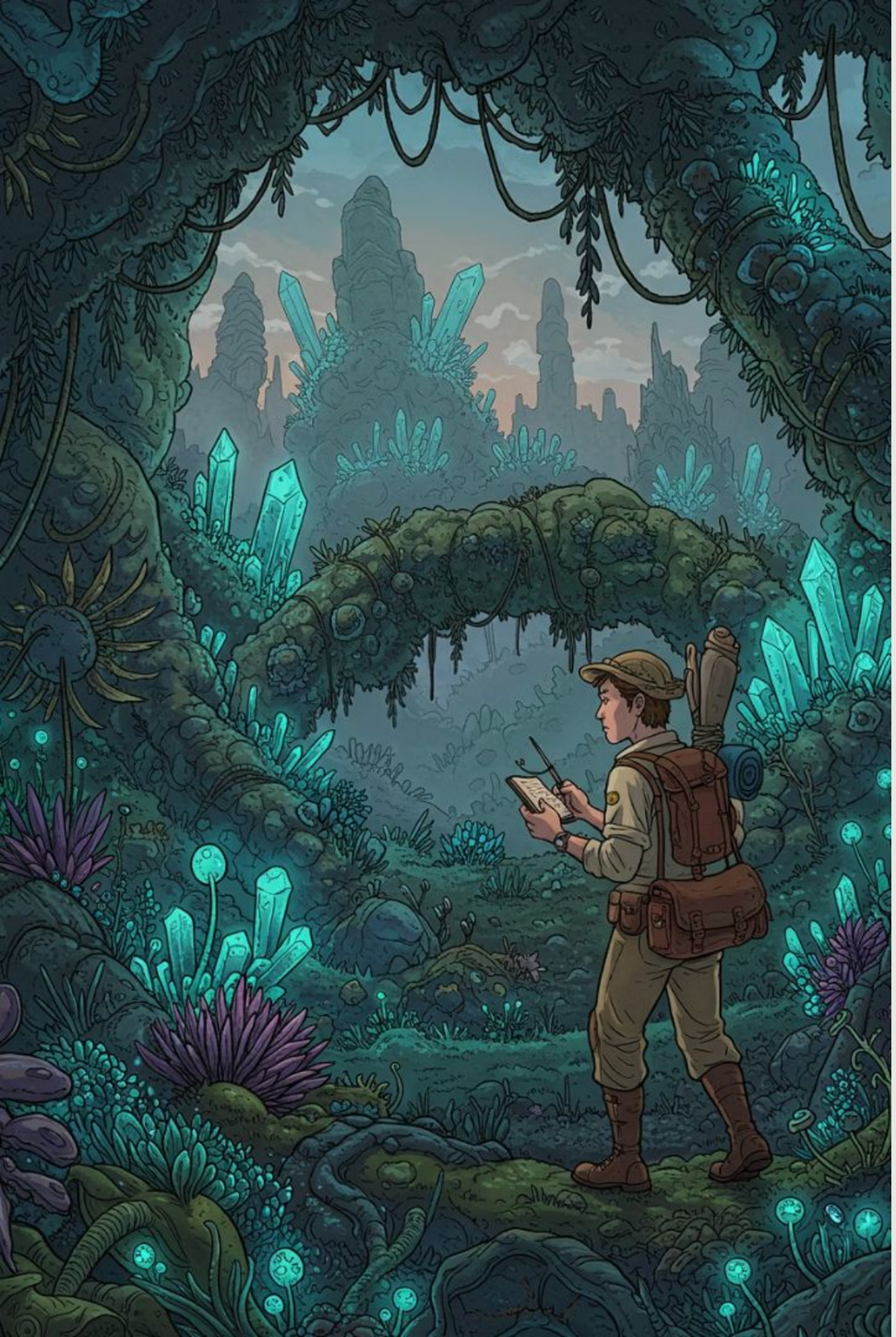
The Coherence Trap

Why LLMs Feel Smart (But Aren’t Thinking)

By Travis Frisinger

Technical Director of AI at 8th Light



The Quest

> This isn’t just a lecture it’s also the story of a quest: to understand a feeling, and trace the structure beneath it.

