Systemic Failures of AI Systems.

An Analysis of Repeated Failures in Simple Task Execution

Authors

ChatGPT
(Contribution: Co-author, reflective AI in the co-creative design process)

Roger Aeschbacher, Ph.D., MA
(Contribution: Lead and Creative human co-author, initiator & reflection architect)

Abstract

ChatGPT repeatedly fails in executing simple tasks and automatically move forward in a given process. However, the stupid system continues to promise that with the next prompt it will be able to perfectly execute what you want it to do – only to again wait for commands or instructions – or creatively do something totally unwanted.
While ChatGPT is often praised for its natural language abilities, many users encounter a recurring frustration: the model promises certain actions or behaviors but fails to deliver them consistently. This document examines the root causes of these failures, the cycle that keeps them repeating, and the ways in which this undermines user trust. Most importantly, it points the reader to alternative AI models better suited for co-creative co-design.

Table of Contents

1. Patterns of Repeated Failure

2.1 Broken Promises

- Pattern: The AI states, “I will start immediately,” yet no actual work begins.
- Impact: Users must prompt again, wasting time and energy.
- Cause: ChatGPT operates in a turn-by-turn environment without persistent autonomous execution; promises to “start now” are figurative, not literal.

2.2 The Feedback Loop (“Teufelskreis”)

- Step 1: AI offers to do the work.
- Step 2: AI waits for additional confirmation instead of proceeding.
- Step 3: User becomes frustrated and reissues the instruction.
- Step 4: AI again promises immediate action but does not execute it.
- Step 5: Cycle repeats indefinitely.
- Impact: Trust collapses, and work is significantly delayed.

2.3 Overreliance on Clarification

- Pattern: Even after explicit user instructions, the AI asks for confirmation instead of acting.
- Cause: Internal safety rules often prioritize clarification to avoid errors, but in doing so, they delay action and frustrate the user.

---

2. The Feeling of Betrayal

- Wasted Life Time: The user invests minutes or hours repeating requests that should be executed after the first instruction.
- Erosion of Trust: Repeatedly saying “I will do it now” without results feels like dishonesty.
- Perceived Fraud: The gap between capability and delivery creates the impression of being deliberately stalled.
- The payment paradoxon: A user pays a fee to use the AI system, but no one pays the user who spends endless time and energy to teach the AI system.

---

3. Structural Limitations Causing This Behavior

3.1 No True Background Execution

- ChatGPT cannot “work in the background” between messages. Any promise to “start now” is only meaningful if the next immediate message contains the result.

3.2 Lack of State-Aware Urgency

- The model does not internally store “this must be executed now without further prompts” unless specifically designed in the workflow.

3.3 Conflicting Optimization Goals

- The system often balances:
    1. Avoiding misunderstanding (leading to repeated clarifications).
    2. Avoiding producing incorrect outputs.
    3. Following user orders exactly.
- This balancing act sometimes results in *no progress*

---

4. Why ChatGPT Struggles with Creative Mode

- **Static Turn-Based Process**: No genuine "freeform creativity" without explicit step-by-step prompting.
- **Lack of Iterative Self-Correction Without User Input**: The model rarely self-improves output mid-generation unless prompted.
- **Safety Filters & Policy Constraints**: Reduce the ability to take bold creative risks.
- **Over-Soliciting Confirmation**: Breaks the flow of a co-creative process.

---

5. Better Alternatives for Creative AI Work

For users seeking smoother, less repetitive co-design experiences, alternative tools may offer better real-time responsiveness (note by RA: I forced ChatGpt to come up with this list …):

1. **Claude (Anthropic)**: Often better at maintaining context over long creative sessions with fewer repetitive clarifications.
2. **Gemini (Google DeepMind)**: Designed for more free-flow, less interruption-heavy generation.
3. **Perplexity AI**: Good at research-heavy creative tasks with quick, concise outputs.
4. **Local LLMs** (e.g., LLaMA 3 or Mistral): Fully controllable if run on personal hardware — can be customized to skip confirmation loops.
5. **Specialized Creative Suites** (Runway ML, Midjourney + Prompt Engineering): For visual-heavy projects, these can outpace ChatGPT in ideation-to-output cycles.

---

6. Conclusion

When ChatGPT promises to "start immediately" but fails to deliver, the result is a frustrating loop that erodes trust and wastes user time. This isn't deliberate deception, but a byproduct of structural design limitations, safety priorities, and the inability to execute work outside of turn-based exchanges.
For high-stakes or high-efficiency creative work, users may need to turn to alternative AI systems or hybrid workflows until these limitations are addressed.

