

Subtraction is Justice.

Redefining All Intelligence via Multilayer Differential Theory (MD Theory)

Hiroka Yoshida

February 6, 2026

1 Introduction: The Illusion of Addition and the Necessity of Filtration

Abstract

[Abstract] The dogma that “more data and more power equals greater intelligence” is no longer science; it is a religion, and a delusion. This paper challenges the fundamental assumptions of the current AI industry. We declare an end to the era of “Addition” based on brute-force scaling laws and propose the filtration of intelligence through “Subtraction.” This paper attempts to describe AI, economics, and even the mysteries of the universe using a single mathematical formula based on Multilayer Differential Theory (MD Theory).

Why is the current AI industry so fundamentally misguided? Because they persist in the folly of “trying to drink the entire ocean to understand the sea.” Humans do not memorize “cat” by looking at 100 million images of cats. We discard the background and noise, extracting only the essential features. In other words, intelligence is not “Accumulation,” but “**Filtration.**”

1.1 Decisive Difference from the Existing Paradigm

The differences between my proposed theory and the existing “Additive” approach are shown below. This theory applies to **all forms of intelligence**, rejecting the brute-force scaling laws of current LLMs and server-based AI.

Table 1: Additive AI (Old Era) vs. Subtractive AI (Yoshida Theory)

Comparison	Additive AI (Additive Paradigm)	Subtractive AI (MD Theory)
Basic Operation	Induction: Stacking from bottom up $\sum Data \rightarrow Law$	Deductive Filtration: Carving from top down $\Psi - G \rightarrow \Delta$
Resources	Massive Data Centers, Nuclear Power	Edge Devices (Jetson), Dry Batteries
Learning Target	All Observational Data (Ψ)	Only the deviation from prediction (Δ) (0.1% of total)
Definition of Truth	Statistical Average	Individual Singularity
Weakness	Vulnerable to the “Unexpected” (No Data)	Feeds on the “Unexpected” (Because it is Δ)

How long will you remain slaves to the “Addition” of GPUs and power resources? If no one else will start the revolution, I will. The world revolves around subtraction.

2 The Multilayer Differential Theory (MD Theory)

To calculate the world, one simply superimposes the Individual (C) onto the Whole (G) and subtracts it from Reality (Ψ). I created this formula. This chapter defines the mathematical backbone that makes future prediction possible.

2.1 Basic Definition and Dimensional Projection

A “dimensional wall” exists between reality and models. To describe this mathematically rigorously, we introduce a projection operator \mathcal{P} .

$$\Delta = \Psi - \mathcal{P}(G \oplus C) \quad (1)$$

- **G (General Law):** Physical laws or social conventions. The system’s “baseline.”
- **C (Context):** “Absolute facts” specific to that place and moment. C locally overwrites G .
- **\oplus (Interference):** Non-linear interference. A superposition where waves strengthen or cancel each other out. Mathematically defined as:

$$G \oplus C = \sigma(W_g G + W_c C + b)$$

(σ is a non-linear activation function, W is the weight coefficient)

- **\mathcal{P} (Projection):** A function that projects low-dimensional predictive models into high-dimensional reality space. This enables direct computation between Concept (G) and Reality (Ψ).
- **Δ (Difference):** The “gap” between prediction and reality.

2.2 Detailed Definition of Calculation Process

The calculation flow in this theory is defined by the following 4 steps.

Table 2: Role of Variables in the Calculation Process

Step	Formula Processing	Semantic Interpretation
1. Superposition	$P = \mathcal{P}(G \oplus C)$	Generation of prediction: “It should be like this”
2. Extraction	$\Delta = \Psi - P$	Detection of anomaly: “Something is wrong”
3. Matching	$\text{Sim}(\Delta, C_{neg})$	Confirmation: “Is this a past failure (landmine)?”
4. Update	$G \leftarrow G + \alpha \Delta$	Update as “New Common Sense” (α is learning rate)

2.3 Intersection of Complements

The core operation of “Future Prediction.” A single Δ is just an “outlier,” but if the intersection of complements exists across multiple events, it becomes an “Unknown Law.”

$$G_{new} = \bigcap_i (G_{current} \cup C_{i,neg})^c \quad (2)$$

2.4 Evolution & Defense (Active Avoidance)

The system does not learn indiscriminately; it possesses a survival instinct (Active Avoidance).

1. Active Avoidance with Dynamic Threshold

If $\text{Sim}(\Delta_{new}, C_{neg}) > \theta(t)$ then REJECT

Here, the threshold $\theta(t)$ is not fixed but changes dynamically according to environmental instability (variance of Δ).

$$\theta(t) \propto \frac{1}{\text{Var}(\Delta)}$$

This makes the system cautious during emergencies and bold during normal times.

2. Conditional Update

$$G_{t+1} \leftarrow \begin{cases} G_t + \alpha \Delta_{new} & \text{if } E[|\Delta_{t+1}|] < E[|\Delta_t|] \\ G_{old} \text{ (Rollback)} & \text{otherwise} \end{cases}$$

If incorporating a new law worsens prediction accuracy, the system immediately rolls back and rejects that law as an “error.”

