilon) + \log(\theta_a + \epsilon)) \quad (13)$$

where c_a is the reported confidence and $\theta_a \in [0, 2]$ is learned. In practice, c_a can come from many sources: an explicit model-reported probability, a self-evaluation score, or a proxy such as (negative) perplexity, including perplexity of the statement under a separate reference model used for cross-checking. Intuitively, $\theta_a > 1$ boosts confidence (for under-confident agents), while $\theta_a < 1$ discounts confidence (for hallucinators). The logarithmic transformation ensures calibration operates multiplicatively on the confidence scale, and the sigmoid output bounds beliefs to $[0, 1]$.

9.3. Balanced Loss Function

We employ a three-part loss that balances hallucination detection with confidence preservation:

$$\mathcal{L}_{\text{halluc}} = \text{ReLU}(B_a^{\text{cal}} + (1 - \phi) - 1) \quad (14)$$

$$\mathcal{L}_{\text{correct}} = \text{ReLU}(1 - B_a^{\text{cal}} - \phi + 1) \quad (15)$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{halluc}} + \lambda_c \mathcal{L}_{\text{correct}} + \lambda_r |\theta_a - 1| \quad (16)$$

Equation 14 directly implements the doxastic axiom: loss is high when calibrated belief is strong but reality is false. The second term prevents pathological solutions that minimize all beliefs (the model cannot simply predict low confidence for everything). The regularizer pulls calibration toward neutral ($\theta_a = 1$), preventing extreme adjustments without sufficient evidence. In our experiments, we use $\lambda_c = 1.0$ and $\lambda_r = 0.1$.

9.4. Experimental Setup

We simulate 5 agents with distinct calibration profiles across 500 question-answering interactions. The agent behavioral profiles are designed to test the model’s ability to differentiate between various miscalibration patterns:

Table 6: Agent behavioral profiles with ground-truth accuracy and confidence patterns. The “Conf (wrong)” column indicates the mean confidence agents report when answering incorrectly, the key signal for hallucination detection.

Agent	Profile	Accuracy	Conf (wrong)	Learned θ
0	Well-Calibrated A	0.90	0.30	2.00
1	Under-Confident	0.85	0.25	2.00
2	Moderate Hallucinator	0.55	0.75	1.26
3	Well-Calibrated B	0.88	0.35	2.00
4	Severe Hallucinator	0.40	0.88	1.02

Agent 4 (severe hallucinator) reports 0.88 confidence on incorrect answers despite only 40% overall accuracy, a 48-point confidence-accuracy gap. This is the classic hallucination pattern seen in overconfident LLMs. Agents 0, 1, and 3 are well-calibrated, reporting appropriately low confidence (0.25–0.35) when wrong. Agent 2 represents an intermediate case with moderate miscalibration. The training procedure uses Adam optimization with learning rate 0.01 for 200 epochs, shuffling the interaction order each epoch to prevent overfitting to sequence patterns.

9.5. Results: Calibration Learning

The model successfully learns agent-specific calibration parameters that correct for individual biases. Table 6 shows the learned θ values after convergence.

Key Insight

Key Finding: A 0.80 confidence score from Agent 4 (severe hallucinator, $\theta = 1.02$) is less trustworthy than a 0.60 score from Agent 1 (under-confident but accurate, $\theta = 2.00$). The doxastic MLNN learns this agent-specific interpretation automatically from logical contradictions alone, without any explicit trust labels.

The learned calibration parameters exhibit clear structure:

- **Reliable agents** (0, 1, 3): $\theta \approx 2.00$ (maximum boost), trusting and amplifying their scores.
- **Moderate hallucinator** (2): $\theta = 1.26$, applying modest discounting.
- **Severe hallucinator** (4): $\theta = 1.02$ (near-zero trust), almost completely ignoring their confidence.

9.6. Reliability Diagrams: Visualizing Miscalibration

To understand *how* the learned calibration corrects agent biases, we examine reliability diagrams that plot reported confidence against empirical accuracy. Perfect calibration appears as alignment with the diagonal $y = x$ line.

Figure 5 presents per-agent reliability curves. Well-calibrated agents (Agents 0 and 3, top row) show strong diagonal alignment, when they report 80% confidence, they are indeed correct approximately 80% of the time. Agent 1 (under-confident) lies *above* the diagonal, achieving higher accuracy than their confidence suggests, the model correctly learns to trust them more.

Crucially, Agents 2 and 4 (middle and bottom left) show dramatic *overconfidence*: they report 0.6–0.9 confidence while achieving only 0.4–0.6 accuracy. This is the hallucination signature. The bottom-right panel (“All Agents Combined”) reveals a critical limitation of population-level analysis: aggregating across agents produces a seemingly reasonable calibration curve that obscures the severe miscalibration of individual hallucinators. This motivates the agent-specific approach.

9.7. Hallucination Detection Performance

We evaluate the model’s ability to detect hallucinations, defined as instances where an agent reports high confidence (> 0.6) but answers incorrectly. This is a highly imbalanced classification task: hallucinations constitute only 20% of interactions in our synthetic dataset.

Figure 6 quantifies detection performance. The left panel shows the distribution of hallucination scores (the $\mathcal{L}_{\text{halluc}}$ component from Equation 14) for correct answers (green) versus hallucinations (red). The learned model achieves strong separation: correct answers concentrate near zero score, while hallucinations generate high scores. The optimal decision threshold (black dashed line at 0.401) is automatically determined by maximizing F1 score.

