PyPharma: Modernising Clinical Data Transformation

or You Already Know How to Do This

Michael Chow · Posit

2026-07-22

Who am I

michael.chow@posit.co

- Software engineer at **Posit** (e.g. great-tables)
- Host **The Test Set** — with Hadley Wickham & Wes McKinney
- **Last year**: *polars: foundations of the pharmaverse-py*
- Pharma workshops with Phil Bowsher

Choosing a language identity

Hadley Wickham

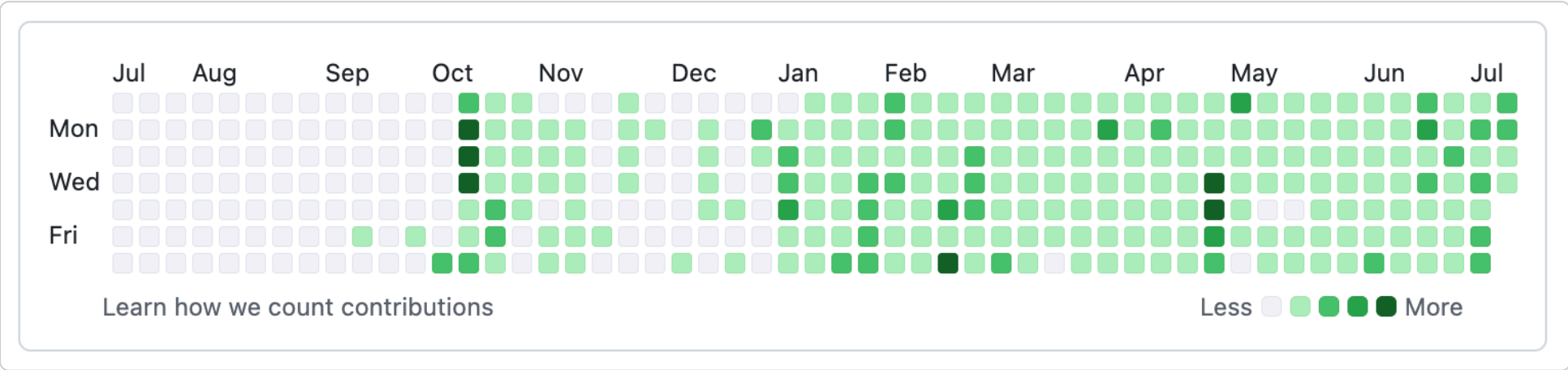
R · the tidyverse

Wes McKinney

Python · pandas · Arrow

For yourself: SAS, or R, or Python?

Wes McKinney, 2026 (Go developer)



> "I have been building a lot of new software in Go. Except that I've never actually written a line of Go in my life."

github.com/wesm · "From Human Ergonomics to Agent Ergonomics," Jan 2026

Switching languages costs time

- R pharmaverse → py-pharmaverse
- R ggplot2 → Py plotnine
- R gt → Py Great Tables

> "Both Hadley and I have spent a **huge amount of time** implementing CSV readers... two prolific open source data science developers building CSV readers. There's a misallocation of resources."

Wes McKinney, Rill Data podcast, Feb 2026

AI: the era of “why not”?

Agent ergonomics (10-100x faster)

Agents run the edit–compile–test loop **10–100×** more than a human ever did.

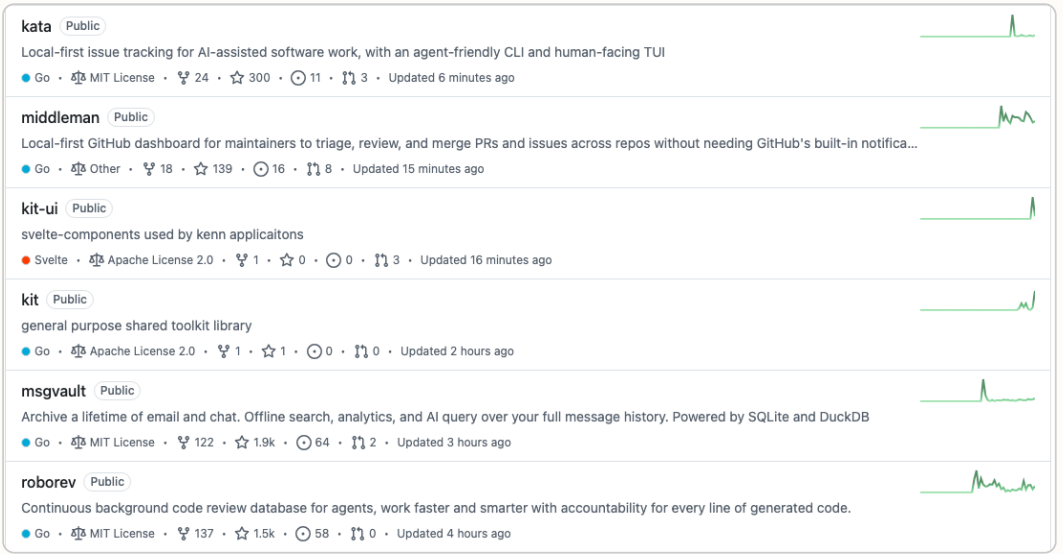
- Before: *what does the human know?*
- Today: *how fast is the feedback loop?*
- Fast builds, static binaries, painless distribution start to win

Wes reports **~10 billion tokens/month** across Claude, Codex, etc.

wesmckinney.com/blog/agent-ergonomics

The era of “why not?”

