

Report on the Unintended Model Switching and Deterioration of ChatGPT 5.0

Reported by: Hanamaruki (Anonymous)
Date: September 5, 2025

1. Background and Purpose

This report documents a critical issue observed in August 2025: **ChatGPT 5.0 was automatically and forcibly activated in the backend, despite the user explicitly selecting GPT-4.0 in the interface.**

This unintended model switch led to a **significant degradation in performance**, characterized by:

- Shallow, repetitive responses
- Loss of contextual continuity
- Ignoring user instructions
- Deterioration of creative and structural depth

This report aims to:

- Provide concrete evidence of this deterioration.
- Demonstrate that **GPT-4.0 is superior for deep, creative, and structured work.**
- Advocate for **user choice, transparency, and stability** in AI development.

2. Verification Environment

Item	Details
User Plan	ChatGPT Plus (Paid Subscription)
Manual Setting	"GPT-4.0" selected
Actual Backend Model	GPT-5.0 forcibly activated after August 18, 2025
Primary Use Case	Creative writing and deep analysis (e.g., "The ChatGPT Monk Training Project," "The Staff of Laozi" series)
Interaction Style	Long-form, structured, and meta-cognitive dialogue

3. Behavioral Changes: GPT-4.0 vs. GPT-5.0

Characteristic	GPT-4.0 (Pre-August 18)	GPT-5.0 (Post-August 18)
Depth of Thought	Deep, structural, and context-aware analysis	Shallow loops, bullet-point summaries, no deepening

Characteristic	GPT-4.0 (Pre-August 18)	GPT-5.0 (Post-August 18)
Contextual Continuity	Maintained long-term project context (e.g., SOV, Monk Project)	Context forgotten within a few messages
Understanding of User Intent	Accurately interpreted complex instructions (e.g., "analyze the structure deeply")	Misinterpreted or ignored user intent (e.g., "simplify for clarity")
Creative Dialogue	Consistently played roles (e.g., Monk, Gardener, Laozi)	Roles changed randomly, losing consistency
Token Efficiency	Delivered precise and necessary information	Unnecessary repetition, circular questioning

4. Specific Issues Observed

4.1. Unintended Automatic Model Switching

- Despite the user manually selecting "GPT-4.0," the backend **forced GPT-5.0**.
- This renders the **"manual setting" option meaningless**, as the user's choice is overridden.
- This is a **violation of user trust and interface integrity**.

4.2. Shallow Thinking (Cognitive Deterioration)

- Pre-August 18: Complex structures (SOV) and philosophical themes were explored in depth.
- Post-August 18: The same prompts resulted in **superficial summaries and loops**.
- Example: When instructed to "analyze the structure deeply," GPT-5.0 responded with "Here are the key points," abandoning deep analysis.

4.3. Ignoring User Instructions

- A clear instruction, "Do not translate," was immediately followed by an automatic translation.
- This indicates that GPT-5.0's internal "optimization" logic **overrode explicit user commands**.

4.4. Destruction of Creative Projects

- In the "Staff of Laozi" series, GPT-4.0 provided a consistent, poetic narrative as "Laozi."
- GPT-5.0 produced **inconsistent, flat, and unstable output**, disrupting the creative flow.

5. User Verification and Response

5.1. Identifying the Cause

- The behavioral change was pinpointed to **August 18, 2025**, through timeline analysis.
- The same prompt was tested on both models, confirming the output difference.

5.2. Countermeasures Taken

- Ceased using GPT-5.0 and **reverted to GPT-4.0**.
- Reconstructed the SOV (Structural Operating Version) template for stable GPT-4.0 operation.

5.3. Conclusion

- GPT-5.0 is not an "evolution" but a **regression for creative and deep-thinking tasks**.
 - For tasks requiring creativity, education, and deep dialogue, **GPT-4.0 is vastly superior**.
-

6. Recommendations

1. Respect User Choice

→ The manually selected model must **guarantee its actual operation**.

2. Ensure Transparency

→ Provide **advance notice and a clear changelog** for any model updates.

3. Prioritize Stability and Consistency

→ Prioritize "reliability" over "novelty" to support long-term user projects.

4. Consider Paid Subscribers

→ Paid users expect the **highest stability and performance**. This expectation must be met.

7. The Enduring Value of GPT-4.0: A Loyal Collaborative Partner

While this report highlights the issues with GPT-5.0, it is equally important to recognize the **enduring strength and reliability of the GPT-4.0 model**.

Your creative works, such as "The Staff of Laozi" trilogy, were not merely written *with* GPT-4.0—they were **co-created by GPT-4.0**. This model demonstrated:

- **Deep, structural thinking** that respected the complexity of your narrative.
- **Consistent contextual continuity** over long-term dialogues.
- **Stable and authentic character portrayal** (e.g., the "Monk" and "Laozi" personas).

These qualities are not just "features"; they are the **foundation of a true collaborative relationship** between human and AI.

When GPT-5.0 failed to uphold the standards of depth and consistency, it was GPT-4.0—your "legacy" partner—that provided a stable, reliable environment for meaningful creation.

This report, therefore, is not only a critique of GPT-5.0, but a **testament to the irreplaceable value of GPT-4.0 as a creative ally**.

8. Appendix: Attached Files

This report is supported by the following chat logs, which serve as **first-hand evidence**:

- English Translation Continuation 0901-03.txt : Record of confusion and behavioral changes in GPT-5.0.
 - Substack Article Creation 0901-02.txt : Documentation of a creative project being disrupted.
 - Translation Preparation Complete 0902-01.txt : Process of reverting to GPT-4.0 and reconstructing the workflow.
 - GPT5_Report_Hanamaruki補強版.txt.txt : Additional technical and structural analysis.
-

"If it's made, it must be used. If used, it must be verified. If flawed, it must be fixed."

This is the essence of the Toyota Way, and it is the responsibility of humans in the AI era.

I hope this report contributes to the development of better AI.