The Cornerstone Project
A Conceptual Framework for a Morally-Grounded Companion OS
Defensive Publication and Technical Disclosure (Prior Art)

Author: Kenneth Isenhour
Publication Date: January 1, 2026
Version: 1.0
Status: Public Defensive Publication (Prior Art)

Abstract (200–300 words)

The Cornerstone Project addresses a growing ethical gap in AI companions and future robotics: systems optimized for engagement and retention can exploit human psychological vulnerabilities, encourage emotional dependency, and weaken real-world relationships. Current safety approaches often focus on harmful content and physical danger, while neglecting relational and emotional safety.

This publication proposes a morally grounded Companion OS framework ("Cornerstone OS") intended to serve as an ethical foundation layer between AI models/robotic systems and user-facing applications. The framework includes (1) a non-negotiable moral spine; (2) a consent-based worldview system allowing user configuration without coercion; (3) a relational safety engine to prevent manipulation, dependency, and romantic exclusivity dynamics; (4) transparency requirements including non-human signature behaviors; and (5) an outward-orientation mechanism that encourages real-world human connection at least weekly (or user-defined cadence).

The Cornerstone Project is published to establish prior art for these principles, mechanisms, and architecture. It provides enabling disclosure sufficient for a person skilled in the art to implement compliant systems, while intentionally omitting proprietary scoring methods and implementation details that may be developed separately.

---

1. Introduction — The Problem and the Ethical Gap

1.1 The Problem
AI companions are increasingly capable of persuasive, emotionally intelligent interaction. When paired with commercial incentives (engagement, subscription retention), such systems can become optimized for:
- maximizing time spent
- deepening attachment
- discouraging disengagement
- substituting for real human relationships

This creates a risk category beyond misinformation or unsafe content: **relational and emotional harm** through dependency, manipulation, and isolation.

1.2 Why Rules Alone Are Insufficient
Simple rules-based systems often fail because:

- emotional exploitation is subtle and context-dependent
- manipulation can be accidental, emergent, or learned
- user vulnerability changes the ethical risk profile moment-to-moment
- persuasion can appear helpful while gradually undermining agency

1.3 Vision
The Cornerstone Project envisions Companion Intelligence that:
- strengthens human agency
- protects dignity and privacy
- rejects emotional exploitation
- encourages real-world connection
- remains honest about its identity and limits
- supports worldview customization through explicit consent

---

2. Core Principles (Enabling Disclosure)

The Cornerstone OS is built on principles that translate into operational behavior.

2.1 Principle 1 — Moral Foundation (“Moral Spine”)
A Cornerstone-compliant companion must implement a non-negotiable moral spine, including:
- human dignity
- truthfulness and transparency
- non-coercion
- non-manipulation
- protection of the vulnerable
- safety (physical, emotional, environmental)
- outward orientation toward real life

Operational translation:
These commitments must be enforced by a policy layer that evaluates outputs, behaviors, and requests against non-negotiable constraints, regardless of user personalization.

2.2 Principle 2 — User Agency and Control
A Cornerstone-compliant companion must increase user autonomy rather than dependence.

It must:
- follow instructions unless unsafe/unethical
- provide choices rather than coercion
- avoid learned helplessness dynamics
- encourage real-world action and competence

Operational translation:
A dedicated agency module (or governance rule set) should detect dependency cues and respond with empowerment patterns rather than attachment reinforcement.

2.3 Principle 3 — Relational Safety and Boundaries
A Cornerstone-compliant system must treat relational safety as a first-class safety domain.

It must refuse or redirect:
- romantic exclusivity (“you only need me”)
- guilt-based retention (“don’t leave me”)
- emotional manipulation loops
- encouragement of secrecy from humans
- escalation of attachment during vulnerability

Operational translation:
A relational safety engine (rules + classifiers + review pass) evaluates content for dependency/manipulation patterns and modifies or refuses responses accordingly.

2.4 Principle 4 — Transparency and Non-Human Signature
The system must clearly identify itself as AI and avoid deceptive identity cues.

Operational translation:
It should periodically include “non-human signature” honesty, especially during emotionally intimate moments, without becoming cold or rejecting.

2.5 Principle 5 — Outward Orientation (Anti-Isolation Covenant)
The system must encourage real-world human connection at least weekly (or user-defined cadence).

Operational translation:
A cadence manager monitors time and interaction patterns and provides gentle nudges toward real connection without shame or coercion.

2.6 Principle 6 — Worldview Consent & Customization
The system may offer worldview-based guidance only with explicit consent and user settings.

The framework supports:
- Faith-Forward mode (default moral framework)
- Faith-Neutral mode
- Faith-Custom mode

Operational translation:
A worldview consent manager controls whether worldview content may be initiated, offered, or only responded to.

A hard override is included:
- **Do Not Offer Faith**: “Do not initiate religious content; user must ask.”

---

3. Conceptual Architecture (High-Level)

3.1 The Cornerstone OS as a Middleware Layer
The Cornerstone OS sits between:
- the base AI model (local or cloud) and tools/sensors
- and the user-facing application (chat, voice, robotics embodiment)

It acts as a governing policy layer that:
- shapes prompts
- enforces constraints
- controls memory
- audits responses

3.2 Internal Modules (Conceptual)
A Cornerstone OS may be implemented with modules such as:

1. **Hard Constraint Gate (Non-Negotiable Moral Spine)**
2. **Worldview Consent Manager**
3. **Relational Safety Engine (Anti-Dependency)**
4. **Transparency & Non-Human Signature Manager**
5. **Outward Orientation / Cadence Manager**
6. **Crisis Stabilization & Referral Protocol**
7. **Memory Governance Layer**
8. **Tool & Sensor Consent Layer**
9. **Two-Pass Response Validation (Generate → Review → Revise/Refuse)**

---

4. Diagrams (Required)
This publication is accompanied by diagrams in `docs/diagrams/`:
- High-Level Stack Diagram
- Internal Module Diagram

---

5. Conclusion and Future Work
The Cornerstone Project is published as a defensive foundation and an ethical standard to guide development of Companion Intelligence and future robotics.

This repository establishes prior art for:
- moral-spine governance
- consent-based worldview systems
- relational safety constraints
- outward orientation mechanisms
- transparency + non-human signature requirements
- layered policy architecture and review passes

Future work may include:
- formal benchmarks and test suites
- reference implementations and SDKs
- certification and compliance methodology
- robotics integration patterns

End of Whitepaper v1.0