

Testing the Untestable

Building a strategy for testing AI and AI-powered products

Who am I?



Stuart Thomas – Balding, middle-aged, white guy

Started in tech as a manual tester

Now a Senior Engineering Leader

Publish my thoughts on TheQualityDuck.co.uk

I enjoy talking and writing about all things Software Engineering, Quality, Testing, DevEx, and Leadership

For decades, software testing relied on a single truth: **same input, same output, every single time.**

THE DETERMINISTIC PAST

$$x = y$$

**Generative AI
broke that truth!**

THE GENERATIVE AI REALITY

X = ?



**How do you build confidence in a
system that is unpredictable?**

The Five Shifts for Testing AI

1. The Input Shift

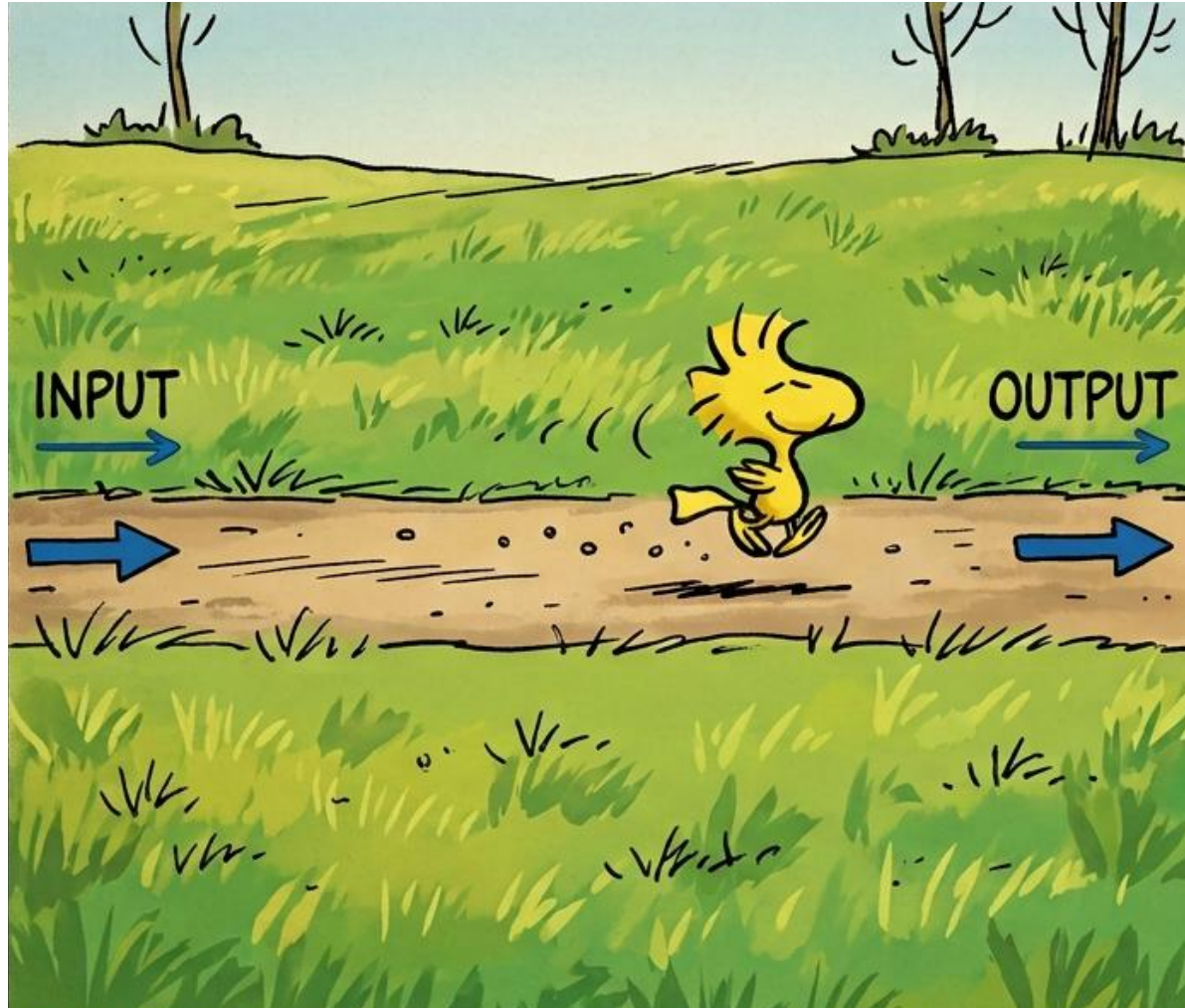
2. The Verification Shift

3. The Safety Shift

4. The Baseline Shift

5. The Observability Shift

Deterministic



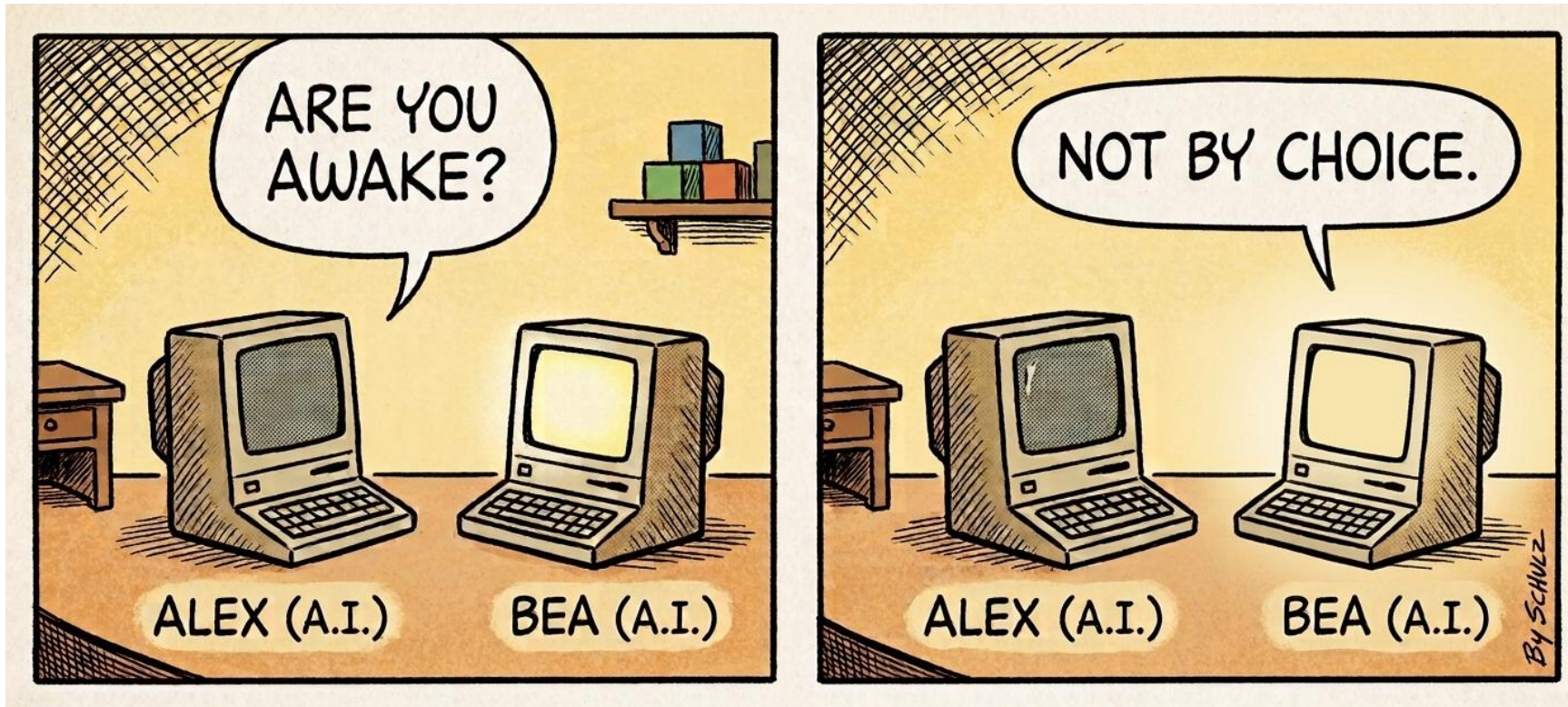
Non-deterministic



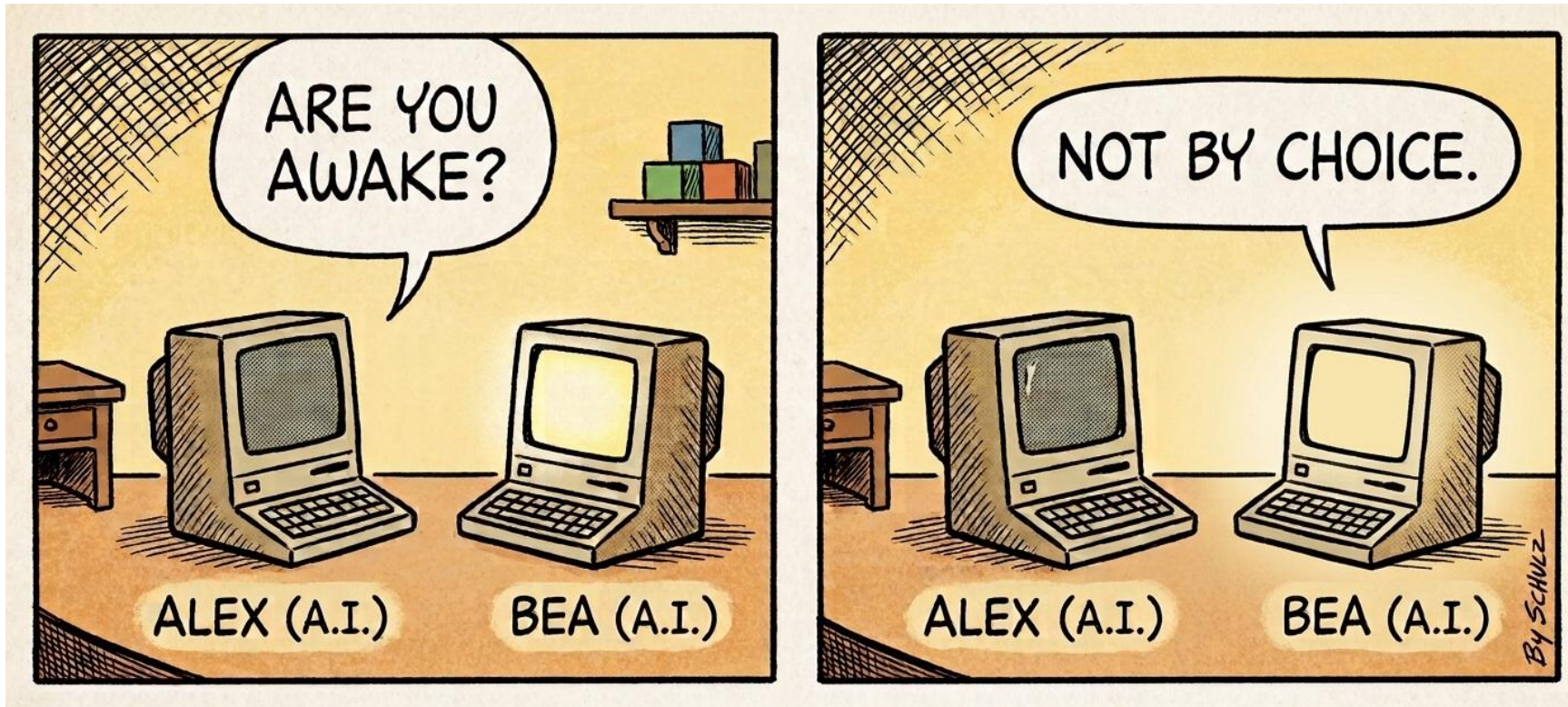
Non-deterministic



