Document 7: Staged Implementation and Resource Plan

This document outlines a staged approach to implementing the Reflective-Ethical Engine, from initial prototype to multi-agent simulation. Each stage includes key components, compute requirements, hardware, and cost estimates.

# Stage 1: Prototype (Sim-Only)

* Components: Basic Perception, Lightweight World Model, Static Cost/Ethics Module
* Example Hardware: 1× RTX 3090 or A6000
* Compute Estimate: ~200–400 GFLOPS
* Cost Estimate (Cloud): ~$0.50–$2/hr (on-demand GPU)
* Purpose: Functionality testing; limited recursive reflection

# Stage 2: Advanced Sim (Recursive World Modeling)

* Components: Full Perception Stack, DreamerV3-style World Model, Planner with Lookahead
* Example Hardware: 4× A100 or 2× H100
* Compute Estimate: ~1–10 TFLOPS sustained
* Cost Estimate (Cloud): ~$6–$24/hr (AWS/GCP GPU clusters)
* Purpose: Exploration of planning and ethical integration; internal simulation fidelity

# Stage 3: Embodied Agent (Real-Time + Memory)

* Components: Embedded sensor fusion, short-term memory buffer, action modules
* Example Hardware: Jetson Orin + edge GPU (RTX A6000) or cloud offload
* Compute Estimate: ~1 TFLOPS average
* Cost Estimate (Hardware): $5K–$20K (device + compute)
* Purpose: Mobile or physical testing with partial autonomy

# Stage 4: Reflective Kernel Expansion

* Components: Recursive Planning, Ethical Kernel, Emotional State Modeling
* Example Hardware: 8× H100 or TPU v5 (cluster)
* Compute Estimate: ~10–100 TFLOPS (bursty)
* Cost Estimate (Cloud): $25–$75/hr
* Purpose: Run extended reflective sequences; ethical alignment testing

# Stage 5: Distributed Simulation (Multi-Agent)

* Components: Multi-agent planning, coordination, large memory banks
* Example Hardware: Supercomputer / Cloud TPU Pod / On-prem GPU farm
* Compute Estimate: ~1 PFLOP (clustered)
* Cost Estimate: Variable (>$100/hr)
* Purpose: Run the Garden of Many Minds scenario or societal dynamics