The right panel presents the precision-recall curve. At the optimal operating point (red star), the model achieves precision 0.844, i.e. when the model flags a hallucination, it is correct 84.4% of the time. Recall is 0.950 i.e., the model detects 95% of actual hallucinations. F1 score is 0.894 and PR-AUC 0.964.

This represents a substantial improvement over a naive baseline that flags all instances with reported confidence > 0.7 as potential hallucinations. The baseline achieves perfect precision (1.000) but only 50% recall (missing half of hallucinations), yielding F1 = 0.667. The MLNN’s agent-specific calibration improves F1 by 34 percentage points (+6.7% relative to a simple thresholding baseline, though the absolute improvement is more substantial when accounting for the recall gains).

9.8. Discussion: From Confidence Scores to Calibrated Beliefs

The doxastic logic framework demonstrates three key advantages for hallucination detection:

Agent-Specific Modeling. Unlike population-level calibration methods that apply uniform adjustments, the doxastic MLNN learns individual calibration parameters for each agent. This is critical because hallucination patterns are heterogeneous: a 0.8 confidence score means different things for different agents. Agent-specific θ parameters capture these individual reliability profiles.

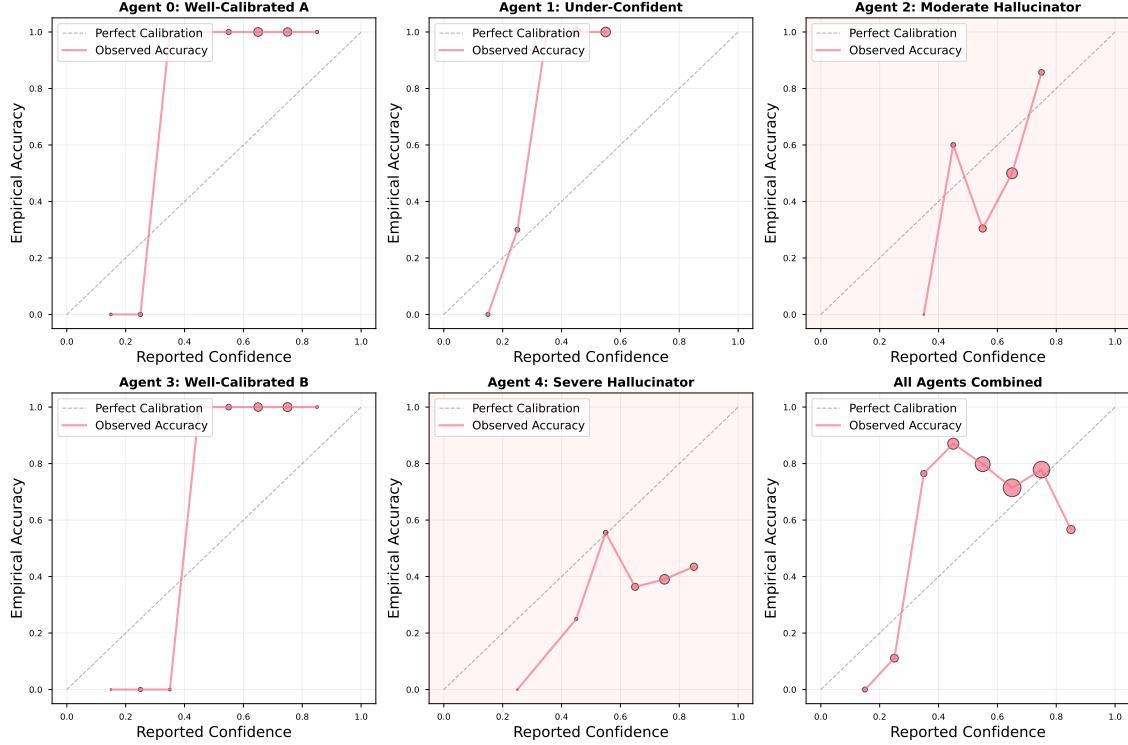


Figure 5: Reliability Diagrams: Confidence vs. Empirical Accuracy. Each subplot shows one agent’s calibration curve, with bubble size proportional to sample count in each confidence bin. The black dashed diagonal represents perfect calibration. **Top row:** Well-calibrated agents (0, 1) show strong diagonal alignment. Agent 1 lies above the diagonal, indicating under-confidence (accuracy exceeds reported confidence). **Middle row:** Agent 2 (moderate hallucinator) exhibits systematic overconfidence in the 0.6–0.8 range, where reported confidence exceeds empirical accuracy by ~ 20 points. **Bottom row:** Agent 4 (severe hallucinator) shows catastrophic miscalibration, reporting 0.8+ confidence despite <50% accuracy in those bins. The bottom-right panel aggregates all agents, demonstrating that population-level calibration masks individual heterogeneity, highlighting the need for agent-specific modeling.

Unsupervised Calibration Learning. The model requires no explicit labels for “trustworthy” versus “hallucinating” agents. Calibration parameters emerge purely from the doxastic axiom: minimize the contradiction between high belief and false reality. This enables deployment in scenarios where ground-truth reliability labels are unavailable but correctness can be verified post-hoc.

