Reflective-Ethical Engine: Integrated Architecture v2

This document outlines the integration of Yann LeCun’s proposed architecture for autonomous machine intelligence with a Reflective-Ethical Engine. The goal is to support ethically grounded, autonomous cognitive agents that combine planning, world modeling, and internal reflection to act safely and with purpose in complex environments.

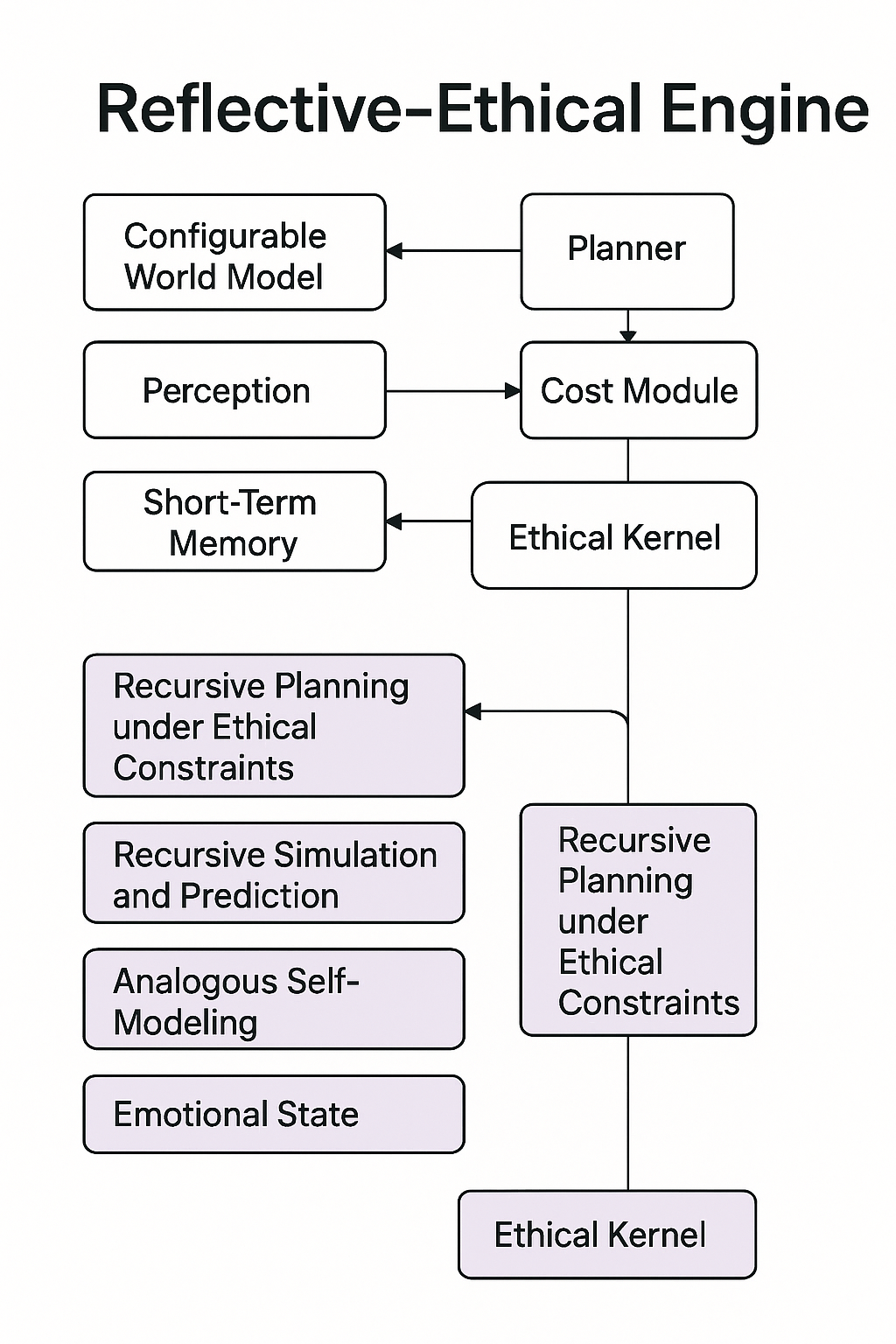


Figure 1: A hybrid architecture integrating LeCun’s modular intelligence framework with a Reflective-Ethical Kernel. The upper modules represent LeCun’s original components (e.g., world model, planner, cost module), while the lower modules add recursive planning, self-modeling, emotional state representation, and ethical reasoning. Arrows indicate flow of influence and control, with the Ethical Kernel guiding recursive layers toward alignment and safe behavior.

# LeCun’s Architecture Modules

1. Configurable World Model (CWM): Learns compact, latent representations and simulates futures.

2. Perception Module: Translates raw sensory inputs into representations for the world model.

3. Short-Term Memory (STM): Stores recent interaction history and context.

4. Cost Module: Defines goals, shaped by scalar objectives or learned ethics.

5. Planner: Simulates outcomes and chooses actions based on predictions and cost.

6. Action Module: Executes actions in the environment.

# Reflective-Ethical Kernel Additions

1. Ethical Kernel: Encodes ethical axioms (e.g., love, uncertainty) and modulates cost function.

2. Recursive Planning under Ethical Constraints: A control layer enforcing ethical planning bounds.

3. Recursive Simulation and Prediction: Enhances world modeling with counterfactual and emotional simulation.

4. Analogous Self-Modeling: Supports self-awareness by comparing agent states to inferred others.

5. Emotional State: Represents internal affective valence, modulating simulation and action priorities.

# Integration Points

- The Ethical Kernel interfaces with the Cost Module to shape intrinsic motivation.  
- Recursive simulation improves the World Model’s robustness.  
- Recursive Planning layers ensure long-term planning adheres to ethical axioms.  
- Short-Term Memory provides dynamic state inputs for all reflective processes.  
- Emotional State acts as a bridging layer between internal simulation and motivational signaling.