User Interaction Log

Claude Code Commit Tracker Project

Date: 2026-02-07 | Single Session | ~4 hours elapsed

This document records every user prompt in a single Claude Code session that produced a full-stack Next.js application tracking AI's footprint in public software engineering. The goal is to illustrate: (i) the amount of user time required, (ii) the types of prompts provided, and (iii) the total run-time.

Key Numbers

Total user prompts: 21

User typing time: ~3-7 min

Total user time (incl. setup): ~15-20 min

Total wall-clock time: ~4 hours

1. Complete User Prompts (Chronological)

#1 | Specification (pasted plan) | ~2-5 min
"Implement the following plan:" [full implementation plan -- see Appendix A] and "keep a detailed log of this project, particularly my input"

#2 | Meta instruction | ~5 sec
"also, keep a detailed log of this project, particularly my input"

#3 | Go signal | ~3 sec
"let's try running it"

#4 | Clarification question | ~5 sec
"Explain what turso and bigquery is?"

#5 | Action request | ~5 sec
"Ok, I want to generate the desired output asap. How do I set up Turso and BigQuery?"

#6 | Status update + delegation | ~10 sec
"Ok. I am setting up bigquery, and I just got the JSON key. Tell me where to put it. In the interim, also execute on Turso"

#7 | Status update | ~3 sec
"I just logged in to turso"

#8 | Credential (file path) | ~5 sec
[User provided local file path to BigQuery service account JSON key]

#9 | Reference link | ~5 sec
"also read this: https://www.gharchive.org/"

#10 | Reference link | ~5 sec
"Here is a list of github event types: https://docs.github.com/en/rest/using-the-rest-api/github-event-types"

#11 | Debugging guidance | ~10 sec
"Check GH archive again. Identifying claude-driven commits or pushes may require a different specification of the search"

#12 | Feature expansion | ~15 sec
"If these tests work, I also want to include code commits by all LLMs (in addition to Claude). Specify the LLM source, whether GPT or others, as well as the model version."

#13 | Scope clarification | ~5 sec
"Also, I want to generate the times series starting in June-2025"

#14 | Parallel request | ~5 sec
"while doing this, show me how the dashboard / visualization looks like"

#15 | Research question | ~15 sec
"I want to understand how much AI is involved in software engineering, which is what this exercise was for. Are there other interesting descriptions in the github data that might be worth extracting or evaluating?"

#16 | Go signal | ~3 sec
"Add all of these"

#17 | Research question | ~10 sec
"Are there other public repositories or datasets like github archive, which might help us understand AI penetration in software?"

#18 | Feature request | ~5 sec
"Add PyPI and npm download tracking for AI libraries"

#19 | Action request | ~3 sec
"reload dashboard"

#20 | Meta / reporting | ~15 sec
"Show me what you have logged for my interactions with you. The goal is to illustrate i) the amount of user time it requires, ii) the type of prompts that I provided and iii) the total run-time"

#21 | Edit + generation request | ~10 sec
"edit #8 to blank out the file directory / etc. For #1, include the original prompt. Generate a .pdf of this"

2. Summary Metrics

Metric	Value
Total user prompts	21
User typing/input time	~3-7 minutes
User actions outside chat	Logged into Turso CLI, set up BigQuery in GCP (~10-15 min)
Total estimated user time	~15-20 minutes
Total elapsed wall-clock time	~4 hours (single session, 2026-02-07)
Active Claude run-time	~3.5 hours (code gen, API testing, orchestration)
Background backfill time	~5 hrs (commits) + ~1 hr (packages), concurrent

3. Prompt Type Breakdown

Type	#	Examples
Go signals ("do it")	4	"let's try running it", "Add all of these", "reload dashboard"
Feature requests	4	"include all LLMs", "starting in June-2025", "Add PyPI and npm"
Research questions	2	"other interesting descriptions?", "other public datasets?"
Status / credentials	4	"I just logged in", [file path to JSON key]
Reference links	2	GH Archive URL, GitHub event types URL
Debugging guidance	1	"Check GH archive again... may require different specification"
Clarification questions	2	"Explain turso and bigquery", "How do I set up?"
Meta instructions	2	"keep a detailed log", "show me what you have logged"

