Why I Believe Open Robotics Makes Us Safer, Not Less Safe

Visibility, physical bottlenecks, and the case against black-box physical AI.

AUTHOR

Steven Palma

AFFILIATION

Hugging Face

PUBLISHED

Jun. 10, 2026

Table of Contents

One question keeps coming up at almost every robotics conference, panel, or networking chat: *“Aren’t you worried that making robotics more accessible will make it easier for ill-intentioned actors?”*

It’s a fair question.

Whenever you say *“make it easier for everyone”*, people start imagining worst-case scenarios. The concern usually sounds like this:

If we open-source robotics, won’t we end up with rogue humanoids, weaponized embodiments, and all sorts of dangerous systems?

I understand where that fear comes from.

Physical AI feels different. Software can stay inside a computer. Robots move through the real world.

After some time working on this question, I’ve reached the opposite conclusion: the safest future for robotics is an open one. The risks are real, and I won’t pretend otherwise. Openness is also how societies learn to understand, audit, and eventually manage powerful technologies.

TLDR; closed vs. open, side by side

Drag the slider from *Closed* to *Open* and watch what changes around the robotics stack: who can see it, who can question it, and how concentrated the power becomes.

From black box to community

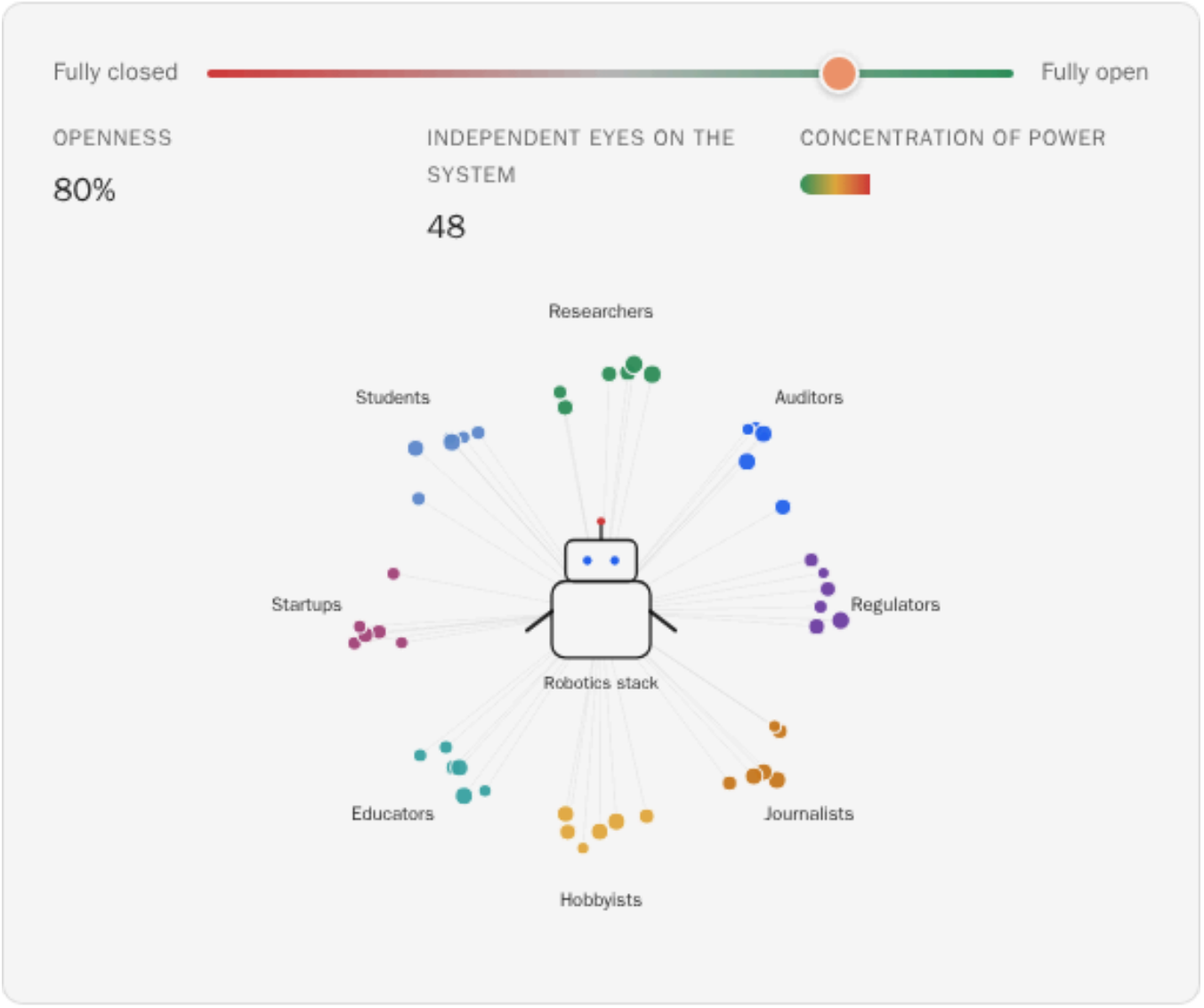

Figure 1 · A closed system reduces to a single point of trust. An open one distributes scrutiny across researchers, regulators, journalists, educators, hobbyists, and the public. Drag the slider to explore.

Every technology can be misused

There's a temptation to frame these discussions as a choice between *"safe"* and *"unsafe."* Reality is messier than that.

> Every meaningful technology in history has had both beneficial and harmful applications: the internet, GPS, cryptography, biotechnology, automobiles, electricity. Robotics will be no different.

The real question is what kind of ecosystem gives us the best chance of catching problems early, building safeguards, and holding people accountable. I don't think the answer is secrecy.

When development happens in the open, researchers can audit systems, policymakers can understand what is actually happening, journalists can investigate, universities can test claims on their own, and the rest of society gets to participate in the conversation. When it happens entirely behind closed doors, the rest of us are forced to trust a handful of organizations that may or may not be getting things right.

Transparency doesn't eliminate risk, but it gives us visibility into it, and visibility is usually where safety starts.

Software isn't the same thing as hardware

A lot of fears around open robotics assume that publishing code is equivalent to distributing robots. It isn't.

One lesson that keeps coming up in security discussions is that physical systems have physical bottlenecks. Try it yourself; click items in each column:

What you can download vs. what you have to build

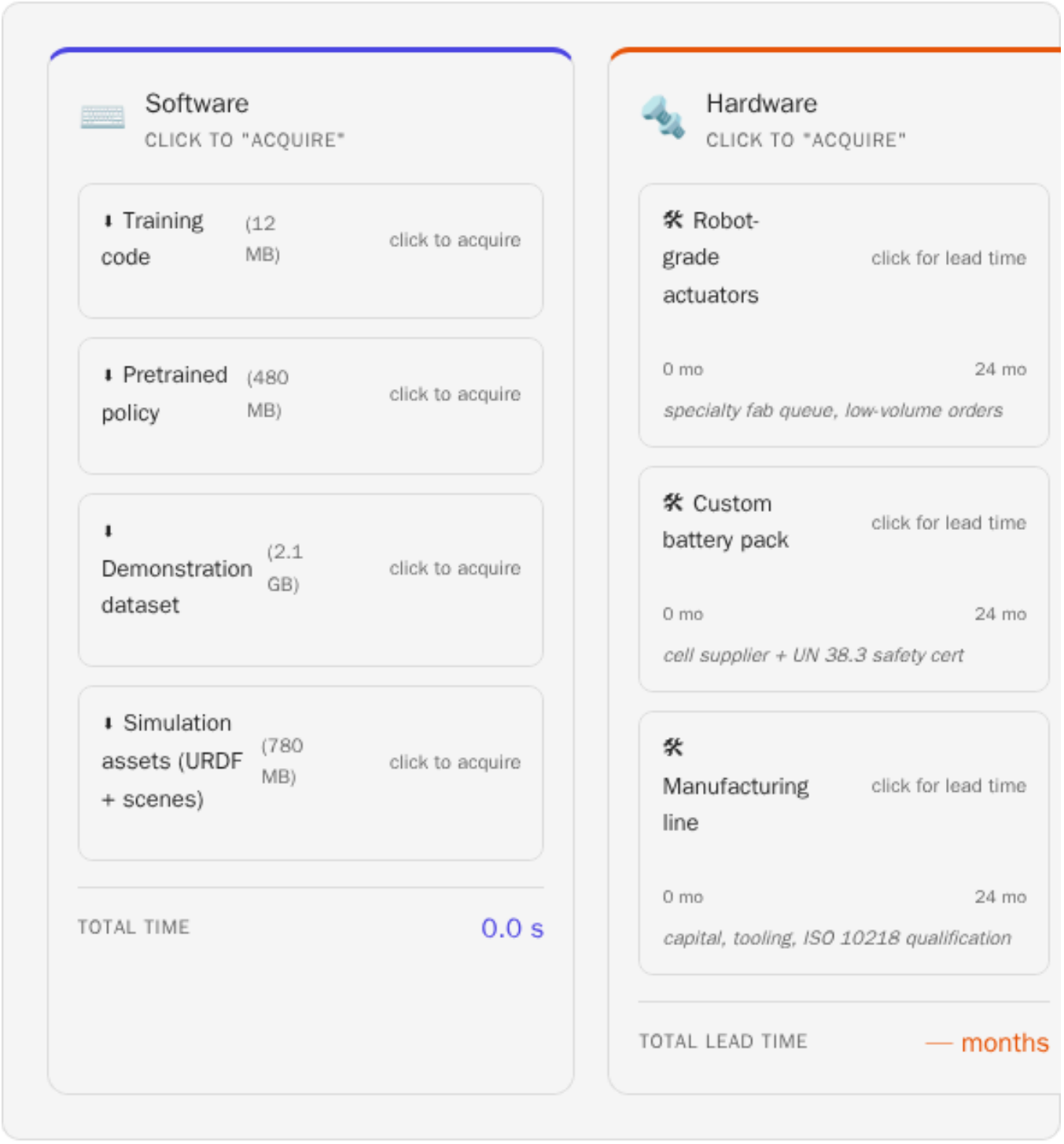

Figure 2 · Software finishes in seconds. Hardware comes with industry lead-time ranges measured in months. Click any item to reveal the real cost.

You can download software instantly. You cannot download actuators, batteries, manufacturing capacity, supply chains, precision hardware, or industrial expertise.

LeRobot doesn't magically give someone a fleet of advanced robots; it gives access to learning frameworks, research tools, datasets, and software infrastructure. The hard parts of building sophisticated physical systems stay hard whether the research is open or not.

Restricting research code doesn't remove the real-world constraints that already govern what people can build.

Biosecurity researchers reached a similar conclusion years ago. The most effective controls on dangerous biology aren't restrictions on publishing; they're screening and logging at the *physical* step of synthesizing genetic sequences [1]. The lesson generalizes. When a technology has real-world chokepoints, that's where safety policy belongs. Restricting information upstream mostly disadvantages the people trying to study the problem.

We've seen this movie before

One reason I'm skeptical of calls for extreme secrecy is that technology history is full of predictions that never materialized [2] [3]. Click any moment on the timeline below to see the original fear and what actually happened.