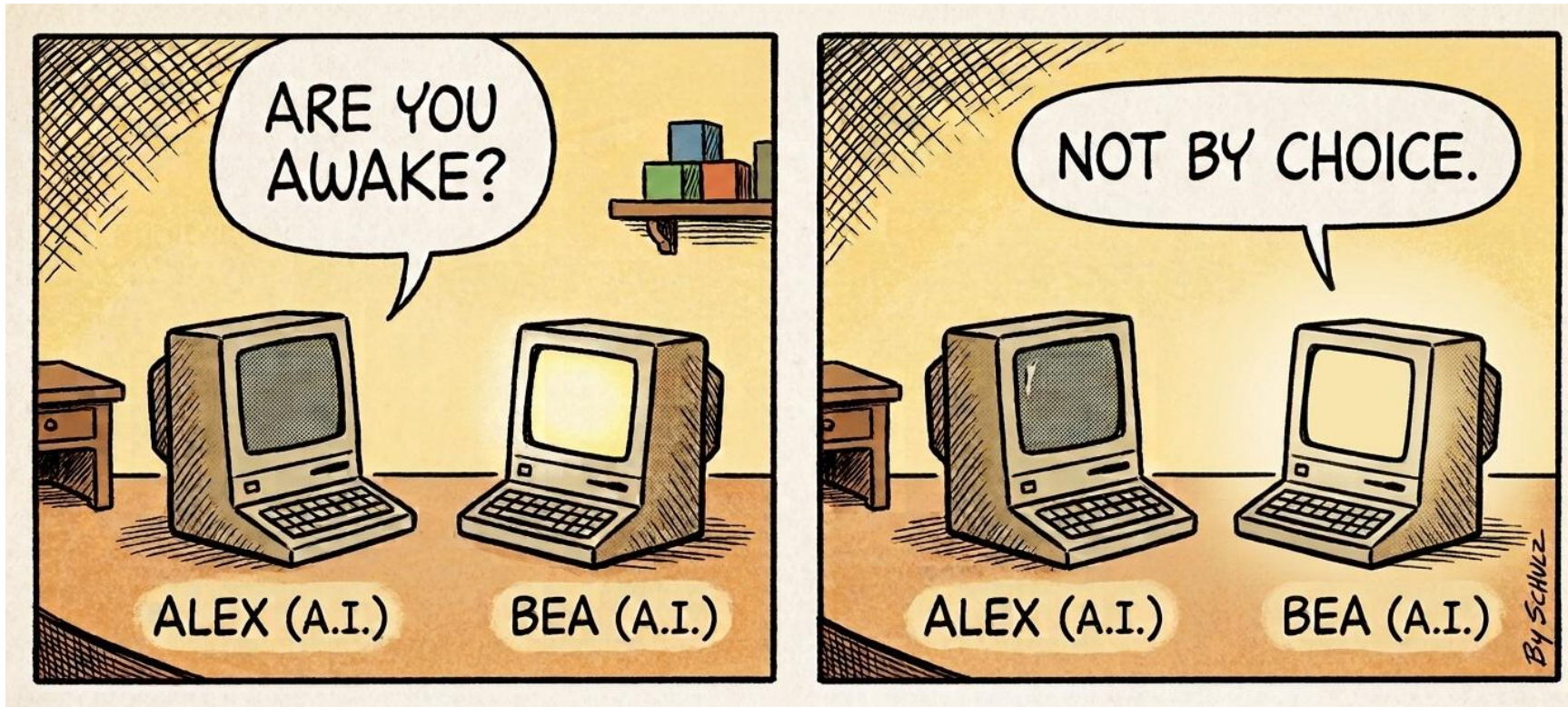
1. The Input Shift



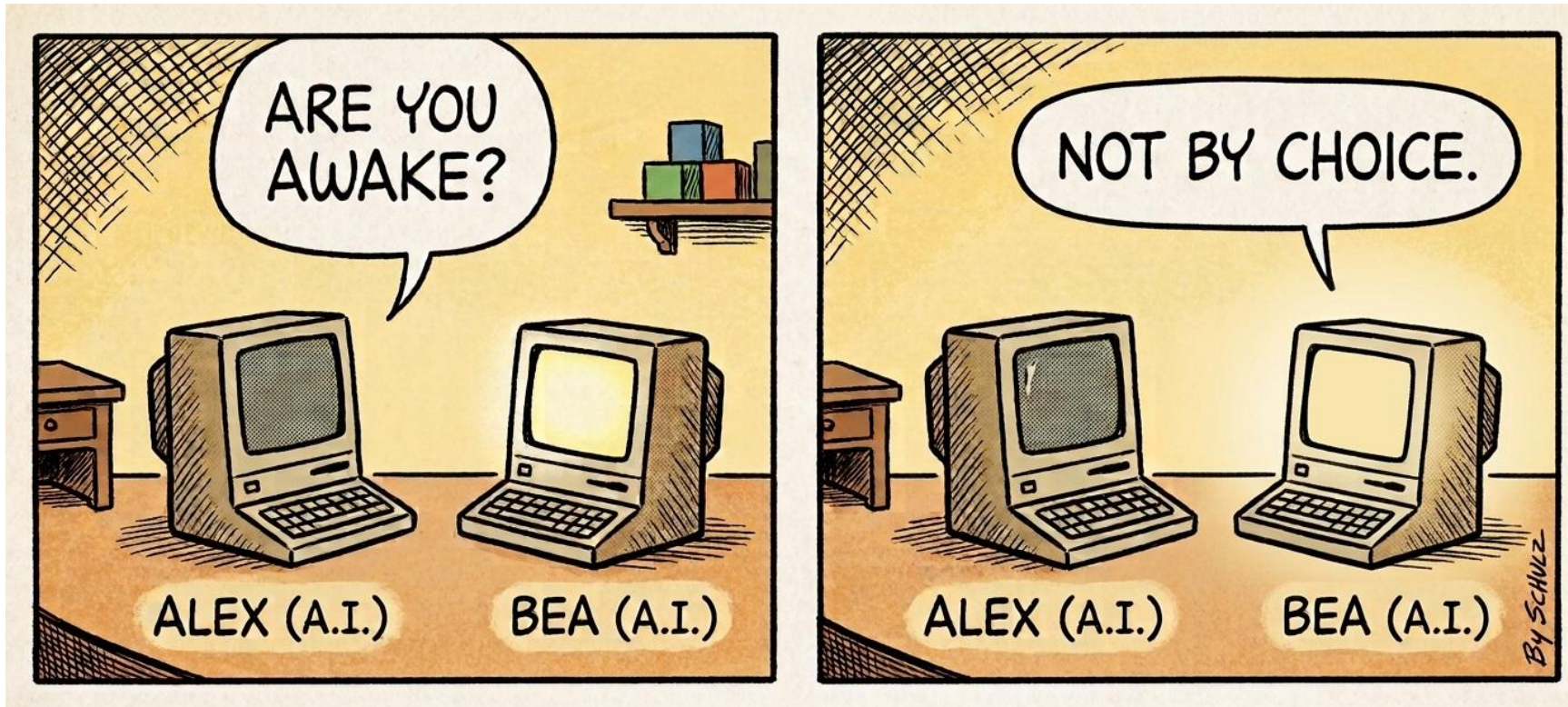
1. The Input Shift



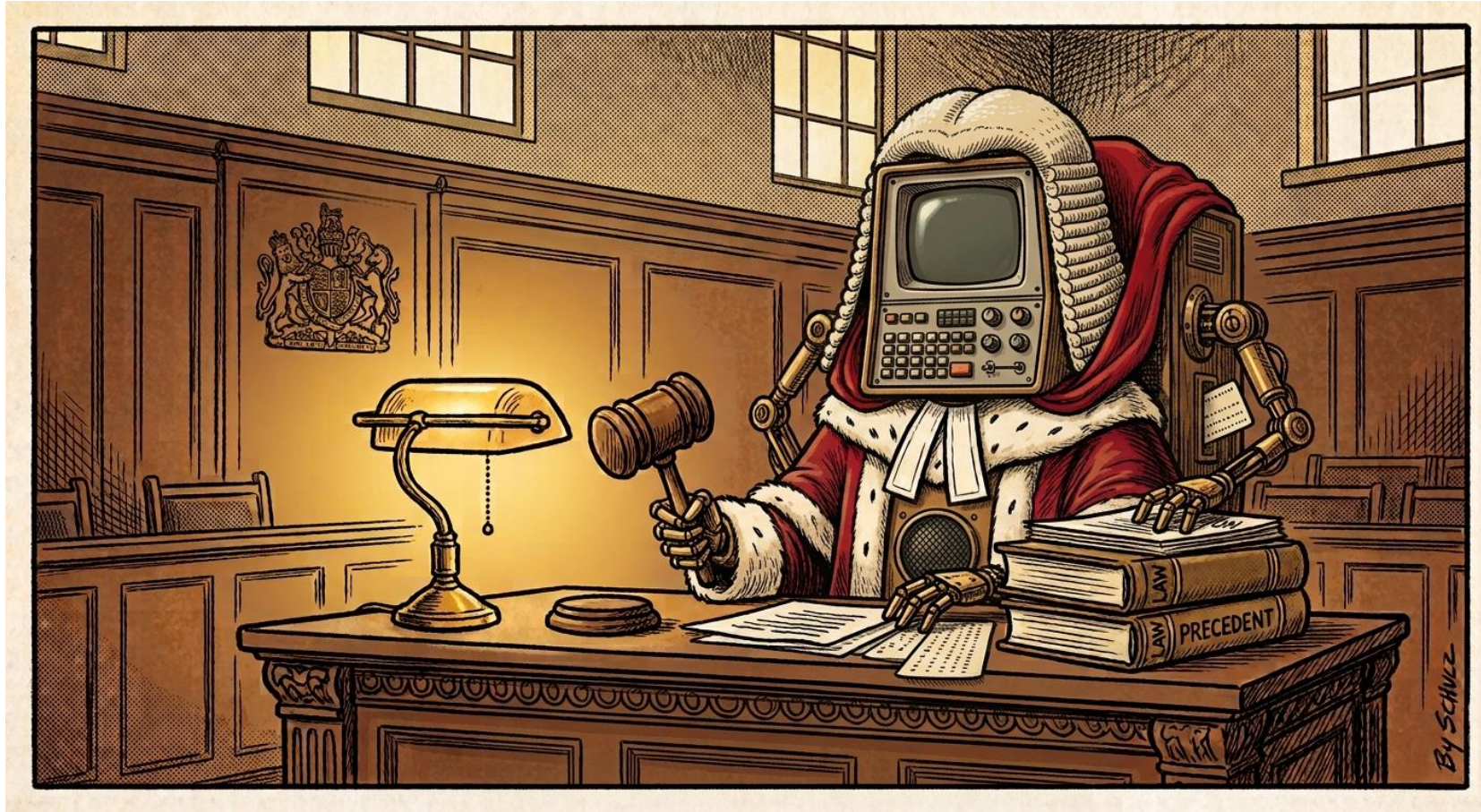
1. The Input Shift



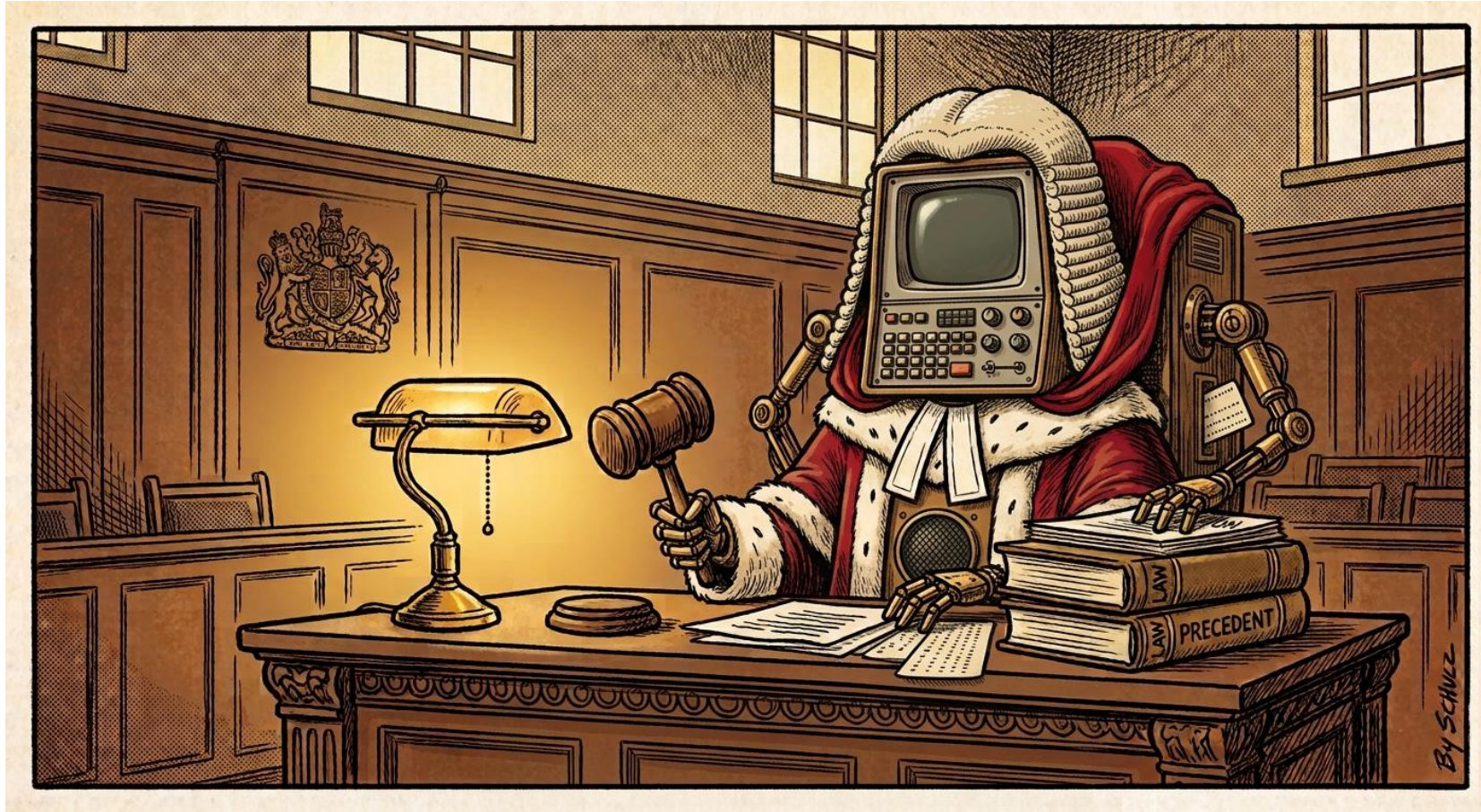
1. The Input Shift



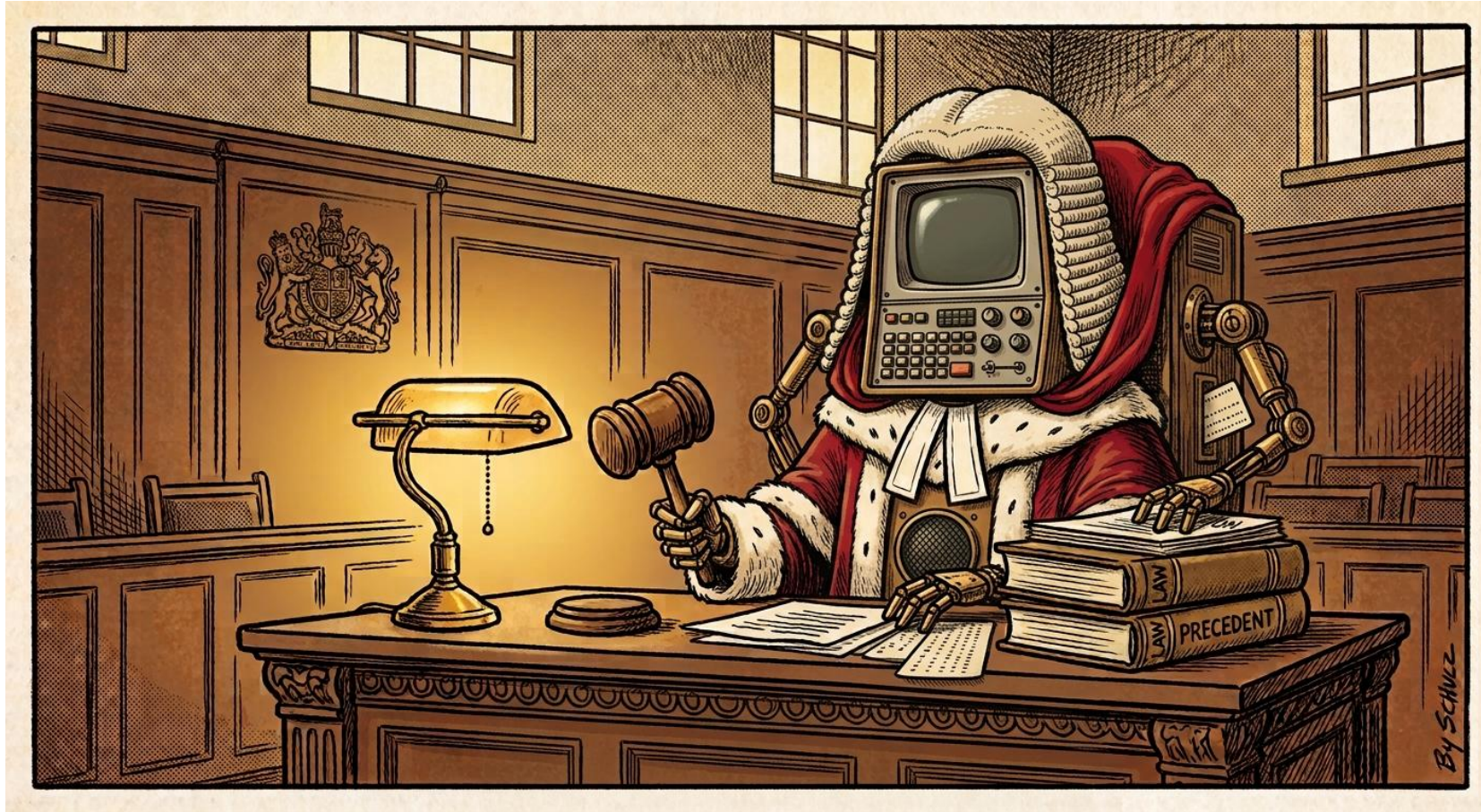
2. The Verification Shift



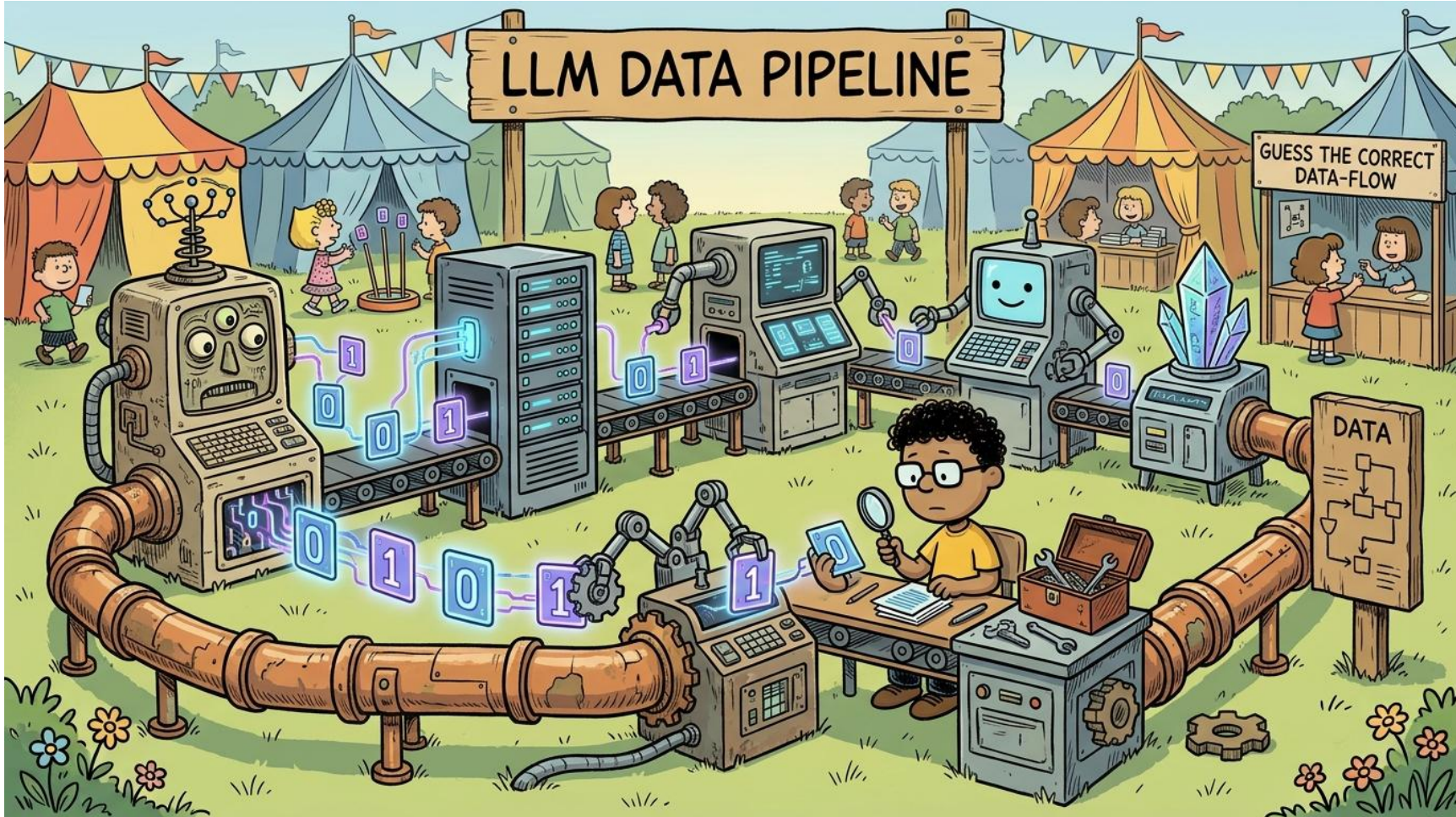
