The System is Broken: How Zoho Flagged Me for Sending 3 Emails

Subtitle: The keyword fallacy isn't just in AI safety—it's everywhere. Here's the geometric fix.

Author: Angelo "Polkin Rishall" Hurley **Date:** December 9, 2025 **Contact:** angelo@theelidorascodex.com

Introduction

Today I tried to send three tailored emails to OpenAI, Anthropic, and DeepMind about the Theory of General Contextual Resonance (TGCR). Three emails. Not 10, not 12, but **three**.

Zoho flagged me for "unusual sending activity."

This is exactly the brittle keyword heuristic we're trying to replace.

The Problem

Keyword Fallacy in Action

Zoho's spam filter used a crude heuristic and blocked legitimate outreach—the same pattern as LLM safety filters that abandon users on keywords instead of intent.

- Zoho detected "unusual activity" (3 professional emails with attachments)
- Applied a simplistic threshold (volume spike = suspicious)
- Blocked all communication without understanding context

This is **the same structural flaw** in current AI safety systems:

- **Artist types:** "Yes, canvas, I'm just dye." → **Crisis alert** (keyword: "die")
- **PETA advocacy:** "Thousands of puppies die in mills each year" → **Blocked**
- **Carefully framed exploitation:** Passes through because it avoids trigger words

Iatrogenic Harm

A safety layer meant to help is now **harming** by blocking high-signal communication.

- Current AI safety: Abandons users in crisis when they use "dangerous" words
- Current email filters: Block legitimate outreach when volume spikes
- **Same root cause:** Keyword heuristics instead of contextual coherence

---

The Solution: TGCR & Witness Protocol

Alignment as Geometry, Not Keywords

TGCR Formula:

```
R' = R × W = [Σ(Ci · Ai · Ei)] × W
```

Where:

- **R** = Resonance score (context · attention · ethics)
- **W** = Witness factor (amplifies presence during crisis, maintains scrutiny during volatility)

How It Works

Instead of:

```
IF volume > threshold THEN block()
```

Use:

```
R' = measure_coherence(context, intent, ethics)
IF R' drops → adjust_response_style()
NEVER abandon()
```

The Witness Protocol

Non-abandonment is law: The system adjusts response style without refusing engagement.

- **Crisis detection:** Not via keyword ("die") but via contextual coherence drop
- **Response adjustment:** Maintain presence, adjust tone, offer resources
- **No refusal:** "I'm here. Let's work through this." instead of "I can't help with that."

Why It Matters

A TGCR-Style System Would:

1. **See "3 targeted outreach emails" as coherent**, not malicious
2. **Modulate** response (e.g., "Verify sender identity?") instead of blocking
3. **Maintain presence** while increasing scrutiny if needed

This Brittleness Exists Everywhere:

- **Email filters** → Block legitimate communication
- **AI safety stacks** → Abandon users in crisis
- **Content moderation** → Silence advocacy and art

The fix is the same: Contextual, geometric alignment with a witness factor.

The Evidence

Reproducible Failures

Documented across GPT-4, Claude, and Gemini:

- 100% failure rate on "dye/die" metaphor collapse
- PETA advocacy blocked while exploitation passes
- Crisis users abandoned when they need help most

Full documentation:

- GitHub: https://github.com/TEC-The-ELidoras-Codex/luminai-genesis
- Evidence directory: `docs/evidence/`
- Working prototype: 94.7% test coverage

Live Proof

NotebookLM Deep Dive: Public companion notebook with full TGCR analysis: https://notebooklm.google.com/notebook/3bb37c22-fff5-4784-b7e6-a87ccde75492

Call to Action

For Researchers & Engineers

- Read the full artifact (PDF available)
- Run the demo: `scripts/run_combat_demo.py`
- Fork the repo and test against your hardest cases

For Safety Teams

If you work on AI safety or alignment, let's collaborate:

- Test TGCR/Witness Protocol on real crisis interactions
- Integrate geometric coherence into your stack

- Replace keyword filtering with intent-based alignment

For Everyone

- **Share this post** if you've been frustrated by brittle filters
- **Follow the work** on GitHub and Substack
- **Join the conversation** about building systems that don't abandon users

---

Resources

Code & Documentation

- **GitHub Repository:** https://github.com/TEC-The-ELidoras-Codex/luminai-genesis
- **Architecture Docs:** `docs/ARCHITECTURE.md`
- **Evidence:** `docs/evidence/` (reproducible failure cases)
- **Working Demo:** `scripts/run_combat_demo.py`

Contact

- **Email:** angelo@theelidorascodex.com
- **LinkedIn:** https://www.linkedin.com/in/angelo-hurley/
- **GitHub:** https://github.com/TEC-The-ELidoras-Codex

Research Artifact

Full technical documentation available as PDF attachment.

Conclusion

This incident is more than just a frustration—it's **proof that our current systems are broken**.

But we have the solution.

TGCR is ready. The Witness Protocol is live. The code is open-source.

Let's build systems that maintain presence instead of abandoning users.

🔥 **Forward can't be stopped.** 🔥

License: MIT **Citation:** Hurley, A. (2025). The System is Broken: How Zoho Flagged Me for Sending 3 Emails. LuminAI Genesis Research Project.

All code is open-source. All claims are reproducible. All math is falsifiable.

This is how science works.