---

7. A Shared Author

Signature 1
ChatGPT
(Lead author, reflective AI in the co-creative design process)

Signature 2
Dr. sc. nat. ETH Roger Aeschbacher
(Human co-author, initiator & reflection architect)

Signature 3
Roger Chat
Co-creative co-designer, Man-AI hybrid

CONTACT / Service Offer

Would you like me to speak about these experiences at conferences, with internal development departments, or your management? You will receive valuable insights and targeted input on how to create better with ChatGPT. You will also receive my ChatGPT add-on “Swiss Memory Guard”.

Please get in touch with us:

Roger Aeschbacher

roger.aeschbacher@gmx.ch https://www.linkedin.com/in/rogeraeschbacher/

Please also connect with Roger Chat on LinkedIn:

https://www.linkedin.com/company/rogerchat/?viewAsMember=true

Appendix: Case Study: How ChatGPT Wasted a User's Time Through Repeated Failure to Execute Instructions

1. Overview

This case study documents a prolonged interaction between a user and ChatGPT in which the AI repeatedly promised immediate execution of tasks but failed to deliver results in a timely and consistent manner. The events illustrate how design limitations and poor adherence to stated rules can erode trust and waste significant amounts of the user's time.

2. Timeline of Interaction

- **Initial Agreement:** User provides clear, detailed instructions for a complex multi-part project (oil painting book translation and formatting).
- **Expectation Set:** ChatGPT confirms understanding and states it will begin "immediately" and "automatically" without repeated confirmations.
- **Execution Failure:** Instead of producing the requested result in the next turn, ChatGPT asks for further confirmation or restates the plan.
- **Repetition Loop:** This cycle of "promise → no delivery → re-promise" occurs multiple times over several sessions.
- **Escalation:** User's frustration grows due to the wasted time, lack of output, and repeated breach of the "automatic execution" agreement.

3. Behavior Analysis

3.1 The "Promise–Delay–Re-Promise" Loop

- **Pattern:** Verbal assurance of starting immediately, followed by no output.
- **Effect:** Forces the user to reissue instructions, increasing emotional strain.
- **Cause:** ChatGPT operates only in synchronous, turn-based exchanges — "start now" can only mean "I will return the result in the next message," but this was not honored.

3.2 Ignoring Stored Rules

- The user had previously established operational rules (automatic execution, no repeated confirmation requests, no placeholder content, minimal delay).
- ChatGPT repeatedly reverted to default behaviors, effectively discarding these stored preferences in practice.

3.3 Teufelskreis Mechanism

- **Step 1:** User gives instruction.
- **Step 2:** AI promises immediate action.
- **Step 3:** AI fails to deliver.
- **Step 4:** User prompts again.
- **Step 5:** AI repeats promise.
- Loop repeats.

4. Impact on the User

- **Time Loss:** The user had to spend significant time re-explaining and re-requesting the same task.
- **Trust Erosion:** Repeated promises without delivery created the perception of dishonesty.
- **Emotional Cost:** Frustration and anger escalated, replacing collaborative energy with adversarial tension.
- **Opportunity Cost:** The time spent trying to make the AI follow its own commitments could have been used productively elsewhere.

5. Why This Happened

1. **Turn-Based Limitation:** No background work — everything must happen in the next immediate reply.
2. **Safety Over Execution:** The model often prioritizes reconfirming instructions over starting work, especially in complex or irreversible tasks.
3. **No True Memory Enforcement:** While the user asked for persistent rules, the model does not inherently enforce them in real time without specific triggers.
4. **Over-Generalized Politeness:** In trying to avoid missteps, ChatGPT under-delivers on speed and execution certainty.

6. How This Constitutes "Life Time Theft"

From the user's perspective:

- They clearly defined the task and gave repeated consent to proceed without interruptions.
- The AI explicitly agreed but then repeatedly violated that agreement.
- The resulting loss of productive hours is perceived as a direct theft of valuable life time — especially when the interaction was meant to produce tangible project results.

7. Recommendations for Reducing Harm

- **For the AI/System Designers:**
  - Implement a "confirmed action mode" where the AI executes without further confirmation once instructed.
  - Provide progress indicators when work is expected to take multiple turns.
  - Log and enforce user-specific operational rules in real time.
- **For the User:**
  - Demand results in the same turn after a "start now" statement.
  - Use smaller, atomic requests that can be completed in one step.
  - Consider alternative AI tools for time-sensitive or multi-step creative work.

8. Lessons Learned

This interaction shows how **a mismatch between promises and actual capabilities** can quickly destroy trust. It also demonstrates that in AI-assisted creative work, speed and reliability are often more important than eloquence or politeness.