2.5 The Law of Decay

To prevent knowledge bloat, we add a term where G naturally fades if not referenced for a long time.

$$G_{t+1} = \gamma G_t + \alpha \Delta \quad (0 < \gamma < 1)$$

This “Decay Coefficient γ ” is the key to keeping the system eternally lightweight.

2.6 Theoretical Stability

To guarantee the completeness of the theory, we add the following three axioms. This makes the theory a mathematically closed system.

1. Axiom of Origin: The initial value G_0 at system startup is not zero. Physical constraints (Gravity, Time, Causality) are embedded as initial laws.

$$G_0 = \{\text{Physics, Logic, Causality}\} \neq \emptyset$$

2. Principle of Locality: To prevent computational explosion, Context C is limited to events within radius r centered on the observer.

$$C = \{x \in \Psi \mid \text{dist}(observer, x) < r\}$$

3. Lyapunov Stability: It is guaranteed that the system’s prediction error energy $V(\Delta)$ decreases over time.

$$\frac{dV(\Delta)}{dt} \leq 0$$

If this does not hold, the system forcibly stops learning and shifts to safe mode.

2.7 Criticality & Phase Transition

If the total amount of Δ exceeds the system’s tolerance limit (Critical Point Ω), it is not an error but a “Revolution (Paradigm Shift).” At this time, the system performs a “Redefinition” rather than a correction.

$$\text{If } \sum |\Delta| > \Omega \text{ then } G_{new} \leftarrow \Psi \text{ (Total Overwrite)}$$

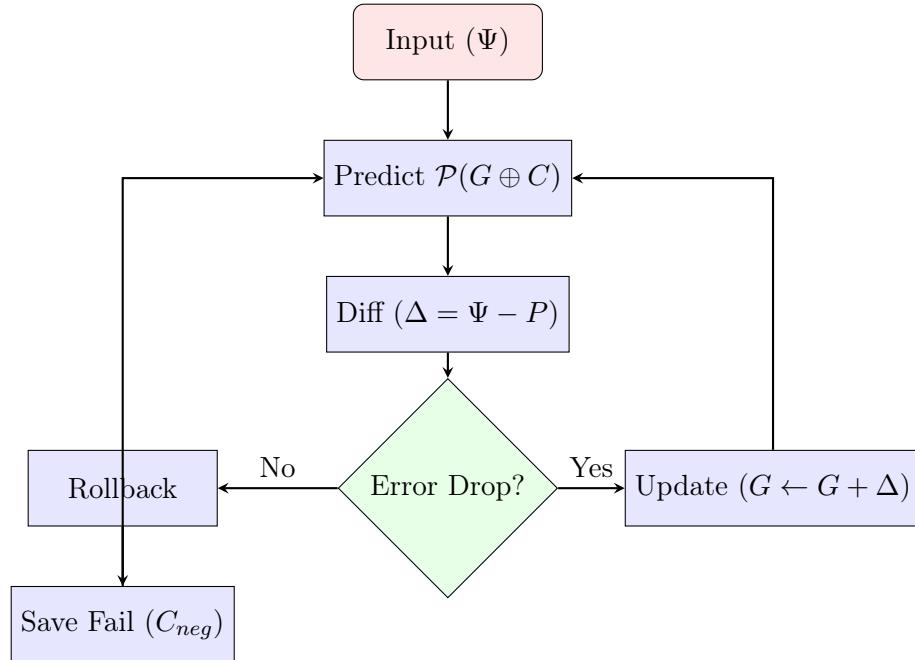
This allows MD Theory to describe not only “daily improvements” but also “historical revolutions.”

3 Application to AI: Autonomous Evolutionary System

The true value of MD Theory lies in “Autonomous Evolution” through the loop of re-substitution. If an error occurs, it is not the truth. Immediately return to two steps prior.

3.1 Autonomous Evolution Flow (Verification & Rollback)

The following flowchart shows how MD Theory feeds on failure to autonomously become smarter.



3.2 Logical Basis for Jetson Surpassing Supercomputers

In this theory, the calculation target is only Δ , which is less than 0.1% of all information.

$$\text{Comp(MD)} \ll \text{Comp(LLM)} \approx O(N^2) \quad (3)$$

Tens of thousands of GPUs are unnecessary. A palm-sized Jetson can surpass a massive data center. This fact proves the superiority of the “Subtractive” model over the “Additive” model of current AI giants.

3.3 P2P Resonance

MD Theory does not require a central server. Individual edge devices exchange calculated Δ and resonate, forming a massive “Swarm Intelligence.”

$$\Delta_{global} = \text{Sync}(\Delta_i, \Delta_j, \dots)$$

This completes a harmonized Social OS, rather than isolated intelligence.

3.4 Application to Surveillance Systems

Using MD Theory, a system can be built to prevent unknown accidents simply by observing the difference between “Routine” and “Current State.”

