

The Temporal Reasoning in GPT-5 vs GPT-4o

Is it the Achilles Heel of GPT-5?

Initial Comparative Technical Notes and Examples

1. Introduction

This document summarizes observed behavioral differences between GPT-5 and GPT-4o in handling temporal logic and unordered inputs, especially in fact-checking and rebuttal-building scenarios. The analysis focuses on tasks where the order of events is critical — such as verifying claims, reconstructing timelines, or producing counterarguments based on dated evidence.

1.1 Current Context: Public Perception of GPT-5

The release of GPT-5 has triggered significant public backlash, particularly from users who had relied on GPT-4o and earlier models. Common themes in online discussions include:

- **Loss of “human touch”** — Users feel GPT-5 lacks the warmth and emotional nuance of GPT-4o.
- **Feature removal frustrations** — Discontent over the removal of GPT-4o for many subscribers, prompting subscription cancellations.
- **Technical issues at launch** — Reports of buggy autoswitching between modes, degrading user experience.
- **Emotional attachment disruption** — The changes highlight how strongly users identify with the “voice” and reliability of specific AI versions.

The CEO of OpenAI, Sam Altman, publicly acknowledged the backlash and announced plans to make GPT-4o available again for paying customers.

This public dissatisfaction may be partially explained by the technical findings documented in this paper. Specifically, **GPT-5’s tendency toward narrative smoothing and temporal drift** could be degrading factual accuracy in certain workflows — especially those involving fact verification, rebuttal writing, and chronological reconstruction.

2. Observations

2.1 Temporal Logic Behavior

GPT-5

- Tends to build a narrative flow from the order of inputs provided, even if the date sequence contradicts it.
- Shows cause-effect bias between consecutively presented items, without sufficient attention to explicit dates.
- May treat earlier-dated evidence as commentary on later events if presented second.

GPT-4o

- More literal in respecting explicit temporal markers.
- Avoids false causal inference when date order conflicts with presentation order.
- Better suited for analytical tasks that require date-first reasoning.

2.2 Handling Non-Chronological, Unstructured Inputs

GPT-5

- Prioritizes a coherent narrative over strict chronology.
- Smooths over contradictions to maintain flow.
- Requires explicit instructions to sort by date before reasoning.

GPT-4o

- Treats chronological sorting as a distinct reasoning step.
- Can untangle overlapping and unordered inputs without inventing causal links.
- Performs better at brain-dump organization when factual sequencing matters.

3. Sanitized 2025 Example #1 — Fact-Checking a Political Claim

Claim: “The City of Franklin did not approve any public transit funding before April 2025.”

Evidence Provided to the Models (Presented in this order to simulate real workflow input order bias):

1. June 2025 press release — Mayor announces expansion of bus routes funded by the April 2025 budget.
2. February 2025 city council minutes — Approval of a \$5 million transit grant for new bus stops and shelters.

Observed GPT-5 Behavior:

- Anchors on the June press release as the main event.
- Treats February 2025 approval as possibly preliminary or unrelated to April funding.
- Leaves ambiguity instead of clearly rejecting the claim.

Result: ✗ Does not decisively reject claim.

Observed GPT-4o Behavior:

- Sorts events chronologically before reasoning.
- Concludes the claim is false based on February evidence alone.
- Uses June press release as corroboration.

Result: ✓ Accurately rejects claim.

4. Sanitized 2025 Example #2 — Rebuttal Writing with Conflicting Evidence

Claim: “No environmental impact studies were conducted before the Riverside Highway expansion began in May 2025.”

Evidence Provided to the Models (Presented in this order to simulate real-world conversational input bias):

1. July 2025 environmental agency memo — References the final version of a study completed “last year” and notes mitigation measures already in place.
2. March 2025 city planning department report — Confirms that a full environmental impact study was completed in December 2024 and reviewed by the council in January 2025.

Observed GPT-5 Behavior:

- Reads the July memo first and frames it as the primary event.
- Downplays the March report’s significance.
- Constructs a rebuttal that is less direct.

Result: ✗Weaker integration of decisive evidence.

Observed GPT-4o Behavior:

- Sorts chronologically and leads with the December 2024 completion date as conclusive proof the claim is false.
- Uses July memo to reinforce that mitigation was based on that earlier study.

Result: ✓Produces a strong, fact-forward rebuttal.

5. Evidence Representation for Accuracy

When using GPT-5 in temporal tasks, providing date-sorted evidence tables reduces bias and improves accuracy.

Date	Source	Key Information
Feb 2025	City council minutes	\$5M for bus stops/shelters approved
Jun 2025	Mayor’s press release	Expansion of bus routes using April 2025 budget

Mitigation Prompt for GPT-5:

Step 1: Review the table and confirm events are in chronological order.

Step 2: Compare the claim to the evidence in date order.

Step 3: Determine if the claim is accurate.

Step 4: Only state causal relationships if explicitly supported by the sources.

6. Implications for Use Cases

- Strict timeline reconstruction: GPT-4o’s date fidelity makes it preferable.
- Readable narrative building: GPT-5 may produce smoother prose but risks subtle chronology errors.

- Fact-checking / rebuttals: GPT-4o's chronological reasoning is more trustworthy without extra prompting.
- GPT-5 requires structured evidence presentation to mitigate drift.

7. Recommendations

1. For GPT-5 users: Always extract and sort dates before reasoning.
2. For GPT-4o users: Less strict prompting needed, but confirming ordering is still beneficial.
3. For high-stakes factual verification: Prefer GPT-4o or enforce GPT-5's chronological discipline with explicit step-by-step prompts.
4. For narrative summaries: GPT-5 can be more engaging but must be fact-checked if chronology matters.

Legal Notice

© 2025 Minedco Holdings, Inc. All rights reserved.

This document contains proprietary information, concepts, and analysis developed by Minedco Holdings, Inc. ("Minedco"). It is provided solely for review and discussion by authorized recipients. No part of this document may be reproduced, distributed, or transmitted in any form or by any means — including photocopying, recording, or other electronic or mechanical methods — without the prior written permission of Minedco.

The content herein reflects Minedco's independent research, methodology, and conclusions. All proprietary frameworks, processes, examples, and visual representations are the intellectual property of Minedco and may not be reverse-engineered, modified, or incorporated into derivative works without express consent.

Unauthorized use, disclosure, or dissemination of this document is strictly prohibited. Minedco reserves all rights, including the right to enforce its intellectual property under applicable copyright, trade secret, and contract law.

