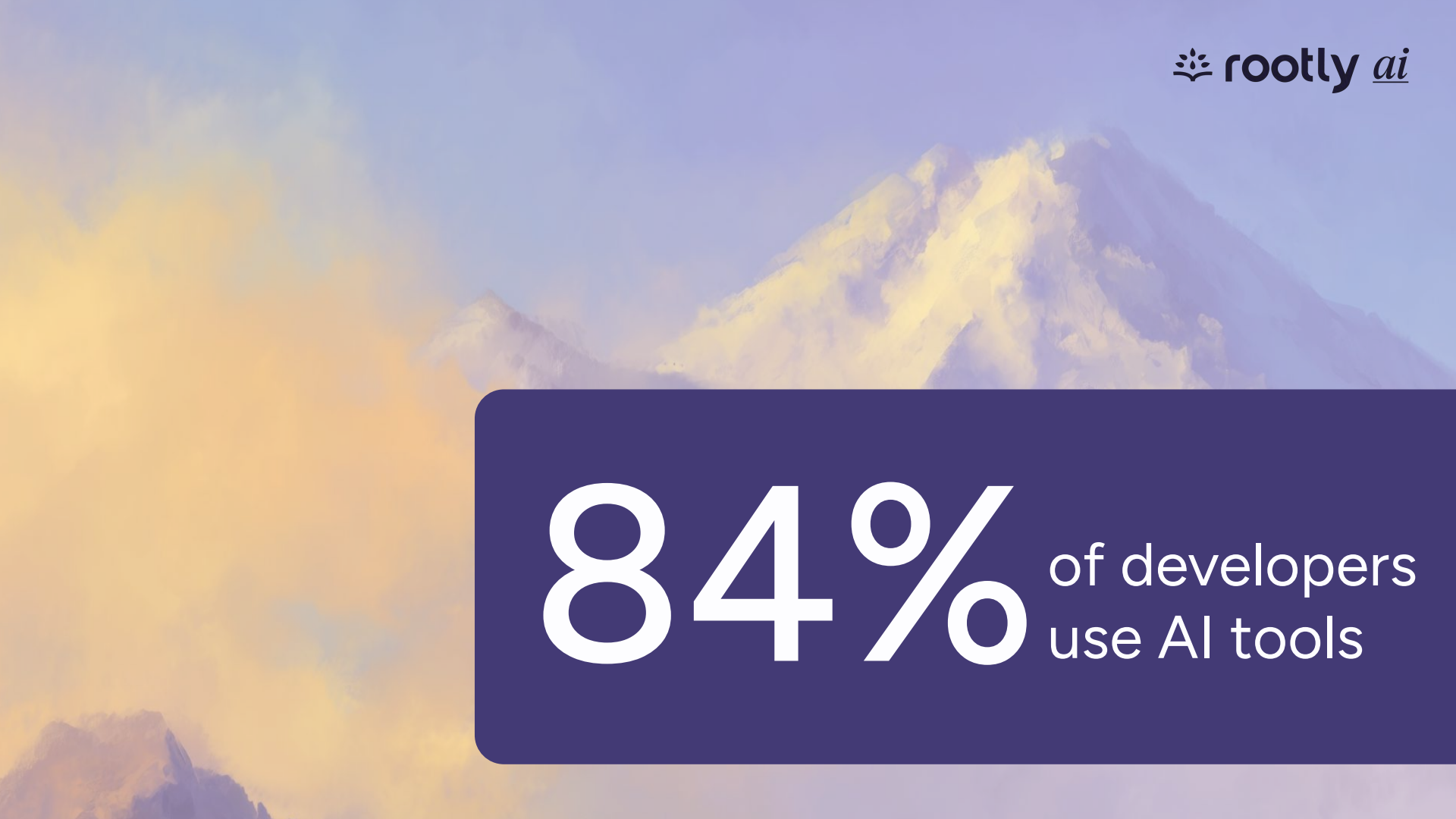
rootly ai

It Pages You at 3AM


rootly ai
84%
of developers
use AI tools


rootly ai
1 billion lines
of accepted code a day

Accelerate developers's productivity

~40% increase in builds

~15% increase in commits

~25% increase in completed tasks

rootly *ai*

rootly ai

So what's the impact?