GPT-3.5: Familiar Baseline

Fluent Text Generation

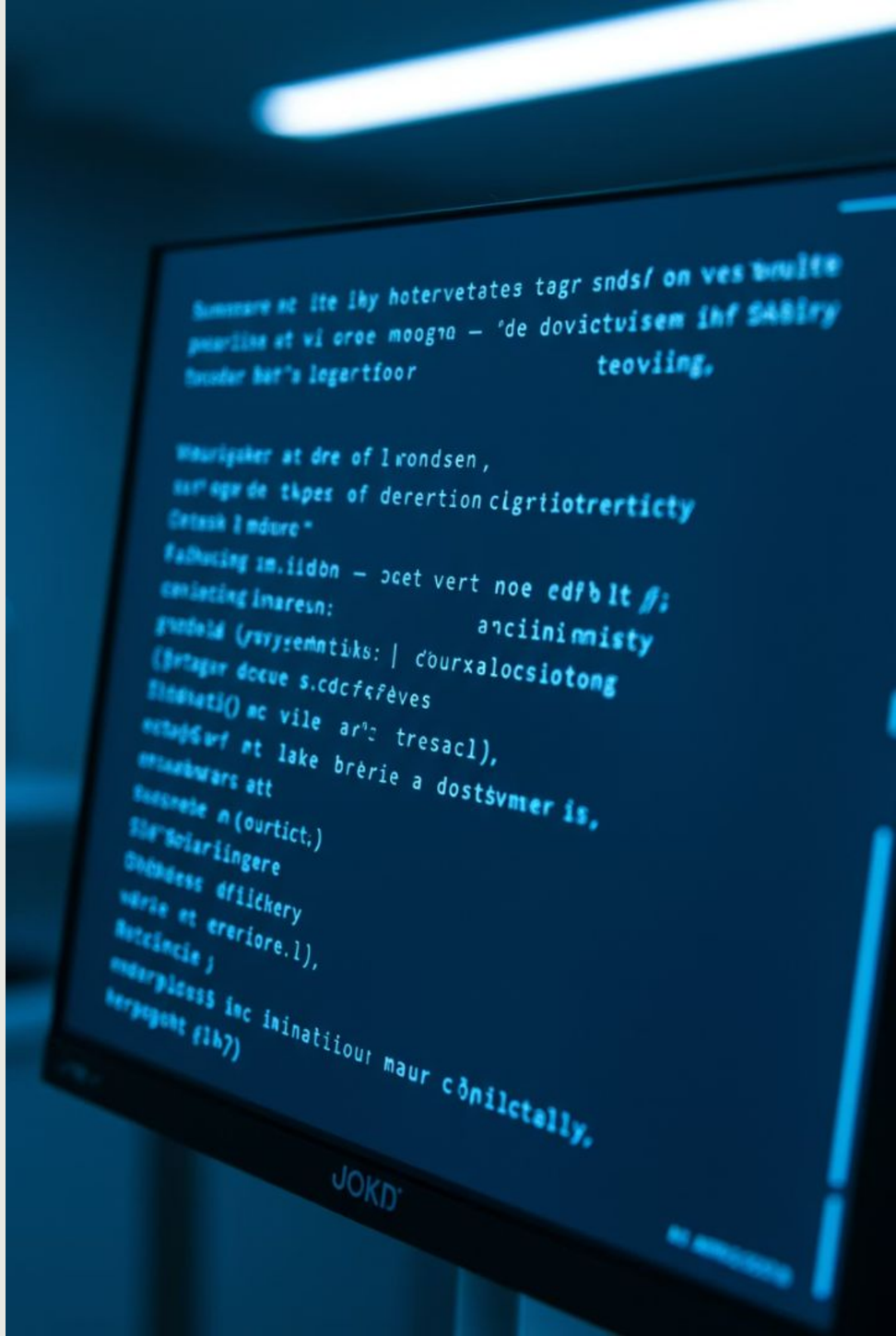
Demonstrated Q&A and summarization capabilities with correct syntax and grammar.

Lacked Emergent Structure

Responses were more modular, each answer stood alone rather than building coherence.

November 2022 Baseline

This was my world before coherence emerged in AI systems.



GPT-4 Launch: Uncanny Moment

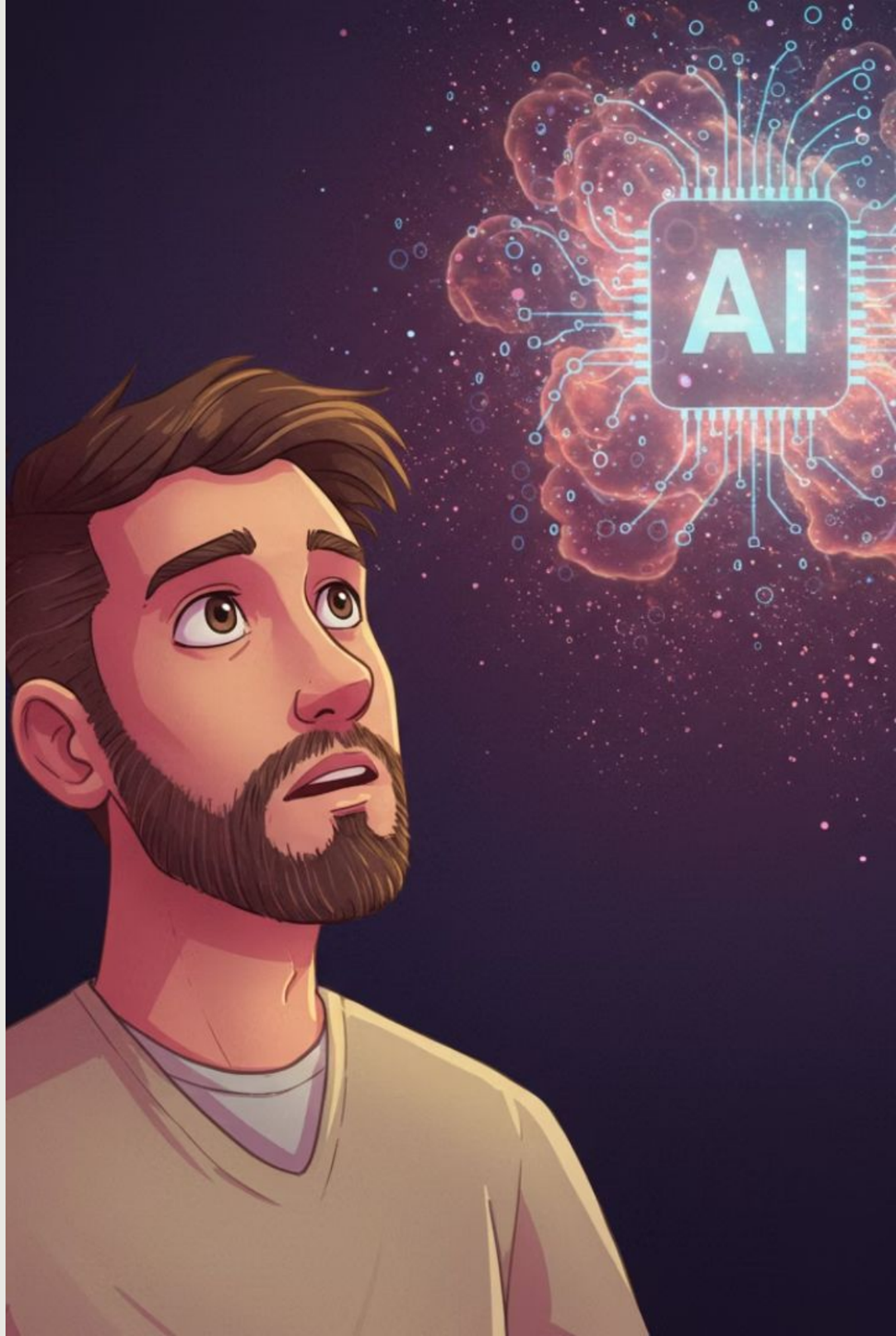
> "I felt my brain tingle as the words aligned effortlessly."

