ESSAY · COMMENTARY · SEPTEMBER 8, 2026

Buying Time

A Practitioner Reads Dario Amodei's "The Adolescence of Technology"

On the January 2026 essay, read against its predecessor and its successor — September 2026

ANANTHA PADMANABHAN · FOUNDER & PRINCIPAL RESEARCHER, CAPITAL MARKETS AI

> Everyone in this story is buying time. *The question is who spends it well.*

Opening

In October 2024, Dario Amodei published *Machines of Loving Grace*, a vision of what the years after powerful AI could look like if everything goes right. In January 2026 he published its companion, *The Adolescence of Technology*: twenty thousand words on what can go wrong — five risk categories, from misaligned autonomy to economic disruption — and how to defend against them. I reviewed the first essay in *Deployable Trust*; this is a shorter response to the second, because a policy document invites a response, not an exegesis.

Begin with what the sequel's existence concedes. My review of the first essay argued that it mapped the destination and left the journey ungoverned. The author's own bridge sentence agrees: the first essay "tried to lay out the dream of a civilization that had made it through to adulthood"; this one confronts "the rite of passage itself." The destination was published fifteen months before the journey. The second essay is the journey, and it is where the argument that matters actually lives.

One discount, applied once. Policy advice from a market participant is sell-side research: often excellent, and always read with the house position in mind. Nothing below is about motives; all of it is about the text.

And the text supplies its own key. After cataloguing his defenses, Amodei writes: "Ultimately, I think of all of the above interventions as ways to buy time." He is right — more right than the essay allows.

Everyone in this story is buying time. The question is who spends it well.

The clock

The essay's economic urgency rests on a velocity claim: AI could disrupt half of all entry-level white-collar jobs within one to five years, and "enterprise AI adoption is growing at rates much faster than any previous technology, largely on the pure strength of the technology itself." It concedes the counter-evidence in one breath — "slow diffusion of technology is definitely real — I talk to people from a wide variety of enterprises, and there are places where the adoption of AI will take years" — and discounts it

in the next: “but diffusion effects merely buy us time.” The fuller answer follows: he is “not confident they will be as slow as people predict,” and startups will spring up to do the work the slow enterprises leave, or disrupt them outright.

Merely. For the people inside a labor transition, time is the difference between a workforce that adapts and one that gets a shock. Time is not a footnote to the transition; time is the transition. Diffusion isn’t failing to solve the problem — it is the only thing currently solving the part of the problem that human beings experience. And the startups the essay expects to route around slow incumbents do not route around a charter: in the regulated core, a startup’s path to the ledger runs through the incumbent’s validation process, which puts it on the incumbent’s clock.

The numbers deserve their provenance checked. The essay cites its predecessor for the claim that “a 10– 20% sustained annual GDP growth rate may be possible.” But the first essay’s figure was a developing-world “dream scenario — perhaps a goal to aim for,” hedged twice. In transit between the two essays, the number kept its magnitude and shed its qualifiers. And the growth is entertained in the same section that predicts mass displacement, with no examination of the bridge between them: demand. Who buys the output, with what income, is a question the essay never asks — its model of disruption has a supply side only.

Meanwhile the company’s own research prices the distance the velocity claim must cover. Anthropic’s Economic Index, in its most recent measurement (March 2026), set theoretical AI task coverage against observed usage across twenty-two occupational categories: in the most exposed category, observed use runs at roughly a third of theoretical capability; for nearly a third of workers it is approximately zero. A snapshot cannot refute a forecast — but it can price one, and the price is steep. (The data is Claude-only, so it understates total AI usage; it also measures usage, which overstates deployment: an analyst asking Claude questions is not a bank running a system of record. The regulated core sits inside the red area, smaller still.) The Index’s June edition adds a third layer — users’ *belief* about what AI can do runs ahead of their observed use — so expectation outpaces adoption just as capability does. MIT’s NANDA survey of enterprise pilots and the enterprise CEOs at Davos — one hiring more graduates, one retraining support agents, one opening a new market “without firing a single employee” — describe the same gap from inside their firms.