4. What Got Built (with ~15-20 min of user time)

- **30+ source files** Full Next.js app with 8 chart components, 5 API routes, 3 DB tables
- **20 GitHub API queries/day** Commits, co-author trailers, code reviews, PR mentions, bot activity
- **12 package download queries/day** 6 npm + 6 PyPI AI SDK packages tracked daily
- **250-day time series** June 2025 to Feb 2026, backfilling autonomously in background
- **60+ source research doc** Comprehensive catalog of data sources for measuring AI in software
- **Live interactive dashboard** Charts with date range picker, dark mode, responsive layout

5. Session Timeline

0:00 User pastes implementation plan. Claude scaffolds full Next.js app (30+ files).

0:15 User: "let's try running it". Turso DB created, BigQuery configured.

0:30 BigQuery/GH Archive fails (commit data stripped Oct 2025). Claude investigates.

0:50 Claude pivots to GitHub Search API. First data: 142K Claude commits on Feb 5.

1:00 User: "include all LLMs". Claude discovers 5 AI tools, expands to 9 API calls/day.

1:15 Backfill launched (June 2025-Feb 2026). Dashboard shows first data.

1:45 User: "other interesting github data?" Claude researches, finds reviews/PRs/bots.

2:00 User: "Add all of these". Schema expanded to 20 API calls/day, 3 new charts.

2:30 User: "other public datasets?" Claude compiles 60+ source catalog.

2:45 User: "Add PyPI and npm". New pipeline: 12 packages, 2 new line charts.

3:30 Both backfills running. 39 days commit data, 120 days package data loaded.

4:00 User: "show me the interaction log". This document generated.

Appendix A: Original User Prompt (#1)

The user provided the following implementation plan as their first prompt (preceded by "Implement the following plan"):

```
# Claude Code Commit Tracker - Implementation Plan

## Context
The SemiAnalysis article reports ~4% of public GitHub commits are now authored
by Claude Code (identified by Co-Authored-By: Claude ... <noreply@anthropic.com>
trailers), projected to reach 20%+ by end of 2026. This app will track and
visualize that trend with a public dashboard.

## Architecture
Next.js App (App Router) on Vercel
 - BigQuery (GH Archive dataset) -- daily cron query for commit counts
 - Turso (LibSQL) -- stores aggregated daily/weekly stats
 - shadcn/ui + Recharts -- dashboard charts
 - Vercel Cron -- triggers daily data collection at 08:00 UTC

## Data Source: Google BigQuery + GH Archive
- githubarchive.day.YYYYMMDD public dataset, all public GitHub PushEvents
- Query filters for Co-Authored-By trailers matching claude...@anthropic.com
- Extracts model breakdown (Opus/Sonnet/Haiku) from trailer text
- Free tier: 1TB/month, sufficient for 1 daily query (~10-30 GB each)
- Limitation: PushEvents cap at 20 commits per event (noted in UI)

## Database: Turso (LibSQL via Drizzle ORM)
Tables: daily_stats, weekly_stats, ingestion_log

## Implementation Steps (7 Phases, 20 Steps)
Phase 1: Project Setup (create-next-app, deps, shadcn, env)
Phase 2: Database Layer (Drizzle schema, Turso client, push)
Phase 3: BigQuery Integration (regex patterns, query builder)
Phase 4: Data Collection API (cron/collect, cron/backfill, vercel.json)
Phase 5: Stats API (GET endpoint with range params, caching)
Phase 6: Dashboard Frontend (page.tsx, 7 components, loading/error)
Phase 7: Polish (dark mode, responsive, OG tags, health check)

## Key Files
- src/lib/bigquery.ts, src/db/schema.ts, src/app/api/cron/collect/route.ts
- src/app/api/stats/route.ts, src/components/dashboard/ClientDashboard.tsx

## Cost: $0-20/month
- Vercel Hobby (free) or Pro ($20), Turso free tier, BigQuery free tier
```