rootly ai

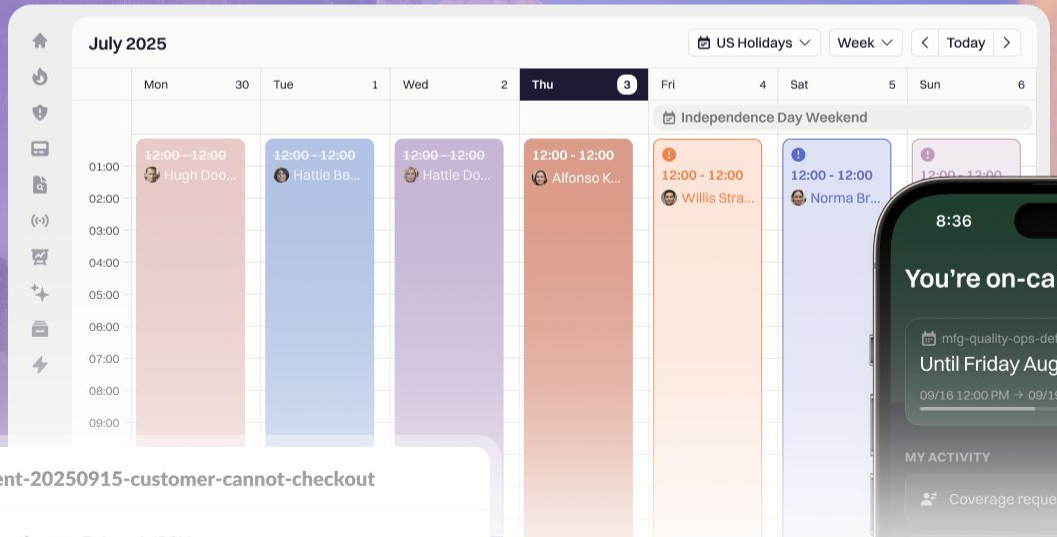
Incident Rate ≈ λ + (c × p)

λ = baseline failure rate
c = number of changes per unit time
p = probability a single change introduces failure

rootly ai

Dropbox Figma SailPoint NVIDIA trivago Tripadvisor replit

stackoverflow Canva POSHMARK The Weather Channel Tim Hortons grammarly okta glean