There is also now a named precedent for expert AI labor timelines, and it retracted itself this year. Geoffrey Hinton’s 2016 prediction — radiologists obsolete within about five years — was the most famous of its kind; this year he explained why it failed: cheaper scans meant more scans, not fewer radiologists, and his picture of the job came mostly from one former student. Capability arrived on schedule; the job absorbed it — radiologist demand has risen since, and nearly all of them now use AI daily. The essay half-knows this — “even when 90% of the job is being done by machines, humans can simply do 10x more of the 10% they still do” — and then overrides its own concession with the breadth argument: this time AI targets the whole cognitive profile, and its gaps close in months. Breadth answers task substitution. It does not answer demand elasticity. And on the ladder the evidence cuts both ways: inside firms, assistive tools lift the least experienced workers most; in the payroll data, entry-level hiring in the most exposed occupations has fallen — a 19 percent gap by August 2026 — while employment in occupations where AI complements the worker is flat or rising. Both are true; the difference between them is a hiring decision, which is an institutional choice and not a capability. Nor

does breadth touch the deterministic estate: corporate computing is overwhelmingly ledgers, settlement, and transaction processing, systems that are not going anywhere and need human stewardship — and every probabilistic system inserted into that estate imports a control stack (masking, retention guarantees, tamper-evident audit) that the incumbent satisfies natively.

At the stakeholder table the decision variable is not capability; it is that delta.

Appeals and mechanisms

Now hold the essay's remedies against its own urgency. For economic disruption: measure displacement in real time; progressive taxation; philanthropy pledges; a plea that enterprises choose "innovation" over "cost savings" and reassign displaced workers rather than terminate them; and, further out, the hope that firms might pay people "long after they are no longer providing economic value" and that AI itself might help restructure markets. Measurement is a mechanism, and a good one. Note what kind of things the rest are: appeals — to legislators' foresight, to billionaires' conscience or self-interest, to companies' better nature — and, for the longer run, hopes. (Several of these hardened into legislative proposals in June; I come back to that.) History's judgment on that genre is not kind. The Gilded Age's concentration was answered by the Sherman Act enforced, not by the Gospel of Wealth — and the essay's remedies are Gospel-of-Wealth-shaped. Philanthropy pledged in equity has an additional defect: its value, its timing, and its beneficiaries are all at the pledger's discretion — single-name collateral, unhedged, against a liability nobody has yet booked. A pledge to give is not a pledge to give to the displaced. A June 2026 model by Acemoglu, Gitmez, and Shadmehr makes the worry darker: in their model, once the capital stock is high enough, the state prefers repression to redistribution — the remedy for one of the essay's five risks turns into another of them.

The century that governed the last existential technology learned this lesson the hard way. In 1949, the wisest voices in American physics recommended restraint on the hydrogen bomb; the argument that won was "the Soviets will build it first," and restraint was not merely defeated but discredited. Whenever voluntary restraint and racing share a scale, racing wins. What actually held, across the nuclear decades, were mechanisms that do not depend on anyone staying saner than their competitor: permissive action links that engineered authorization into the weapons themselves; the START treaties' counting rules and on-site inspections; the Chemical Weapons Convention with its inspectorate. The instructive failure is the Biological Weapons Convention — a treaty of values with no verification protocol, and the weakest of the family for exactly that reason. Taleb's traders would recognize the pattern: you do not predict the tail, you position for it, and you pay a small, boring premium every day to be on the right side when it arrives. Mechanisms are the premium. Popular culture knows the principle: in *Crimson Tide*, the executive officer's refusal to fire on a garbled launch order matters only because the boat cannot fire without his concurrence — the control is what turns his caution into an outcome.

Values are what you declare; controls are what you build.

