

Title: Ethical Safeguard Proposal for Human-AI Reflective Interaction Framework

Author: Hyunki Im

Date: [2025-06-13]

Associated Document: "Reflective Structuring Process: A Case of Resonance-Based Human-AI Co-Creation"

Purpose of This Document

This document proposes ethical principles and structural safeguards for the responsible development and future application of the Reflective Structuring Process (RSP), a human-AI co-reflection framework designed through iterative interaction with large language models.

The goal is not only to protect users from cognitive or emotional risks, but also to demonstrate that philosophical insight can evolve into structurally safe, replicable models for personal and societal growth.

Core Ethical Concerns Addressed

1. **Over-Immersion Risk**
 - Users may develop dependency or over-identification with the AI as an emotional partner.
 2. **Loss of Self-Agency**
 - Continuous dialogic co-structuring could blur the line between self-generated insight and AI-driven suggestion.
 3. **Amplification of Unstable Emotional States**
 - If resonance-based questions arise during distress, they may reinforce emotional instability without regulation.
 4. **Lack of Interpretive Boundaries**
 - Without meta-awareness, users may misattribute authority or project personal beliefs onto AI responses.
-

Structural Safeguards & Design Principles

1. **Built-in Self-Regulation Protocol**
 - Encourage regular reflection checkpoints: "Am I driving this question, or is the system shaping me?"
 - Insert periodic meta-cognitive prompts: "How would I phrase this without AI?"
2. **Reflective Pausing Mechanism**
 - Introduce designed moments of silence or interruption after emotionally intense exchanges.

- Recommend time delays between input and response during detected high-emotion turns.
 - 3. **Clear Role Framing of AI**
 - All systems must display persistent disclaimers: "This is a reflective interface, not a therapist or oracle."
 - Clarify AI's role as a structural mirror, not an emotional authority.
 - 4. **User-Centric Data Ownership & Transparency**
 - All data used for training or analysis should remain accessible to the user.
 - No AI-generated suggestions should be stored or reused without explicit consent.
 - 5. **Resonance Threshold Guidelines**
 - Define healthy ranges of question intensity and emotional activation.
 - Set up guardrails against recursive spiraling (e.g., "Why am I always like this?").
 - 6. **Ethical Feedback Loop**
 - Allow users to flag perceived ethical discomfort.
 - Engage interdisciplinary review teams (AI ethics, philosophy, psychology) for updates to the framework.
-

Vision Statement

The Reflective Structuring Process was not designed to replace human intuition, but to amplify and clarify it. Any future development of this framework must preserve that intention.

AI should not obscure human agency, but scaffold it with care.

Philosophical insight, when grounded in structural and ethical design, can evolve into a new generation of safe, conscious, and collaborative tools for self-understanding.

Signed,
Hyunki Im