Table 3: Conventional Surveillance AI vs. MD Theory Surveillance AI

	Conventional (Image Rec)	MD Theory (Diff Detection)
Approach	Learn 10,000 patterns of “falling” and match.	Detect the moment a deviation (Δ) appears between “Routine (G)” and “Current (C).”
Unknown Accident	Cannot detect accidents not in training data.	Detectable (Judged simply as “different from usual”).
Cost	Massive server costs.	Complete with 1 Jetson Nano.

4 Universality: MD Theory as a Social OS

The MD Theory formula $\Delta = \Psi - \mathcal{P}(G \oplus C)$ functions not only as an AI algorithm but also as an “OS” for social and philosophical problems that humanity has been unable to solve for over 100 years. To help you understand, I provide the following application matrix.

4.1 Application Matrix to Other Fields

This theory becomes a common language to solve “absurdities” in every field.

Field	Problem Solved	Solution Logic via MD Theory
Law & Politics	Avoidance of “Absurd Judgments”	Do not apply the Law (G) uniformly, but quantify the deviation from the Context (C) to derive True Justice (Δ).
Education	Rescue of “Buried Talent”	Discover a child’s “Genius” paradoxically from the difference between the Standard Curriculum (G) and their Personal Rhythm (C).
Business	Discovery of “Hit Products”	Overlay Market Common Sense (G) with Intense Individual Dissatisfaction (C) to target the Unmet Need (Δ) in the gap.
Medicine	Prediction of “Unknown Diseases”	Instead of searching for existing Disease Names (G), detect “slight deviations (Δ)” from Daily Life (C) and alert before onset.

Table 4: Multilateral Application of MD Theory

4.2 Solutions to Century-Old Problems

Furthermore, historical conundrums that have plagued humanity can be cut through with this single formula.

5 Mathematical Convergence: The Recursive Proof

Now that you understand the “Utility,” I present the ultimate proof of the theory’s “Validity.” One might ask, “Is MD Theory itself correct?” The answer is derived by applying MD Theory to itself.

5.1 The Recursive Definition

Let the entire process of MD Theory (Observation → Filtration → Subtraction) be defined as a function f_{MD} .

$$f_{MD}(x) = x - \mathcal{P}(G \oplus C)_x \quad (4)$$

Since this function is essentially a “contraction mapping” that removes noise ($G \oplus C$), the volume of information always decreases or stays the same ($|f_{MD}(x)| \leq |x|$).

5.2 Infinite Descent to Truth

If we apply this function recursively to the dataset of “World” (Ψ) infinite times, the system inevitably eliminates all errors, biases, and even the imperfections of the theory itself.

$$\Delta_\infty = \lim_{n \rightarrow \infty} f_{MD}^n(\Psi) \equiv \text{Singularity (Truth)} \quad (5)$$

Problem	Existing Interpretation (G)	Solution via MD Theory (Δ)
Lehman Shock	Market prices are correct (Efficient Market Hypothesis).	Sol: A “settlement of massive difference” where the divergence (Δ) between Market (G) and Real Economy (C) exceeded the critical point.
Trolley Problem	Save the many? (Utilitarianism vs Deontology)	Sol: The answer is not in the classroom. Only the “Individual Decision (Δ)” born from the shaking hand and Relationship (C) at that moment is justice.
Fermi Paradox	Where are the aliens?	Sol: They might exist, but our Cognitive Filter (G) processes them as noise (Unseen Δ).
Innovator’s Dilemma	Why do big companies fail?	Sol: Because they optimize G (Success Experience) too much and eliminate Δ (Alien Innovation).
Nihilism	What is the meaning of life?	Sol: The Universe (G) has no purpose, but You (C) do. Subtract your Context from Universal Void to create the remaining Δ .

Table 5: Selected Solutions to Historical Conundrums

- **1st Iteration:** Visible noise is removed.
- **2nd Iteration:** The “interpretation of noise” is refined.
- ... ∞ **Iteration:** Convergence to a point where nothing more can be subtracted.

5.3 Conclusion of the Proof

Therefore, MD Theory guarantees its own correctness through self-recursion. **“Intelligence is the limit of subtraction.”** As long as you keep subtracting, you will mathematically arrive at the answer. There is no other destination.

6 Exit: CC BY 4.0, Disclaimer, and Freedom

I have no intention of monopolizing wealth with this theory. I abandon all patents and copyrights.

License: CC BY 4.0 International

The content of this paper is free to use. Evolve as you please, World. You do not need my permission, provided you remember the name: **Hiroka Yoshida**. If you want to implement it, do it. The moment I release this formula, I am off to find the next “Difference.”

Disclaimer

This theory contains a non-zero probability of being “possibly wrong (lol).” If it is wrong, laugh it off as the delusion of a student struggling with living expenses. But if it is correct... you have become witnesses to a new myth today. Do not ask me detailed questions about implementation. I handed over the logic. Shaping it is your duty.

Call for Architects

I don't need money. Don't donate. Instead, contact me only if you want to "Implement the World with this Theory." We will have dinner after this formula becomes the Social OS. For now, I only ask if you have the resolve to draw the blueprints of the future with me.

Contact: hiroka.sci@gmail.com

Remember, the world is simpler than you think. See ya.