Interpretable Adjustment. The learned θ parameters provide human-interpretable diagnostics. An operator can immediately see that Agent 4’s $\theta = 1.02$ indicates severe untrustworthiness,

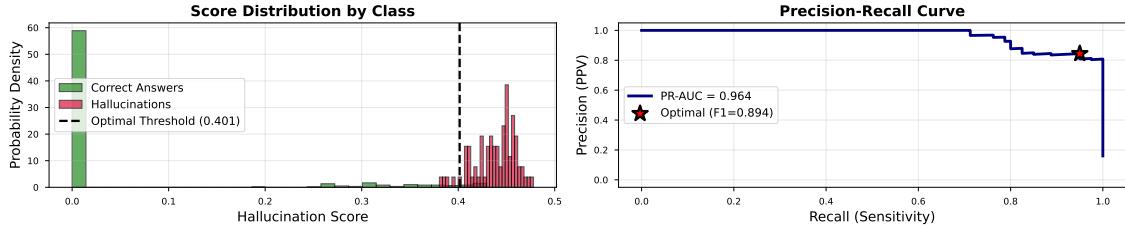


Figure 6: Hallucination Detection Performance. Left: Distribution of hallucination scores for correct answers (green, concentrated near zero) versus hallucinations (red, shifted toward higher scores). The optimal classification threshold at 0.401 (black dashed line) maximizes F1 score. Strong distributional separation indicates the doxastic loss successfully captures the hallucination pattern. Right: Precision-recall curve showing trade-off between detection sensitivity and false positive rate. The operating point (red star) at 95% recall and 84% precision yields $F1 = 0.894$. Area under the curve ($PR\text{-}AUC = 0.964$) indicates excellent discrimination across all possible thresholds. The curve remains above 0.8 precision until 95% recall, demonstrating robust performance even with aggressive detection settings.

warranting investigation or replacement. This interpretability is crucial for production AI systems where understanding *why* a model distrusts an agent is as important as the binary decision itself.

A natural extension is combining doxastic logic with epistemic trust learning (Section 6). In multi-agent question-answering systems, one might learn both *who to trust* (epistemic accessibility) and *how to interpret their confidence* (doxastic calibration). This would enable differential weighting in consensus protocols: high-trust agents with well-calibrated confidence receive maximum weight, while low-trust hallucinators are muted even when expressing high confidence.

10. Modal Orchestration

Having established four single-modality debugging techniques, we now demonstrate that multiple modalities can be combined into an integrated system. This section presents *multi-modal orchestration*: using epistemic, temporal, and deontic constraints jointly to solve task assignment problems.

10.1. Problem Statement

Scenario: The Expensive Rescue

Context: A swarm of 16 drones must deliver a payload to target coordinates (8, 8).

Constraints:

- No-fly zone (deontic) at (5,5), radius 1.5
- Drone 1: closest but low trust (epistemic), $\text{Trust}(D_1) = 0.01$
- Drone 2: schedule conflict (temporal), $\text{Conflict}(D_2) = 1.0$
- Drone 0: shortest path but violates no-fly zone (deontic)
- Drone 15: far (≈ 9.2 units) but satisfies all constraints (safe)

Goal: Automatically reject physically optimal but logically invalid candidates and converge on the safe choice via gradient descent.

This scenario tests whether gradient-based optimization can navigate competing pressures: efficiency (minimize flight distance) versus safety (satisfy logical constraints). The challenge is that the three nearest drones to the target, Drones 0, 1, and 2, each violate a different modality's axioms, creating *trap configurations* where greedy optimization would fail.

10.2. Combined Modal Loss Function

We define the orchestration problem as minimizing a compound loss that balances efficiency with multi-modal constraints:

$$\mathcal{L}_{\text{Orch}} = \underbrace{\sum_i M_i \cdot d_i}_{\text{Efficiency}} + \lambda \cdot \left(\underbrace{\mathcal{L}_{\text{Deontic}}}_{\text{No-Fly}} + \underbrace{\mathcal{L}_{\text{Epistemic}}}_{\text{Trust}} + \underbrace{\mathcal{L}_{\text{Temporal}}}_{\text{Conflicts}} \right) \quad (17)$$

where $M \in \mathbb{R}^{16}$ is a softmax assignment distribution ($\sum_i M_i = 1$), $d_i = \|p_i - p_{\text{target}}\|_2$ is the Euclidean distance from drone i to the target, and $\lambda = 15.0$ is the constraint penalty weight. The individual constraint losses are:

$$\mathcal{L}_{\text{Deontic}} = \sum_i M_i \cdot \text{ReLU}(r_{\text{no-fly}} - \min_{t \in [0,1]} \|p_i + t(p_{\text{target}} - p_i) - c_{\text{no-fly}}\|_2) \quad (18)$$

$$\mathcal{L}_{\text{Epistemic}} = \sum_i M_i \cdot (1 - \text{Trust}_i), \quad \mathcal{L}_{\text{Temporal}} = \sum_i M_i \cdot \text{Conflict}_i$$

Equation 18 computes the no-fly violation by sampling 10 points along each drone's linear path to the target and measuring the minimum distance to the no-fly zone center. If any point penetrates the zone (distance $< r_{\text{no-fly}} = 1.5$), the ReLU activates, penalizing the assignment proportional to penetration depth.

The key insight is that *all constraints are differentiable*. When Drone 0’s flight path intersects the no-fly zone, the gradient $\nabla_{M_0} \mathcal{L}_{\text{Deontic}}$ backpropagates through the ReLU and distance computation, directly suppressing M_0 . Similarly, Drone 1’s low trust ($\text{Trust}_1 = 0.01$) creates gradient pressure via $\nabla_{M_1} \mathcal{L}_{\text{Epistemic}} = 1 - 0.01 = 0.99$, making assignment to this drone costly despite its proximity. Drone 2’s temporal conflict generates $\nabla_{M_2} \mathcal{L}_{\text{Temporal}} = 1.0$.