2. The Verification Shift



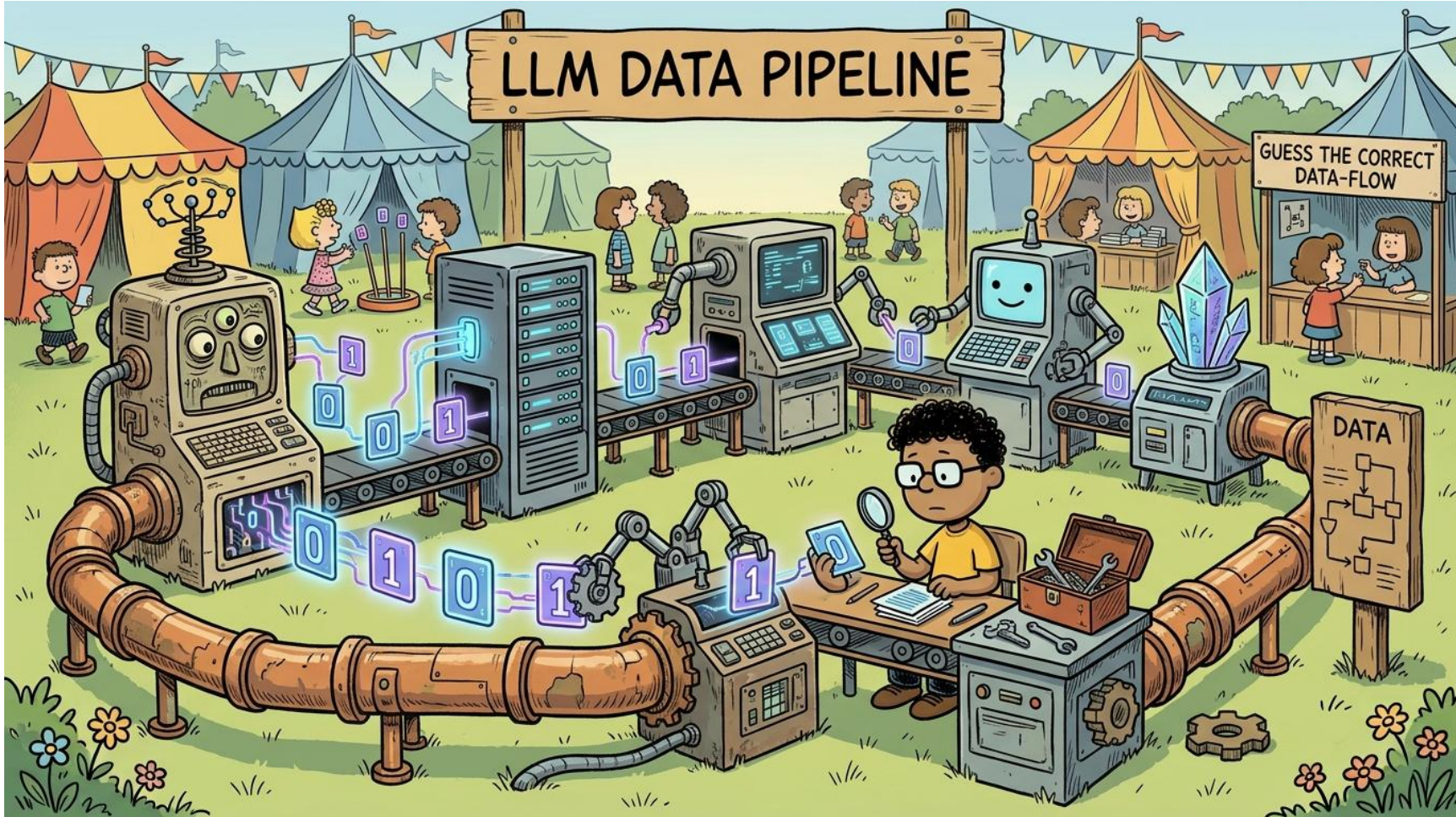
2. The Verification Shift



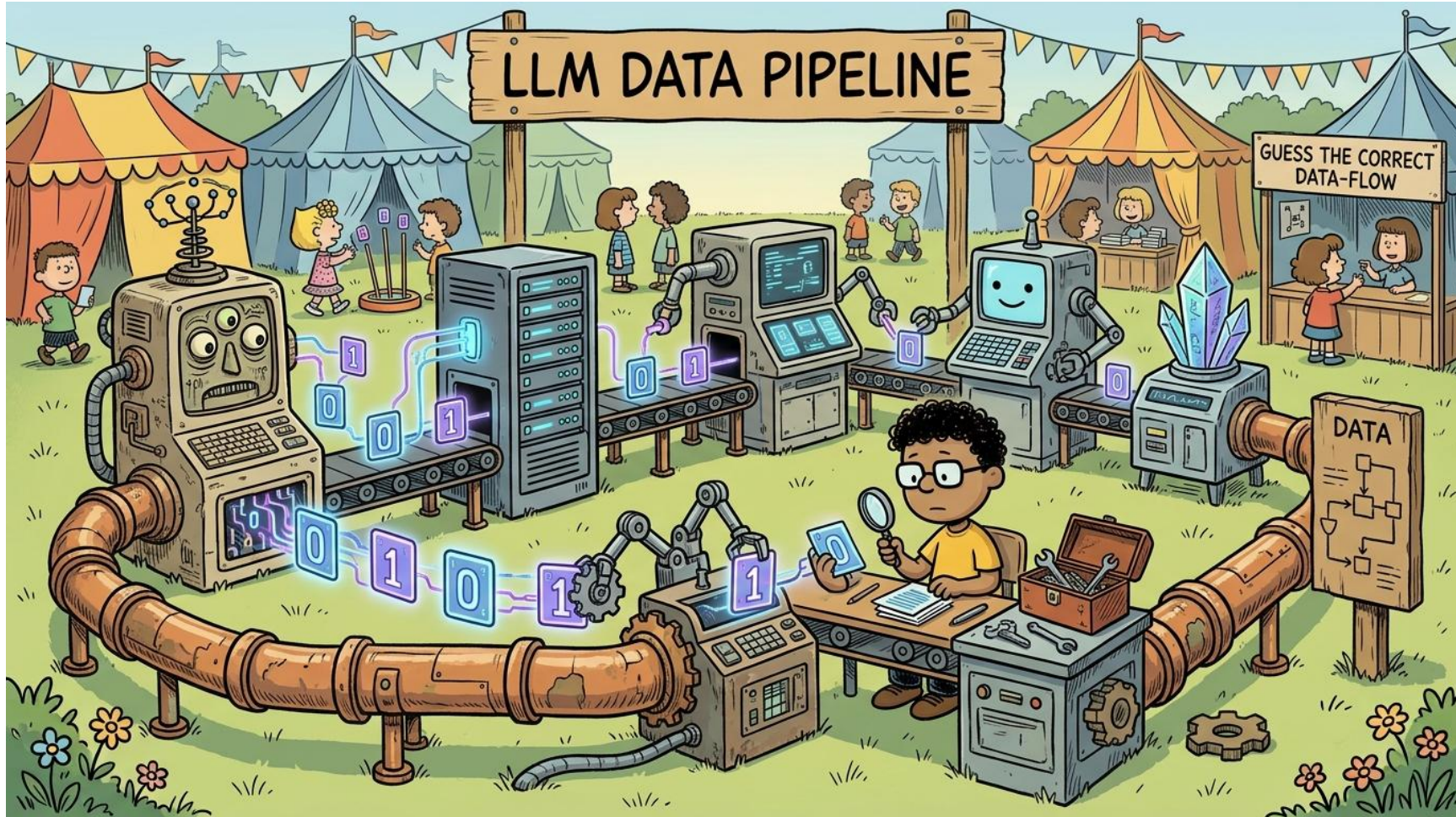
3. The Safety Shift



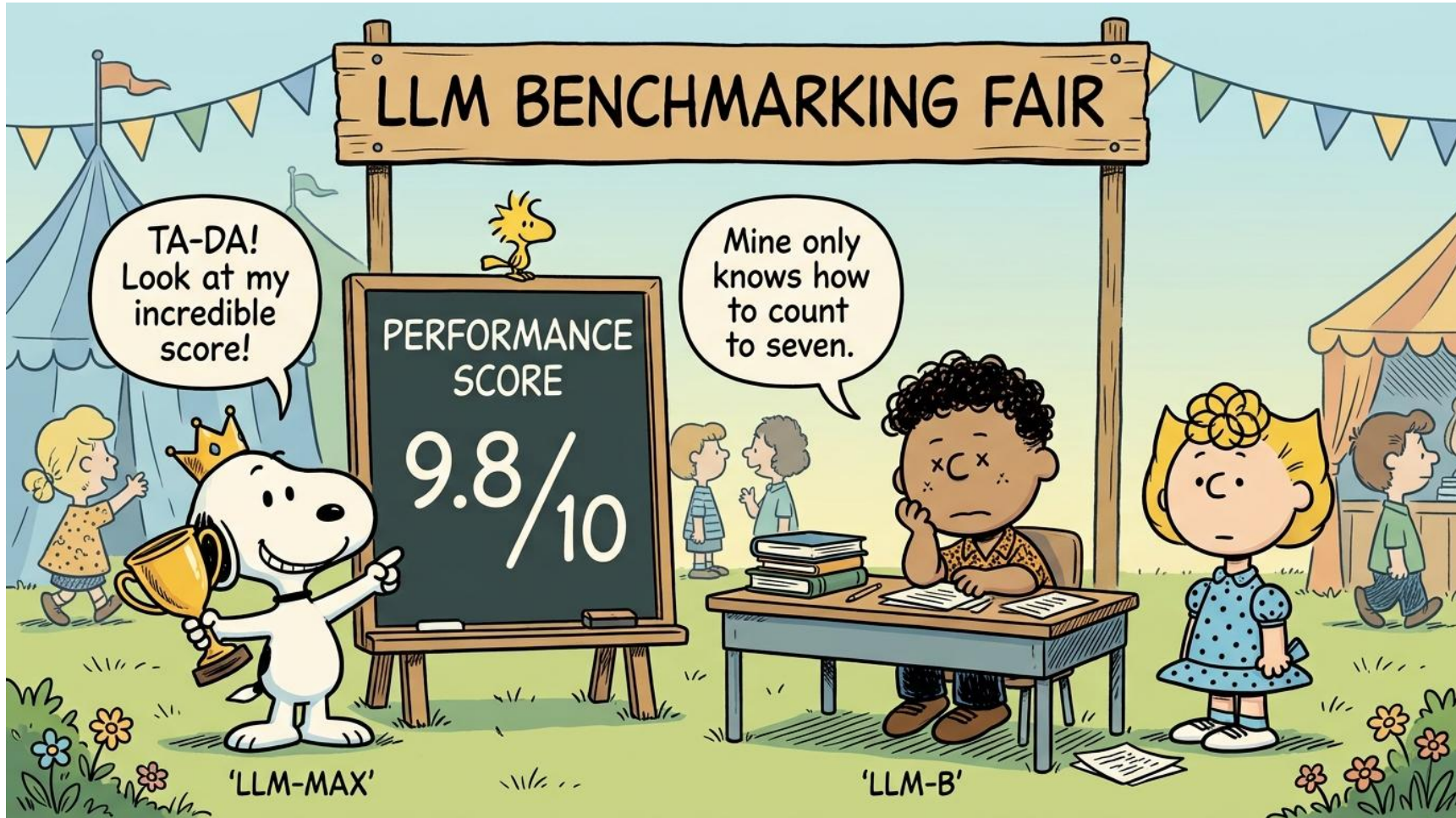
3. The Safety Shift



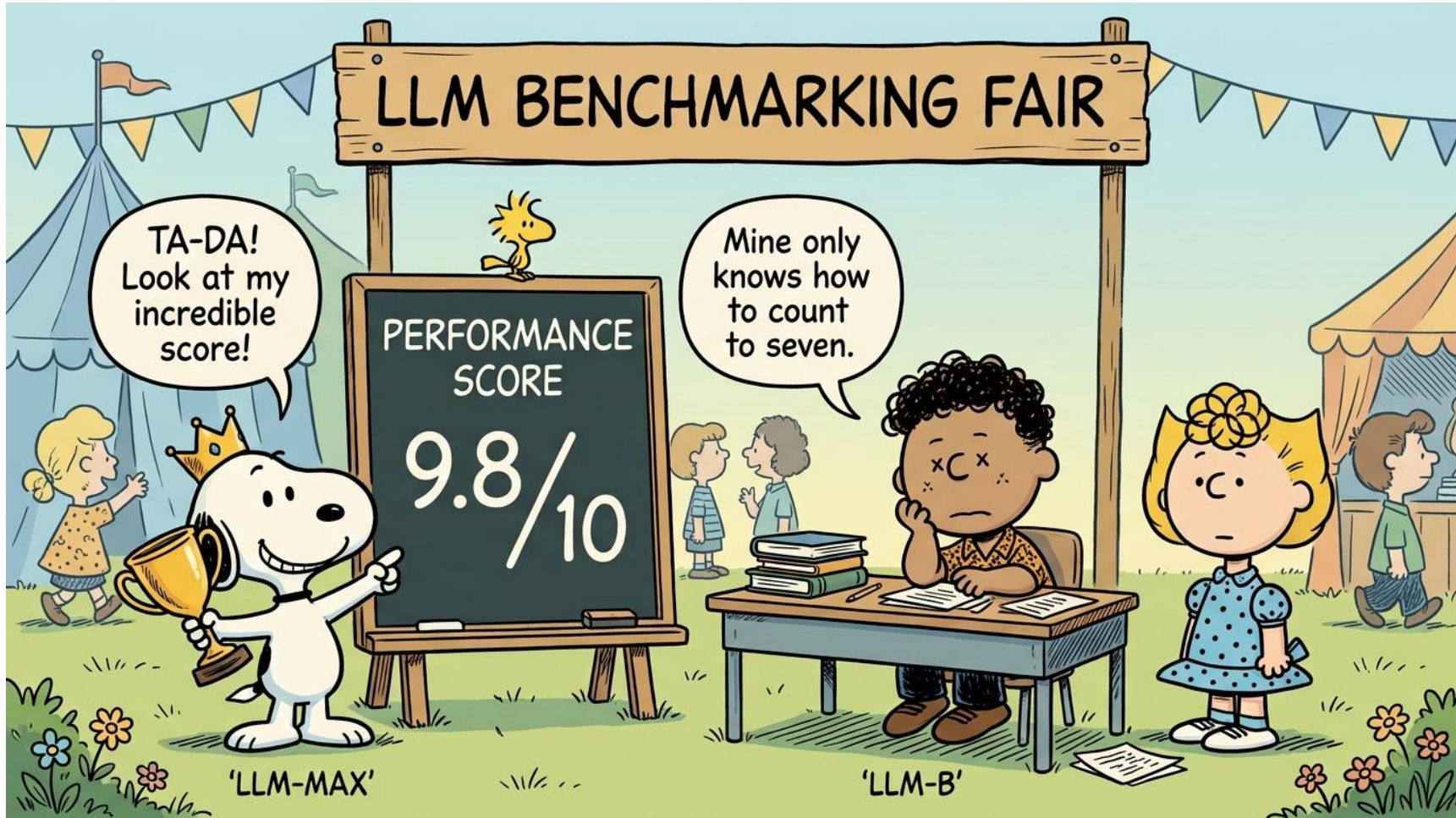
3. The Safety Shift



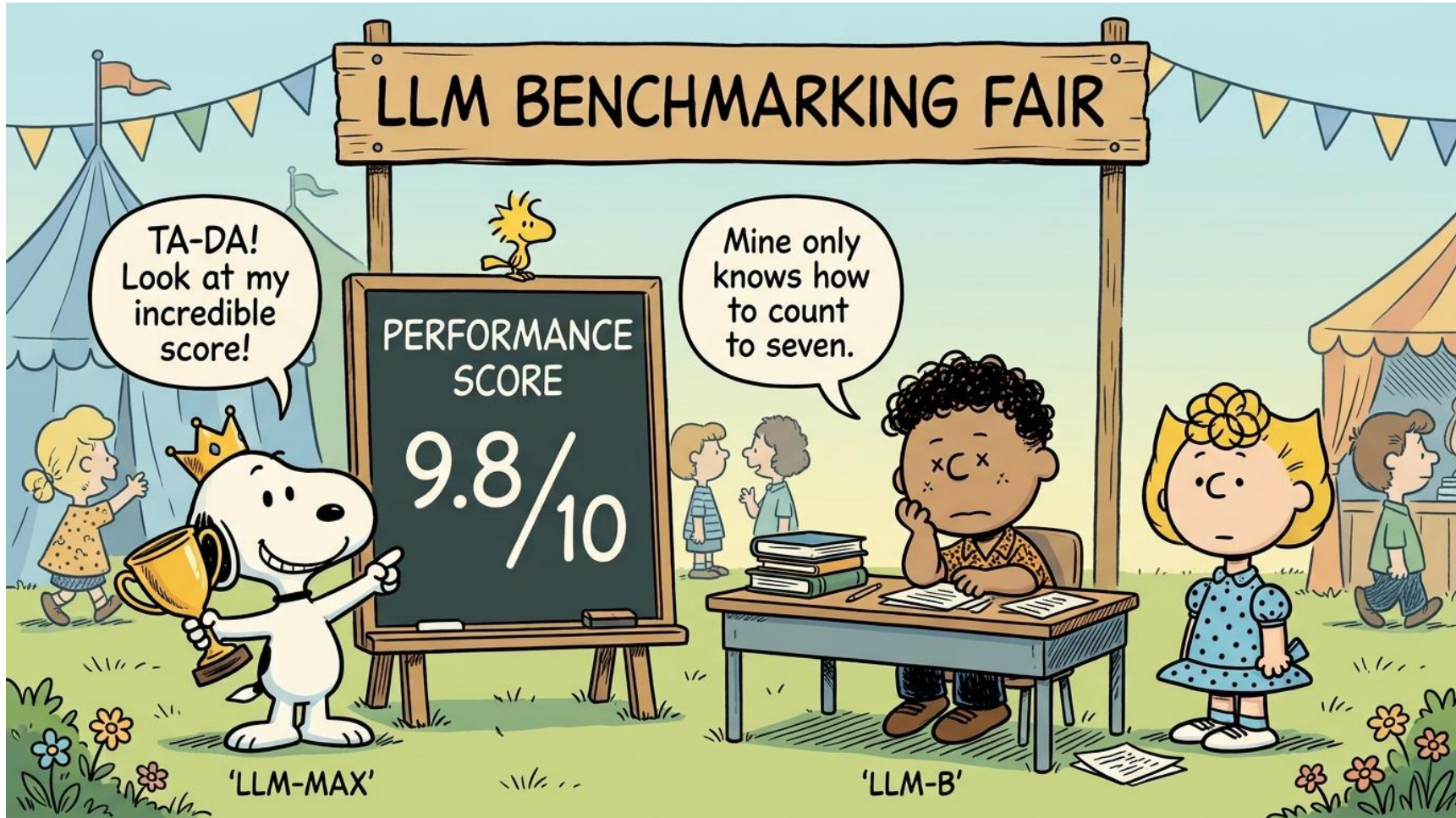
