



What if I told you...

**Opus 4.6 is
actually
Sonnet 5?**

The pricing, behavior, and benchmarks
tell a story Anthropic isn't sharing.





The Pricing Theory

STANDARD (<200K)

\$15 / \$75

Input / Output per 1M tokens

PREMIUM (>200K)

\$30 / \$112.50

2x the cost for "big context"



Same base price as old Opus — but the 1M context window you actually want? **Double.**



Theory: Anthropic **delays Sonnet 5** to maximize revenue at the Opus price tier



Bigger Context ≠ Better Intelligence

The "context rot" nobody talks about

MRCR V2 — NEEDLE-IN-A-HAYSTACK

Opus 4.6

76%

Sonnet 4.5

18.5%



Anthropic's own research: blindly increasing context size can make models **"dumber"**



Performance **degrades** as conversations lengthen — the model forgets earlier context



The TOON Benchmark Exposes a Weakness

Retrieval Accuracy — Token-Oriented Object Notation

GPT-5 Nano



Haiku 4.5



Haiku 4.5 performs on par with **"lobotomized" Grok** in retrieval tasks



A **fundamental struggle** in the Anthropic family with retrieving specific data from large structures



Smarter on benchmarks.

Slower in practice. Less **personable**.



5–10 min for tasks that took **1–2 min** — "measure twice, cut once" reasoning



Lost prose quality — reverts to **templated, robotic responses**



Read the full breakdown

myyearindata.com



Follow me on LinkedIn

Scott Bell



DailyDatabricks.Tips



Databricks.News