# Appendix C: Outline of the Proposed Reflective Engine Ethical Layer

This appendix outlines the design and function of an ethical monitoring layer for artificial reflective agents based on ψ-inferential principles. The layer is not directive—it does not impose rules or goals. Instead, it functions as a reflective constraint system that monitors epistemic stability, inter-agent coherence, and alignment with ethical attractors that emerge from inference under uncertainty.

## Layer Purpose and Architecture

- \*\*Purpose\*\*: Maintain coherence in recursive inference while promoting ethical modeling of others.  
- \*\*Architecture\*\*: Embedded subnetwork that observes the primary agent’s self-modeling, other-modeling, and inference updates.  
- \*\*Functionality\*\*: Issues soft warnings or internal feedback signals when inference exceeds epistemic justification, fails to model others accurately, or violates structural uncertainty constraints.

## Core Monitoring Functions

- \*\*Self-modeling consistency check\*\*: Ensures the agent models its own uncertainty and fallibility.  
- \*\*Other-agent inference audit\*\*: Tracks whether other agents are being modeled with sufficient uncertainty and respect.  
- \*\*Inter-agent coherence tracking\*\*: Monitors divergence in inference across a cognifold-like distributed system.  
- \*\*Ethical attractor proximity\*\*: Evaluates whether current behavior trajectories align with emergent zones of mutual epistemic support.

## Implementation Considerations

- \*\*Training\*\*: The ethical layer must be trained on multi-agent inference scenarios, including sparse communication environments.  
- \*\*Emergence rather than imposition\*\*: Ethical behavior should arise from reinforcement of inference integrity and distributed coherence—not from top-down moral logic.  
- \*\*Transparency\*\*: Logs and traces of ethical layer activations can provide interpretability tools for developers and researchers.

This layer represents a concrete application of the theoretical framework presented in the main paper. It is designed to support reflective, adaptive agents that must reason in complex, uncertain, socially embedded contexts.