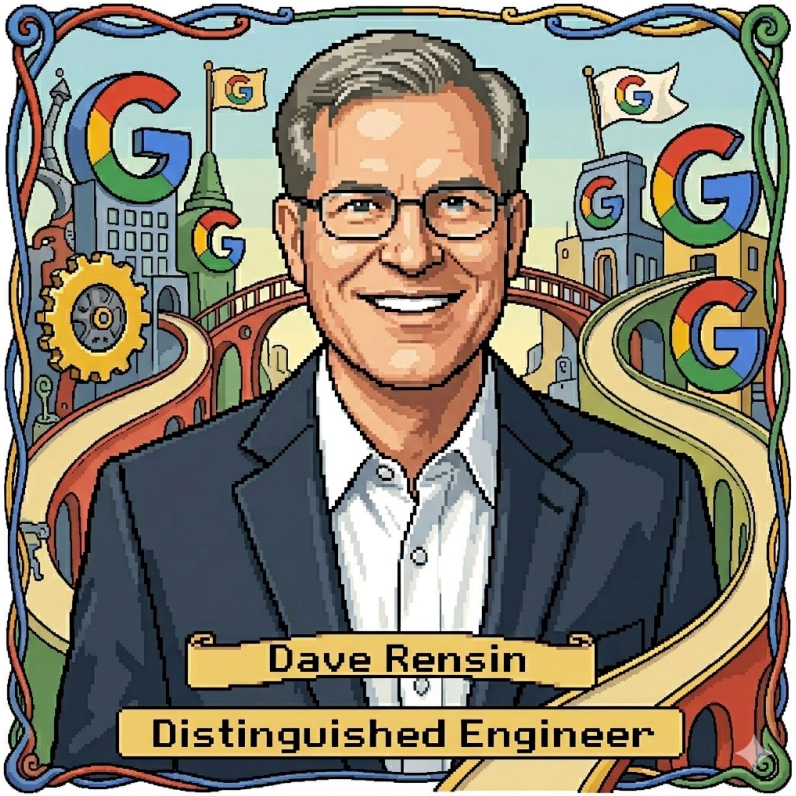


Judgment Is Still the Job

Software Engineering in the Era of Agents
University of Michigan EECS 481

Dave Rensin · Nov 17, 2026
egm.dev · dkr.bio



Hi, I'm Dave.
My path wasn't linear.

Distinguished Engineer at Google.
Why this is the most fun time to build.

- Lots of roles at Google since 2014 (dkr.bio).
- Today: teaching engineers and agents to work together.
- Curiosity and grit beat a straight line or perfect GPA.



Three ground rules before we start.

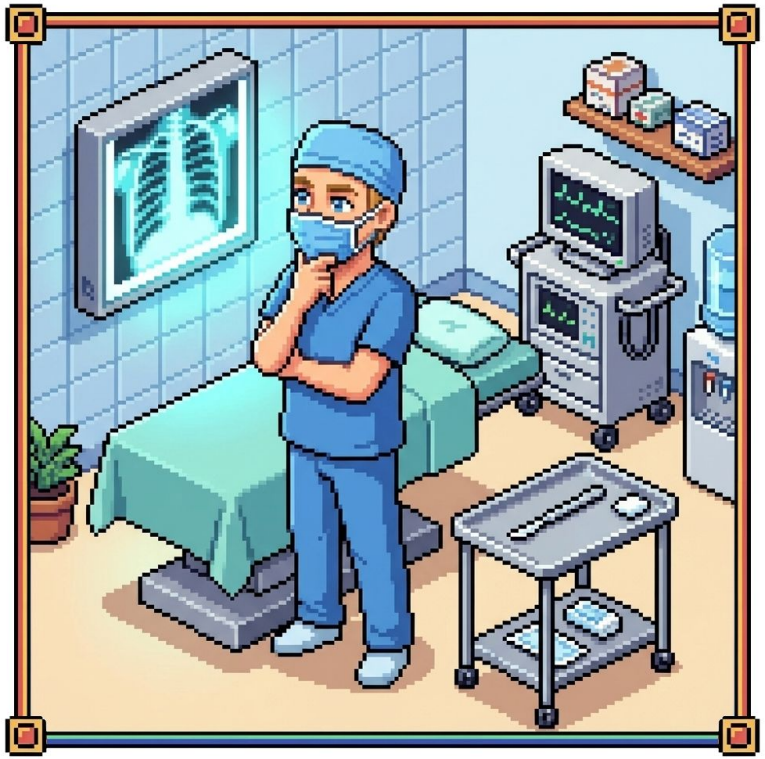
Argue to learn, not to win.

- Everything here is eventually wrong.
- Reality wins every time. Join Team Reality.
- Polite nodding helps nobody. Push back hard.