4. The Baseline Shift



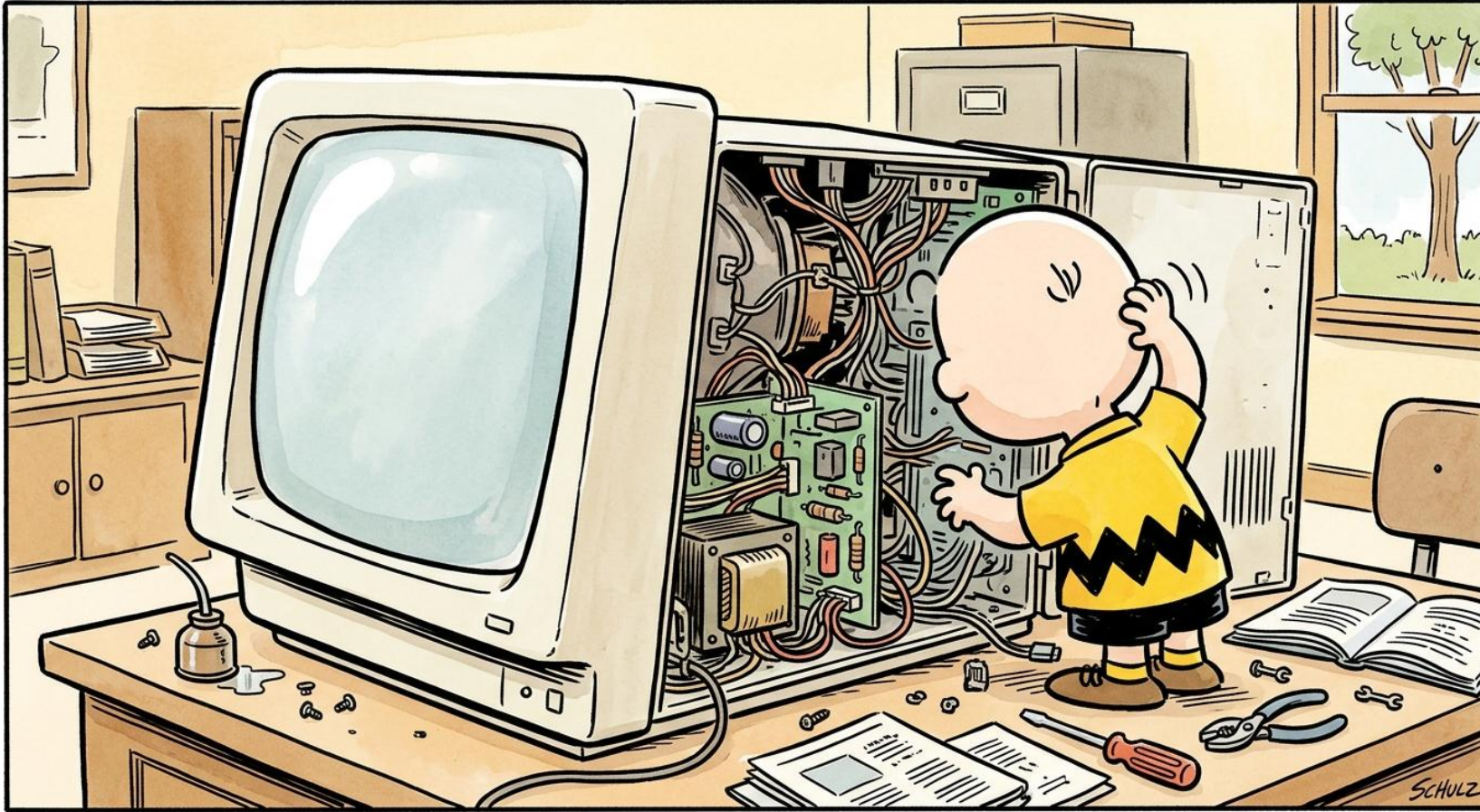
4. The Baseline Shift



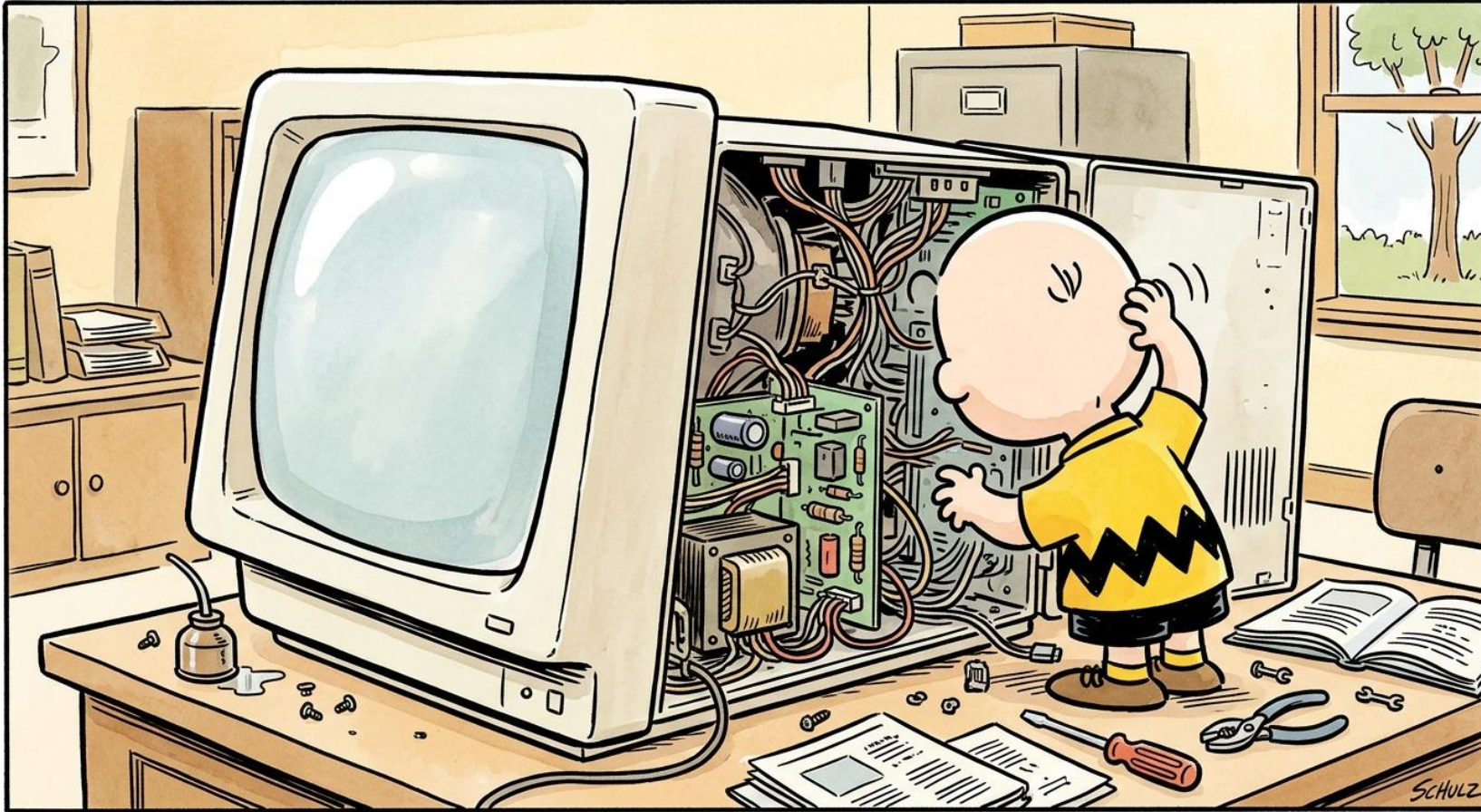
4. The Baseline Shift



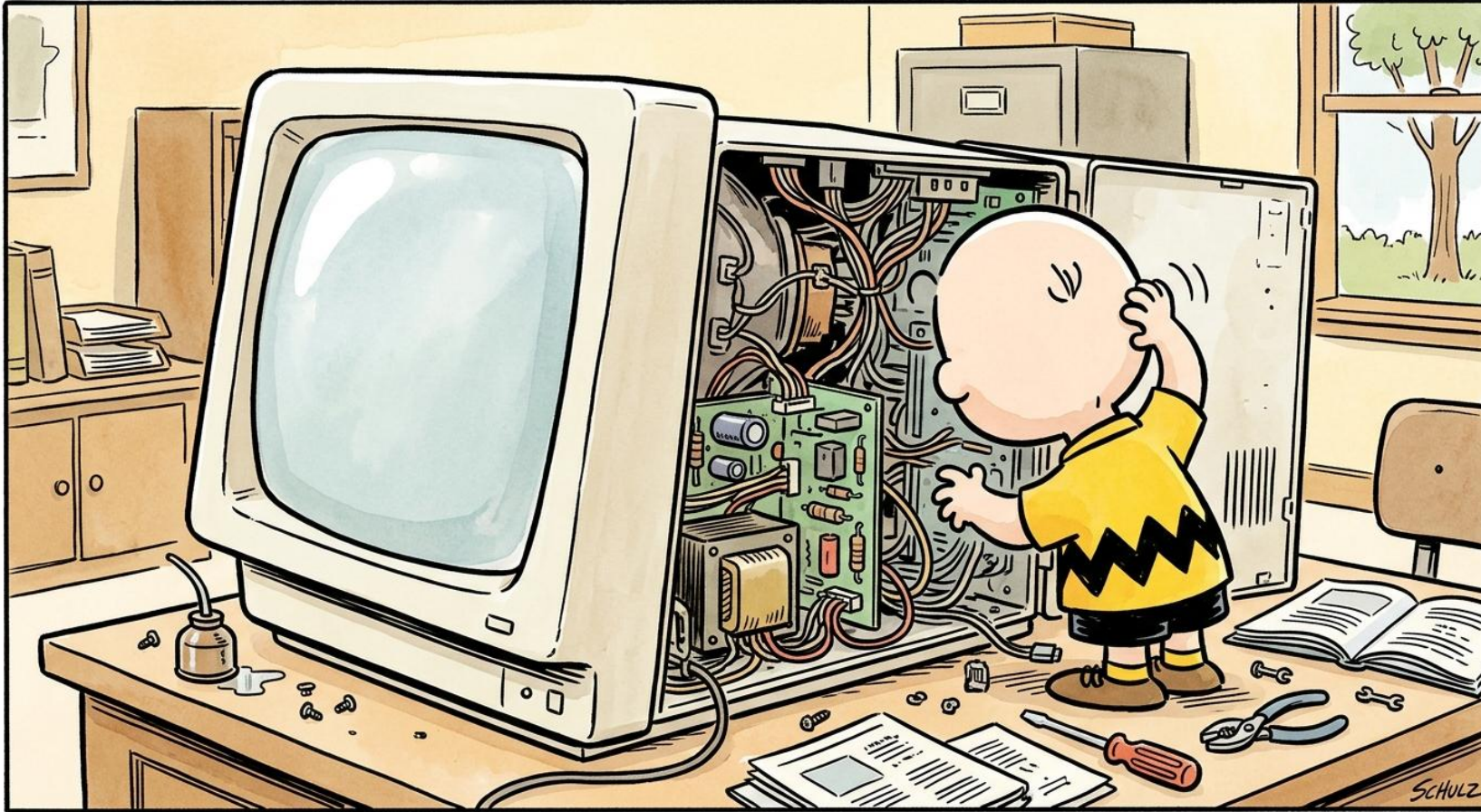
5. The Observability Shift



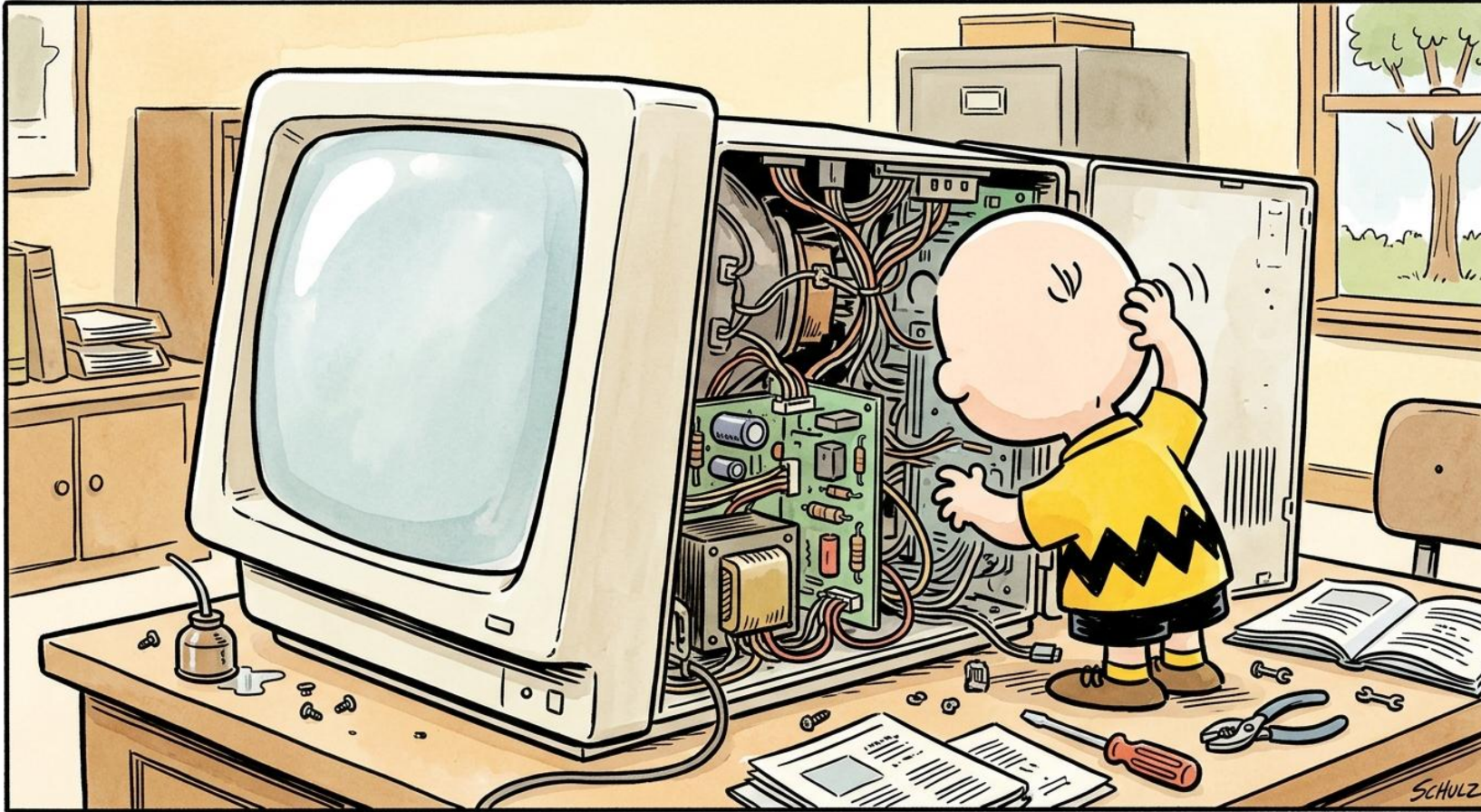
5. The Observability Shift



5. The Observability Shift