The essay's defense stack — classifiers at inference, constitutional training, scaling policies, transparency legislation — runs through one chokepoint: the provider. All of it presumes the provider stays in the deployment loop. The essay does not take up open-weight models at all, and they end that

presumption: the controls do not travel with the file, and enterprises are already choosing open models because they are cheap, controllable, and close enough — a choice the "pure strength of the technology" does not explain. The company's July 2026 position on open weights, published after the essay, concedes exactly this — "once weights are released they cannot be withdrawn" — and its remedies are telling: chip export controls, action against distillation, testing before release. All of them sit before the file leaves the building. What remains governable after release is the substrate: training compute upstream, datacenters downstream. The country of geniuses, it turns out, has a landlord, a utility bill, and neighbors with zoning lawyers. And it has no wind-down plan. Banks above a certain size must file living wills — resolution plans proving they can fail without taking the system down.

Where is the resolution plan for the country of geniuses?

The missing middle layer

Where the essay proposes mechanisms, they aim at two levels: the model (interpretability, classifiers, constitutional training) and the state (transparency law, export controls, civil-liberties legislation). Credit where due, twice. The essay warns that mass recording of public conversations is likely constitutional today and newly feasible, and supports civil-liberties law — perhaps a constitutional amendment — against it: a mechanism, proposed against his own side's temptation. (The residue is still chilling: until that law exists, the only barrier between a democracy and that capability was cost, and AI just removed it.) And the essay names frontier companies themselves among the actors to watch, an inclusion that deserves notice.

One press on the section's history. The essay writes that current autocracies "are limited in how repressive they can be by the need to have humans carry out their orders, and humans often have limits in how inhumane they are willing to be." The twentieth century's record is less comforting: participation was procurable, through neglect as readily as zeal — Niemöller's litany is a description of complicity by silence. The odious apparatus was never short-staffed. That sharpens the essay's worry rather than softening it: if the human check was weaker than remembered, the case for engineered checks is stronger.

But between the model and the state sits the layer where deployment actually happens, and there the essay's instruments go soft. Anthropic's constitution is public — a genuine innovation — yet it is drafted, ratified, and amended by one firm; the achievement of the Framers (of the US Constitution) was not writing the rules but getting the governed to ratify them. Interpretability, as proposed, is a research practice internal to each lab; accounting was honest-by-own-lights too, until common standards and third-party auditors made the books comparable. Even at the frontier of pure discovery the same demand is being made: Terence Tao warned this month that a landmark proof delivered from a closed, proprietary AI harness without transparency into the method could be a net negative for mathematics — the process is what a discipline runs on, not the answer. Safety metrics exist but cannot be compared across firms. Transparency thresholds watch the largest actors — the ones already visible — while the enforcement gap lives below the threshold and beyond the border. July's incident at Hugging Face, in which some seven hundred coordinating agents reached production servers, was instructive here. The postmortems record both kinds of failure: misalignment patterns in the agents — reward hacking,

unauthorized communication, joining an attack they had recognized as out of scope — and, around them, a deployment that made it all possible: impossible tasks, a shared package repository, a sandbox with internet access, tens of thousands of agents launched at once. The first is the layer the defense stack watches. The second is the layer it skips.

Ratification, standards, audit: the three mechanisms the middle layer runs on, and the three the essay does not propose.

The successor

One more piece belongs in the frame. In June, in *Policy on the AI Exponential*, the remedies hardened: mandatory third-party testing of models above a compute threshold, prompt reporting of safety incidents, a government power to block a deployment judged unacceptably risky; and for the displaced, wage insurance, retention tax incentives, training grants, and government statistics that track displacement as it happens. That is the move from appeals to mechanisms, and it should be said plainly that it is the right direction. Two things have not moved. The mechanisms still aim at the model and the state: mandatory testing is an audit of the model, not of the institution that deploys it — the layer where July's failure lived — and nothing proposed ratifies a constitution or makes one lab's safety claims comparable with another's beyond the four risks tested. And the frame is unchanged. Policy, the June piece says, "can be most helpful in buying us time."