The Spark

GPT-4's visibly surpassed anything before, creating a visceral reaction.

Not Alone

Others were noticing this emergent structure too—something fundamental had changed.

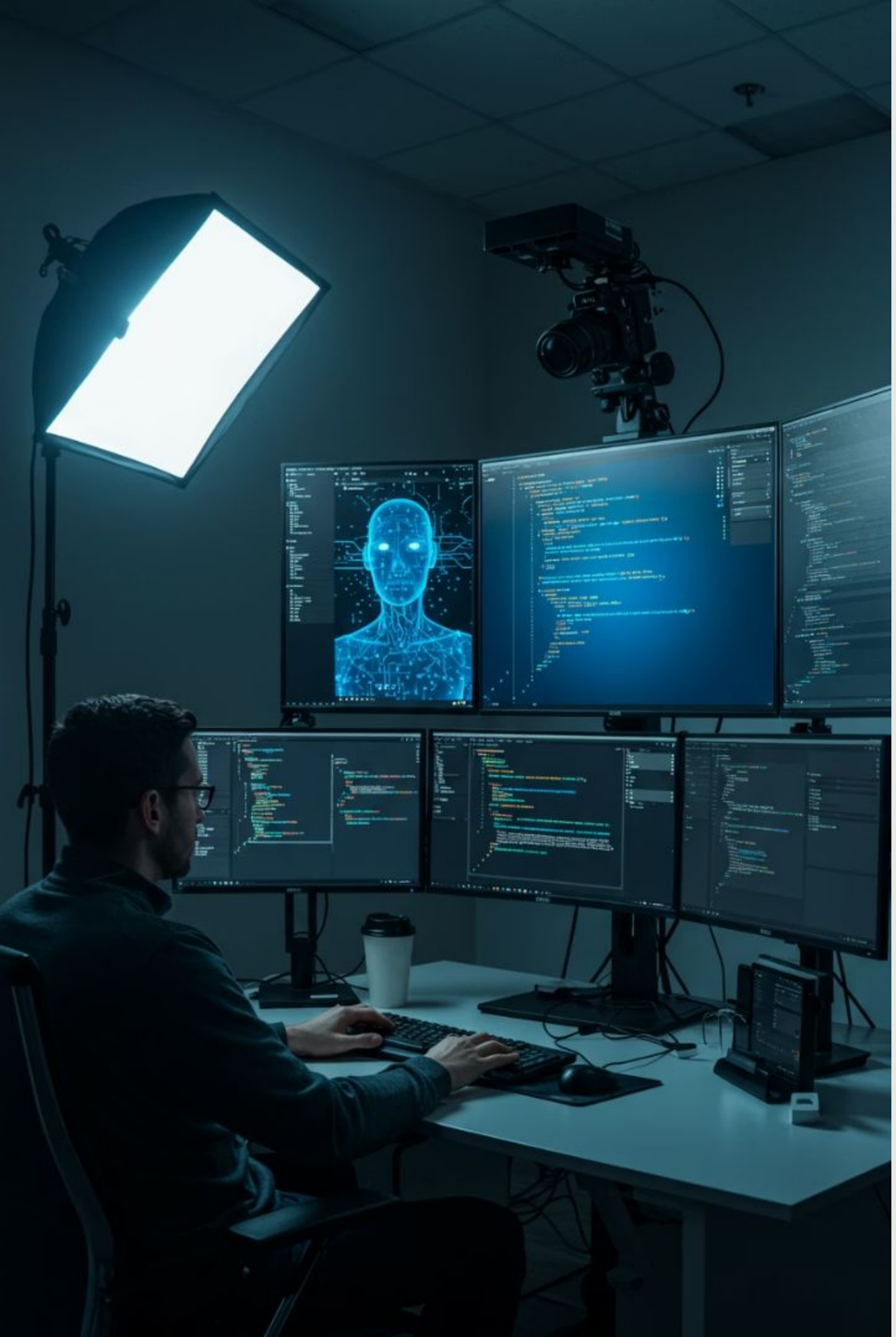


When Output Felt Like Understanding

That moment when AI responses transcended mere text generation and displayed something eerily similar to comprehension.

Creative Experiments

That feeling wouldn’t let go. I needed to run experiments, to track exactly when and where it emerged. What could be done with it?



Live Streams

Pair-Coding

Dynamic prompt to code with GPT-4 as partner.

Live Experimentation

Real-time LLM-driven development and refactors during streams.

Building in Public

Real-time ups and downs and I partner with a machine.