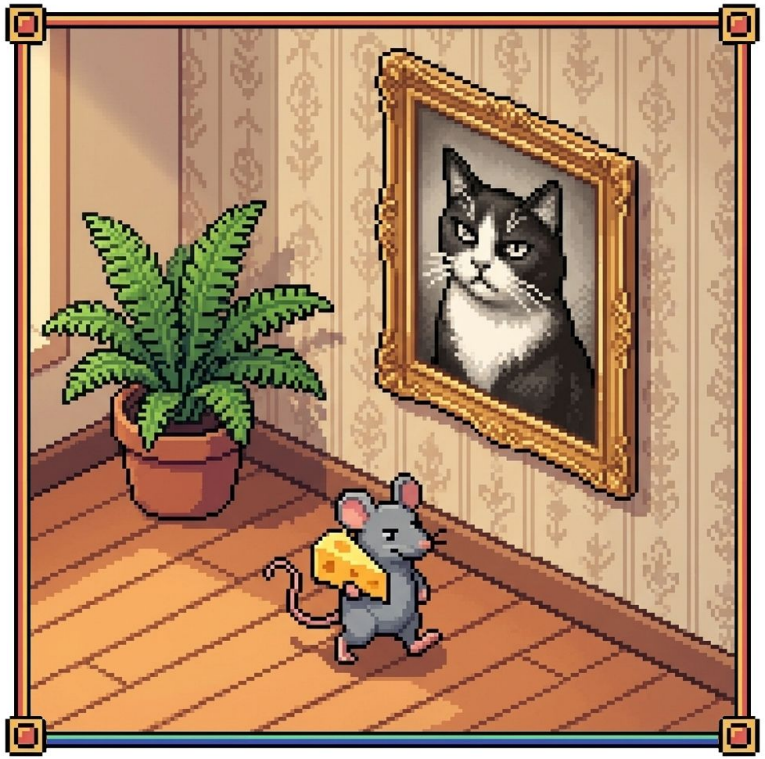
Jevons' Paradox: undefeated since 1865.

Cheaper code unlocks the other 95%. AI needs more good engineers, not fewer.



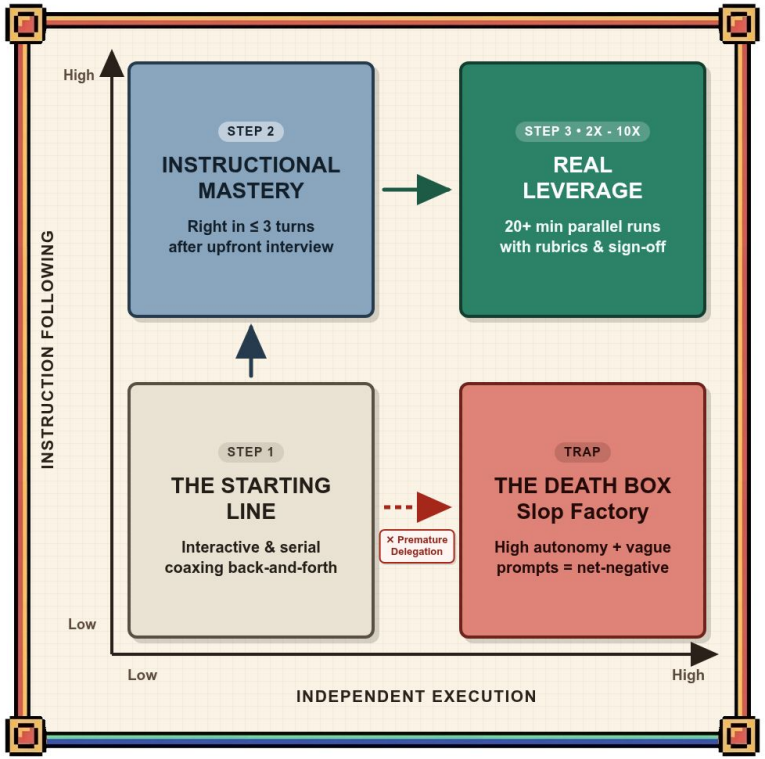
Surgeons aren't paid to just hold scalpels.

Going 10x faster in the wrong direction isn't productivity.



A photo of a cat can't catch a mouse.

Text that sounds like judgment isn't judgment.



Go up before you go across.

Skip straight to autonomy and you build a Slop Factory.

egm.dev · github.com/google/elephant-goldfish-model



No more "solo coders." We are all managers now.

Directing silicon looks like managing people - just ten times faster.

- Give clear goals, constraints, and real guardrails.
- Stop letting the agent agree just to make you happy.
- Inspect the result yourself before you ship it.

What 107 EECS 481 Reports Showed in HW2

Anonymized class summary: planning and fresh eyes worked; 9 manual steps took too long.

91.3%

Planning Beat One-Shot

No Code + proposal.md surfaced libpng and pngtest limits before writing a single line.