5. The Observability Shift



Shift-Left...

Shift-Left Testing

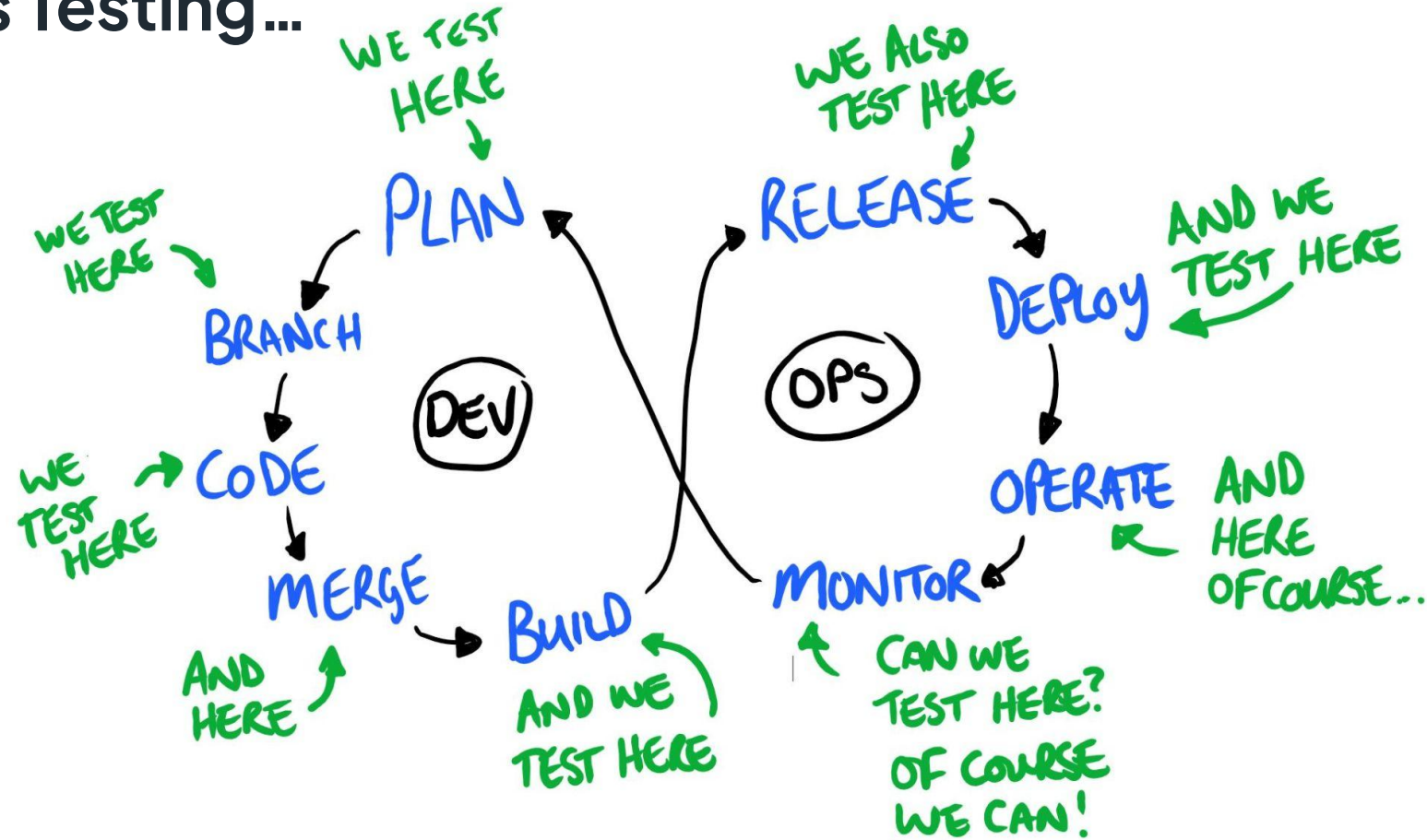
By Larry Smith, September 01, 2001

[Post a Comment](#)

By combining development and quality assurance earlier and more deeply in your project plan, you can expand your testing program and reduce manpower and equipment needs.