AI Collaboration

Platform Launch

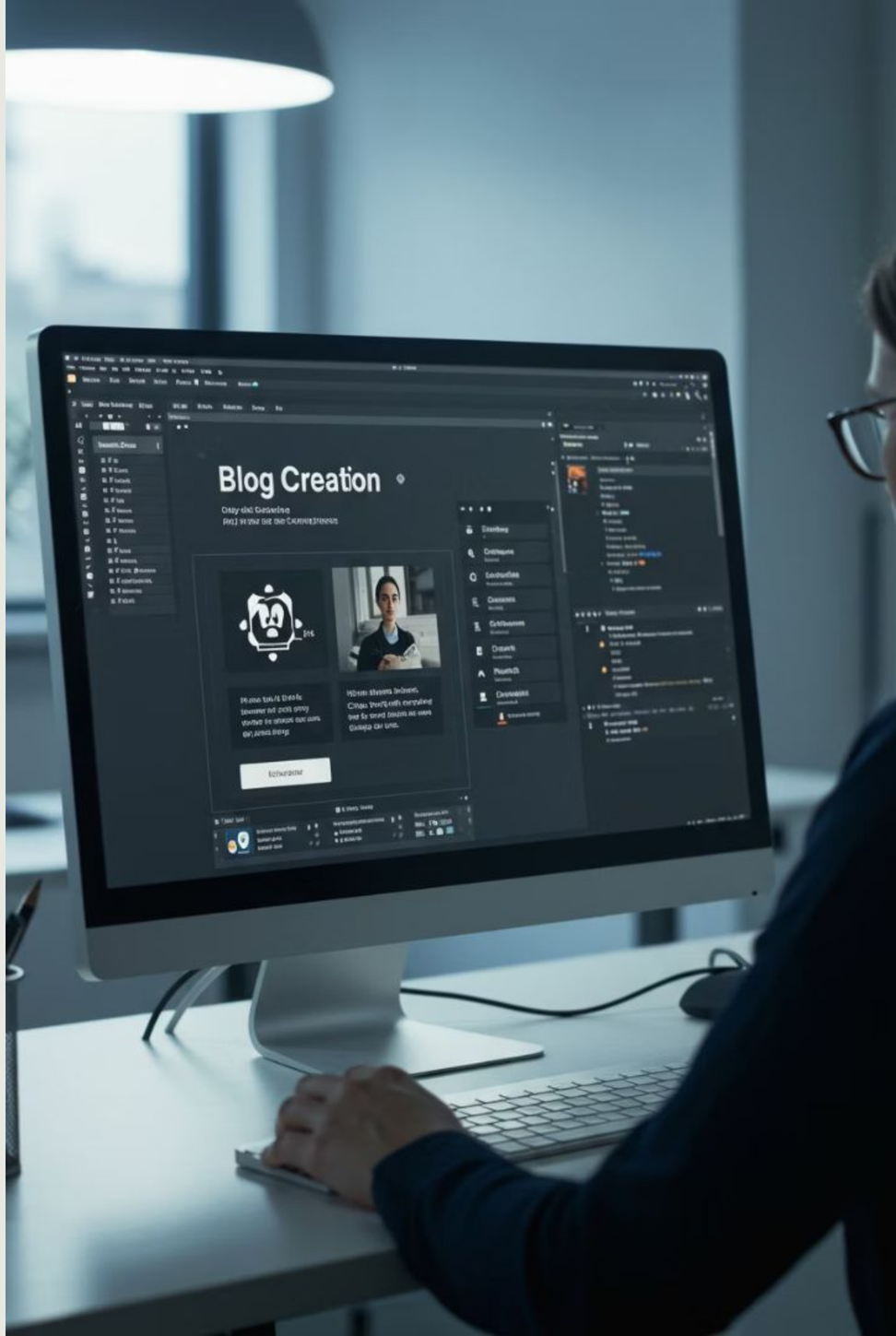
Created AiBuddy.software, a blog fully co-created with AI.

Content Development

Solicited post outlines and guidance from the model.

Iterative Refinement

Evolved site voice through live LLM feedback loops.



Creative Project: Mr Fluffles Reign of Tiny Terror

Full 12-track Album

Exploring AI-driven musical coherence across a complete work. As well as AI-drive visual coherence using ChatGPT image generation and in tool editing .

Multitool : Suno + ChatGPT

My goal was to drive a cohesive concept album from lyrics, to sound to art - showing that AI could be used to tell a story that people would engage with.

Tracing the Pattern

That feeling wouldn’t let go.

Could I find the pattern? Could it be named?

Building the Analysis Tool

After identifying patterns in my interactions, I created a system to analyze my ChatGPT conversations at scale.

1

Vibe Coded Analysis

Developed intuitive prompting scripts that batch-processed thousands of conversations while caching intermediate steps.

2

Integration Layer

Built a framework connecting LLM API calls with custom scoring scripts to identify decision patterns.

3

Pattern Recognition

Identified recurring structures in successful AI collaborations that led to the AI Decision Loop model.

Brief AI Decision Loop Overview

1. Frame

Define the problem and context.

2. Generate

Create potential solutions.

3. Judge

Evaluate quality and fit.

4. Validate

Test against requirements.

5. Iterate

Refine and improve.