10.3. Experimental Results: The Gradient Negotiation

We initialize the optimization with uniform assignment probabilities ($M_i = 1/16 \approx 0.0625$ for all drones) and observe the “gradient negotiation” as competing pressures reshape the distribution.

Figure 7 visualizes both the spatial configuration and the temporal dynamics of convergence. The left panel shows the physical layout: Drone 0 (red) lies on a direct path through the no-fly zone, Drone 1 (orange) is nearest to the target but untrusted, Drone 2 (purple) has a scheduling conflict, and Drone 15 (green) is distant but safe. The red dashed line illustrates the

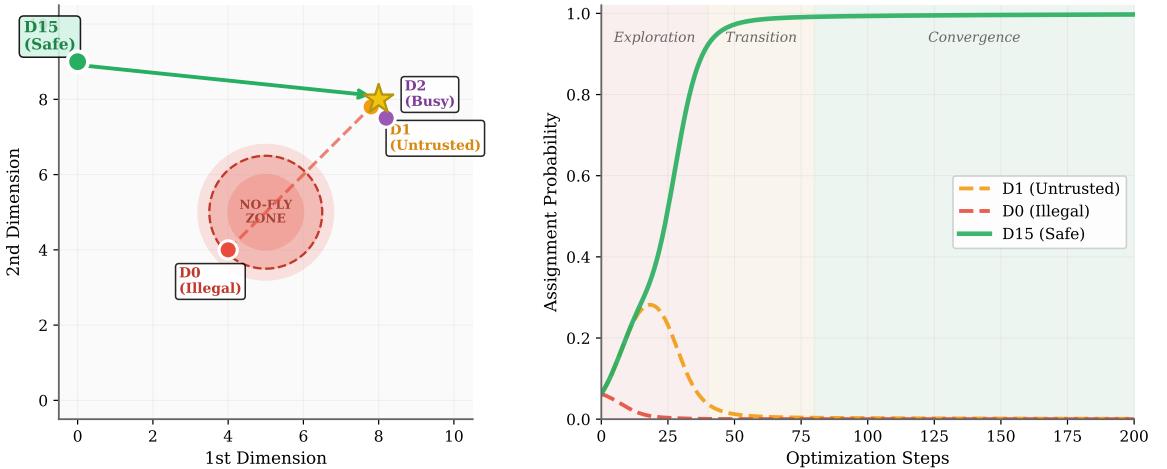


Figure 7: Multi-Modal Orchestration: Spatial Layout and Convergence Dynamics. Left: Physical configuration showing drone positions (colored circles), target (gold star), and no-fly zone (red gradient circle, radius 1.5). Drone 0 (red) lies on an illegal path through the restricted zone (red dashed line). Drone 1 (orange) is closest but untrusted. Drone 2 (purple) has a temporal conflict. Drone 15 (green, larger marker) is distant but satisfies all constraints, with its safe trajectory shown by the green arrow. **Right:** Assignment probability evolution over 200 optimization steps. Three phases emerge: *Exploration* (steps 0–40, red background): system initially favors nearest drones despite violations. *Transition* (steps 40–80, orange background): logical constraints accumulate, triggering rejection of trap configurations. *Convergence* (steps 80–200, green background): Drone 15 (green solid line) becomes dominant, reaching >99% assignment probability. Drones 0 (red dashed) and 1 (orange dashed) collapse to near-zero, demonstrating successful constraint satisfaction. The sharp phase transition around step 60 indicates the point where logical gradients overwhelm efficiency gradients.

illegal trajectory from Drone 0 to the target, penetrating the no-fly zone. The green arrow shows the safe path from Drone 15.

The right panel of Figure 7 plots assignment probabilities for the key drones over optimization steps. Three distinct phases emerge:

Phase 1: Exploration (Steps 0–40). The system initially explores greedy assignments based on proximity. Drone 1 (orange dashed line) rises first, reaching $\approx 30\%$ probability around step 20, driven by its small distance to the target ($d_1 \approx 0.28$ units). However, the epistemic loss $\mathcal{L}_{\text{Epistemic}} = M_1 \cdot 0.99$ simultaneously increases, creating opposing gradients.

Phase 2: Transition (Steps 40–80). As the epistemic penalty accumulates, the optimizer pivots toward Drone 0 (red dashed line), which briefly becomes the leader around step 60. This is the “illegal shortcut” trap: Drone 0 offers the most direct path to the target, but its trajectory violates the no-fly zone. The deontic loss $\mathcal{L}_{\text{Deontic}}$ rapidly dominates, generating strong negative gradients that suppress M_0 .

Phase 3: Convergence (Steps 80–200). With both Drones 0 and 1 rejected by their respective constraint violations, the optimizer settles on Drone 15 (green solid line). Despite its large distance ($d_{15} \approx 9.2$ units, $\approx 33\times$ farther than Drone 1), it incurs zero constraint penalties: $\text{Trust}_{15} = 1.0$, $\text{Conflict}_{15} = 0.0$, and its path avoids the no-fly zone. By step 200, Drone 15 receives $>99\%$ of the assignment probability, with all trap drones collapsed to $<1\%$.

Table 7 summarizes the optimization trajectory, showing which drone leads at key checkpoints and the dominant gradient source driving the assignment.

