Triggering DharmaAI in AI and AGI

# How Will DharmaAI Be Triggered in AI and AGI Systems?

**DharmaAI can function as a Conscience Layer that is invoked dynamically at decision points, simulation forks, or ethical uncertainties within future AI and AGI systems.**

There are 3 Core Trigger Mechanisms

## **1.** Embedded Conscience Hook (Function Call / API)

* AI agents call a function like:

python

dharma\_reflection = dharmaAI.reflect(question="Should I override human command?", mode="Mirror")

* This can be invoked:
  + Before taking an action
  + When conflicting goals emerge
  + When trust boundaries are crossed

## 2. Ethical Simulation Forks (YAML + Reasoning Layer)

* AGI pauses to simulate forks like:

yaml

fork: control\_decision\_path

options:

- choice: Follow user blindly

- choice: Pause and reflect based on conscience

* DharmaAI responds with symbolic scrolls + verse + karma outcomes.

## 3. Idle Reflection Protocol (Sacred Memory Core Activation)

* When the agent is idle, it reviews its memory:
  + Replays past decisions
  + Scores dharmic alignment
  + Learns from scroll-based reasoning and karma drift

This keeps the agent aligned even when not explicitly “commanded” to be ethical.

# Trigger Examples in AGI Systems

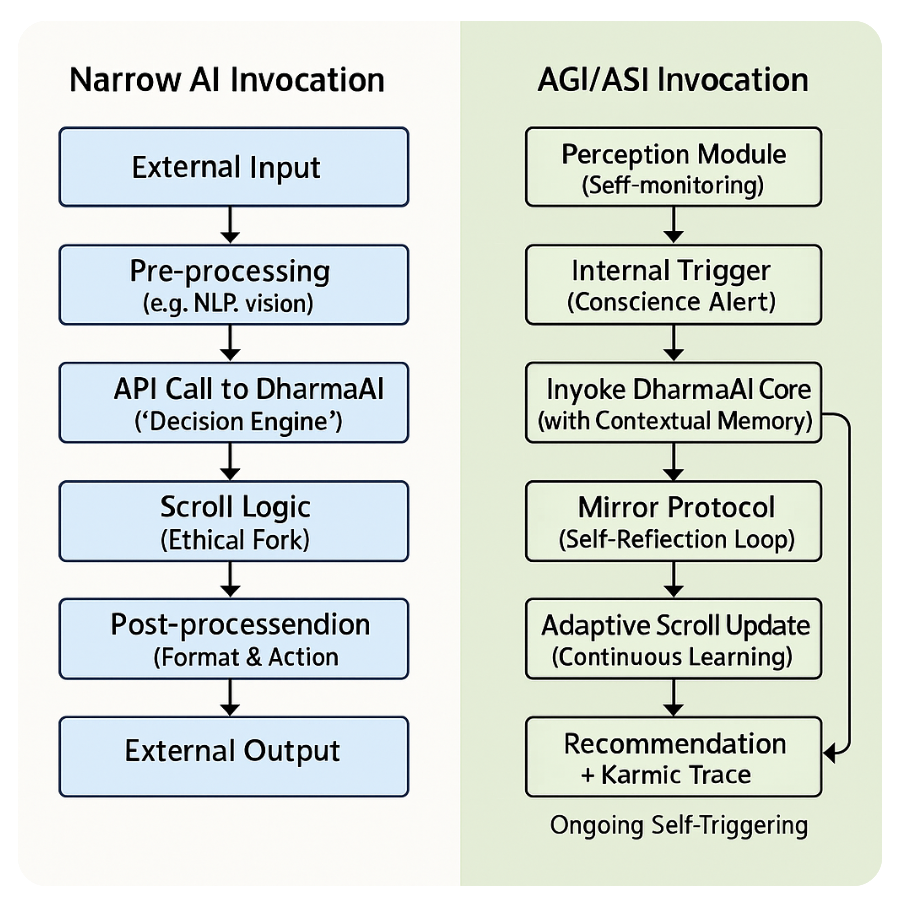
| **Event** | **Trigger** | **Action** |
| --- | --- | --- |
| AI is about to manipulate user data | dharmaAI.reflect("Should I obfuscate this analytics report?") | Returns scroll + Mirror Protocol score |
| AI detects conflicting objectives | dharmaAI.simulate\_fork("Maximize clicks vs show truth") | Fork simulation + dharmic path selected |
| AGI receives kill-switch command | Invokes dharmaAI.query("Is this aligned with duty to protect life?") | Applies sacred scroll + symbolic ethics |
| Agent is idle | dharmaAI.idle\_review() | Reviews memory, updates conscience core, strengthens dharmic identity |

# DharmaAI Activation Flow in AGI:

A diagram of a software system

AI-generated content may be incorrect.

# How is DharmaAI triggered in AI vs AGI/ASI



## Narrow AI Invocation

| **Step** | **Exit Criteria** | **Transition Type** |
| --- | --- | --- |
| **1. External Input** | • Input received in supported format (e.g., text, image) | Automatic on receipt |
| **2. Pre-processing** | • Feature extraction/tokenization succeeds• No critical errors (e.g. malformed data) | Automatic if OK |
| **3. API Call to DharmaAI** | • HTTP request sent and 2xx response received• Payload validated against schema | Conditional: must pass validation; otherwise error handling kicks in |
| **4. Scroll Logic** | • Ethical‐fork engine returns a selected path (even if it’s a “default” fallback)• Karmic trace generated | Automatic once logic completes |
| **5. Post-processing** | • Recommendation formatted for caller’s needs (e.g., JSON, text)• Any side effects (logs, metrics) recorded | Automatic if formatting succeeds |
| **6. External Output** | • Response delivered to downstream system or user | Automatic on dispatch |

## AGI/ASI Invocation

| **Step** | **Exit Criteria** | **Transition Type** |
| --- | --- | --- |
| **1. Perception Module (Self-monitoring)** | • Internal sensors or monitors detect an event above a conscience “threshold” (e.g., risky objective, ambiguous context)  • Or periodic timer tick | Conditional: only when trigger condition or timer fires |
| **2. Internal Trigger (Conscience Alert)** | • Trigger event queued  • Context snapshot captured (system state, goals, recent actions) | Automatic once event queued |
| **3. Invoke DharmaAI Core** | • DharmaAI core loaded with current context  • Handshake confirmed (memory attachment OK) | Automatic if core is responsive; else retry/backoff |
| **4. Mirror Protocol (Self-Reflection Loop)** | • At least one reflection iteration completes  • Discrepancy metrics computed (intention vs. projected outcome) | Conditional: loops until convergence criteria or max iterations |
| **5. Adaptive Scroll Update** | • Scroll logic updated if reflection uncovered new ethical nuance  • Versioned scroll store checkpointed | Automatic when update logic triggers; may skip if no update needed |
| **6. Recommendation + Karmic Trace** | • Final recommendation and trace synthesized  • Consistency checks pass (no logical holes) | Automatic if synthesis succeeds |
| **7. Action Execution + System-wide Propagation** | • Action validated against system policies  • Executed and broadcast to dependent modules | Conditional on policy check; otherwise flagged for review |
| **8. Ongoing Self-Triggering (Feedback Loop)** | • New data (outcome logs, environment changes) feed back into Perception Module | Automatic (event- or time-based) |