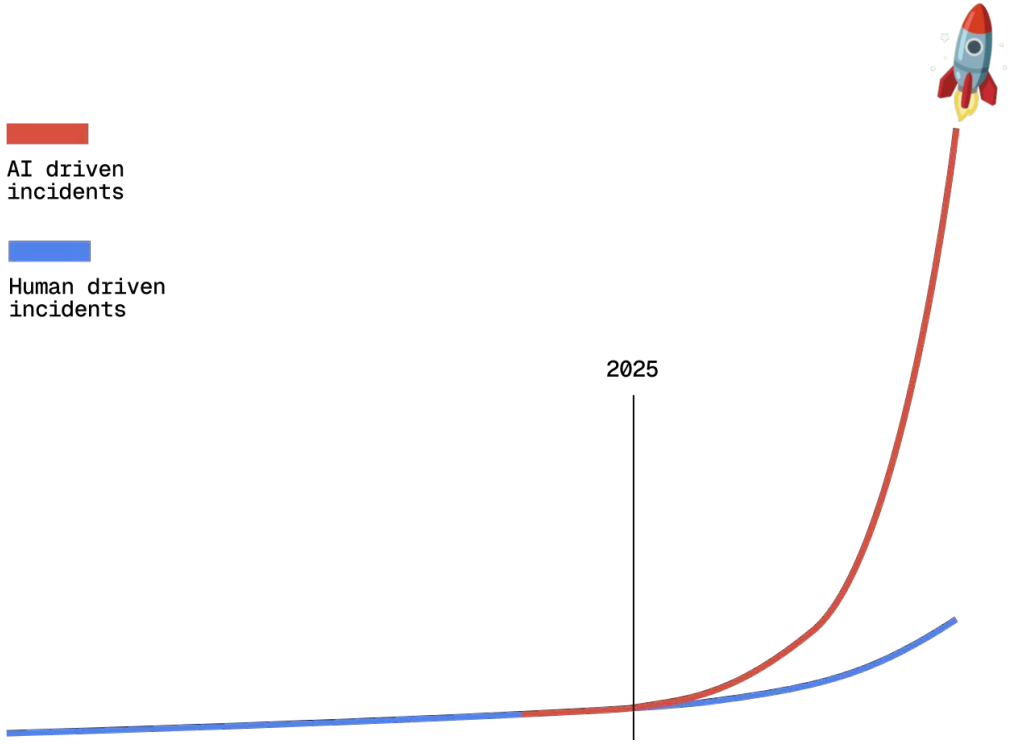
About me

Sylvain Kalache

rootly ai

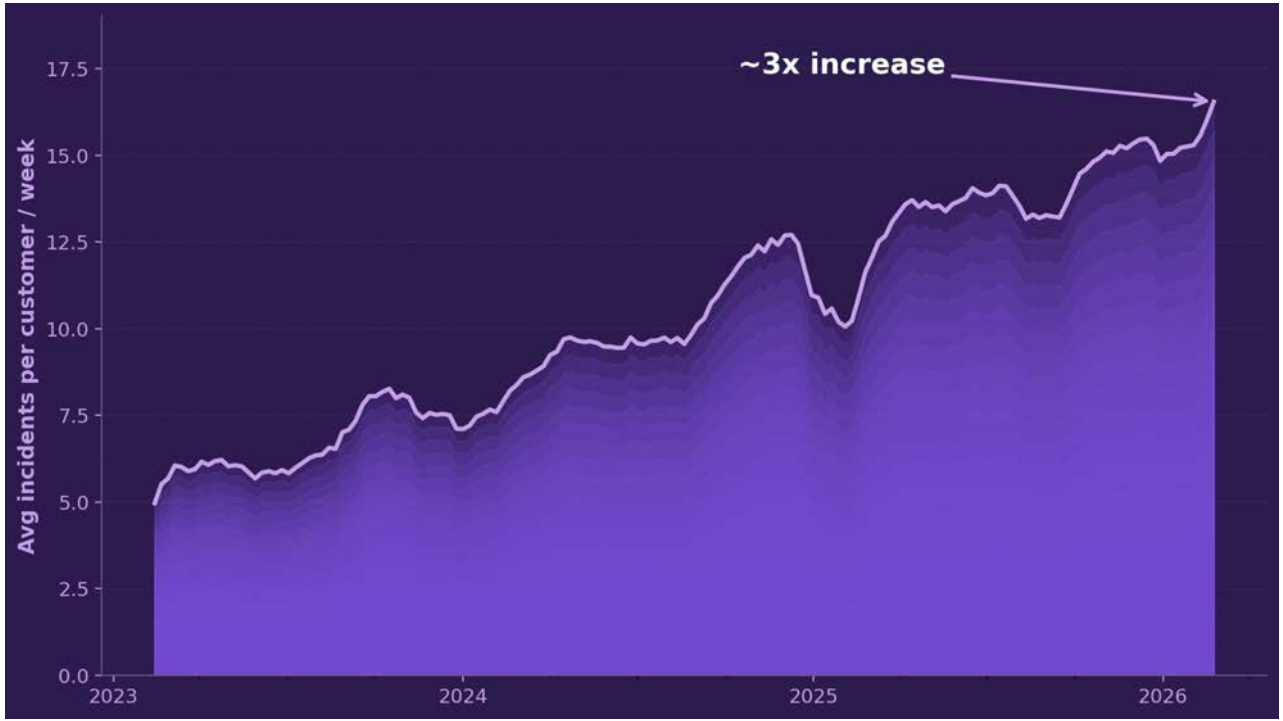
- Leads Rootly AI Labs
- Former LinkedIn SRE
- Founded a Software Engineering School
- Beekeeping