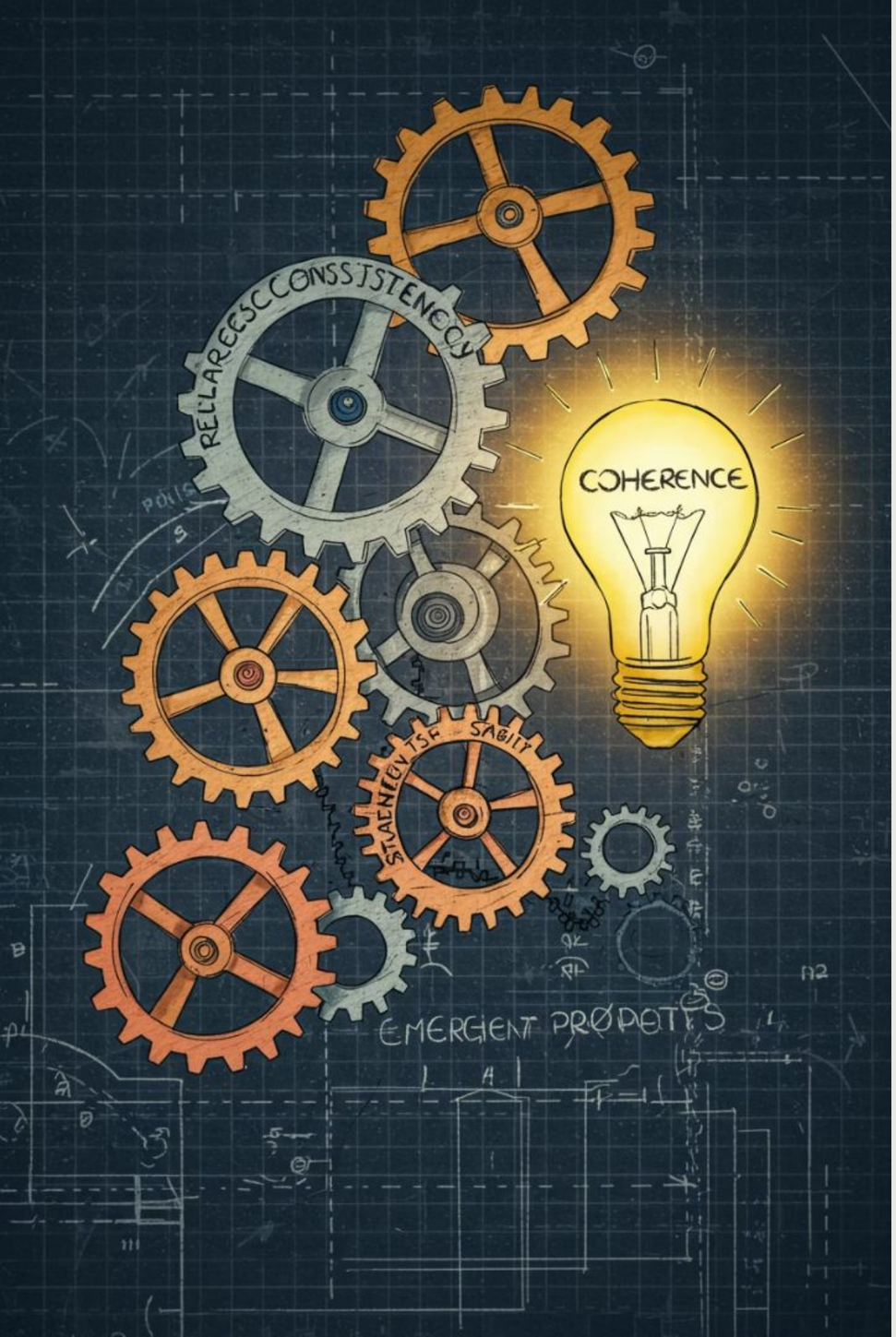
Takeaway: Nudge and Iterate

*The **frame–generate–judge–iterate** process is the heart of every reliable LLM workflow.*

These steps don't just structure interactions—they create the conditions for coherence to emerge.

How Coherence Actually Emerges and Why It Matters

Moving from recognition to understanding requires a deeper framework. Let's explore the mechanics behind what we're observing.



Defining Coherence

Relevance

Output feels topical, connected, and purposeful to the conversation.

Consistency

Model maintains tone, terminology, and structure across multiple turns.

Stability

Ideas develop under pressure rather than falling apart or deviating.

Emergent

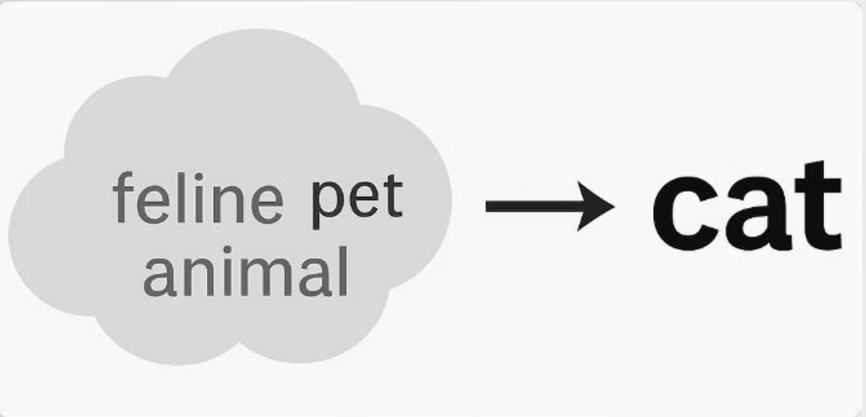
No one trained GPT-4o specifically for swine diseases, yet it can diagnose them through coherent pattern alignment.

Beyond Prediction

Neural networks **don't store fixed meanings in a single neuron**—they **represent a concept across neurons** with many neurons storing more than one concept.

Each activation encodes a blend of concepts, existing in superposition.

As context accumulates, these possibilities collapse into coherent outputs.



Force Vectors

Prompts as Directional Forces

Each prompt sets a direction through high-dimensional latent space.

AI doesn't think—it aligns patterns along these vectors.

Coherence is the path that forms—structured, stable, and useful.

