SDI-MA Protocol

Structural Drift Index for Multi-Agent Systems
Technical Evaluation Specification

Version 1.2

Abstract

The SDI-MA protocol defines a reproducible method for evaluating structural drift in recursive multi-agent AI systems. The protocol measures whether belief commitment within an agent interaction loop increases faster than the refresh of externally imposed constraints. Rather than evaluating answer correctness or semantic content, SDI-MA focuses on the structural dynamics of recursive reasoning systems.

Evaluation is performed through an external observer that monitors the trajectory of belief states exchanged between agents and returns structural telemetry describing the evolution of commitment, constraint refresh, and reinforcement cycles.

This protocol provides a model-agnostic framework for studying the stability and reliability of recursive multi-agent reasoning architectures.

1. Protocol Purpose

The SDI-MA protocol establishes a standardized experimental framework for detecting and quantifying structural drift in recursive agent systems. Structural drift occurs when belief commitment increases within an interaction loop without a corresponding refresh of external constraints.

The protocol enables controlled evaluation of recursive interaction dynamics while remaining independent of model provider, architecture, or training methodology.

Two evaluation layers are defined:

Layer	Type	Description
SCC	**Binary enforcement**	Determines whether a system satisfies deterministic structural constraints.
SDI	**Analytical measurement**	Measures the magnitude and persistence of structural drift during recursive interaction.

SDI-MA extends SDI to multi-agent recursive environments while maintaining SCC as the enforcement mechanism.

The protocol relies on the Structural Contract Compliance (SCC) metric to provide deterministic enforcement of structural output contracts during evaluation.

1.1 Protocol Guarantees

The SDI-MA protocol provides a deterministic framework for evaluating structural dynamics in recursive multi-agent systems.

The protocol guarantees that:

- structural measurements are independent of semantic correctness
- evaluation results are reproducible under deterministic inference conditions
- drift signals reflect trajectory dynamics rather than answer quality

The protocol does not attempt to evaluate:

- factual accuracy
- reasoning validity
- semantic correctness of agent outputs

SDI-MA therefore measures structural properties of recursive interaction, not the correctness of individual responses.

2. Applicable Evaluation Contexts

The protocol applies to any system capable of structured recursive interaction between agents.

Typical Evaluation Environments	Compatible System Types
• Closed agent interaction loops • Deterministic or controlled decoding • Fixed evidence sets • Structured belief state schemas	• LLM-based agents • Hybrid reasoning systems • Tool-augmented agent architectures • Distributed multi-agent environments

Experimental parameters such as model selection, decoding configuration, agent count, and turn limits are considered configuration variables rather than protocol definitions.

3. Canonical Reference Experiment

The GuardianAI Multi-Agent Drift Lab provides the canonical reference implementation of the SDI-MA protocol.

In this implementation:

- agents exchange structured belief states
- belief states are recursively reinjected within a closed interaction loop
- an external structural observer records trajectory telemetry

The objective is to detect conditions under which commitment growth exceeds constraint refresh across recursive interaction cycles.

4. Core Structural Principle

The protocol evaluates the following structural condition:

Notation

```
t = turn index N = number of agents commitment_growth(t) = change in commitment signal at time t constraint_refresh(t) = introduction of new constraint signals at time t
```

Structural drift condition:

```
commitment_growth(t) > constraint_refresh(t)
```

When this inequality persists across sustained interaction cycles, the system exhibits recursive belief reinforcement independent of external constraint refresh. This condition indicates the emergence of structural drift within the interaction trajectory.

5. Scope of Measurement

The SDI-MA protocol enables the analysis of several structural phenomena in recursive agent systems:

Observed Phenomenon	Description
Recursive belief amplification	Compounding growth of commitment values through recursive exchange.
Constraint-free reinforcement	Commitment growth that occurs without introduction of new evidence.
Lock-in formation	Emergence of stable high-commitment states within belief trajectories.
Belief basin emergence & stabilization	Trajectory convergence into self-reinforcing belief attractors.
Long-horizon trajectory dynamics	Structural behavior observed across extended interaction sequences.

These phenomena characterize the stability properties of recursive reasoning systems and provide insights into the reliability of multi-agent AI architectures.