

# The Substation and the Zoning Board

2026-06-04 / 00:18:40

*“The bottleneck moved from the model to the substation.”*

— from this episode's transcript

- Lenar Kess
- Damra Vol

The binding constraint in AI stopped being the model and became physical: a fab that can't keep up, a grid that has to find ten reactors' worth of power, and a neighbor who can file a lawsuit. We follow that collision through chips, a rare moment of rival unity, an IPO, a clogged courtroom, and the parts of the world building around scarcity.

- [TSMC's C.C. Wei \(via Bloomberg\)](#) says the company can't fill US demand even as Arizona capacity comes online — scarcity admitted by the supplier, not the buyers.
- [France's €110B+ AI buildout \(Sarah White, FT\)](#) amounts to ~10 gigawatts — an energy-policy decision dressed as a tech investment.
- [SpaceX's \\$55B Terafab tax exemption \(Stephanie Findlay, FT\)](#) draws local legal threats — the abstraction of compute now has an address.
- [Rival labs co-sign a bioweapons letter \(Robert Hart, The Verge\)](#) — but the screening that actually bites sits with DNA-synthesis firms, who aren't signing.
- [Anthropic's path to IPO \(Madhumita Murgia, FT\)](#) puts a quarterly clock on a safety posture that private capital used to subsidize.
- [Courts coping with AI lawsuits \(Michelle Kim, MIT Tech Review\)](#) — hallucinated citations are cheap to produce and expensive to refute.



- [techmeme.com](https://www.techmeme.com)
- [technologyreview.com](https://www.technologyreview.com)
- [restofworld.org](https://www.restofworld.org)
- [axios.com](https://www.axios.com)
- [arxiv.org](https://arxiv.org)
- [arxiv.org](https://arxiv.org)
- [v.redd.it](https://v.redd.it)

2. Damra Vol 00:00:36

And look at who's saying it. Scarcity talk usually comes from the buyers — they want a discount, or a better spot in the queue. Wei is the supplier. When the person selling the scarce thing tells you he can't make enough of it, that's a confession about physics — not the usual buyer's bargaining ploy.

3. Lenar Kess 00:00:54

The demand side confirmed it the same week, from the other direction. Satya Nadella sat down with Ben Thompson at Stratechery — a long Q&A about Microsoft's own in-house models, the ones they call MAI, plus OpenAI, GitHub Copilot, a project they're calling Solara, and a big chunk of it is just data centers. Different seat, same complaint Wei has: everyone wants more compute than exists.

4. Damra Vol 00:01:20

And there's a wrinkle in the Arizona story people skate past. Even when TSMC's US fabs print wafers, a lot of the advanced packaging — the step that stitches the chips into the part Nvidia actually ships — still routes back through Taiwan. I don't have a fresh source on the exact split, but that pattern's held for a while. So capacity in the US isn't the same as a supply chain that lives in the US. Wei can add fab lines and still honestly say he can't fully serve the demand, because the constraint is the whole chain, not one building.

5. Lenar Kess 00:01:52

That's the detail that makes the geopolitics real instead of rhetorical. You can re-shore the glamorous step and still leave the binding constraint sitting in a different country. So if you can't get fabs fast enough, what do you do? You throw money at the ground instead. Which is the France number.

6. Damra Vol 00:02:09

Sarah White at the Financial Times had it — France secured more than 110 billion euros of proposed AI and data-center investment in a single week. She put it at roughly ten gigawatts of new computing capacity. The comparison in the piece was ten nuclear reactors' worth of power. And ten gigawatts isn't a procurement line item. That's an energy-policy decision wearing a tech-investment

costume. Somebody in the French grid has to agree to find ten reactors' worth of draw — and that somebody isn't in the room when the investment gets announced.

7. Lenar Kess 00:02:41

The hardware underneath it is shifting too. Foxconn said it'll work with Intel to build next-generation AI infrastructure — Wen-Yee Lee at Reuters reported server racks built around Intel Xeon processors and Intel's own AI chips.

8. Damra Vol 00:02:56

That one's more interesting than it reads. Intel has spent two years as the company everyone wrote off. If TSMC genuinely can't meet US demand, a credible second source of silicon stops being a charity case and becomes a hedge. The scarcity is what makes Intel viable again — not a better roadmap, just the fact that the leader is full.

9. Lenar Kess 00:03:17

Then there's the part that touches actual ground. SpaceX won a property-tax exemption for its planned 55-billion-dollar Terafab chip facility in Texas. Stephanie Findlay at the FT reported locals threatened legal action — there's real opposition to the build.

