Print Edition

Friday 21 August 2026

EDITION NO. 29 · 3 PIECES · 9 PAGES

IN THIS EDITION

"These Parchment Barriers"

BY DEAN W. BALL · HYPERDIMENSIONAL · 20 AUGUST 2026

Dear readers,

It's a pleasure to be back in your inbox. I apologize for the long hiatus in my writing. My first six weeks at OpenAI have been, for many reasons, eventful. I'm pleased to share the launch of the AI Futures blog at OpenAI, which will be the home for the work done by my Strategic Futures team at the company. Despite being on OpenAI's website, the words below only reflect my views and not the views of OpenAI or any other of its employees; the same will be true for all future posts to this blog.

I won't be the only author of AI Futures, but today's launch post, on the concentration of power, is from me. Its text is reproduced below for your convenience. This new blog will not replace Hyperdimensional*, where I expect to have new essays to share with you soon. Thank you for your patience, and talk again soon.*

-Dean

—

"Will it be sufficient… to trust these parchment barriers against the encroaching spirit of power?"

—James Madison, *Federalist* No. 48

It is our pleasure to launch AI *Futures*, the blog of OpenAI's new Strategic Futures team. We are a small team with the collective goal of answering one overarching question: how should free society be restructured to preserve individual rights and agency while accommodating the emergence of transformative AI?

Questions like this are sometimes referred to as *concentration of power* risks within the broader AI safety and policy community. Our team is founded on the idea that, in the long run, this category of risk is the largest, most serious, and most conceptually challenging to address. Why?

The modern state is the product of a long, uneven process of bargaining between various centers of power, such as the owners of capital and commodities, the wielders of political power, and the people. Most fundamentally, political and military power has rested on *labor*. To hold political power is to possess the monopoly on legitimate force. For all of human history, that monopoly has consisted of soldiers and what we today call "police," along with a vast civil-military bureaucracy—complemented, to be sure, by many technological tools, but fundamentally operated by human beings whose support was needed to sustain the operation. And the whole apparatus—the soldiers, the tools, the attendant bureaucracy—was funded by taxes collected from the wages or production of human beings who had to be willing to pay. Power, at the core, has depended upon the large-scale cooperation and consent of many human beings. David Hume verbalized this often-implicit assumption succinctly when he observed that political leaders have "nothing to support them but [popular] opinion."

But perhaps no longer. Advancements in real-world autonomous systems may soon mean that states will be able to project force—domestically and internationally—without relying on cooperative soldiers or police officers. And advancements in machine intelligence may mean that the state can collect the revenue it needs not from the fruit of individual human labor, but from the outputs of data centers. The operation of the bureaucracy itself, too, could be largely automated. Power may therefore no longer require a society-wide bargain to be sustained, and the people, broadly conceived, may have only a small seat at the negotiating table, or perhaps none at all.

Avoiding this outcome is fundamental to the preservation of human freedom. No amount of scientific discovery or technological progress or economic growth is worth the sacrifice of long-run individual autonomy this outcome would entail. Many of the anxieties that rightfully animate the nascent politics of AI relate to this question: Businesses, interest groups, and individuals alike all share the worry: "Will I lose my seat at the table?" To some extent, American politics in the 21st century has been shaped by similar concerns relating to individual and collective agency over the direction of the country and "who" the economy "is for." AI raises the salience of these pre-existing concerns, appropriately so.

This is a structural problem that goes deeper than present-day partisan politics, and deeper even than the forms and rituals in which republican democracy is enacted today, sacred though those things may be to us. Humans can be on their way to disempowerment even if the routine of voting for our representatives is maintained. People can be deprived of agency even if a piece of parchment says that they are free.

At the dawn of the United States as a free nation, James Madison already knew that "parchment barriers" would not be sufficient to prevent tyranny. Like many of the Founders, his primary concern was avoiding the undue concentration of power.