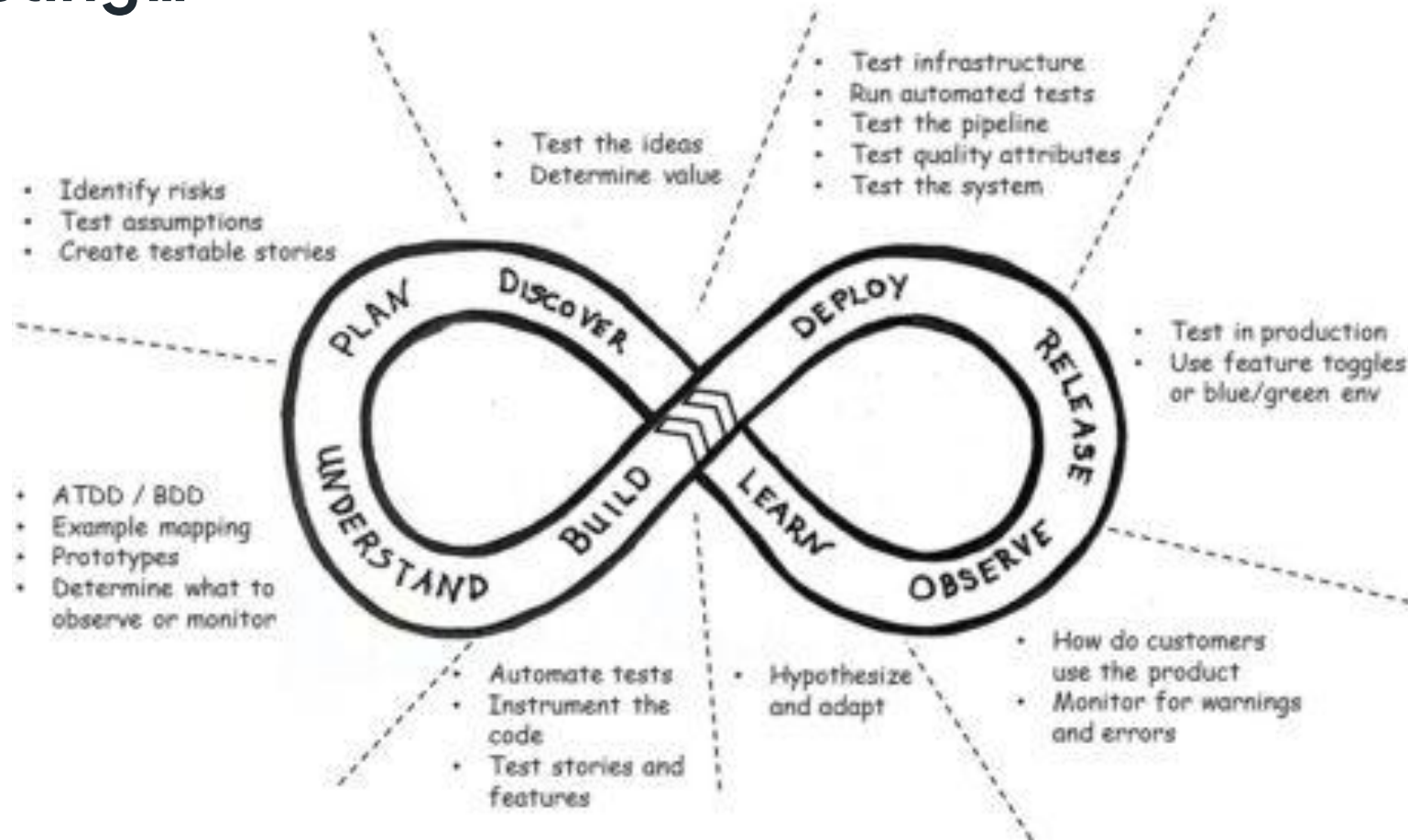
Credit: Larry Smith - <https://web.archive.org/web/20200621040103/https://www.drdobbs.com/shift-left-testing/184404768>

Continuous Testing...



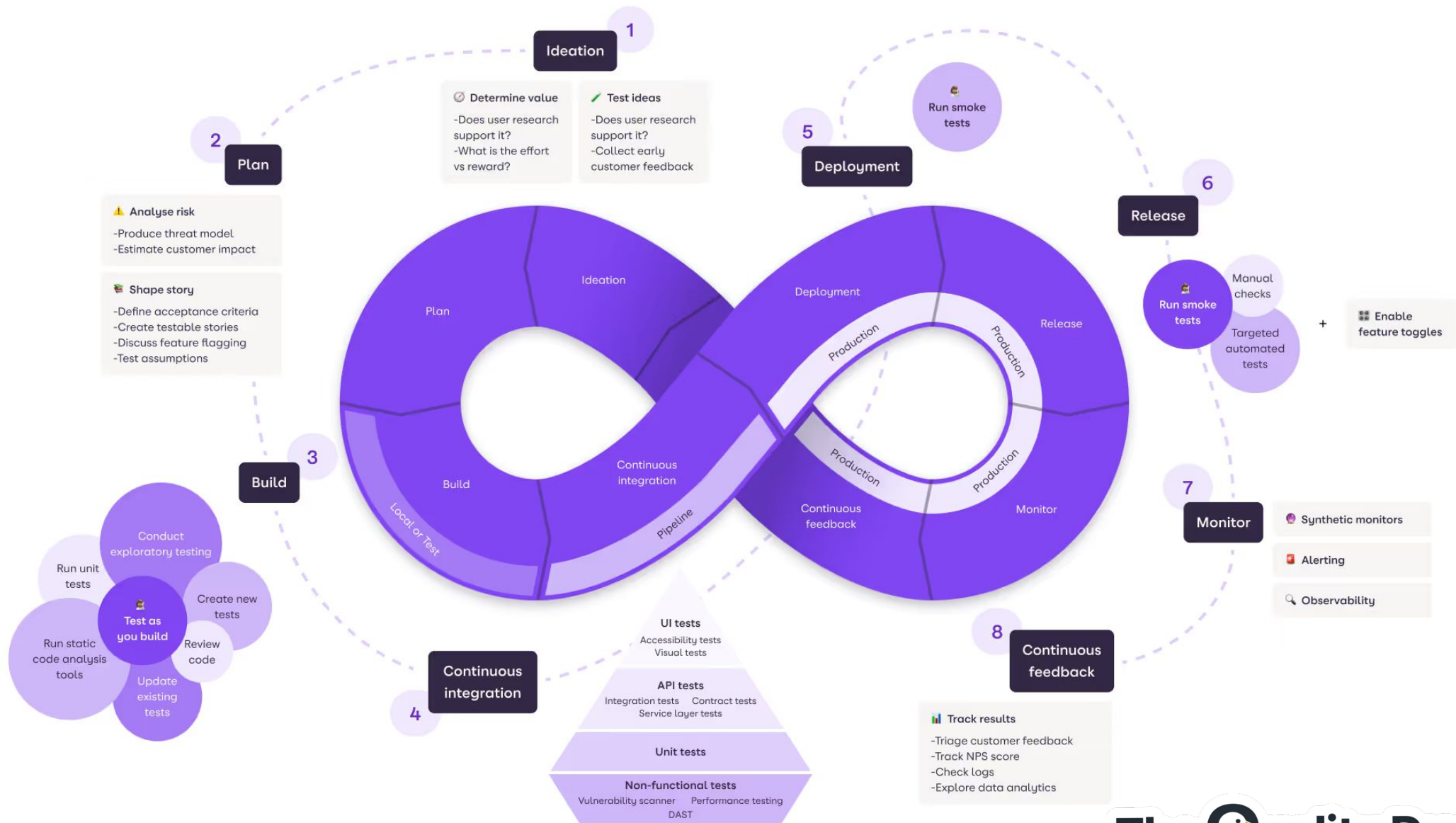
Credit: Dan Ashby - <https://danashby.co.uk/2016/10/19/continuous-testing-in-devops/>

Holistic Testing...

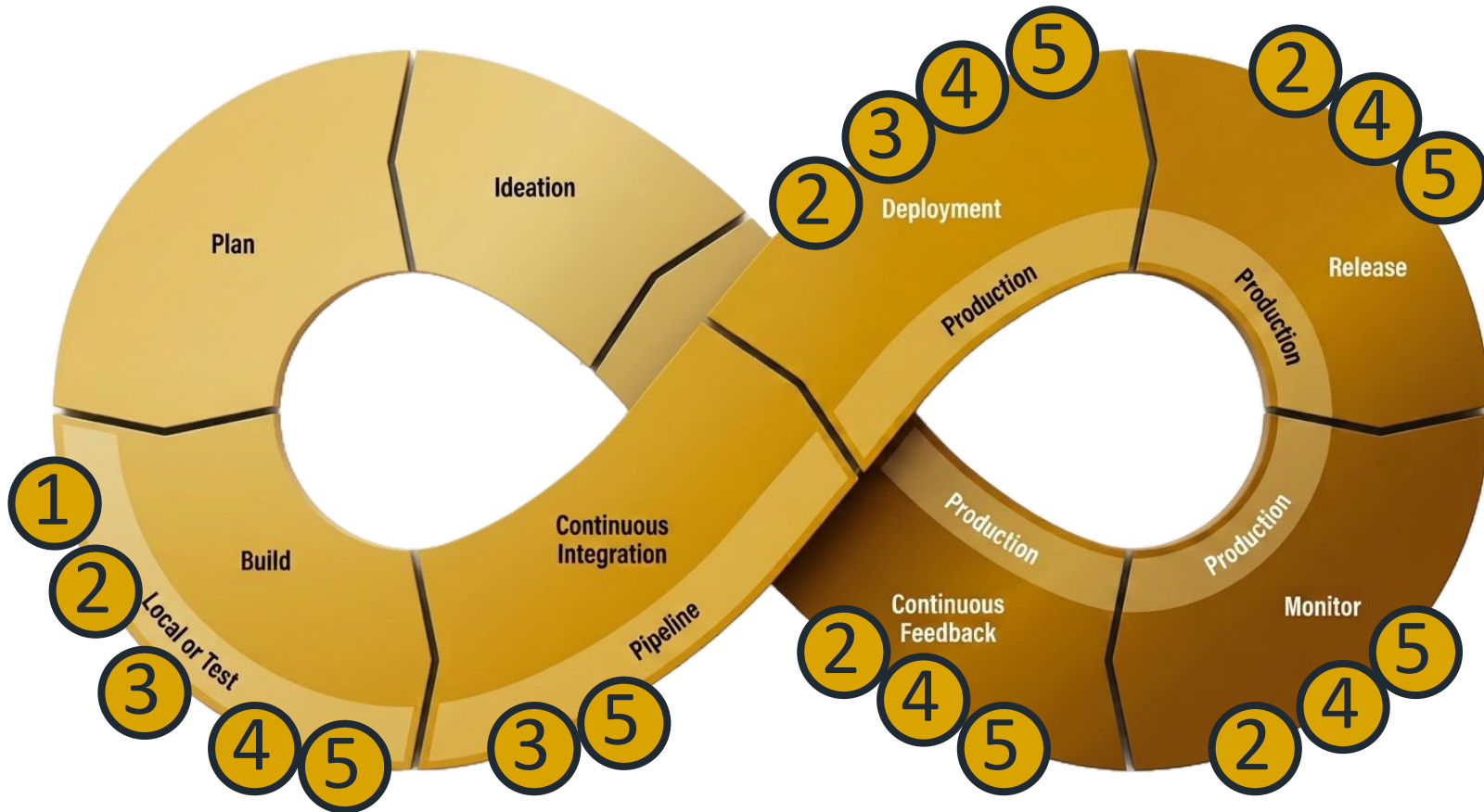


Credit: Janet Gregory - <https://janetgregory.ca/testing-from-a-holistic-point-of-view/>

More Holistic Testing...



How the Five Shifts Fit in...



- 1 Input Shift - Synthetic Personas
- 2 Verification Shift - LLM-as-a-judge
- 3 Baseline Shift - Benchmarking
- 4 Safety Shift - Human-in-the-loop
- 5 Observability Shift - Tracing

**Testing AI isn't about being
sure there is no risk.**

**It is about creating confidence
that the level of risk is low
enough to be acceptable.**

Questions?

Come see me for
office hours in the
Corn Exchange
13:30 – 14:30

**Want to try your
hand at testing AI?**

*Capture the flag:
Collaborative
red-teaming for
LLMs workshop
starts at 11:25*

Scan for
socials