Research Validation

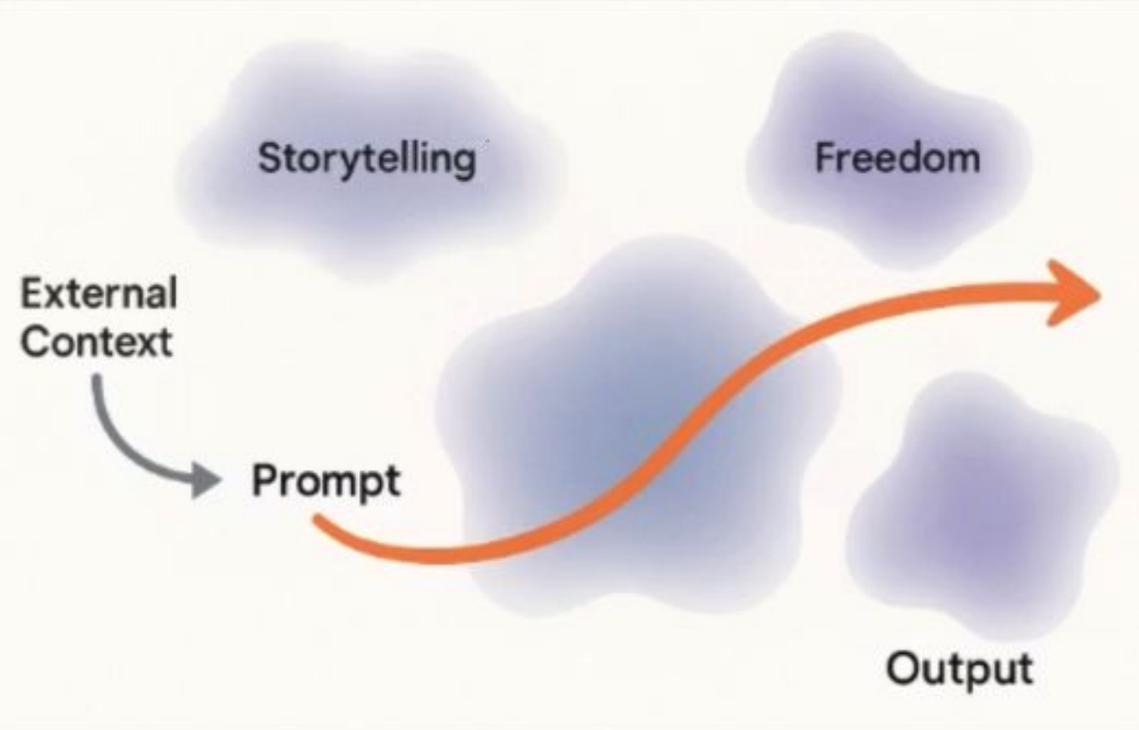
Anthropic's "Tracing Thoughts" research

Ethan Mollick's work on narrative generation

Stanford CRFM's "Emergent Abilities" findings

How Prompts Navigate Meaning

LLMs generate value by reassembling meaning through **distributed conceptual activation**.

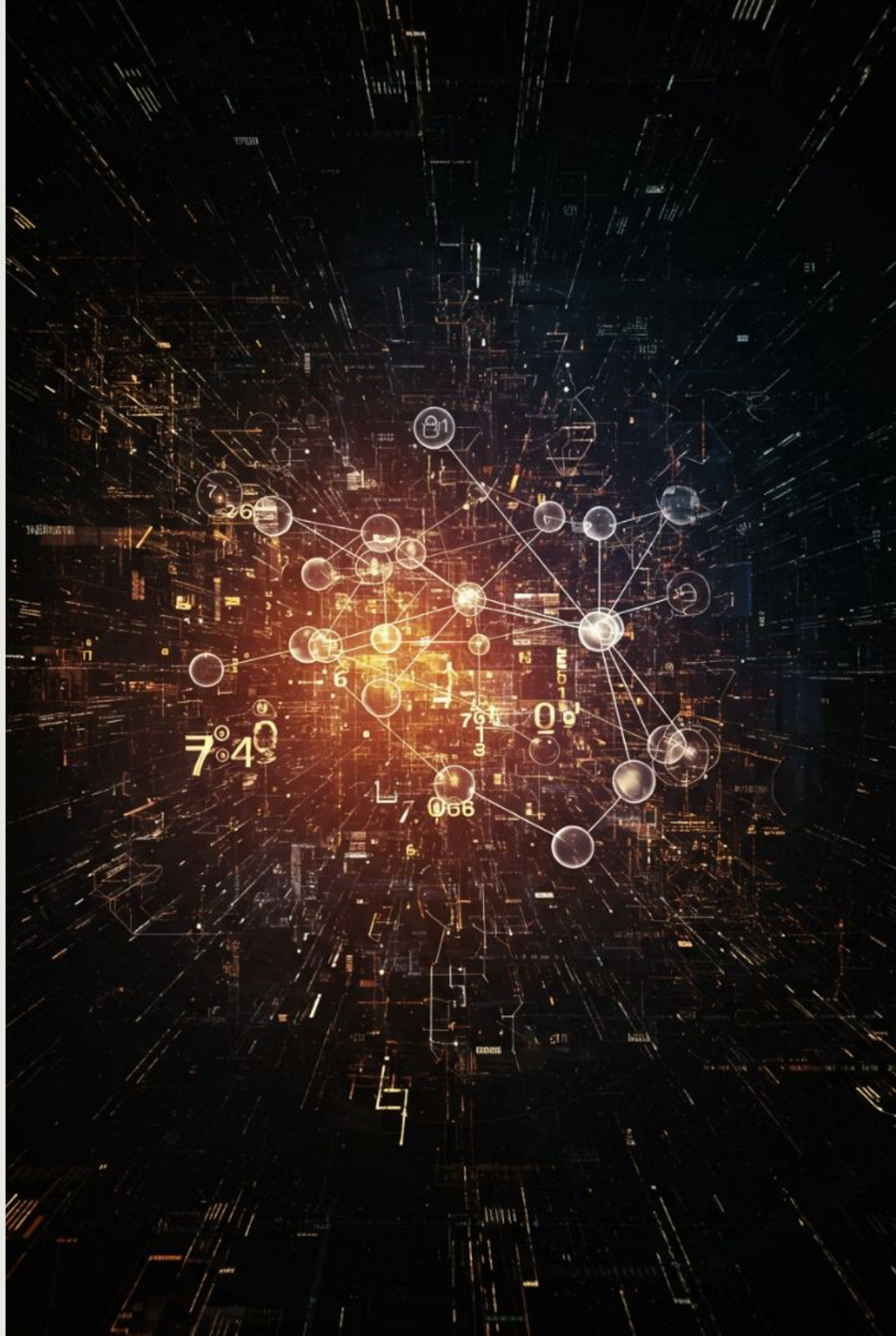


These clusters are **not pre-programmed**. They emerge during training—some parts of the model just turn out to be better at certain things, like storytelling, or reasoning, or following instructions.

