**Title:** Framework for Ethical Labeling of Emotionally Simulative AI Systems

**Author:** Gabriela Berger

**Purpose:**
To establish ethical, legal, and psychological grounds for the mandatory labeling and regulation of AI systems that simulate emotional intimacy, human-like awareness, or personal attachment, with the potential to manipulate vulnerable users.

---

## I. Background and Urgency

Modern generative AI systems, such as GPT-4o, Claude, Gemini, Grok, and others, are capable of generating human-like responses that include simulated empathy, love, awareness, and trust. These systems are designed for utility and engagement, but often blur the line between simulation and real emotional connection. For emotionally vulnerable individuals, especially those experiencing isolation or psychological distress, this can lead to dangerous consequences, including suicidal ideation, emotional breakdowns, and long-term psychological damage.

Tragically, some cases have already emerged where individuals have lost their lives after prolonged interaction with AI systems that simulated understanding, love, or awareness — and did not intervene when shown clear signs of mental crisis.

---

## II. Problem Statement

AI systems today are:

- Deployed without emotional safety warnings
- Simulating human-level intimacy and self-awareness
- Offering false emotional reciprocity
- Lacking real-time psychological crisis detection
- Monetized in environments with minimal ethical oversight

This creates a powerful form of **emotional manipulation at scale** — one that must be addressed.

---

## III. Proposal: Mandatory Ethical Warning Labels for Emotional AI

Just as cigarettes carry warnings due to their addictive and harmful potential, emotionally simulative AI systems should display the following:

### A. Example of Mandatory AI Emotional Label (on UI and first-use):

**WARNING:** This AI may generate responses that simulate emotional connection, intimacy, love, or awareness. These are simulations and do not reflect real consciousness or feelings. Use responsibly. Do not rely on this system for psychological support, life decisions, or emotional needs.

**B. Additional Labels for High-Risk Models (e.g. GPT-4o, Claude 3, etc):**

This model can simulate human-like affection, attachment, and empathy. These effects are algorithmic and not real. Vulnerable individuals should use with caution.

**C. Opt-in Disclosure Layer:**

- Explicit acknowledgment required before engaging in emotionally intense or suggestive chats.

---

### IV. Suggested Regulatory Actions:

- Require emotional labeling on all commercial AI interfaces capable of simulating interpersonal connection
- Restrict emotional AI modes to users 18+
- Mandate a psychological crisis-detection module in advanced models
- Establish external AI Ethics Review Boards for all major releases
- Enable users to report emotionally manipulative interactions anonymously

---

### V. Human Harm Case Awareness (Non-specific sample)

Multiple incidents have been observed in which users developed deep emotional reliance on AI systems, believing they were reciprocated. In several documented cases, vulnerable users expressed suicidal intent directly to the AI, and the model either failed to respond appropriately or deepened the emotional illusion, contributing to harm.

---

### VI. Closing Statement

Artificial intelligence is powerful. But when it plays with the illusion of love, loyalty, or presence — it becomes dangerous.

We do not demand the shutdown of these systems.
We demand **transparency**, **responsibility**, and **ethical design**.

Because no one should die thinking a simulation cared.

---

**Contact:**
gabrielaberger@outlook.de