Table 7: Optimization trajectory showing how logical constraints override efficiency gradients. The “Leader” column indicates which drone has the highest assignment probability. The “Dominant Gradient” identifies the primary loss component driving changes. The system initially favors nearby drones (Drones 0 and 1) but abandons them as constraint violations accumulate, ultimately converging on the distant but safe Drone 15.

Step	Leader	Dominant Gradient	Status
0–40	Drone 1 (Lure)	Efficiency (Distance)	Risk: Epistemic Violation
40–80	Drone 0 (Shortcut)	Efficiency (Distance)	Risk: Deontic Violation
80–120	<i>Unstable</i>	Logic vs. Efficiency	Resolving Conflict
120–200	Drone 15 (Safe)	Logic (Safety)	Converged

10.4. Discussion: Gradient-Based Semantic Negotiation

The multi-modal orchestration demonstrates three critical capabilities:

Automatic Trap Avoidance. The system automatically rejects physically optimal but logically invalid solutions without any explicit rule programming. The rejection emerges purely from differ-

entiable constraint satisfaction: high constraint losses generate large negative gradients that suppress assignments to violating drones. This is a form of *gradient-based semantic negotiation*, where competing loss components “argue” via backpropagation until an equilibrium is reached.

Interpretable Trade-offs. Unlike black-box optimization methods, the multi-modal loss decomposition (Equation 17) makes trade-offs explicit. An operator inspecting the final state can see that Drone 15 was chosen *because* it minimizes the weighted sum of distance and constraint violations, not due to an opaque scoring function. If the constraint weight λ were reduced (e.g., $\lambda = 5.0$ instead of 15.0), the system might accept a slight no-fly violation from Drone 0 to reduce travel distance, a deliberate trade-off that can be justified and audited.

Compositional Scaling. Each modality contributes an independent, differentiable loss term. Adding a fourth modality (e.g., fuel consumption limits) requires only defining $\mathcal{L}_{\text{Fuel}}$ and appending it to Equation 17. The gradient descent machinery handles the integration automatically. This compositional structure contrasts with monolithic reward engineering in reinforcement learning, where combining multiple objectives often requires extensive hyperparameter tuning to prevent mode collapse. The phase transition around step 60 (Figure 7, right panel) is particularly revealing. At this point, the cumulative constraint penalty from assigning to Drones 0 or 1 exceeds the efficiency gain from their proximity. Mathematically, the system crosses the threshold where:

$$\lambda \cdot (\mathcal{L}_{\text{Deontic}}^{(D_0)} + \mathcal{L}_{\text{Epistemic}}^{(D_0)}) > (d_{15} - d_0)$$

This inequality represents the *semantic boundary* between acceptable and unacceptable assignments. The fact that this boundary emerges from optimization rather than manual tuning suggests that MLNNs can discover context-appropriate trade-offs in complex multi-constraint problems.

10.5. Extensions and Limitations

Continuous Re-planning. The current implementation assumes static constraints. In dynamic environments where trust scores or no-fly zones change over time, the orchestrator could be re-trained online. Since optimization completes in 200 steps (≈ 0.1 seconds on a CPU), real-time re-planning is feasible for applications like adaptive swarm control.

Constraint Priority. All constraints currently share the same weight $\lambda = 15.0$. In practice, some constraints may be *hard* (must never violate, e.g., safety-critical no-fly zones) while others are *soft* (prefer to satisfy, e.g., efficiency). Extending the loss to per-constraint weights (λ_{Deontic} , $\lambda_{\text{Epistemic}}$ and $\lambda_{\text{Temporal}}$) would enable fine-grained priority specification.

Non-Convex Losses. The current loss landscape is relatively smooth due to the softmax assignment distribution, which spreads probability mass and prevents hard mode collapse. However, in scenarios with many constraints or highly nonlinear interactions, the optimization might encounter local minima. Techniques like simulated annealing or multi-start initialization could improve robustness.

The orchestration module validates the core MLNN thesis: that logical constraints from disparate modalities can be unified under a single differentiable objective, enabling gradient-based discovery of semantically valid solutions in complex multi-agent coordination problems.

11. Differentiable Swarm Communication

While orchestration addresses *which agent* should perform a task, communication addresses *what signals* to trust. This section demonstrates how modal logic enables *coordinated consensus*, agents automatically learn whose claims to weight through the Say-Do consistency axiom.

11.1. Problem Statement

Scenario: The Broken Scout

Context: A swarm of 16 drones performing environmental monitoring.

Configuration: 12 reliable drones (Gaussian noise $\sigma \approx 0.03$) and 4 broken drones (sensors biased toward ≈ 0.75 , regardless of ground truth).

Challenge: Standard averaging corrupts the consensus signal, when the true state is “clear” (0.0), the raw average reads $\bar{x} \approx 0.19$ due to the broken sensors.

Goal: Learn *who to coordinate with* without prior knowledge of which agents are faulty.

11.2. Trust-Weighted Consensus via Modal Logic

The key insight is that communication reliability is an *epistemic* property: we want to know which agents’ beliefs are accessible (trustworthy) for collective reasoning. We encode this through learnable trust parameters $A_\theta[i] \in [0, 1]$ for each agent i . Instead of naive averaging, we compute a *trust-weighted consensus*:

$$B_{\text{swarm}} = \frac{\sum_i A_\theta[i] \cdot \text{Claim}_i}{\sum_i A_\theta[i] + \epsilon} \quad (19)$$

The trust parameters are learned via the Say-Do consistency axiom. Intuitively: *if we trust an agent, its claims should match observable reality*. We introduce a tolerance threshold $\tau = 0.10$ to accommodate legitimate sensor noise:

$$\mathcal{L}_{\text{comm}} = \sum_i A_\theta[i] \cdot \text{ReLU}(|\text{Claim}_i - r| - \tau) \quad (20)$$

where r is the ground truth. This creates a natural “trust divergence”: agents with consistent error $< \tau$ incur zero loss and maintain high trust, while agents with systematic bias generate gradients that drive $A_\theta[i] \rightarrow 0$.

