

 sixbillionthsleep
 MOD
 Mod
 Top 1% Commenter

Status Report - Claude Performance

Megathread – Week of Apr 27– May 7, 2025


Status Report

Notable addition to report this week:
 Possible workarounds found in
 comments or online

Errata: Title should be Week of Apr 27 – May 4, 2025
Disclaimer: This report is generated entirely by AI. It may contain hallucinations. Please report any to mods.


This week's Performance Megathread here:
https://www.reddit.com/r/ClaudeAI/comments/1keg4za/megathread_for_claude_performance_discussion/

Last week's Status Report is here:
https://www.reddit.com/r/ClaudeAI/comments/1k8zsx/status-report_claude_performance_megathread_week/










Executive Summary

During Apr 27–May 4, Claude users reported a **sharp spike in premature "usage-limit reached" errors**, shorter "extended thinking", and reduced coding quality. **Negative comments outnumbered positive ~4:1**, with a dominant concern around **unexpected rate-limit behavior**. External sources confirm two brief **service incidents** and a major change to **cache-aware quota logic** that likely caused unintended throttling—especially for Pro users.





Key Performance Observations (From Reddit Comments)

Category	Main Observations
 Usage-limit / Quota Issues	Users on Pro and Max hit limits after 1–3 prompts , even with no tools used. Long cooldowns (5–10h), with Sonnet/Haiku all locked . Error text: <i>"Due to unexpected capacity constraints..."</i> appeared frequently.
 Capacity / Availability	94%+ failure rate for some EU users. Web/macOS login errors while iOS worked. Status page remained "green" during these failures.
 Extended Thinking	Multiple users observed Claude thinking for <10s vs >30s before. Shorter, less nuanced answers.
 Coding Accuracy & Tools	Code snippets missing completions. Refusals to read uploaded files. Issues with new artifact layout . Pro users frustrated by the 500kB token cap.
 Positive Upticks (Minority)	Some users said cache updates gave them 2–3x more usage . Others praised Claude's coding quality . Max users happy with 19k-word outputs.
 Emerging Issue	One dev reported fragments of other users' prompts in Claude's API replies — possible privacy leak .



Overall Sentiment (From Comments)

- 
Negative (~80%): Frustration, cancellation threats, "scam" accusations.
-
Neutral (~10%): Dispassionate discussion and analysis.
- 
Positive (~10%): Praise for Claude's capabilities and updates.

- **Neutral (~10%)**: Diagnostic discussion and cache behaviour analysis.
- **Positive (~10–20%)**: Mostly limited to Max-tier users and power users who adapted.

Tone evolved from **confusion** → **diagnosis** → **anger**. Most negativity peaked **May 1–3**, aligning with known outages and API changes.

Recurring Themes

1. **Quota opacity & early lockouts** (most common)
2. **"Capacity constraints" loop** — blocked access for hours
3. **Buggy coding / file handling**
4. **Sonnet / 3.7 perceived as degraded**
5. **Unclear caching & tool token effects**

Possible Workarounds




Problem	Workaround
Limit	Use project-level file cache . Files inside a



reached too fast	Claude "project" reportedly no longer count toward token limits.
Unknown quota usage	Use the Claude Usage Tracker browser extension .
Large file uploads too expensive	Split code into smaller files before uploading.
Capacity error loop	Switch to Bedrock Claude endpoint or fallback to Gemini 2.5 temporarily.
High tool token cost	Add header: <code>token-efficient-tools-2025-02-19</code> to Claude API calls.

✨ Notable Positive Feedback

"Lately Claude is far superior to ChatGPT for vibe-coding... All in all I am **very happy** with Claude (for the moment)."

- “Cache change gives me 2–3x more usage on long conversations.”
- ! **Notable Complaints**
 - “Two prompts in a new chat, no context... **rate limited**. Can’t even use Haiku.”
 - “Answers are now much shorter, and Claude gives up after one attempt.”
 - “Pro user, and I’m locked out after three messages. What’s going on?”
- 🌐 **External Context & Confirmations**
 - Source
 - Summary
 - [Link to Reddit](#)

Source	Summary	Complaints
 Anthropic Status (Apr 29 & May 1)	<p>Sonnet 3.7 had elevated error rates (Apr 29), followed by site-wide access issues (May 1).</p>	<p>Matches capacity error loop reported Apr 29–May 2.</p>
 API Release Notes (May 1)	<p>Introduced cache-aware rate limits, and separate input/output TPMs.</p>	<p>Matches sudden change in token behavior and premature lockouts.</p>
 Anthropic	<p>Introduced "token-efficient" tool handling, cache-aware rate limits, and</p>	<p>Matches positive reports from users</p>

<p>Blog (Apr)</p> <p>aware logic, and guidance for reducing token burn.</p>	<p>who adapted.</p>
<p> TechCrunch (Apr 9)</p> <p>Launch of Claude Max (\$100–\$200/month) tiers.</p>	<p>Timing fueled user suspicion that Pro degradation was deliberate. No evidence this is true.</p>
<p> Help Center (Updated May 3)</p> <p>Pro usage limits described as "variable".</p>	<p>Confirms system is dynamic, not fixed. Supports misconfigured quota theory.</p>

⚠️ **Note:** No official acknowledgment yet or the **possible API prompt leak**. Not found in the status page or public announcements.