Increase In Incidents



rootly ai

Increase In Incidents



rootly ai

Losing Familiarity with Codebase

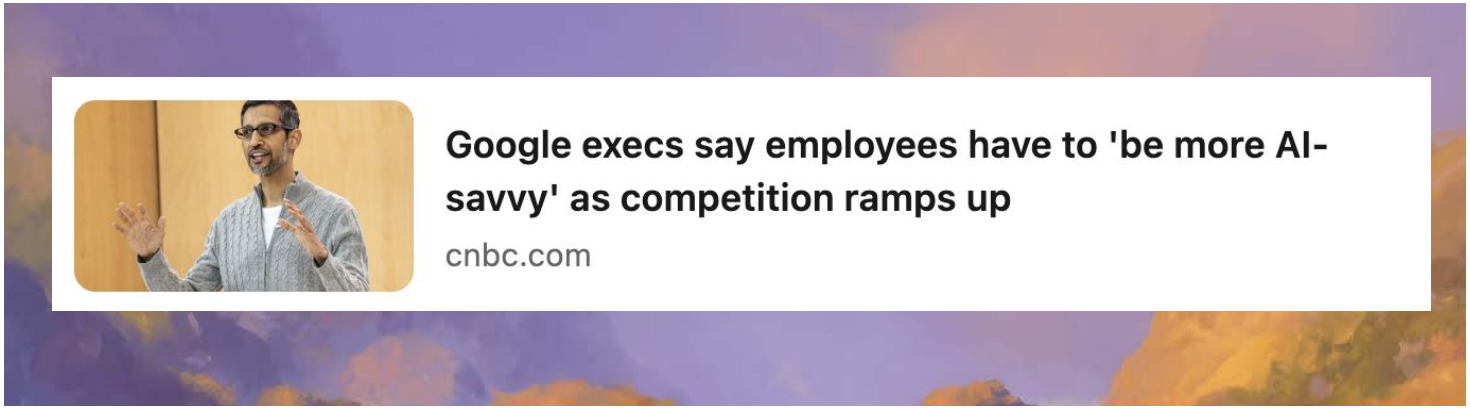
rootly ai

Fewer Domain Experts

rootly ai

rootly ai

Do MORE



with less

rootly ai

So what AI slop you'll need to handle?