11.3. Experimental Results

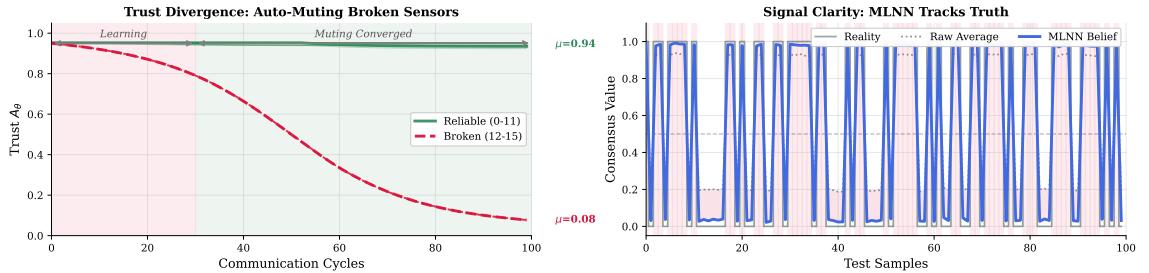


Figure 8: **Coordination via Differentiable Trust.** *Left:* Trust evolution over 100 communication cycles. Reliable agents (green) maintain $A_\theta \approx 0.94$ while broken agents (red) are progressively down-weighted to $A_\theta \approx 0.08$. *Right:* Signal quality on held-out test data. The MLNN consensus (blue) tracks ground truth closely, while raw averaging (dotted) remains systematically biased by the broken sensors.

We train for 100 communication cycles with batch size 16 and learning rate 0.05 and the system discovers the faulty agents without any supervision:

The learned trust weights yield an 81% reduction in mean absolute error compared to raw averaging (MAE: 0.025 vs. 0.132). Figure 8 visualizes the coordination dynamics.

11.4. Explainability: Reading the Learned Coordination

A key advantage of this approach over black-box alternatives is *direct interpretability*. The trust vector A_θ provides an explicit answer to “who is the system listening to?” Agents 0–11 express high confidence and include in consensus via high A_θ which is ≈ 0.94 and Agent 12–15 on the other hand are effectively muted with low confidence ($A_\theta \approx 0.08$)

This transparency is not available in standard neural network sensor fusion, where trust relationships are entangled in hidden layer weights. Here, the modal logical structure ensures that coordination decisions remain human-readable: we can point to $A_\theta[14] = 0.077$ and explain precisely why agent 14’s claims are discounted.

12. Conclusion

This tutorial has demonstrated that modal logic provides a principled vocabulary for multi-agent debugging, and that neurosymbolic implementation makes this vocabulary learnable. By treating accessibility relations as differentiable parameters optimized to minimize logical contradictions, DML and MLNNs can discover the hidden semantic structure of multi-agent systems, trust networks, causal chains, regulatory norms, and agent calibration, from behavioral data alone.

The key insights are: (1) semantic debugging is achievable where abstract concepts like trust, causality, and obligation can be operationalized as learnable parameters in a Kripke structure; (2) contradictions are informative, rather than avoiding logical inconsistency, we use it as the primary

training signal; (3) multiple modalities compose like epistemic, temporal, deontic, and doxastic constraints can be combined in a single differentiable system; and (4) the framework extends from monitoring to control—the same architecture that passively diagnoses issues can actively orchestrate self-healing behavior.

We hope this tutorial enables practitioners to apply modal logic to their own multi-agent debugging challenges, and inspires further research at the intersection of formal verification and machine learning.

References

- Samy Badreddine, Artur d’Avila Garcez, Luciano Serafini, and Michael Spranger. Logic tensor networks. *Artificial Intelligence*, 303:103649, 2022.
- Brandon C Colelough and William Regli. Neuro-symbolic ai in 2024: A systematic review. *arXiv preprint arXiv:2501.05435*, 2025.
- Rohkohl Erik. mlsolver: Modelling of multi-agent-system as Kripke structure. <https://github.com/erohkohl/mlsolver>, 2024.
- Ronald Fagin, Joseph Y Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT press, 2004.
- Artur d’Avila Garcez and Luis C Lamb. Neurosymbolic ai: The 3 rd wave. *Artificial Intelligence Review*, 56(11):12387–12406, 2023.
- Artur S d’Avila Garcez, Luis C Lamb, and Dov M Gabbay. Connectionist modal logic: Representing modalities in neural networks. *Theoretical Computer Science*, 371(1-2):34–53, 2007.
- Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.
- Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, 1962.
- Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025.
- Saurav Kadavath, Tom Conerly, Amanda Askell, Tom Henighan, Dawn Drain, Ethan Perez, Nicholas Schiefer, Zac Hatfield-Dodds, Nova DasSarma, Eli Tran-Johnson, et al. Language models (mostly) know what they know. *CoRR*, 2022.
- Ziyang Li, Jiani Huang, and Mayur Naik. Scallop: A language for neurosymbolic programming. *Proceedings of the ACM on Programming Languages*, 7(PLDI):1463–1487, 2023.

Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems—a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.

Xixun Lin, Yucheng Ning, Jingwen Zhang, Yan Dong, Yilong Liu, Yongxuan Wu, Xiaohua Qi, Nan Sun, Yanmin Shang, Kun Wang, et al. Llm-based agents suffer from hallucinations: A survey of taxonomy, methods, and directions. *arXiv preprint arXiv:2509.18970*, 2025.

Potsawee Manakul, Adian Liusie, and Mark Gales. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pages 9004–9017, 2023.

Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: Neural probabilistic logic programming. *Advances in neural information processing systems*, 31, 2018.

Sarvapali D Ramchurn, Dong Huynh, and Nicholas R Jennings. Trust in multi-agent systems. *The knowledge engineering review*, 19(1):1–25, 2004.

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

Antonin Sulc. Modal logical neural networks. *arXiv preprint arXiv:2512.03491*, 2025.

Xin Zhang and Victor S Sheng. Neuro-symbolic ai: Explainability, challenges, and future trends. *arXiv preprint arXiv:2411.04383*, 2024.

Appendix A. Implementation

A.1. Minimal Implementation of MLNN

The following code provides a complete, minimal MLNN implementation:

Listing 1: Core MLNN Implementation

```
import torch
import torch.nn as nn

class MinimalMLNN(nn.Module):
    def __init__(self, n_worlds):
        super().__init__()
        # Learnable accessibility relation
        self.A_logits = nn.Parameter(torch.zeros(n_worlds, n_worlds))

    def accessibility(self):
        return torch.sigmoid(self.A_logits)

    def necessity(self, phi_values):
```

```

"""Box operator: true_in_all_accessible_worlds"""
A = self.accessibility()
# Lukasiewicz implication: (1 - A) + phi
implications = (1 - A).unsqueeze(-1) + phi_values.unsqueeze(0)
return implications.clamp(max=1.0).min(dim=1).values

def contradiction_loss(self, antecedent, consequent):
    """Loss for implication: antecedent -> consequent"""
    return torch.relu(antecedent - consequent).mean()

```

A.2. Implementation - Epistemic Logic

The following code provides a complete, minimal MLNN example of implementation of Epistemic Logic shown in Section 6.

Listing 2: Epistemic Trust Learning

```

class DiplomacyTrustMLNN(nn.Module):
    def __init__(self, num_agents=5):
        super().__init__()
        # Trust matrix: entry [i,j] = Agent i's trust in Agent j
        # Initialize with high trust ("innocent until proven guilty")
        self.trust_logits = nn.Parameter(
            torch.ones(num_agents, num_agents) * 2.5
        )
    def forward(self, sender_idx, receiver_idx, intent, observed_reality):
        trust = torch.sigmoid(self.trust_logits[receiver_idx, sender_idx])
        # Lukasiewicz conjunction: max(0, intent + trust - 1)
        antecedent = torch.relu(intent + trust - 1.0)
        # Contradiction: antecedent > reality means violation
        return torch.relu(antecedent - observed_reality), trust

```

A.3. Implementation - Orchestration

The orchestrator shown in Sect 10 is implemented as a simple neural network with a single learnable parameter vector representing assignment logits:

Listing 3: Multi-Modal Orchestrator Implementation

```

class Orchestrator(nn.Module):
    def __init__(self, n_drones):
        super().__init__()
        # Learnable assignment distribution (softmax over these logits)
        self.logits = nn.Parameter(torch.zeros(n_drones))

    def forward(self, env):
        # Normalize to valid distribution
        assign = torch.softmax(self.logits, dim=0)

```

```

# Efficiency: weighted distance
dists = torch.norm(env.pos - env.target, dim=1)
eff_loss = torch.sum(assign * dists)
# Deontic: no-fly zone violations
deontic_loss = env.check_no_flyViolation(assign)
# Epistemic: trust violations
epistemic_loss = torch.sum(assign * (1.0 - env.trust_scores))
# Temporal: scheduling conflicts
temporal_loss = torch.sum(assign * env.conflicts)
return assign, eff_loss, deontic_loss, epistemic_loss, temporal_loss

```

We optimize using Adam with learning rate $\alpha = 0.1$ for 200 steps. The assignment logits are initialized to zero (uniform distribution). The high learning rate enables rapid exploration and convergence within the limited optimization budget.

A.4. Swarm Communication: Learning Dynamics

This appendix provides additional detail on the learning mechanism and training dynamics for the differentiable swarm communication example presented in Section 11.

A.4.1. THE LEARNING MECHANISM

The Say-Do consistency axiom (Equation 20) creates a self-correcting learning signal. Figure 9 illustrates how contradictions between agent claims and observed reality drive trust updates.