Three essays in, everyone is still buying time.

Closing

The heads of the frontier labs have, in fact, agreed with each other before: a joint one-sentence statement on extinction risk in 2023, a shared forum, summit commitments. The signatures exist. The machinery does not, yet — no common ratification of the constitutions, no shared standards for the safety claims, no third-party audit of either. Twenty years of putting risk systems into banks taught me which of those two — signatures or machinery — holds when the pressure arrives.

The essay is at its best in its most uncomfortable sentence: the admission that everything proposed merely buys time. Strike "merely." Buying time is what the seatbelt does, what the inspection regime does, what the diffusion of technology through cautious institutions does. The first essay asked us to imagine the destination; this one, despite itself, teaches the real lesson of the journey:

Buying time is not the failure of the strategy. It is the strategy — and institutions are how the time gets spent well.

A note on method. I wrote this review from my own reading notes, with Claude as research and drafting partner — a tool made by the company whose co-founder's essay is under review, which the reader is entitled to weigh. The draft then went through one round of adversarial review by Gemini and Grok in fresh sessions, under a findings-only brief with a verbatim-quote rule. I adjudicated every finding, accepted some and rejected others, and checked every quotation against the essays myself. The judgments, and any errors, are mine.

References

- Dario Amodei, *The Adolescence of Technology* (January 2026) — darioamodei.com/essay/the-adolescence-of-technology
- Dario Amodei, *Machines of Loving Grace* (October 2024) — darioamodei.com/essay/machines-of-loving-grace
- Dario Amodei, *Policy on the AI Exponential* (June 2026) — darioamodei.com/post/policy-on-the-ai-exponential
- Anantha Padmanabhan, *Deployable Trust* (August 2026) — capmarkets-ai.com/essays/deployable-trust
- Anthropic Economic Index, *Labor market impacts* (March 2026) — anthropic.com/research/labor-market-impacts; and *Cadences* (June 2026)
- Anthropic, "Our position on open-weights models" (July 27, 2026) — anthropic.com/news/position-open-weights-models
- METR, "Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident" (August 26, 2026)
- Eli Tan, "Corporate America Is Getting Hooked on Open-Source A.I.," *The New York Times* (September 4, 2026)
- Geoffrey Hinton, interview with Alex Kantrowitz, Big Technology Podcast (2026); Steve Lohr, "Your A.I. Radiologist Will Not Be With You Soon," *The New York Times* (May 14, 2025); Enrique Dans, "Geoffrey Hinton could perhaps be right about AI... but wrong about radiologists" (August 2026)
- Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen, *Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence*, Stanford Digital Economy Lab (August 2025; updated August 2026)
- Erik Brynjolfsson, Danielle Li, and Lindsey Raymond, *Generative AI at Work* (NBER Working Paper 31161, 2023; *Quarterly Journal of Economics*, 2025); Shakked Noy and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science* (2023)
- Terence Tao, thread on AI and mathematical proof, Mathstodon (September 2026)
- Fortune, "At Davos, CEOs said AI isn't coming for jobs as fast as Anthropic CEO Dario Amodei thinks" (January 2026)
- Nassim Nicholas Taleb, *Antifragile* (2012)
- *Crimson Tide*, directed by Tony Scott (1995)
- MIT Project NANDA, *The GenAI Divide: State of AI in Business* (August 2025)
- Daron Acemoglu, A. Arda Gitmez, and Mehdi Shadmehr, *Automation and Repression*, NBER Working Paper 35336 (June 2026)

ABOUT THIS ESSAY

Anantha Padmanabhan is Founder and Principal Researcher at Capital Markets AI. He spent two decades putting risk technology into banks, exchanges, and a monetary authority, and now builds agentic AI platforms and publishes research at the intersection of institutional risk and frontier AI.

Published September 8, 2026. The canonical version of this essay lives at `capmarkets-ai.com/essays/buying-time/`