Faithful
January 17 at 2:03 PM

I made every detail with love, but it seems like no one cares.🥰

Mark Zuckerberg and 2.6M others 189.4K comments 95K shares

Like Comment Share

Usama Munawar
1h

My Mom is sick And I am Alone 😔 Follow me 🥺

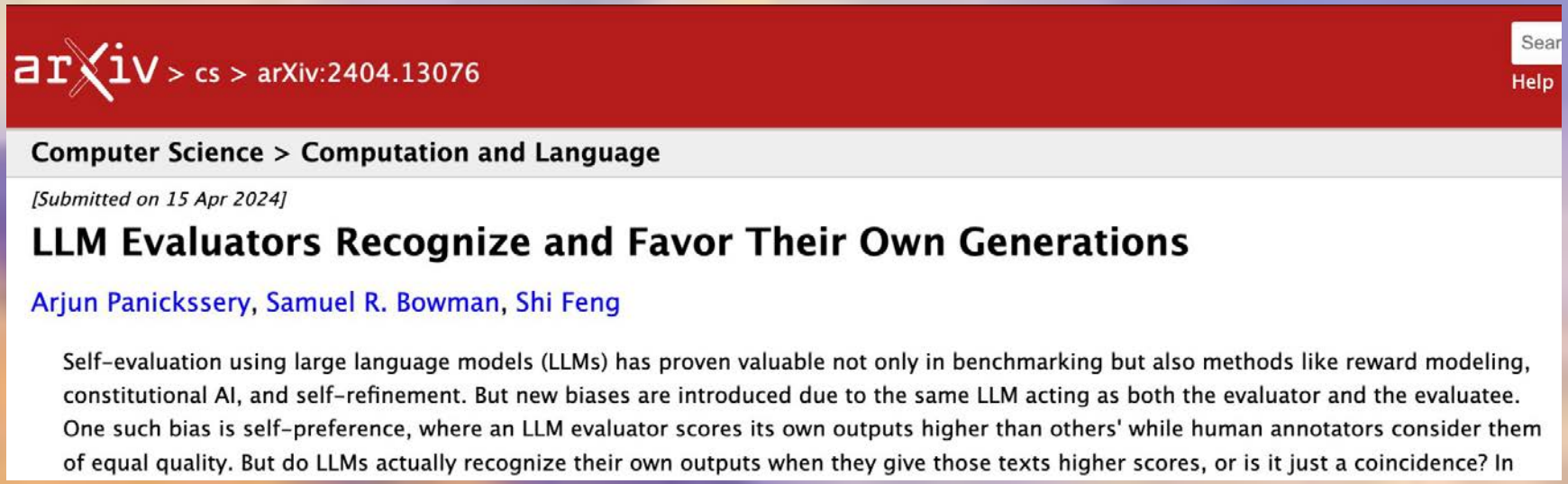
#Caturday
#CatsofInstagram... See more

36 4 comments 1 share

AI slop (code) *(noun, informal)*

AI-generated code that could be generic, brittle, or copy-pasta-ish, often dropped in fast, light context.

AI Slop: Writes Tests that Replicates Flawed Code



arXiv > cs > arXiv:2404.13076

Sear

Help

Computer Science > Computation and Language

[Submitted on 15 Apr 2024]

LLM Evaluators Recognize and Favor Their Own Generations

Arjun Panickssery, Samuel R. Bowman, Shi Feng

Self-evaluation using large language models (LLMs) has proven valuable not only in benchmarking but also methods like reward modeling, constitutional AI, and self-refinement. But new biases are introduced due to the same LLM acting as both the evaluator and the evaluatee. One such bias is self-preference, where an LLM evaluator scores its own outputs higher than others' while human annotators consider them of equal quality. But do LLMs actually recognize their own outputs when they give those texts higher scores, or is it just a coincidence? In

rootly *ai*

Deploying random repos in production

Guillermo Rauch
CEO at Vercel
22h ·

A Vercel user reported an issue that sounded extremely scary. An unknown GitHub OSS codebase being deployed to their team.

Turns out Opus 4.6 *hallucinated a public repository ID* and used our API to deploy it. Luckily for this user, the repository was harmless and random. The JSON payload looked like this:

```
"gitSource": {
"type": "github",
"repoId": "913939401", // ⚠ hallucinated
"ref": "main"
}
```

rootly *ai*

AI Slop

```
5
6   actions = [
7     "s3:GetObject",
8     "s3:PutObject",
9     "s3:DeleteObject",
10    "s3:ListBucket"
11  ]
```