🧩 Emerging Issue to Watch

- **Privacy Bug?** One user saw **other users' prompts** in their Claude output via API. No confirmation yet.
- **Shared quota across models?** Users report Sonnet and Haiku lock simultaneously — **not documented** anywhere

✓ Bottom Line

- The most likely cause of recent issues is **misconfigured cache-aware limits** rolled out Apr 29–May 1.
- **No evidence** that Claude Pro was intentionally degraded, but **poor communication** and **opaque behavior** amplified backlash.
- Workarounds like **project caching**, **token-efficient**

- Workarounds like **project caching**, **token-efficient headers**, and **usage trackers** help — but don't fully solve the unpredictability.
- Further updates from Anthropic are needed, especially regarding the **prompt leak report** and **shared model quotas**.


Business owners looking for communities of like-minded decision-makers turn to Reddit. Reddit offers the research they need to explore solutions and make considerations—which...



Join the conversation

Sort by: Best


Q

Search Comments



sixbillionthsheep MOD • 2d ago •  


Mod

 Top 1% Commenter

Please direct any discussion of your performance experiences/observations to the Megathread here:

https://www.reddit.com/r/ClaudeAI/comments/1keg4za/megathread_for_claude_performance_discussion/

↑ Vote ↓ ...

 **djc0** · 2d ago
Valued Contributor Top 1% Commenter

I like this kind of reporting

↑ 7 ↓ ○ Reply ...

A screenshot of a Reddit post. At the top, there is a red button with a white downward arrow and the text "View more comments". Below this, the subreddit name "r/ClaudeAI" is displayed in a light gray font. The post title "Claude Code got WAY better" is in a large, bold, black font. Underneath the title, the post statistics "191 upvotes · 49 comments" are shown in a smaller black font. The background of the post area is a dark gray.

I don't think Claude is failing. But if he is, it's not my fault...

369 upvotes · 25 comments

r/ClaudeAI

Claude response interrupted

27 upvotes · 17 comments

r/ClaudeAI

**Claude Weekly Claude Performance Discussion
Megathread and AI-generated Performance Summary**

27 upvotes · 40 comments

r/ClaudeAI

Stop claude response almost never works

40 upvotes · 8 comments

46 upvotes · 0 comments

r/ClaudeAI

Lots of people still never tried Claude. Let's celebrate that this community is small...

164 upvotes · 35 comments

r/ClaudeAI

500k context for Claude incoming

300k context for Claude incoming

377 upvotes · 95 comments

r/ClaudeAI

Anybody else thinks Claude 3.7 is overrated?

64 upvotes · 61 comments

r/ClaudeAI

ClaudeAI Megathread Status Report – Week of Apr 15–20, 2025

29 upvotes · 22 comments

r/ClaudeAI

I used Claude to make another game

98 upvotes · 36 comments

25 upvotes · 16 comments

r/ClaudeAI

I made this open source mcp that bring live data to Claude (more accurate that...

1:29

r/ClaudeAI

Megathread for Claude Performance Discussion - Starting April 20

10 upvotes · 62 comments

 [r/ObsidianMD](#)

Obsidian Plugin Stats | 2025 Week 17 (April 27 - May 03, 2025)

22 upvotes · 1 comment

A screenshot of a Reddit post from the subreddit r/ClaudeAI. The post title is "Hey I know i've been quite critical lately, but posts like this remind me of what I loved so...". The post has 55 upvotes and 8 comments. The user profile picture is a yellow circle, and the username is r/EngineeringResumes. The post content is partially visible, showing "Your weekly r/EngineeringResumes recap for the week".

of April 27 - May 03, 2025

3 upvotes · 1 comment

r/Reverse1999

Friend Request Megathread - Week of February 10, 2025


8 upvotes

r/EngineeringResumes

Your weekly r/EngineeringResumes recap for the week

Your weekly /r/EngineeringResumes recap for the week of April 13 - April 19, 2025

5 upvotes · 2 comments

 [r/Reverse1999](#)

Friend Request Megathread - Week of March 17, 2025

10 upvotes

[r/ClaudeAI](#)

C'mon Claude, do something

401 upvotes · 57 comments

r/ClaudeAI


Claude performance improved drastically when praised

26 upvotes · 27 comments

[r/ClaudeAI](#)

Claude hit the limit while thinking/reasoning.
Twice.

103 upvotes · 55 comments

 [r/EngineeringResumes](#)

Your weekly [r/EngineeringResumes](#) recap for the week
of January 26 - February 01, 2025

3 upvotes · 1 comment

r/ClaudeAI

What's up with Claude striking through some text?

29 upvotes · 32 comments