Predicted catastrophes vs. what actually happened

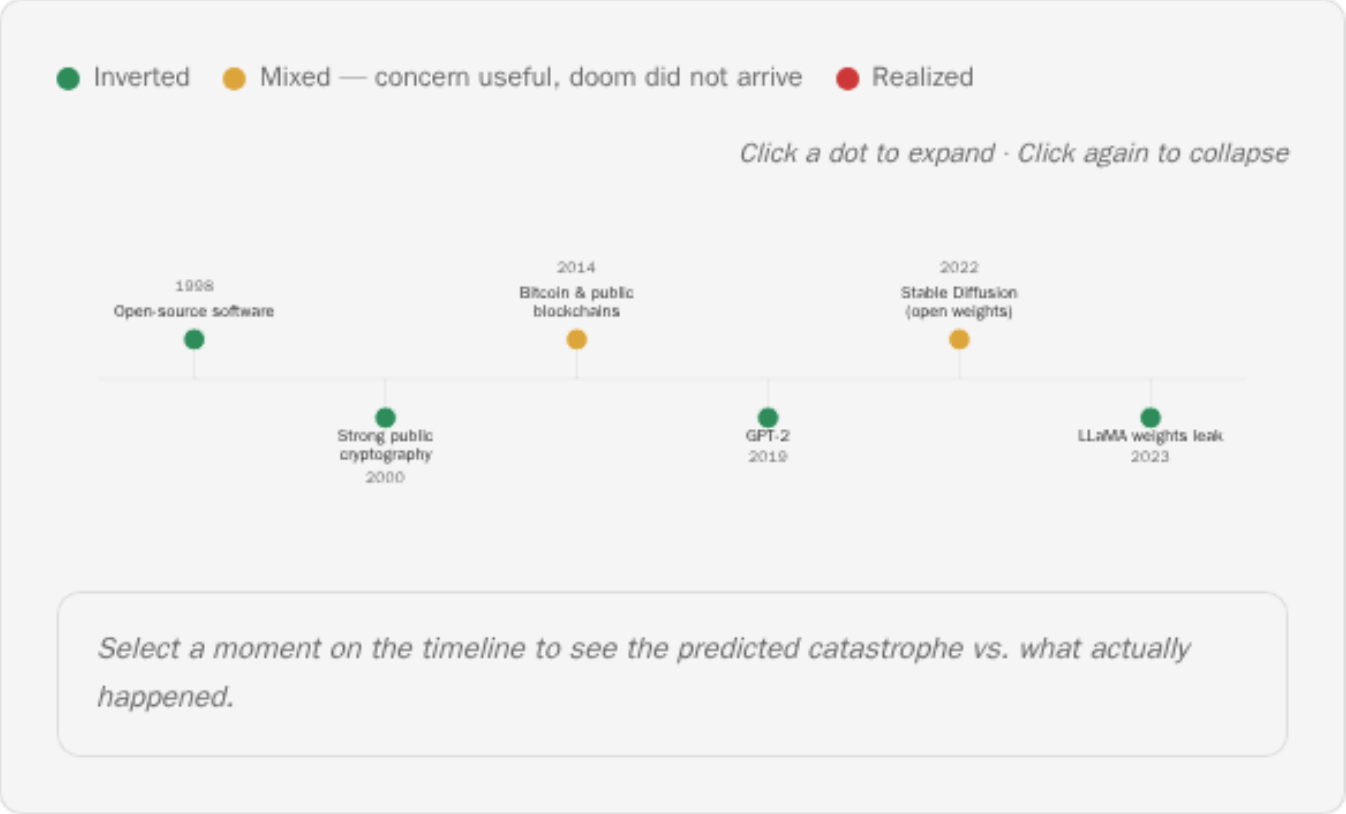

Figure 3 · A non-exhaustive timeline of technologies whose openness was predicted to cause disaster. Green = the prediction was inverted; amber = mixed (the concern was useful, the doom did not arrive).

That doesn't mean every concern was irrational. Some pushed people to think seriously about safety. History repeatedly shows, though, that open ecosystems often hold up better precisely because they attract scrutiny from thousands of independent contributors.

What about export controls, dual-use research, and weapons systems?

Who openness is actually for

Something tends to get lost when this turns into a pure safety debate: robotics is a science before it's a product.

The science part comes with rules: reproducibility, peer review, the ability for someone outside the original lab to confirm a result. None of that works when model weights, datasets, and training code sit behind a corporate firewall.