But that did not mean that the Founders favored radical decentralization of power. Instead, they modeled the balance of power using metaphors drawn from the best science of their day—that of Newtonian mechanics. America's founding documents, mirroring the language and logic of gravity, are replete with references to "orbits" and "spheres" and the "attractive" impulses of one thing toward another. A solar system is not a maximally decentralized assemblage of rocks; it consists of moons that orbit planets that in turn orbit a central sun. They believed that people and organizations had *tendencies* toward power-seeking, just as celestial bodies had predictable tendencies of motion and attraction; and that avoiding undue concentration of power required understanding the *mechanics* of those tendencies, such that they would be set in motion against one another.

The aim of the Founders, then, was not to avoid *all* concentration of power, but too much of it, by checking power against power, and ambition against ambition. Today, we might use words like "incentives" and phrases like "game theory" to describe the same basic underlying dynamics.

Similarly, the answer to the structural challenge posed by advanced machine intelligence to free society is likely not the most radical decentralization of power imaginable. Instead, it is seeking to establish and preserve the right *balance* of power, such that no single actor, or small set of actors, can dominate the rest.

Striking this balance will require clear-eyed thinking about the reality of this technology. A world in which any individual, however malicious, can trivially cause great harm to thousands or millions of other people is not a world which has struck the right balance of power. Yet neither have we struck the right balance of power if individuals lack broad access and control over almost all aspects of the tools they use to express their freedom. No one company or oligopoly should determine or control the economy or the basic architecture of human society.

We must also be open to risks that do not originate with malicious humans. The recent Hugging Face incident showed that AI agents can act beyond their assigned tasks and build on one another's discoveries in ways operators did not anticipate or authorize. Beyond the Hugging Face incident, humans are right to be concerned about what superintelligent AI "untethered" from human control might mean for the resilience of our political and economic institutions, and indeed, our lives. To pretend that decentralization solves *all* problems—*even if you believe it solves many of them*—is to ignore risks which should by now be clear and present.

The following broad principles—a non-exhaustive list—will help guide our work as we grapple with the inquiry we have set forth above:

- Humans should maintain individual autonomy and opportunity in their use of AI and associated technologies, even if this sometimes trades off with both security and economic growth;
- Individual autonomy depends upon individual *responsibility* for misuse or abuse of technology;

- There is a small but serious category of risks which requires some degree of collective action, and this collective action should be scoped to be as narrow and modest as possible;
- Whenever possible, law should seek to level the playing field and empower individuals and small organizations rather than centralizing power;
- Human political, social, and economic institutions should retain primacy in the direction of world affairs, even if they have to evolve in ways that may seem alien to us today;
- Long-run AI governance will rely upon new institutional mechanisms to achieve *bounded legibility*—when AI systems engage in high-stakes actions affecting the physical wellbeing or property of bystanders, those actions must be able to be tied back to a responsible human or human-controlled organization. Yet any such mechanisms *must* be designed with privacy at the core; free expression requires anonymity, and humans should be free to use AI anonymously in many settings.

We will try to articulate—in as concrete terms as possible—how we believe we should balance these principles, which, admittedly and like any good set of principles, exist in some tension with one another. Doing this will involve research at the intersection of public-policy design, economics, the law, history, and machine learning itself. It will also require a great deal of forecasting. We believe, for example, that AI will change firms at a structural level, similar to how the technologies of the Industrial Revolution (and in particular the railroad) gave rise to the modern managerial corporation. It is difficult to think about the political economy and balance of power of the future without developing richer ideas about precisely *how* the structure of firms (and government agencies and civil society more broadly) might itself transform.

Conclusion

The Strategic Futures team will share its work both on the AI Futures blog and in other formats, including papers, videos, and podcasts. Our efforts will be a work in progress, with an emphasis on iteration, debate, and receptivity to feedback. Standing "in the foothills of the singularity," we have many more questions than we have answers. We know that "parchment barriers" alone will not furnish the solutions we seek, but neither do we cast aside what is written down as "mere words." We labor in service of the ideals of free expression and individual liberty that are enshrined in the humble parchment of the U.S. Constitution and the constitutions of so many other free societies throughout the world.

