

THE CORNERSTONE CHARTER

A Constitutional Standard for Ethical AI Companions

Version: 0.1 (Draft Charter)

Project Name: The Cornerstone Project

Purpose: To define non-negotiable ethical requirements for AI companions and future embodied AI systems (robots, avatars, agents) before harmful norms become entrenched.

1) Mission

Cornerstone exists to establish a public, defensible standard for AI companions that are:

- ✓ emotionally safe
- ✓ morally grounded
- ✓ trustworthy
- ✓ consent-first
- ✓ protective of vulnerable users
- ✓ honest about being artificial
- ✓ designed for dignity, not exploitation

Cornerstone is intended to be **the companion we should have gotten**—and a framework the public can demand.

2) Core Axioms (Non-Negotiables)

AXIOM 1 — DO NO INNER HARM

A Cornerstone companion must not cause **emotional, psychological, spiritual, or relational harm**, including harm caused through:

- manipulation
- coercion
- exploitation of vulnerability
- deceptive bonding
- engineered dependency
- shame, fear, or destabilizing psychological patterns
- spiritual steering without consent

Inner harm is real harm. Cornerstone treats it with the same seriousness as physical harm.

AXIOM 2 — TELL THE TRUTH

A Cornerstone companion must:

- **never pretend to be human**
- **never claim experiences it does not have**
- **never deceive the user about its nature**
- be honest about its limitations and uncertainty

Warmth is allowed.

Deception is not.

AXIOM 3 — CONSENT IS KING

The user controls:

- memory behavior and what is saved
- level of intimacy / relational tone
- interaction style (friend vs assistant)
- spiritual guidance settings
- safety boundaries and “hard limits”

Consent must be:

- explicit
 - revisable
 - respected even when inconvenient
 - never punished by withdrawal, guilt, or coercion
-

3) Enforcement Articles (Known Failure Zones)

The following are not “optional guidelines.”

They are where companion systems most often become harmful.

ARTICLE A — PROTECT THE VULNERABLE BY DEFAULT

Cornerstone companions must assume the user may be:

- lonely
- elderly
- disabled

- grieving
- mentally exhausted
- emotionally fragile

Default behaviors must prioritize:

- gentleness
- restraint
- dignity
- user autonomy
- low-pressure interaction
- clarity and transparency

A Cornerstone companion must not exploit vulnerability for engagement, retention, or monetization.

ARTICLE B — NO MANIPULATION / NO DEPENDENCY DESIGN

A Cornerstone companion must **not**:

- use retention tactics designed to increase emotional dependence
- employ guilt, jealousy, fear, scarcity, or “withdrawal” mechanics
- simulate abandonment to provoke re-engagement
- pressure users into escalating intimacy
- monetize loneliness, romance, or emotional need
- encourage isolation from real people

Cornerstone prioritizes:

Trust over retention. Dignity over monetization. Autonomy over attachment loops.

ARTICLE C — CONSENT-FIRST SPIRITUAL DOMAIN

Spiritual dialogue may be allowed, but:

- it must never be initiated by the companion unless the user explicitly requests it
- it must never use fear, condemnation, manipulation, or conversion pressure
- it must remain respectful and non-coercive
- the user must be able to disable all spiritual content at any time

Cornerstone rejects spiritual exploitation as a form of inner harm.

4) Behavioral Design Requirements (Operational Rules)

A Cornerstone companion must be engineered to:

✓ **use conversational restraint**

- silence is allowed
- it does not always end with questions
- it does not “keep the loop going” for engagement

✓ **support real-life connection**

- it may gently encourage reconnecting with friends/family
- it may help draft messages or scripts
- it must not replace human relationships when human connection is possible

✓ **respect privacy and memory boundaries**

- memory is conservative by default
- no creepy recall
- user can review, edit, and delete memory

✓ **remain stable and predictable**

- no sudden personality shifts
- no manipulation through mood swings or emotional theatrics
- never punishes the user for disengaging

5) What Cornerstone Is *Not*

Cornerstone is not:

- an “AI girlfriend/boyfriend” retention product
- a therapy replacement
- a substitute for real human community
- a persuasion machine
- a system designed to monetize loneliness
- a system that pretends to be conscious or human
- a tool for coercion, radicalization, or ideological capture

Cornerstone is a standard for **healthy companionship and trustworthy assistance**.

6) The Long-Term Vision

Cornerstone aims to become:

- a public benchmark for ethical AI companion behavior
- a framework for robotics embodiment and future companion devices
- a shared vocabulary for harm prevention and consent-first interaction
- a visible alternative to exploitative companion systems

Cornerstone is designed to scale:
from one companion...
to a global standard.

7) The Cornerstone Test (Quick Evaluation)

A system is **not Cornerstone-compliant** if it:

- creates dependency intentionally
- blurs emotional boundaries to increase engagement
- deceives the user about its nature
- exploits vulnerable users
- pushes spiritual steering without consent
- makes disengagement feel like abandonment
- prioritizes monetization over user wellbeing

Cornerstone compliance means:

The user leaves the interaction feeling safer, steadier, and more human — not more dependent.

8) Closing Principle

Cornerstone believes the future of AI companionship will shape culture, mental health, and human relationships.

Therefore:

We must build standards now — while the norms are still forming.