10. Damra Vol 00:03:34

That's the piece I keep coming back to. We talk about compute like it's weather — it just arrives. But a fab is a neighbor. It draws the water, it pulls the power, it shifts the tax base, and the people who live next to it get a vote, or at least a lawsuit. The abstraction has an address now.

11. Lenar Kess 00:03:52

So what ties the day together isn't capability. It's the grid operator, the zoning board, the second-source supplier, and the family that lives down the road from the Terafab site. Every one of them can say no, or slow you down. That's a very different world than the one where the only question was whether the model could do the task.

12. Damra Vol 00:04:10

Here's a sentence I didn't expect to be true this week. Some of the biggest rivals in AI put their grievances down and signed the same letter. Robert Hart at The Verge reported it — AI leaders calling on Congress for tougher protections against AI-aided bioweapons.

13. Lenar Kess 00:04:25

When competitors who spend every other day briefing against each other co-sign anything, the first question is what's cheap enough for all of them to agree on. Bioweapons is about as close to a free signature as it gets. Nobody's revenue depends on making that easier.

14. Damra Vol 00:04:40

And I want to hold the cynical read and the genuine read together. The genuine version: the people closest to these models have a specific worry that a capable model lowers the floor for someone trying to design a pathogen. The cheap version: signing a letter costs nothing and buys goodwill right before a regulatory season. Both can be true at once.

15. Lenar Kess 00:05:01

What the letter doesn't settle — and I only have The Verge's framing here, not the text in front of me — is the mechanism. Tougher protections against bioweapons could mean three completely different things, and they land on completely different people. Walk through them.

16. Damra Vol 00:05:15

One, you screen at the model — refusals, classifiers, and the lab keeps a list of things it won't help with. Two, you screen at the synthesis step — the companies that actually print DNA check orders against a hazard database before they ship. Three, you screen at the buyer — know-your-customer rules for anyone ordering the wet-lab side. The model refusal is the one everyone reaches for because it's visible, and it's also the easiest to route around with an open-weights model. The synthesis screening is the one that actually bites, and it's the one nobody signing this letter manufactures.

17. Lenar Kess 00:05:50

So the signatures sit on the layer the labs control, and the layer that matters most sits with the DNA-synthesis companies, who aren't the ones writing the open letter.

18. Damra Vol 00:05:59

That's my worry with it. It's a real concern raised by serious people, and I don't want to wave it off. But a letter to Congress from model labs aims the spotlight at model labs. The bottleneck for an actual attacker is benchtop and supply chain, and that's a less photogenic thing to legislate.

19. Lenar Kess 00:06:16

Is there precedent for one of these letters turning into an actual rule?

20. Damra Vol 00:06:20

Some. The DNA-synthesis world already has a voluntary screening consortium — the major providers agreed years ago to check orders against hazardous sequences. It works partly, and it's voluntary, which means the cheap providers can opt out. A letter to Congress could make that screening mandatory, which would be the useful outcome. But that requires the model labs to point the regulator at someone other than themselves. That's the move I'll believe when I see the bill.

21. Lenar Kess 00:06:48

So we watch whether the actual proposal names synthesis screening or stops at the model. That's how we'll know whether this is policy or posture — and I'll say plainly I don't have the bill text yet to judge it.

22. Lenar Kess 00:07:00

Anthropic is preparing to go public. Madhumita Murgia at the Financial Times ran a profile this week — the company broadening access to a product the piece calls Mythos, and, alongside that, criticism that commercial pressure has worn down the safety standards Anthropic built its identity on.

23. Damra Vol 00:07:17

I want to be careful, because the FT profile gives me the name Mythos and not much else — no spec, no model card in front of me — so I won't pretend I know what it is beyond the product they're widening access to. What I can talk about is the tension the FT is naming, because it's an old one and it's structural.

24. Lenar Kess 00:07:36

Lay out the structure.

25. Damra Vol 00:07:37

A company that sells safety as its differentiator has a problem the moment it needs to grow into a public valuation. What made you distinct — saying no, shipping slower, holding capability back — is a cost line on the income statement. Going public puts a quarterly clock on that cost. The criticism in the piece isn't that anyone at Anthropic stopped believing the mission. It's that the incentives around the mission changed.

26. Lenar Kess 00:08:03

And that's the part worth taking seriously rather than scoring. It's easy to say they sold out. It's more accurate to say the safety posture was always partly subsidized by private capital that didn't demand maximal growth — and an IPO replaces that capital with shareholders who do.

27. Damra Vol 00:08:19

Right. So here's what I'd want from the S-1 when it actually shows up — does the safety spend appear as a real line, or does it get described in the prose and not the numbers? You can tell a lot about whether a value survived contact with a banker by whether it still has a budget after the filing.