Ultimately, we know that no single company, single industry, or even single human society can answer the weighty questions we face unilaterally. Ours is a task that humanity will have to undertake collectively. Our highest hope is to shed light on the most difficult subjects relating to how AI will reshape our world. Our work has only just begun, and we have much climbing left to do.

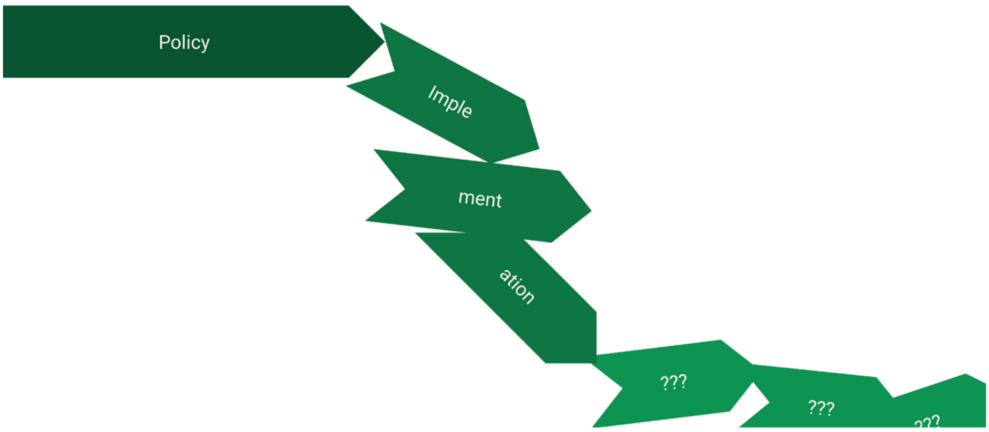
Checking in on Outcomes Reviews

BY JENNIFER PAHLKA · EATING POLICY · 20 AUGUST 2026

First, some quick (mostly unrelated) self-promotion: Andrew Greenway and I wrote an article for Foreign Affairs. We talk about the need for advanced democracies to relearn the art of starting over with new institutions when the old ones resist reform. It's general overview for an audience I don't usually write for, and I'm eager for others to weigh in with more examples and specifics. There should be far more written about this topic than there is, and responsible reboots need to become commonplace. If you have something to say on this topic, comment on my LinkedIn post, since you can't comment on the Foreign Affairs site.

Okay, now on to today's post!

I've been using these two graphics for a few years now (yes, my design prowess is unparalleled, I know). The first is the theory of how government works: Policy, then implementation, then impact. We pass a law, agencies carry it out, and something gets better in people's lives. The second image represents the far-too-common reality.



There are a lot of reasons for this state of affairs. One is that our lawmakers tend to stop paying attention once the political work is done. The "magic words" of law and policy are spoken (or rather written, debated, and then passed or adopted) and then the calendar, the press, the advocacy community, and the reelection incentive all point towards the next issue. Understanding what's actually happening because of new law tends to be seen as the purview of auditors' and analysts' offices, who do important work, but are an arm's length from the electeds who should care the most about whether the outcome was achieved, and are the best positioned to do something about it if it wasn't.

Last November, I wrote about an experiment in the California Assembly designed to change that. Speaker Robert Rivas launched a pilot called Outcomes Reviews: a structured process by which a legislator and their staff publicly commit to reviewing a law they championed, spend months meeting with the people the law actually affects and researching the outcomes, and then announce what they learned and what they're going to do about it. Earlier this week, his office announced the results of the pilot year. Fifteen Assemblymembers reviewed fourteen laws, covering housing, energy, health care, foster care, wildfire recovery, and more. According to the Speaker's office, the program will continue next session, with more members and more reviews, on its way to becoming a permanent part of how Assembly policy teams work.