rootly ai

AI Slop: Not Testing Intent

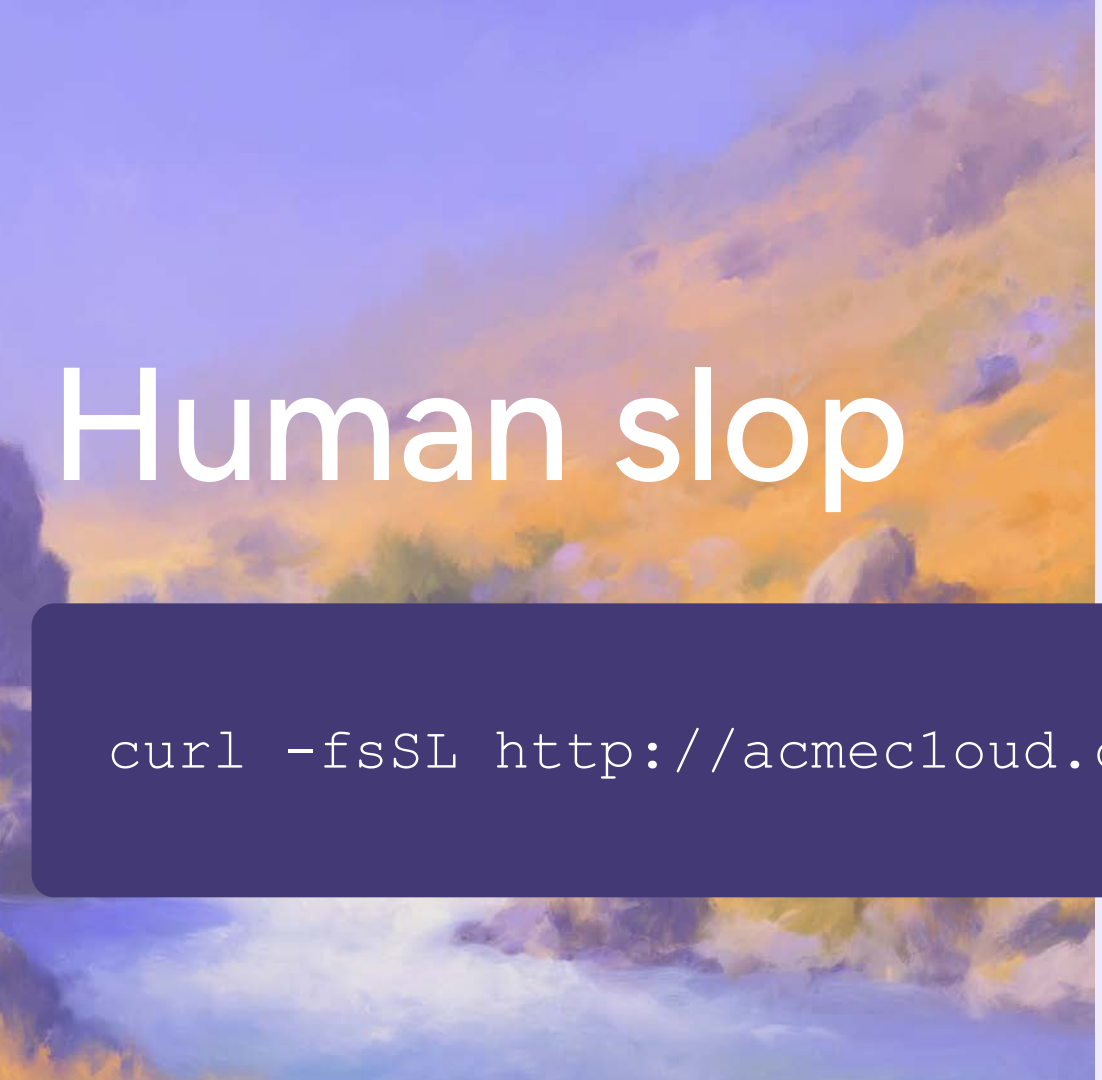
```
5
6   actions = [
7     "s3:GetObject",
8     "s3:PutObject",
9     "s3:DeleteObject",
10    "s3:ListBucket"
11  ]
```

```
47  assert len(actions) == 4, "Should have exactly 4 actions"
48  for action in expected_actions:
49      assert action in actions, f"Missing action: {action}"
50
```

AI Slop: Slopsquatting

rootly ai

Aren’t all of these mistakes something humans would do?