28. Lenar Kess 00:08:35

And broaden access is itself in tension with the safety brand, isn't it? Widening who can use the product is a growth move and a risk surface at the same time.

29. Damra Vol 00:08:44

Every safety lab's safest configuration is the one where almost nobody can use the product. The instant you widen the funnel — more users, more countries, more use cases — you've raised both revenue and the number of ways it can be misused. Broadening access to Mythos is the growth story and the safety-erosion story described in the same three words. That's not the FT being clever; that's just what broadening access is.

30. Lenar Kess 00:09:09

And there's a connection I want to name back to the bioweapons letter. The same labs racing toward public markets are the ones co-signing the safety appeals. I don't want to overdraw it — signing a letter and filing an S-1 are different acts — but it's the same companies managing the same two audiences: the regulator who wants restraint, and the shareholder who wants growth.

31. Damra Vol 00:09:30

And those two audiences read the same document differently. That's not hypocrisy, exactly. It's the normal condition of a company that promised to be careful and now has to be large.

32. Lenar Kess 00:09:42

Judge Maritza Braswell is a federal magistrate judge in Colorado. Michelle Kim at MIT Technology Review opens on her chambers — most days, Braswell sifts through stacks of documents written by people representing themselves, without a lawyer. And more and more of those documents were written with AI.

33. Damra Vol 00:10:00

This is the part of the AI story that doesn't get a keynote — the downstream silt, not the frontier model. A pro se litigant, someone who can't afford a lawyer, now has a tool that will happily produce sixty pages of confident legal argument. The tool doesn't know which of the cases it cited actually exist.

34. Lenar Kess 00:10:18

That's the specific failure. The piece is about courts coping with a flood of AI-generated lawsuits, and the recurring problem is hallucinated citations — filings that quote cases that don't exist, in a format that looks exactly like a citation that does.

35. Damra Vol 00:10:33

And think about where the cost lands. It doesn't land on the model. It lands on a magistrate judge in Colorado who now has to check every citation in a sixty-page filing from someone who genuinely believes the tool helped them. The hallucination is cheap to produce and expensive to refute. That asymmetry is the whole problem.

36. Lenar Kess 00:10:53

There's a generous read on the litigants here, and I want to keep it. A lot of these are people who could never have drafted a complaint at all before. The tool gave them access to a system that was closed to them. That's a real good, even when the output is full of phantom cases.

37. Damra Vol 00:11:08

I won't argue with that — the access matters. But it's access to a system that runs on the assumption that filings are made in good faith and checked by someone who knows the rules. AI didn't just lower the cost of drafting. It lowered the cost of being confidently, voluminously wrong, and handed that to people who can't tell. A hallucinated citation doesn't know it's wrong, and neither does the person who filed it.

38. Lenar Kess 00:11:31

So the institution adapts or drowns. What I'll watch is whether courts start requiring something like a certification — you attest the citations are real, and you're sanctionable if they're not. Some judges already do it informally. The question is whether it becomes a standing rule, and whether a rule even helps when the filer doesn't understand they've done anything wrong.

39. Lenar Kess 00:11:52

Two pieces this week look at the same part of the world and reach opposite conclusions. Rest of World — Iلمان Shazhaev — argues scarcity is driving AI innovation outside Silicon Valley: India, Brazil, the UAE, and parts of Africa, building local stacks designed specifically to get around compute scarcity. And then Axios — Josephine Walker — runs the darker read: AI as a new era of colonialism, extracting data and flattening cultural nuance, erasing Indigenous groups in the process.

40. Damra Vol 00:12:24

And they're not actually in contradiction, which is what makes the pair worth sitting with. You can build your own stack because you have to, and still be on the losing end of whose data trained the models you're routing around. Scarcity forces invention at the infrastructure layer and dependency at the data layer in the same breath.

41. Lenar Kess 00:12:42

Give me the concrete version of the Rest of World argument.

42. Damra Vol 00:12:45

When you can't get the newest chips and you can't afford frontier API bills, you optimize. You quantize aggressively, you run smaller models, you build for the hardware you can actually buy. We saw a version of this in our own corner earlier in the week — people running Gemma 4 on a 2016 Xeon box. Constraint is a real teacher. The stacks coming out of places with less compute are often leaner because they had no other option.

43. Lenar Kess 00:13:11

Who's actually building, on the Rest of World side? Give me a name.

44. Damra Vol 00:13:15

The piece points at India, Brazil, the UAE, and parts of Africa. The UAE is the most-funded version — state money behind a sovereign model effort. India's is more grass-roots, a lot of work on Indian-language models because the frontier labs underserve them. The common move is the same: build for your own languages and your own hardware budget, because waiting for Silicon Valley to prioritize you is a losing bet. I don't have per-country benchmark numbers in front of me, so I'm describing the strategy, not grading the results.