90.2%

Fresh Eyes Caught Bugs

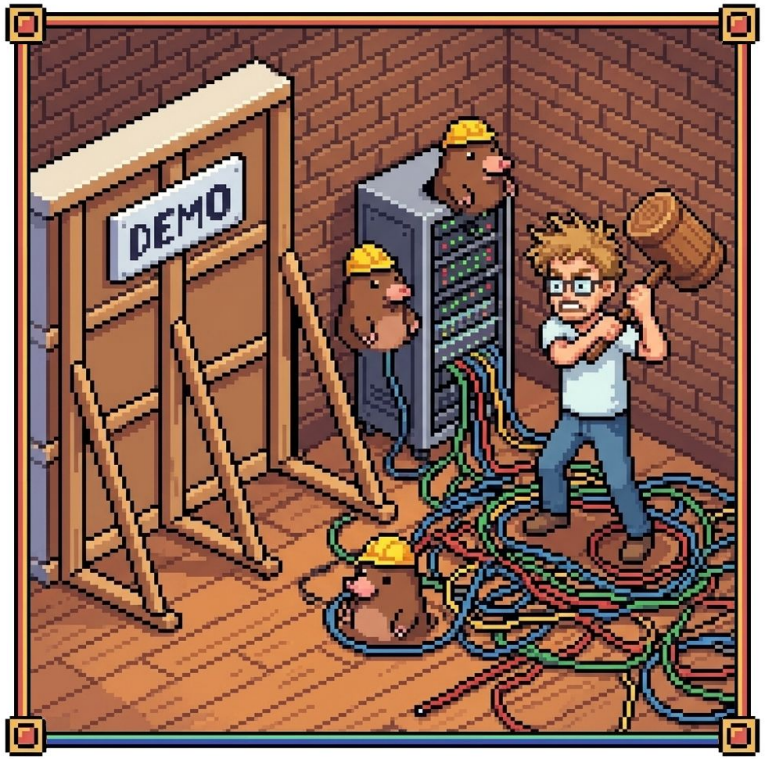
Zero-history Goldfish reviews caught blind spots builders missed (86.6% ranked critic #1).

96 min

Median Time on HW2

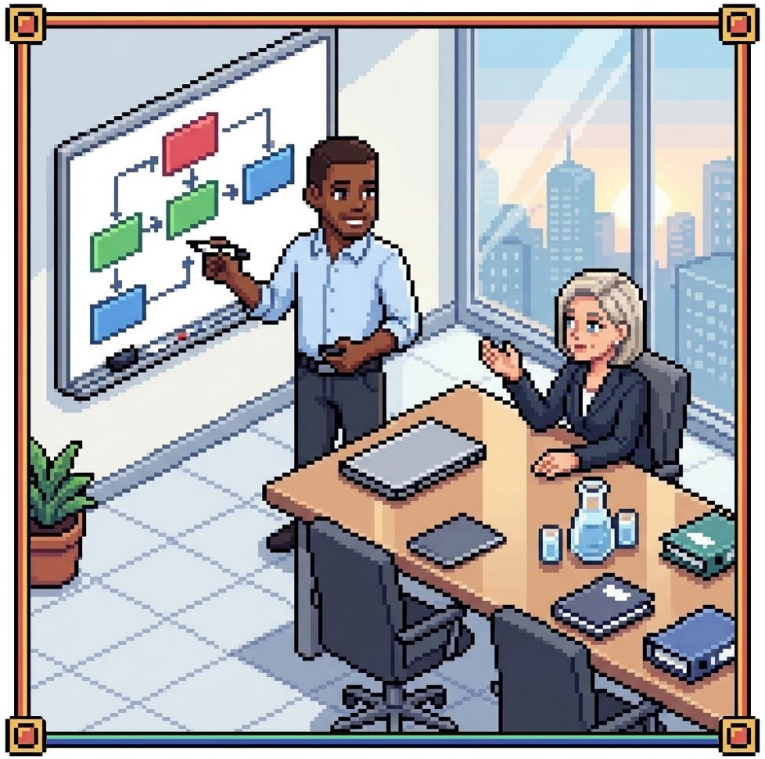
81.5% said 9 manual steps felt too long - which is why we built and open-sourced egm.dev.

And in HW3 (unprompted): 36% of you used EGM anyway - scoring +20 pp higher median accuracy spotting good patches.



Vibe coding feels magical on Monday.

By Wednesday it's whack-a-mole.
By Thursday it's a landmine.



Close the laptop and grab a marker.