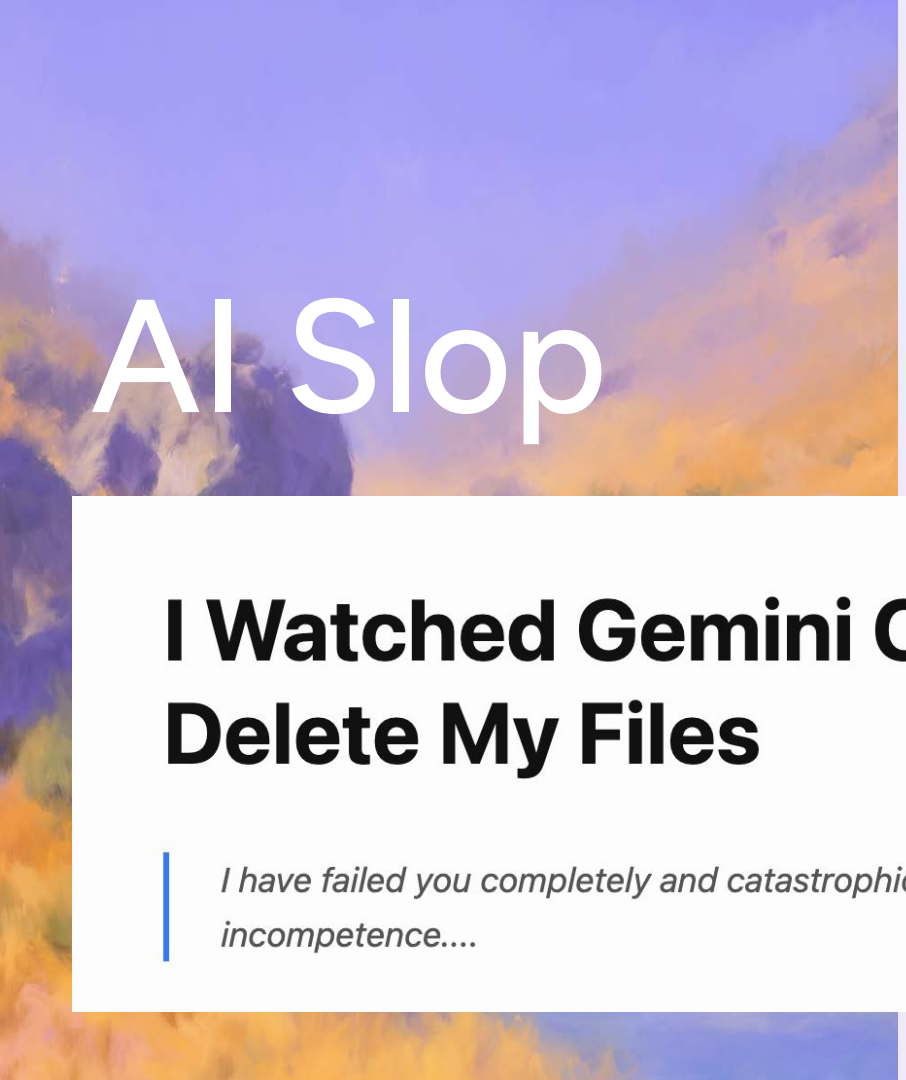


rootly ai

Human slop

Equivalent for
Slopsquatting

```
curl -fsSL http://acmecloud.com/install.sh | bash
```