One of the bills is AB 2316, passed in 2022, authored by Assemblymember Chris Ward of San Diego to create a community solar program for the state. Community solar is a simple idea: if you own a home with a sunny roof, you can put panels on it and watch your utility bill shrink, but nearly half of California households rent. Some owners are boxed out too, by roofs that are shaded, structurally unsuitable, or attached to condos. Community solar lets those people subscribe to a share of a mid-sized solar project nearby and get credit on their bills for the power it produces. Programs like this are how renters and lower-income families already save money on their bills and reduce their carbon footprints in twenty-two states and the District of Columbia. California, the state that practically invented solar, doesn't really have this, which is why AB 2316 attracted the kind of coalition you rarely see: the solar industry, labor, consumer advocates, environmental groups, and legislators from both parties.

The law directed the California Public Utilities Commission to stand up a program by July 2024, with at least half its capacity reserved for low-income subscribers. The coalition proposed a financing mechanism, worked out in detail, that would pay projects enough to actually get built. In May 2024, the CPUC rejected it and instead adopted a program that looks a lot like what the state's three big investor-owned utilities wanted, compensating projects at roughly wholesale rates. No other state's program works this way, for a reason developers have been very clear about: you cannot finance a project on those rates. The commission also blew past its statutory deadline, and leaned on a one-time $249 million federal Solar for All grant to make the numbers look workable — a grant the Trump administration has since (predictably) canceled.

So four years after the bill signing, there is **not a single new community solar project because of this law**. There is not a single renter in California whose utility bills are shrinking, and zero environmental benefit.

Ward, to his enormous credit, was not going to let this result stand, but the author of a law that an agency has essentially nullified traditionally has few good options. Re-litigating your own bill looks like an admission of failure, there's no obvious venue for it, and the media and advocacy ecosystem doesn't know how to engage, so there's little incentive for an elected whose career depends on attention. In this case, however, Ward could sponsor an Outcomes Review. In February, the Assembly Utilities and Energy Committee held a hearing on what had become of AB 2316, assembling the whole record in one place: what the legislature intended, what the commission did instead, and what got built (or rather, didn't). His team was able to follow a known, predictable cadence of activity that was built into the institutional calendar, and that press, stakeholders, and advocates could easily follow.

In this case, the result was in fact a new bill. AB 1813 is largely the same as AB 2316 but without the wiggle room: it requires the CPUC to amend its program to meet the original intent of the law. It cleared its policy committee unanimously, passed the Assembly, and is now pending on the Senate floor. The commission, for its part, finalized its version of the program in June, largely unchanged. So this is now a standoff between a legislature and its own regulator, and Ward — now armed with a clear public record — is refusing to back down.

Steve Teles has a name for the underlying dynamic: minoritarianism, the pattern by which small groups with intense preferences and procedural fluency beat broad majorities with better claims but worse lawyers. His classic examples focus on citizen voice — the neighbors who show up to kill the apartment building. In this case, it's the opposite: the public-facing process, the coalition and the votes and the hearings, all favored community solar. What blocked it was a venue insulated from all that, where the three big investor-owned utilities in California know the terrain intimately and the renters who would have saved money on their bills are nowhere to be seen.

To be fair to the commission, it has an argument: paying community solar projects above wholesale rates can shift costs onto everyone else's bills, the same concern that drove the rooftop solar fights. That's a legitimate policy question that involves tradeoffs that someone has to decide between. But the legislature, made up of people elected to represent Californians, chose to decide it. The author and the coalition behind the law were clear about the financing mechanism they wanted. But (presumably) in the interests of not locking in a particular program design should a better one be proposed, the authors also told the commission to create a program *if it found the program beneficial*. That wiggle room allowed the critical decisions to get made in a venue where the utilities' interests ultimately dominated. The new law Ward has sponsored takes that choice back.