A closed robotics paper is closer to a press release than a contribution to the science. The compute and data divide between industry and academic labs widens accordingly, until university programs can no longer meaningfully evaluate the systems they're being asked to teach.

That has a cost, and it's absorbed by a specific group: students, researchers, educators, agricultural engineers, occupational therapists, factory operators, teachers, and the next generation learning from them. That's exactly the mission LeRobot exists to serve: by publishing the models, datasets, learning frameworks, and reference hardware we let anyone reproduce, audit, and build on the work. The most recent example being our release of a blueprint for building a real-world t-shirt folding robotic learning system: https://huggingface.co/spaces/lerobot/robot-folding

I don't want the physical world controlled by black boxes

This is actually the scenario that worries me more. Imagine the robots in our homes, workplaces, farms, and hospitals all controlled by systems nobody outside a few corporations can understand. Software is proprietary, decision-making is opaque, updates are mandatory, and everything runs through a centralized service.

⚠️ That creates enormous concentrations of power, and single points of failure [4]. If a central service goes down, changes policy, disappears, or simply makes a mistake, entire downstream systems break.

There's also a quieter cost. Robots see things software doesn't. A vacuum maps your floor plan. A kitchen assistant hears your conversations. A care robot watches an elderly relative. In a closed regime, that ambient data flows by default into corporate databases under terms users can't meaningfully inspect.

The other assumption behind secrecy is that hiding information makes everyone safer. In practice, while it might *feel* safer, it usually just makes defenders less informed. Bad actors don't wait for permission before collaborating; they don't need public acceptance to experiment.

The people most constrained by secrecy are researchers, regulators, auditors, educators, and independent safety experts. The result is an information asymmetry: the entities building physical AI know far more about its real behavior than the public institutions responsible for governing it. Courts, agencies, and independent reviewers end up evaluating systems they can't actually see. When those institutions can't see, self-reporting becomes the only signal, and self-reporting has a poor track record.