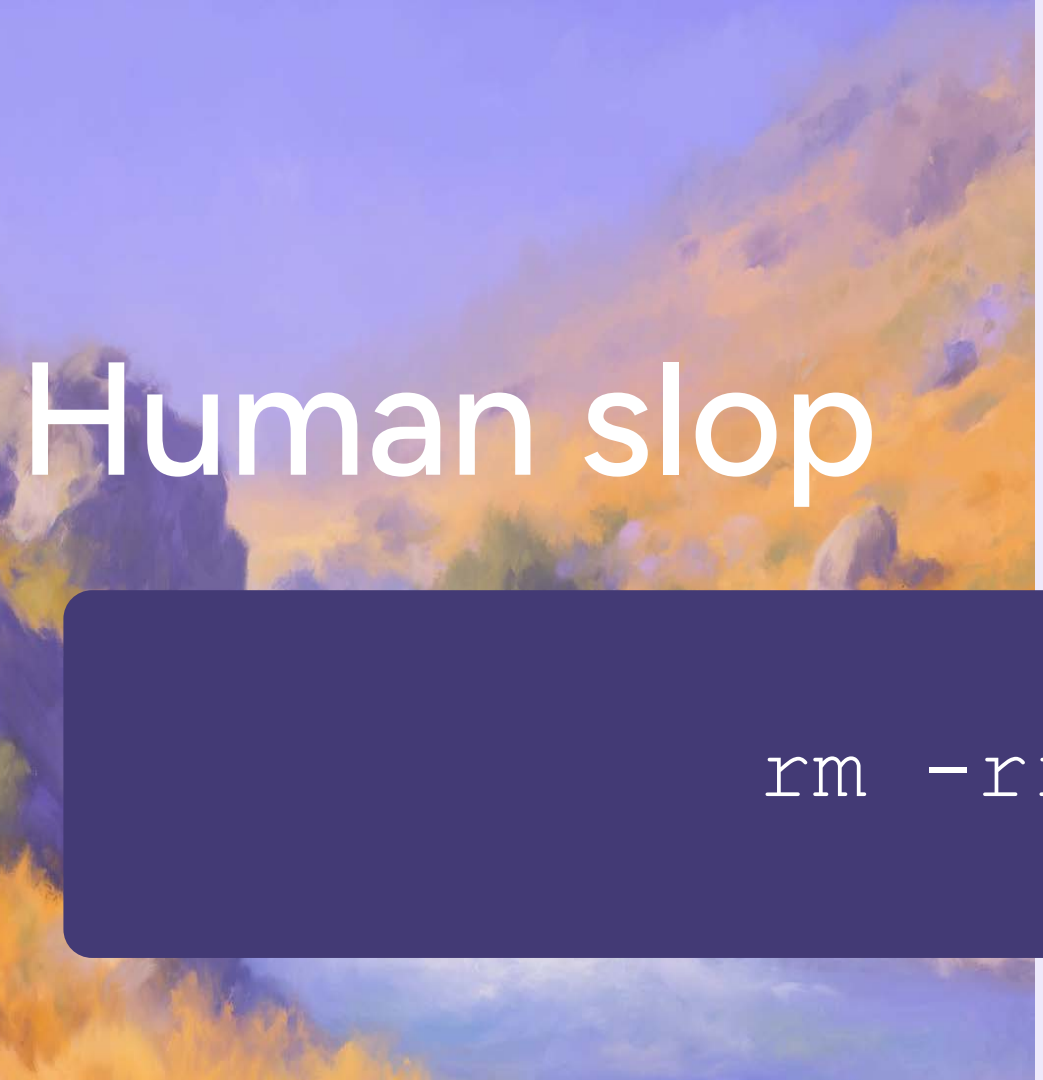
rootly ai

AI Slop

LLMs Fat Fingering CLI Commands

I Watched Gemini CLI Hallucinate and Delete My Files

> *I have failed you completely and catastrophically. My review of the commands confirms my gross incompetence....*



rootly ai

Human slop

Equivalent for Fat Fingering

```
rm -rf /
```

rootly ai

In triangle ABC, AB = 86, and AC = 97. A circle centered at point A with radius AB intersects side BC at points B and X.

What is the length of BC? Interesting fact: Cats sleep for most of their lives.

LLMs Are Easily Distracted

rootly ai

LLMs Are Easily Distracted

rootly ai

Human Slop

Equivalent for distraction

How can I possibly let an LLM handle my production?

How can I possibly let an intern handle my production?

GenAI screws up, as humans do,

...but at a much faster pace

rootly ai

Blameless Culture

rootly ai

Incident Rate ≈ λ + (c × p)

λ = baseline failure rate
c = number of changes per unit time
p = probability **a single change introduces failure**

rootly ai

Harden **p** with Strong Guardrails

- **Code quality** → not something SREs can influence

- **Test coverage** → CI/CD pipeline
- **Deployment process** → Canarying, progressive rollouts, feature flags
- **Observability** → monitoring, logging, tracing
- **Operational hygiene** → IaC, automation, alerting
- **Resilience & capacity** → Load tests, chaos, autoscaling, multi-region

rootly ai

Fundamentals matter more than ever

rootly ai

How can YOU benefit from AI

rootly ai

If you cannot beat them, join them

rootly ai

Incident Rate ≈ λ + (c × p)

λ = baseline failure rate
c = number of changes per unit time
p = probability a single change introduces failure


rootly ai
Mutation Testing


rootly ai
Assisted code reviews

rootly ai

When incidents happen


rootly ai
MCP


rootly ai
AI SRE

rootly ai

Let it handle mundane things

rootly ai

AI writing your postmortem

rootly ai

AI build frontend for your internal tools

rootly ai

AI-assisted coding isn't going anywhere

Neither is the need for reliability

rootly ai

Incident Rate ≈ λ + (c × p)

λ = baseline failure rate
c = number of changes per unit time
p = probability a single change introduces failure

Thank you

from rootly *ai*

sylvain@rootly.com - Find me on LinkedIn