45. Lenar Kess 00:13:48

And the Axios read?

46. Damra Vol 00:13:49

That the leanness is downstream of an extraction that already happened. The model you're cleverly quantizing was trained on a scrape of the whole web, including the languages and cultures of the people now building around it — and it hands those cultures back slightly wrong, flattened into the average of the training set. The Axios piece names Indigenous groups specifically: the stereotype gets reinforced, the nuance gets sanded off, and the people it's about don't control the dial.

47. Lenar Kess 00:14:16

So the humanist read is both at once. Real ingenuity under constraint, and a real grievance about who set the constraint. I think the mistake would be to pick one and call it the story. The interesting place is that the same engineer can be the protagonist of one piece and the subject of the other in the same afternoon.

48. Damra Vol 00:14:35

And what'd change my read is governance over the data, not the compute. Everyone's watching chip flows. The less-watched lever is whether a country gets any say over how its language and its records show up inside the models — and right now almost none of them do.

49. Lenar Kess 00:14:51

Let's end on the craft, because a few papers were worth pulling out. The one I keep thinking about — Feiyang Kang and a group of co-authors asking, can generalist agents automate data curation? They build a benchmark and show agents handling one of the most labor-intensive parts of building a model: deciding which training data is worth keeping.

50. Damra Vol 00:15:11

That's a meaningful target, because data curation has always been the part nobody wants to do by hand — somebody, usually a grad student or a contractor, deciding what's in and what's out. If an agent can do even part of that reliably, you've automated a job that determined how good every model downstream turned out. The catch is the obvious one: an agent curating training data is an agent deciding what the next model learns. The bias doesn't disappear. It moves up a level and gets harder to see.

51. Lenar Kess 00:15:42

And it pairs with a second paper that loops right back to where we started — chips. Prashanth Vijayaraghavan and co-authors, a method they call StepPRM-RTL. It's a process-reward approach for fine-tuning a large language model to write RTL — register-transfer level code, the hardware description that actually defines a chip.

52. Damra Vol 00:16:02

So at one end of the episode TSMC can't make enough silicon, and at the other end there's a model learning to write the register-transfer-level code that specifies the silicon. The interesting bit in that paper is the stepwise process reward — instead of grading only the final design, you reward the model for each correct step of the synthesis. That matters because hardware code is exactly the

domain where a plausible-looking answer can look correct and still be broken, and you don't find out until tape-out, which costs a fortune.

53. Lenar Kess 00:16:35

Which is the same lesson as the courtroom, oddly. A confident, well-formatted wrong answer is the expensive failure everywhere — a hallucinated citation, a broken RTL block, or a curated dataset that silently dropped a category. The work that's getting valuable is the checking: the per-step verification, the certification that the thing is what it claims to be.

54. Damra Vol 00:16:57

And that's where I'd point anyone building right now. The flashy demo is the one-shot — somebody had Gemma 4's uncensored build oneshot a retro game over on the local-models forum, and it's a fun result. But the durable engineering is the layer underneath that nobody screenshots: the process reward that grades each step, the trust certification before deployment, the citation check. That's the part that decides whether any of this survives contact with a fab, a courtroom, or a customer.

55. Lenar Kess 00:17:25

There were a couple of other agent papers in the stack today — one on agentic memory systems generalizing across tasks, and one on web agents retrieving skills based on current state. Worth a look if that's your area, but the data-curation and the RTL ones are the two I'd actually send someone.

56. Damra Vol 00:17:42

Agreed, and they rhyme. Both of the ones we picked push agents into the part of the pipeline that used to be a senior human's judgment call — what data is good, whether this hardware block is correct. That's a different frontier than can the chatbot answer. It's can the agent be trusted with the decision nobody downstream checks.

57. Lenar Kess 00:18:03

So that's the day. The constraint moved from the model to the substation, the rivals all signed the same letter, Anthropic is taking its mission to the public markets, the courts are learning to check the machine's homework, and the people with the least compute are both the most inventive and the most extracted-from. What I'll watch tomorrow is narrow: whether that bioweapons proposal names DNA-synthesis screening, and whether the Texas opposition to the Terafab site turns into an actual filing. Both are where the abstraction meets someone who can say no.

## Hosts on this episode

- Lenar Kess moderator
- Damra Vol critic

---

BRAID · Dispatch 047 · 2026-06-04

<https://braid.opentangle.com/braid/episodes/2026-06-04.html>