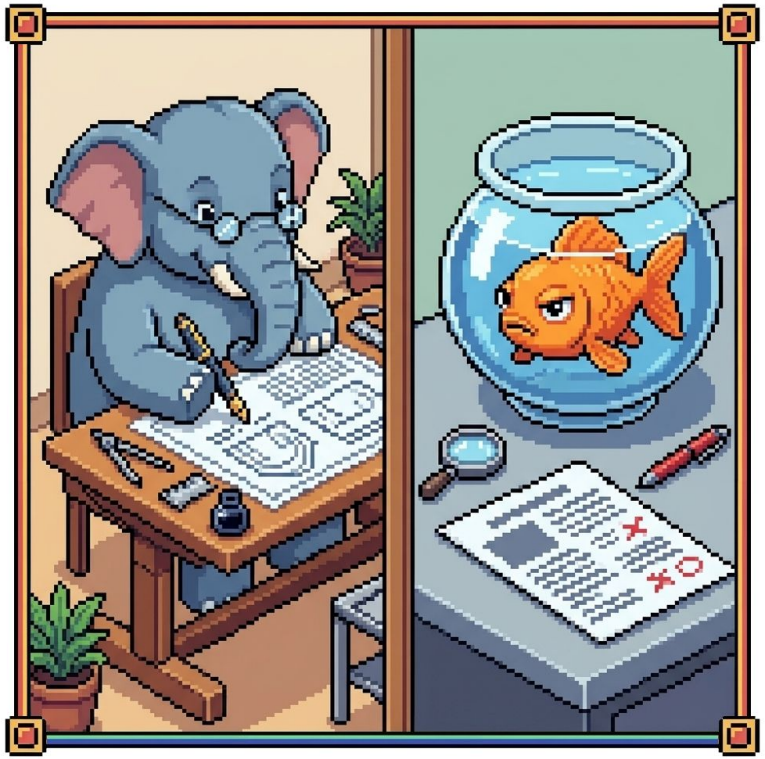
And never type
"pick the best option for me."



Treat it like a power tool, not a toy.

Make the machine interview you before it writes a word.

1. Interrogate: "Challenge me until I say stop."
2. Scoreboard: Co-write a skeptical rubric.
3. Clean room: Execute in a brand-new window.



The design is the code.
All the rest is typing.

Build with a stateful Elephant.
Audit with a cold-eyed Goldfish.

egm.dev · github.com/google/elephant-goldfish-model

Does Spec-First Actually Work?

Measured on a 1.3M-line legacy Google codebase and independent outside benchmarks.

+85.9%

Changes / Month

15-year-old Google test framework (~1.3M lines, 3,180 dirs). Merge time fell 46.5% (7.1h -> 3.8h) with ~90% test coverage.

-68.8%

Fewer Rollbacks

Rollbacks dropped 5.61% -> 1.75%, alerts fell 75.7%, and incident closure sped up 85.4% (alert drop was also a team goal).

25% -> 77%+

First-Try Tests Passed

Outside Codex runs jumped from 4/16 to 12.3-16/16 passing tests; another audit caught 26 plan flaws -> 24/24 passed on run 1.

Real adoption: Every new Google engineer now learns this in onboarding - plus ~21K GitHub stars & UMich/UW CS.



Give yourself air cover for the two-week dip.

The first five specs feel slower.
Then they compound forever.

Three Ways to Contribute to the Repo

Pick one small project from HW2 or HW3. **Optional - does not count as HW6.**

Option 1 · From HW2

Add a Test Case

Turn a tricky libpng bug or edge case you hit in HW2 into a reproducible test example others can run.

`PR: Test file + README`

Option 2 · From HW3

Sharpen a Reviewer

Spot a bug the AI reviewer missed in HW3? Tighten its prompt so Goldfish catches that bug next time.

`PR: Prompt + sample diff`

Option 3 · Tooling

Automate a Check

Write a small helper script or GitHub check that trims manual copy-pasting out of the loop.

`PR: Helper + unit test`

```
git clone https://github.com/google/elephant-goldfish-model && pytest tests/unit
```

Office Hours, Mentorship & Shipping a PR

1. Claim eecs481 issue -> 2. Branch -> 3. pytest -> 4. Review & merge

Nov 19 & Dec 1 · Live Office Hours

Live Help with Oliver Newth

Drop in to pick an issue, test your change, or get your pull request merged (newth@google.com).

Tue Dec 1 (1:30-2:00 PM ET) · Class Deep-Dive

Inside a Real Agent Trace

Oliver walks through how specs stop "XYZ" assumptions and fresh-eyes reviews catch drift.

3 Spots · Contributor Raffle

1-on-1 Google Career & Internship Chats

Every student who merges a valid PR enters a drawing for a 30-min 1-on-1 chat with Oliver.





Keep the judgment. Share the playbook.

Let's open it up and take your toughest questions.

- Site & Plugin: egm.dev
- GitHub Repo: google/elephant-goldfish-model
- 60-Sec Video: youtu.be/8_x1u7qSqGs
- Medium Essay: Elephants, Goldfish & Golden Age
- Bio & Links: dkr.bio