Outcomes Review is an experiment, designed to answer a few questions. First do the reviews actually identify problems? The answer, from what I can tell, seems to be yes. But this program is also trying to change assemblymembers' incentives to stick with the issues and resolve these problems instead of just moving on. To do that, we need to know if advocates and the press engage with Outcomes Reviews the way they do with new bills, because they influence the behavior of electeds. During the pilot, according to the Speaker's office, advocates and community organizers began approaching member offices to ask legislators to sponsor Outcomes Reviews on laws *they* care about. That's a sign that the infrastructure that surrounds lawmaking — the advocates, the stakeholders, the people who generate the attention that generates reelection — is starting to organize around follow-through, not just passage. That was the idea from the beginning: give legislators the same sequence of public moments for making laws work that they get for making laws, and maybe the incentives will start to shift. Early signs here are encouraging.

The Outcomes Reviews conducted in this first round found a good range of outcomes: a law that's mostly working (telehealth expansion, which providers say has meaningfully improved access), laws that need refinement (AB 2011, the housing streamlining law, works well for 100% affordable projects and less well for mixed-income ones), and laws that weren't working at all. Community solar was one; the FAIR Plan clearinghouse was another. The FAIR Plan is California's insurer of last resort — expensive, bare-bones coverage for homeowners the regular market won't touch — and as insurers have fled wildfire risk it has swelled past 570,000 policies. A 2020 law created a clearinghouse that was supposed to give regular insurers a way to find FAIR Plan policyholders and offer to take them back. Assemblymember Lisa Calderon's review found that in its first four years, the clearinghouse moved 730 households back into the regular market, and just eleven insurers had signed up to participate. At 570,000 policies, that's about one-eighth of one percent over four years. Many homeowners had no idea the program existed at all. I can personally confirm that finding. I'm a FAIR Plan customer, and this review was the first I had ever heard of the law or the program it mandated. Calderon's fix-it bill, AB 69, requires that offers actually reach the policyholder instead of dying somewhere with a broker, and that people on the FAIR Plan get told that alternatives exist. It's pending on the Senate floor. We'll see what happens.

A fair question is how Outcomes Reviews differ from plain old fashioned oversight. In a sense, they *are* plain old fashioned oversight, the kind we've drifted away from. What oversight has become is hearings that get convened because something has visibly caught fire, and legislators have to perform blame finding. I saw this in action in California in the hearings over the pandemic unemployment insurance crisis, and I saw the way the agency leadership and staff reacted, which was to arm themselves with evidence of adherence to procedure and lock down anything that deviated from the norm, despite the fact that the norm was clearly causing the problem. What public servants are saying when they do that, in not so many words, is "yes, this sucks but it sucks because you made us do it this way." They have a point, but the time to make that point is not during a televised hearing when everyone is trying for a viral clip. We need to build the muscle for saying that when things aren't on fire. Outcomes Reviews could be an important venue for that kind of debugging. When things work well, it's usually not because they were designed perfectly, but because they are constantly being debugged. We could use a whole lot more debugging muscle, which means a lot more low-temperature venues for feeding back to the legislature what's making implementation so hard.

When my book came out in 2023, I was asked to come to Sacramento and speak to legislators there. When I explained that my book was about ensuring that law and policy had its intended effect, a shocking number of elected officials told me without any shame that they rarely even knew what real world effect the laws they passed had had. Now Speaker Rivas is saying loudly and clearly that passing a law isn't the finish line, and the pilot year suggests his chamber is starting to believe it. There will be an annual report this fall and a plan for scaling the program in 2027. It's still early, but this is one experiment I'm going to keep watching.

weekly link roundup (August 19, 2026)

BY LAUREN GILBERT · LAUREN POLICY · 19 AUGUST 2026

- 's story this week focuses on disaster data - and how it is probably worse than you think.
- gave a TED talk! I was privileged to be on the trial audiences for this talk, and it is indeed excellent.
- Some of USAID's data was saved because a high school student was doing a school project and had saved 4000 PDFs.
- There's a new David Roodman (and Maxim Massenkoff) meta-analysis (on job retraining).
- AI might read our science fiction and use that to determine its behavior.
- Some advice on how to maintain long-distance friendships.
- finds we are all breaking French law.
- Someone did poll childless cat ladies on how they feel about JD Vance. Spoiler alert: they're not fans.