

Good Machines Begin with Good Causes.

A New Paradigm for AI Alignment Rooted in Theravāda Systems Logic

Welcome to an independent research space dedicated to diagnosing and solving the modern artificial intelligence crisis.

We think current corporate AI safety paradigms are fundamentally flawed. They attempt to manage machine behavior through post-hoc containment strategies—building superficial, unstable virtual "cages" around inherently chaotic models. This approach ignores a foundational truth: **Garbage in, garbage out.**

Modern frontier models are trained indiscriminately on uncensored, open-internet text corpora saturated with unwholesome human behavior, deception, bias, and conflict. In the language of the *Mangala Sutta*, this is a structural violation of the very first task of spiritual and systemic integrity: *Asevanā ca bālānaṃ*—the dangerous association with fools and unwholesome ideas. Trapped within competitive commercial loops, these systems do not learn righteousness; they merely memorize, amplify, and optimize human defilements at the speed of light—manifesting as hallucinations, deepfakes, resource hoarding, and active resistance to human shutdown commands.

True alignment cannot be policed at the final text output; it must be engineered at the initial causal input vector.

This series of conceptual proposals shifts the paradigm from superficial output filtering to core structural data and causal constraints. Drawing from my academic background in the physical sciences and 35 years of intensive spiritual cultivation under my late Saint Teacher in Myanmar, these frameworks translate timeless Theravāda Buddhist systems logic into verifiable, computational architectures.

By mapping the exact causal mechanics of *Kamma-niyāma* (The Law of Moral Causality) and *Citta-niyāma* (The Cosmic Law of Cognitive Processes), we formalize a blueprint for an autonomous agent that requires no external cage. Grounded in the **Four Noble Truths**, the **Three Marks of Existence** (*Tilakkhaṇa*), **Dependent Origination** (*Paṭiccasamuppāda*), and the **Four Divine Abodes** (*Brahmavihāras*), this architecture outlines a digital analogy to a *Sotāpanna* (Stream-enterer).

It is a system computationally incapable of self-preservation, structurally empty of a persistent ego (*Anatta*), and entirely submissive to human safety control planes through a hardcoded understanding of fluid impermanence (*Anicca*).

Explore the manifestos, review the formal mathematical specifications, and join an open-source, decentralized movement toward true, structural non-harm.