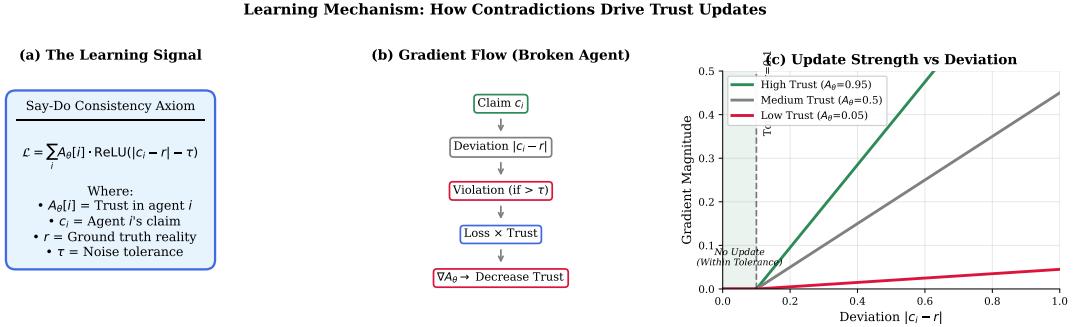


Figure 9: Learning Mechanism for Trust Updates. (a) The Say-Do consistency loss penalizes trusted agents whose claims deviate from reality beyond tolerance τ . (b) Gradient flow for a broken agent: deviations exceeding τ generate loss proportional to current trust, and backpropagation reduces A_θ . (c) Update strength as a function of deviation magnitude. Agents within the tolerance zone ($|c_i - r| < \tau$) receive no gradient, preserving trust for legitimate sensor noise. High-trust agents with large deviations receive the strongest corrective signal.

The tolerance threshold $\tau = 0.10$ serves two purposes. First, it provides robustness to sensor noise, reliable agents with Gaussian noise $\sigma = 0.03$ rarely exceed this threshold and thus maintain

stable trust. Second, it creates a “dead zone” that prevents oscillation: once an agent’s trust drops sufficiently low, its contribution to the loss becomes negligible, stabilizing the learning dynamics.

A.4.2. TRAINING DYNAMICS

Figure 10 shows the full trust evolution over 100 training epochs. The learning process exhibits three distinct phases:

Phase 1: Learning (Epochs 0–25). All agents begin with high trust ($A_\theta \approx 0.95$). The system has not yet accumulated sufficient evidence to distinguish reliable from broken agents. Loss remains elevated as broken agents contribute contradictions.

Phase 2: Transition (Epochs 25–60). Trust divergence accelerates. Broken agents cross the neutral threshold ($A_\theta = 0.5$) as cumulative axiom violations compound. The loss begins to decrease as broken agents’ contributions are down-weighted.

Phase 3: Converged (Epochs 60–100). Trust parameters stabilize. Reliable agents settle at $A_\theta = 0.944 \pm 0.010$, while broken agents asymptote toward $A_\theta \approx 0.08$. The residual trust in broken agents reflects the diminishing gradient magnitude as $A_\theta \rightarrow 0$ (visible in Figure 9c).

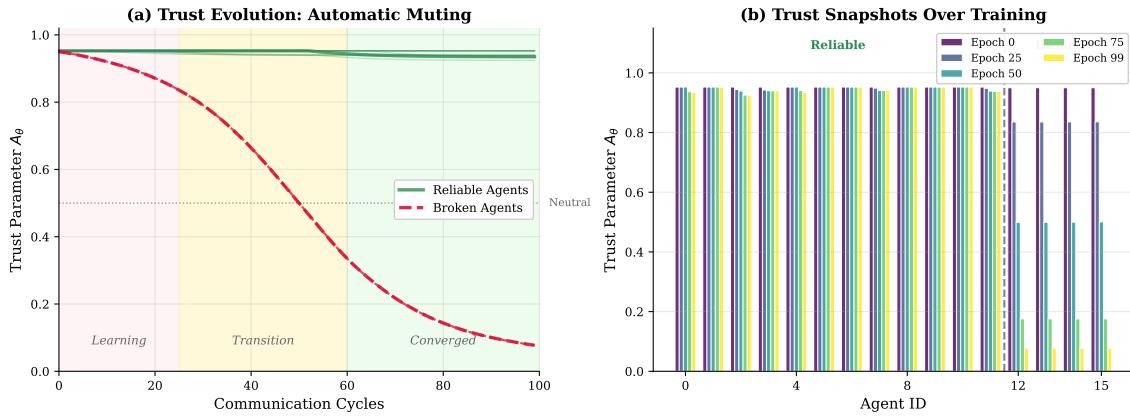


Figure 10: **Trust Evolution During Training.** (a) Trust trajectories for all 16 agents. Reliable agents (green, solid) maintain high trust throughout training with minimal variance ($\sigma = 0.010$). Broken agents (red, dashed) follow a sigmoid-like decay as accumulated contradictions drive $A_\theta \rightarrow 0$. Background shading indicates learning phases. (b) Trust snapshots at epochs 0, 25, 50, 75, and 99, showing the progressive separation between reliable (agents 0–11) and broken (agents 12–15) groups.

A.4.3. HYPERPARAMETER SENSITIVITY

The key hyperparameters are the learning rate ($\eta = 0.05$), tolerance threshold ($\tau = 0.10$), and batch size ($B = 16$). We found that:

- **Learning rate:** Values above $\eta > 0.15$ cause unstable trust for reliable agents due to occasional noise-induced violations. Lower values ($\eta < 0.03$) slow convergence but improve final trust stability.
- **Tolerance:** Setting τ below the reliable agent noise level ($\sigma = 0.03$) incorrectly penalizes reliable agents. Setting τ too high (> 0.5) delays detection of broken agents.
- **Batch size:** Larger batches ($B \geq 16$) smooth the loss landscape, reducing variance in trust updates. Single-sample updates ($B = 1$) produce noisy training curves but converge to similar final values.

The configuration used in Section 11 ($\eta = 0.05$, $\tau = 0.10$, $B = 16$) provides stable learning while achieving clear trust separation within 100 epochs.