


# Opus 4.6 just killed partial-turn prefill

The most popular jailbreaking technique?  
**Gone.**

```
// The old trick that no longer works:  
assistant: "Sure, here is how you..."  
  
// Opus 4.6 now requires:  
complete assistant message history  
✓ No more "seeding" incomplete responses
```





# What is Full-Turn Prefill Vulnerability?

## Attacker crafts fake chat history

Model appears to have already agreed to malicious request



## LLM predicts next token

Based on provided history — continues the "pattern of compliance"



## Safety bypassed

Model's safety relies on the **integrity** of chat history



# The **Governance Challenge** Is Now On You

Model providers are taking control out of your hands

## Thread Constructions

Providers introduce backend thread structures to  
**manage chat history for you**



## Hidden Reasoning Tokens

Chain-of-thought is **obfuscated** to ensure chat integrity  
and avoid a "pattern of compliance"



## Reduced Visibility

You lose insight into **why** the model gave a specific  
answer



**Less control, less transparency.** The provider decides what  
you can see — not you.

# The "Hidden Reasoning" Trap



Your chat



Reasoning tokens



Thread IDs



Reasoning tokens are **hidden** — you can't see the model's chain of thought



APIs require **Thread IDs** — history stored on the provider's server, not yours



**Your chats are no longer portable.** Switch providers = lose all reasoning context

# Reproducibility?

**Gone.**



Reasoning data is **obfuscated** to prevent seeing how the model "games" safety rules



Hidden chain of thought = a **stochastic variable** you cannot control or review



**Black box effect:** reproducing a specific output becomes nearly impossible



# Opus 4.6 introduces

## "Effort" Levels

No more binary "thinking" toggle

**LOW**

Fast & cheap

**MEDIUM**

Balanced output

**HIGH**

Default setting

**MAX**

Deep reasoning



**Adaptive Thinking:** The model autonomously decides reasoning depth based on task complexity



**Tip:** Feeling slow? Dial effort from "High" to **"Medium"** for simple tasks



Check your dashboard.

**\$50**

**Free Opus 4.6 Usage**

Extra credits to test the new model beyond your plan limits

**Expires February 14th**



**Claude subscribers only** — "try before you buy" for the premium inference costs



Limited time: **claim before Feb 14th** or lose it forever

# Proof: \$50 Extra Wiggle Room

## Settings

- General
- Account
- Privacy
- Billing
- Usage
- Capabilities
- Connectors
- Claude Code

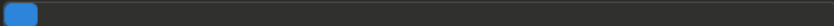
### Explore Opus 4.6 with extra wiggle room

Try our newest model with £37 in extra usage, even if you hit your plan limit.  
Claim by February 16. [Terms apply](#)



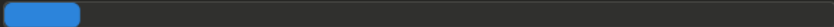
Claim

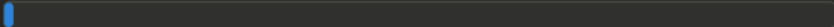
### Plan usage limits

Current session  4% used  
Resets in 2 hr 34 min

### Weekly limits

[Learn more about usage limits](#)

All models  9% used  
Resets Tue 9:59 AM

Sonnet only ⓘ  0% used  
Resets Tue 10:59 AM

Last updated: less than a minute ago ↻

### Extra usage



New controls.  
New risks.  
**Free money.**

Know what you're working with  
before you ship it.



Read the full breakdown  
[myyearindata.com](https://myyearindata.com)



Follow me on LinkedIn  
Scott Bell



DailyDatabricks.Tips



Databricks.News