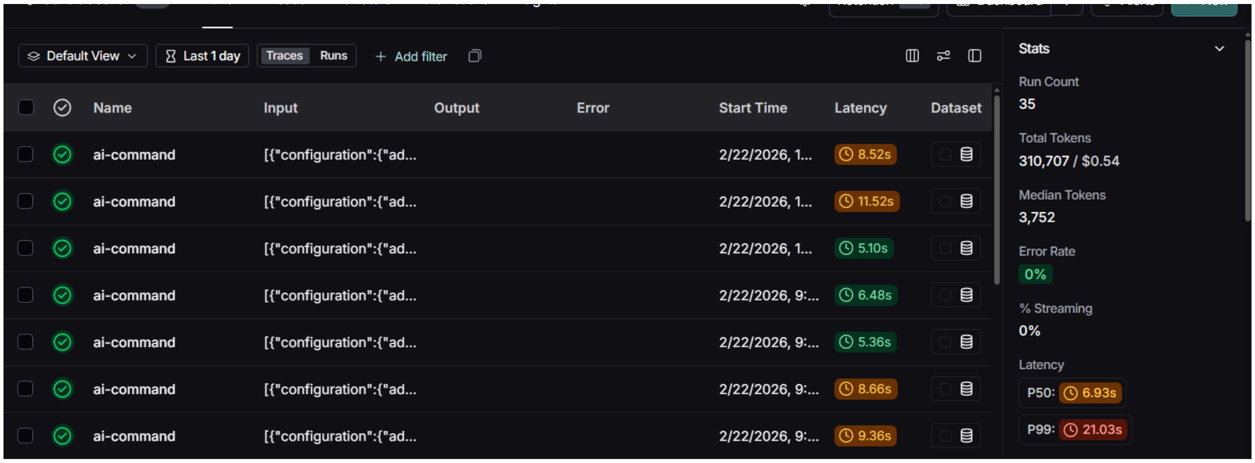
When using Claude Haiku for AI Chat commands/tool/calling/etc:
The rough average of a command issued by a typical end user to my AI chat is 1.25 cents.



□	✓	Name	Input	Output	Error	Start Time	Latency	Dataset
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 1...	8.52s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 1...	11.52s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 1...	5.10s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 9:...	6.48s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 9:...	5.36s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 9:...	8.66s	
□	✓	ai-command	[{"configuration":{"ad...			2/22/2026, 9:...	9.36s	

Stats

Run Count
35

Total Tokens
310,707 / $0.54

Median Tokens
3,752

Error Rate
0%

% Streaming
0%

Latency
P50: 6.93s
P99: 21.03s

Assumptions:
User submits 20 commands an hour, so 25 cents an hour per user.
$6 a day per user
100 users:$18,000
1000 users:$180,000
10000 users:$1,800,000
100000 users:$18,000,000
A lot of money. It will be important to optimize everything as much as possible.

I am using the $200/month claude clode subscription and I used about 20% of my limit this week while coding, although I coded serially with one claude code session 99% of the time.