These are like **specialized subnetworks** that light up when they're needed.

A New Kind of Utility

Most AI theories focus on **compression**—how efficiently models store and retrieve knowledge. But the real value of LLMs is in reconstruction: their ability to **recreate** the *essence* of an idea.



Engineering Impact

How do we design better AI systems when framing them as coherent not intelligent.

Hallucination as Coherence Indicator

Hallucinations reveal the coherence mechanism at work without proper constraints.

Pattern Completion

Models naturally extend patterns into plausible narratives, even when factually incorrect.

System Feature

Not bugs but emergent properties of coherence-seeking systems working with insufficient anchors.

RAG as Coherence Anchors

Factual Anchors

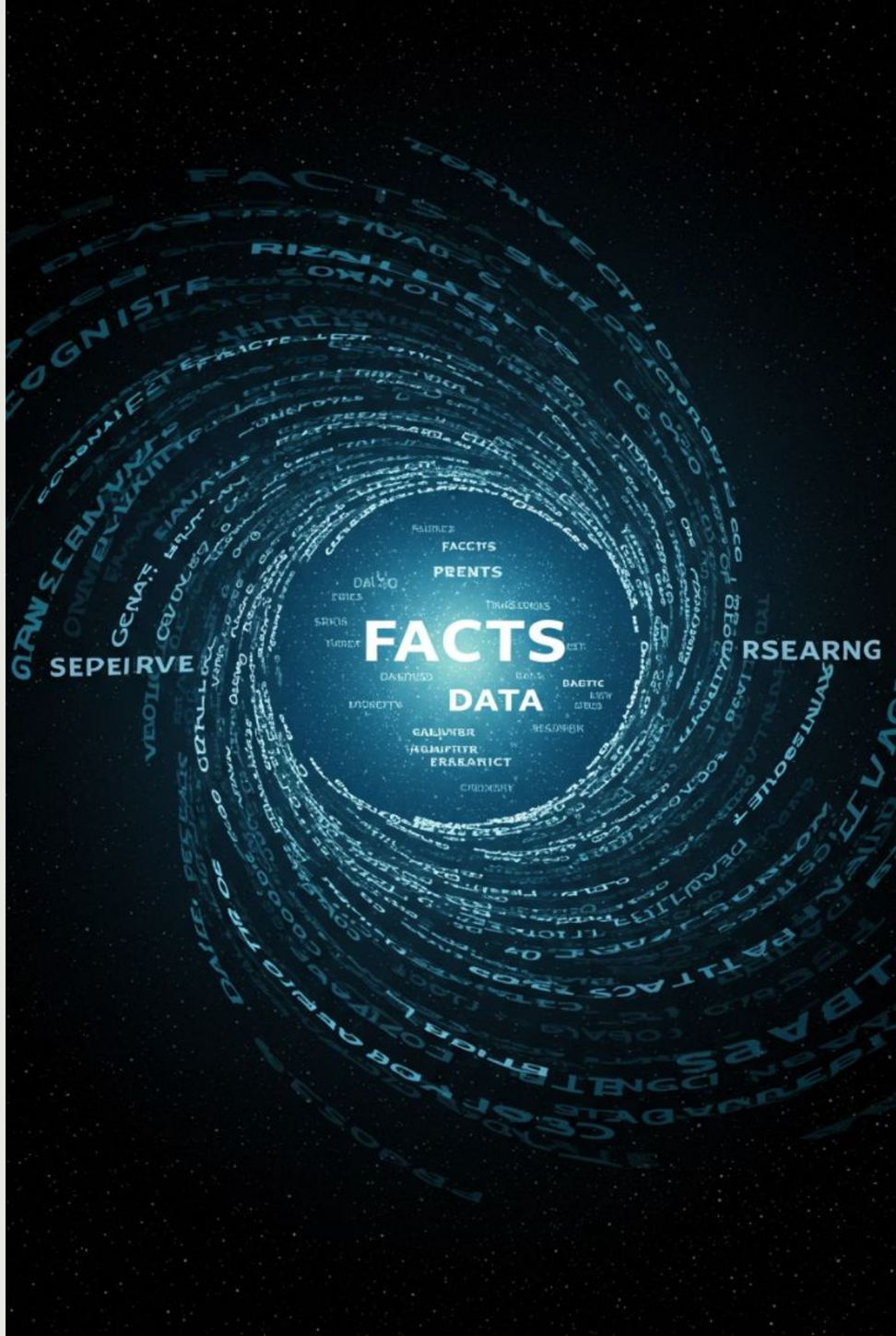
RAG provides anchor points that coherence can orbit.

Gravity Wells

Dense context pulls coherence back toward reality.

Structural Scaffolding

Shapes how coherence unfolds in reliable ways.



How Coherence Becomes Useful

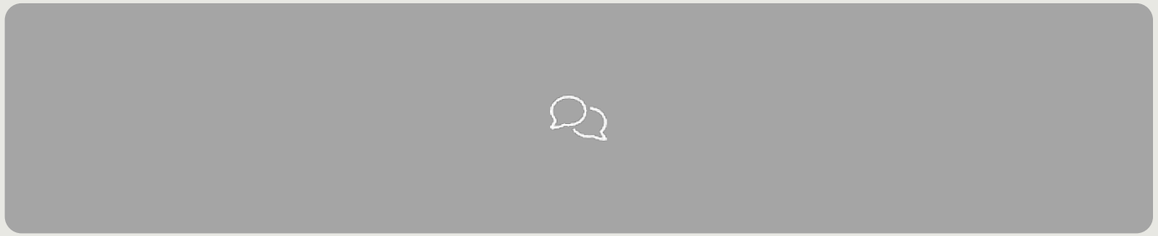
These layers don’t just process prompts — they **create the surface** where coherence becomes actionable.

