LOCAL TELEMETRY / ZERO JUDGMENT

Agentic Swear Jar.

PROMPTS IN NUMBERS ✦ 100% LOCAL

SCREENSHOT

A field report from the moments when code fought back.

The harness is a machine. No feelings were hurt.

Combined | Claude Code | Codex CLI

PROMPTS SCANNED

4,441

203,798 words, streamed locally

TOTAL SWEAR HITS

79

0.39 per 1,000 words

SPICY PROMPTS

1.4%

61 prompts contained ≥1 hit

LONGEST CLEAN RUN

292

consecutive prompts

Harness face-off

Normalized by prompt count, because volume alone is a lousy referee.

Claude Code

1.7

hits / 100 prompts · 2,764 prompts total

VERSUS

Codex CLI

1.91

hits / 100 prompts · 1,677 prompts total

Pressure over time

Monthly share of prompts containing at least one match.

When the code bites

Raw matches by local hour of day.

AN UNNECESSARY EXPERIMENT / EXCELLENT CHARTS

Apparently, I swear at coding agents.

So I analyzed every prompt I sent to Claude Code and Codex CLI. **Locally. With zero AI.**

4,444	79	1.4%
PROMPTS SCANNED	MATCHES FOUND	SPICY PROMPTS

CLAUDE CODE VS. CODEX CLI

Which harness gets the spicier prompts?

Claude Code

1.70

matches / 100 prompts

1.52% of prompts contained a match
2,764 prompts total

vs

Codex CLI

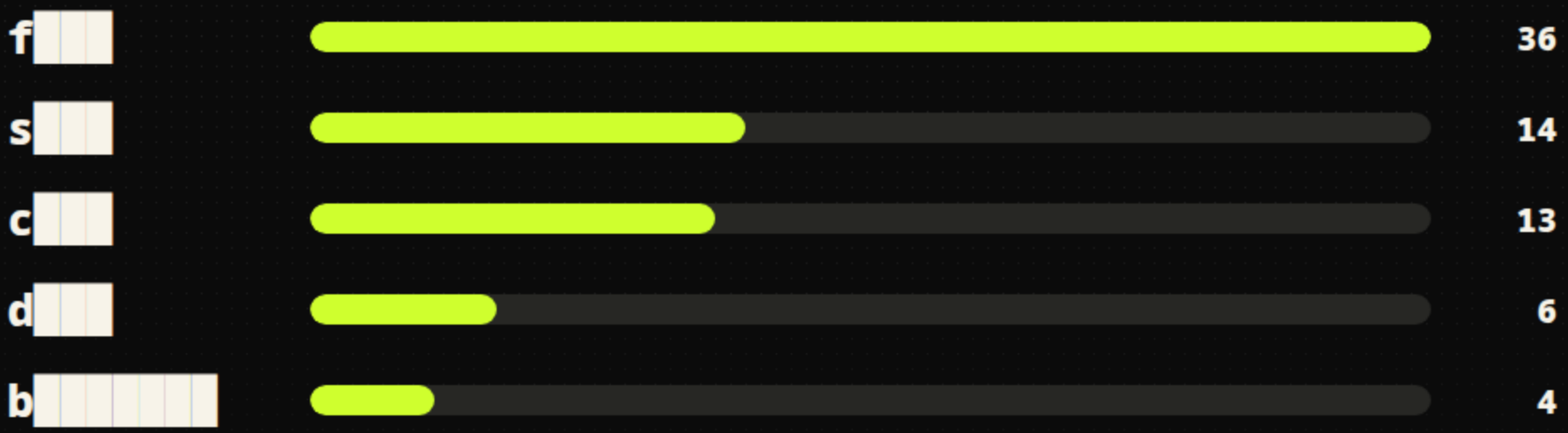
1.90

matches / 100 prompts

1.13% of prompts contained a match
1,680 prompts total

> **Plot twist:** Codex gets more spicy words per prompt. Claude sees them across more prompts.

What frustration looks like.



292
LONGEST CLEAN RUN

14:00
PEAK LOCAL HOUR

0.39
HITS / 1K WORDS

BUILT WITHOUT A CLASSIFIER

No AI was needed to analyze the AI work.

01

Stream local JSONL histories

→

02

Regex tokens + hash-map lexicon

→

03

Aggregate private, redacted stats

203,913 words analyzed in one pass. No prompts uploaded. No raw text retained. Observed 2026-04-10 → 2026-08-11.

Would you run this on your agent history?

`github.com/marcelpetrick/AgenticSwearJar`