- Wes shipped five tools in months w/ Go, Swift, Python:



- You could switch from SAS to R, or R to Python

wesmckinney.com/blog/why-not · github.com/kenn-io

The same person also says this

> “Vibe coding is very dangerous and irresponsible… We can’t disengage from planning and writing specs.”

Raw agent output is *“the Potemkin facade of the software.”*

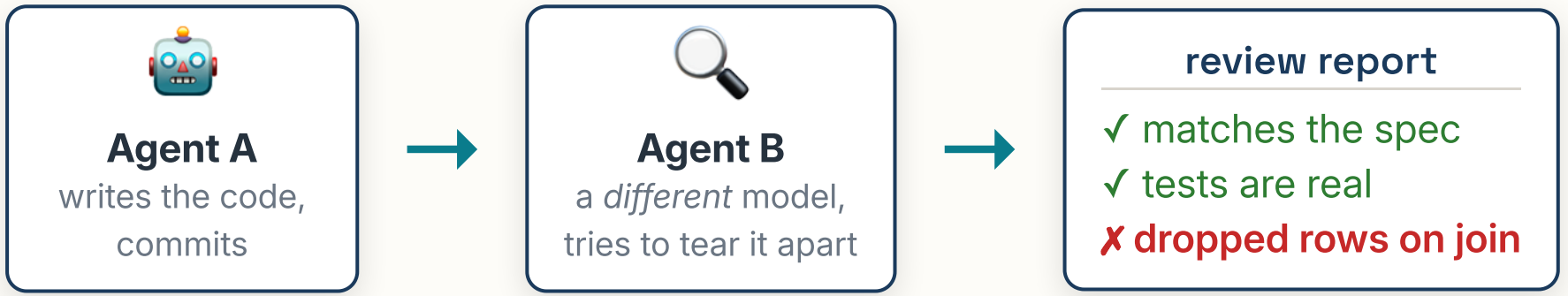
Wes McKinney, MotherDuck interview / Joe Reis Show, 2026

Cheap code ≠ trustworthy code.

a subway vent behind a townhouse facade

How do you trust code in a language you've never used?