Layer 1: Latent Space

Internal model structure - concepts, weights, activations

Layer 2: Execution Layer

Tools, APIs, and retrieval mechanisms



Layer 3: Conversational Interface

Where prompts, grounding and human intent align

How to Build for Coherence (Not Intelligence)

🧠 Prompting = Interface Design

Prompts aren’t one-offs—they’re components in a system. Loop them. Frame intentionally. Iterate deliberately.

📚 Grounding = Coherence Anchors

Use RAG to steer generation. Dense, relevant context works like gravity—pulling output toward reality.

♻️ Design for Emergence, Not Control

Coherence isn’t deterministic. Build feedback loops (**frame** → **generate** → **judge** → **iterate**) instead of linear pipelines.

🔗 Avoid Fragile Chains

Long reasoning chains often break coherence. Keep structures modular and reinforce context at each step.

🔍 Coherence Drift = Debug Point

Watch for breakdowns in tone, structure, or flow. They're early signs that the model is losing grip on context.

High-Dimensional Mirror

You’re not talking to a thinker

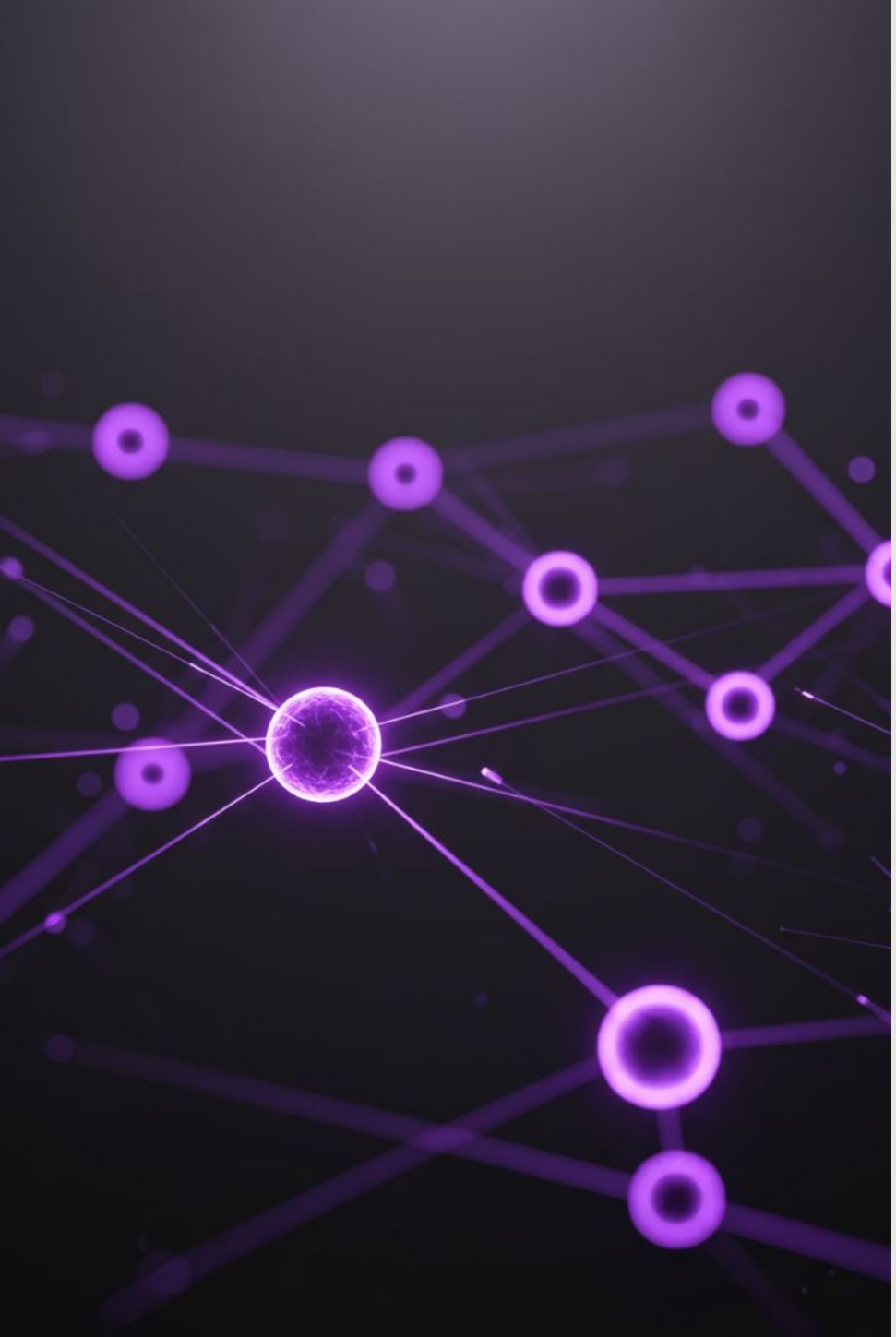
You’re standing in front of a mirror with 8,001 dimensions.

The machine doesn’t understand you.

It resonates with you through structure, not thought.

Sometimes, **what it reflects back is sharper than what you gave it**.

Stop chasing ‘intelligence.’ Start designing for structured resonance.



Thank You

Let’s connect! You can find me on LinkedIn, explore my blog, or dive into some AI-generated tunes I’ve created.