Who can audit: closed vs. open, same canvas

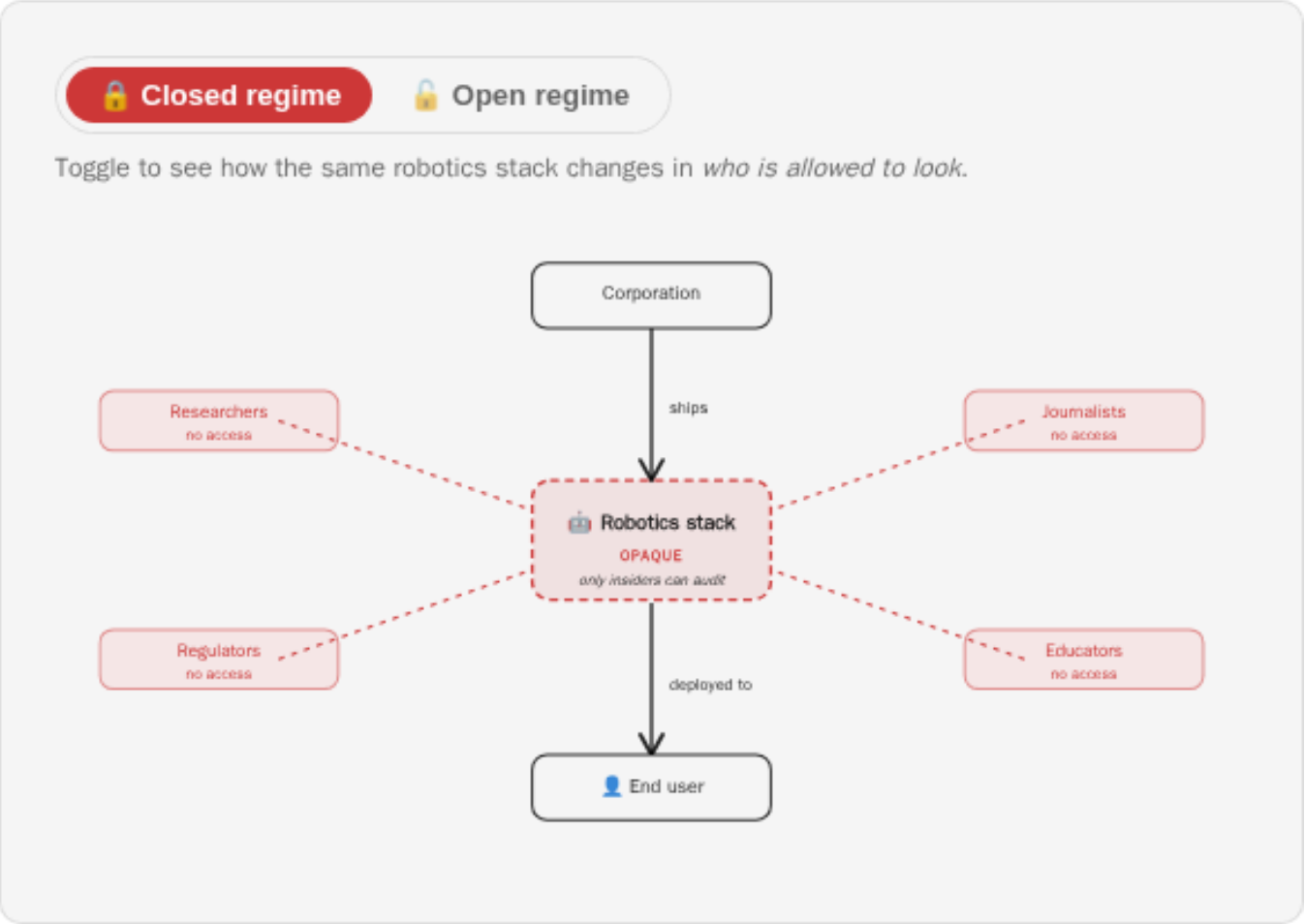

Figure 4 · Toggle between regimes. In the closed view, the stack is opaque and external observers are blocked. In the open view, the same stack becomes inspectable and the same observers gain explicit roles.

Open ecosystems create more eyes on the problem, more testing, more criticism, more chances to find vulnerabilities before they become serious. Cryptography is the clearest precedent: *"trust me, my closed cipher is unbreakable"* lost to *"here is the algorithm, try to break it, and we'll patch what you find."* No company, no matter how talented, can compete with the combined attention of a global community, and I suspect the same dynamic will play out for robotics.

The future I want to see

When I think about open robotics, I don't picture armies of rogue machines. I picture:

- Students building their first robot.
- Researchers sharing breakthroughs instead of duplicating effort behind closed doors.

- Agricultural automation, home assistance, elder care, scientific research, manufacturing, and education becoming accessible to far more people.
- Broader economic opportunity, with more people able to build, adapt, and benefit from the technology.

Most importantly, I picture a future where the people shaping robotics aren't limited to a handful of companies. A future where the technology is visible, understandable, and open to scrutiny.

Because if robotics is going to become part of everyday life, then society shouldn't experience it as a black box. We should be able to see how it works, participate in how it evolves, and help decide where it goes next.

> For me, that's the responsible path.

Citation

For attribution in academic contexts, please cite this work as

```
Steven Palma (2026). "Why I Believe Open Robotics Makes Us Safer, Not Less Safe".
```

BibTeX citation

```
@misc{palma2026_why_i_believe_open_robotics_makes_us_safer_not_less_safe,
  title={Why I Believe Open Robotics Makes Us Safer, Not Less Safe},
  author={Steven Palma},
  year={2026},

}
```


Footnotes

1. https://www.science.org/doi/10.1126/science.ado1671 ↑
2. https://dl.acm.org/doi/abs/10.1145/3593013.3593981 ↑
3. https://crfm.stanford.edu/open-fms/ ↑
4. https://huggingface.co/blog/cybersecurity-openness ↑