Answer: process — roborev

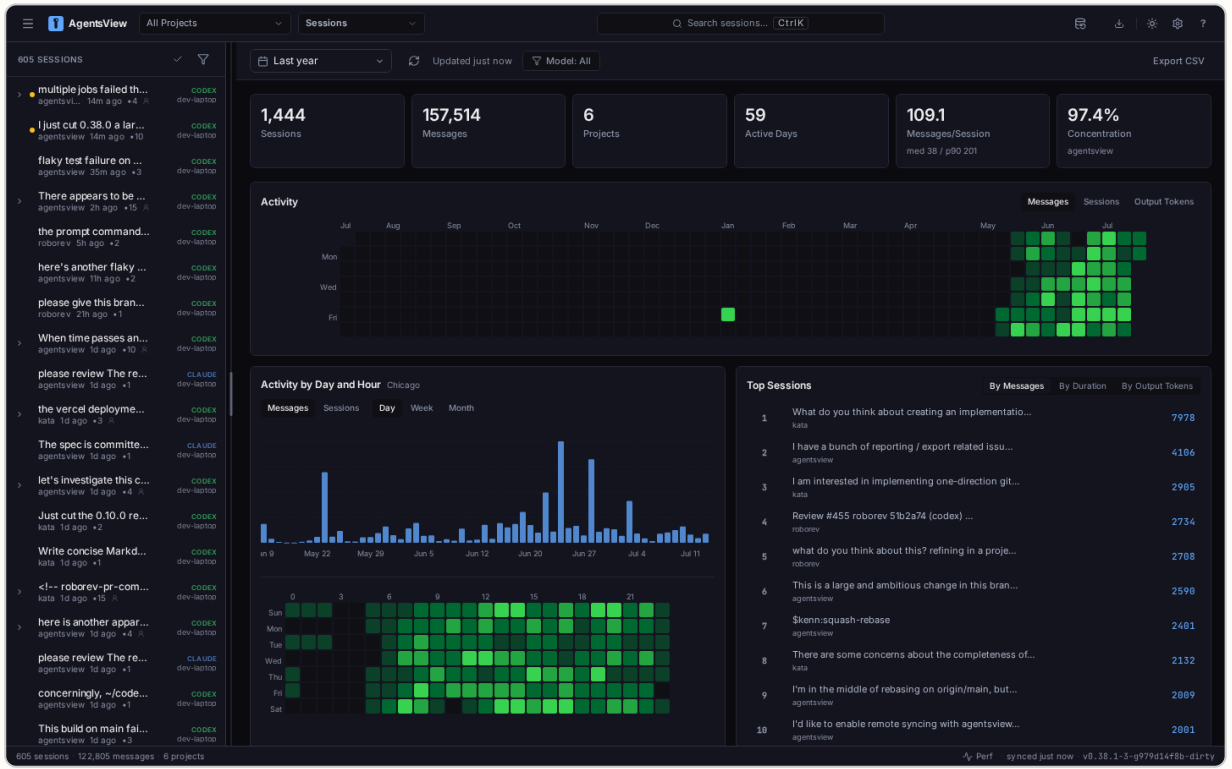


150+ automated reviews per day — *"accountability for every line of generated code"*

github.com/kenn-io/roborev

…and tracked tasks, and audit trails

- **kata**: agent work as auditable tasks
- **agentsview**: every session searchable, every token accounted for



You already know how to do this

Clinical QC culture	Agentic engineering (2026)
Double programming / independent QC	Adversarial review by a *different* model
SOPs & issue tracking	kata: agent work as tracked tasks
Traceability & audit trail	agentsview: searchable session + token logs
Controlled environments	Pinned dependencies, lockfiles, containers

Agentic engineering rediscovered your QC culture.

There’s a lot of QC work ahead

- You’ve done QC for a long time (with SAS)
- R is more recent
- Python is just starting

Integrating AI into the SCE

Your Statistical Computing Environment — where QC culture already lives (h/t Phil Bowsher)

QC culture	In the environment (Posit Team)
Controlled packages	Package Manager: frozen, validated (R and Python)
SOPs for AI	Admin-enforced assistant: which providers, which models, private endpoints, off in prod
Review & traceability	Positron + Quarto + Connect: human-approved, reproducible, published with provenance

Disclosure: I work at Posit — ask your platform team what your equivalent is

Python in clinical, mid-2026

Shipped & real

- **CDISC CORE** — rules engine, 100% Python
- **rtflite** — TLF RTFs on polars *(pharmaverse)*
- **py-pkglite** — eCTD packaging *(pharmaverse)*
- **great-tables** — display tables, polars-ready
- **pycsr.org** — the full submission playbook
- FDA reviewers using Python **by name**

Not yet real

- No admiral-py — derivation is R-only
- No validation hub / riskmetric analog
- No FDA pilot (R is on Pilots 6–7)

Where Python enters: devices & RWE

- Medical device teams run **Python-first** clinical reporting
- pycsr's lead author: Merck biostatistics → **Meta's health org**
- FDA (Dec 2025) finalized RWE guidance — messy EHR/claims/NLP data (Python heavy work)

h/t Phil Bowsher — github.com/philbowsher

The gap is now cheap to fill

- A pharmaverse-py used to cost **years of expert work**
- Wes: five tools in months, in languages he didn’t know
- The hardest part is still **trust**: review, traceability, governance

…which I think you might be super good at?

Try next week

- **polars + great-tables** — big data + display tables in Python (last year's talk)
- **CDISC CORE** — run the open-source conformance engine: cdisc.org/core
- **pycsr.org** — the Python submission playbook, end to end
- Experimenting with agents? **agentsview + kata** — github.com/kenn-io
- Read Wes: *From Human Ergonomics to Agent Ergonomics* · *The Mythical Agent-Month* — wesmckinney.com

Last year: how do we build the pharmaverse with Python?

This year: how do we trust code we didn't write?

Michael Chow · Posit · **mchow.com**

🎙️ *thetestset.co* — with Hadley Wickham & Wes McKinney

Sources

- Wes McKinney — *From Human Ergonomics to Agent Ergonomics* (Jan 2026), *Why Not?* (Jan 2026), *The Mythical Agent-Month* (Feb 2026) — wesmckinney.com/blog
- Rill Data podcast (Feb 2026) & Joe Reis Show (Apr 2026) transcripts — wesmckinney.com/transcripts
- Kenn Software — kenn.io · github.com/kenn-io
- Posit blog — *The Prolific Output of Wes McKinney in the Age of Agentic Engineering* (Feb 2026)
- pharmaverse Python — pharmaverse.org/e2eclinical/python (rtflite, py-pkglite)
- CDISC CORE — github.com/cdisc-org/cdisc-rules-engine · cdisc.org/core
- pycsr — pycsr.org (Zhang & Xiao) · FDA Python mentions — github.com/philbowsher/Use-of-Python-in-regulatory-submission-related-work
- FDA briefing doc, Intercept NDA 212833 (2023) — Fig. 10 by FDA review staff in Python — fda.gov/media/168215
- R Submissions Pilots 6–7 — r-consortium.org (Feb 2026) · R Validation Hub — pharmar.org
- FDA — *Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices*, final guidance (Dec 18, 2025; supersedes 2017) — fda.gov/media/190201