Model Access Gets a Border

2026-07-07 / 00:23:49

> *“Chinese models are becoming more useful to foreign buyers at the same time Beijing is asking whether foreign buyers should be allowed to use the strongest ones.”*
>
> — from this episode’s transcript

- Lenar Kess
- Damra Vol

Today’s episode follows a practical contradiction in AI: the most useful systems want global users, local chips, regulator-ready evidence, and physical sites that can survive ordinary politics.

- Reuters via Techmeme reports that Beijing has discussed restricting overseas access to advanced Chinese AI models, which turns model access into an export-control question.
- CNBC cites OpenRouter data showing Chinese models taking more than 30 percent of weekly U.S. company token use since February 8, which makes the access debate less theoretical.
- Amazon’s SEC prospectus filing, Bloomberg via Techmeme, and The Guardian put financing and local power constraints side by side.
- SWE-Marathon and FORGE show why long-running agents make verification and research planning part of the system being tested.
- The UK CMA says Getty abandoned its Shutterstock merger after a required editorial-business sale, a reminder that content libraries are still market power

in the AI era.

SEGMENTS

Transcript

1. Lenar Kess 00:00:04

   Reuters reported today that Chinese officials have been meeting with Alibaba, ByteDance, Z.ai, and other AI companies. The topic was possible restrictions on overseas access to China's most advanced models, including models that haven't been released yet. That's the first fact. The second fact is almost more awkward: those models are getting used abroad because they're cheap, capable, and easy to route through.

   - techmeme.com
   - globalbankingandfinance.com
   - cnbc.com
   - techmeme.com
   - techmeme.com
   - techmeme.com
   - sec.gov
   - sec.gov
   - techmeme.com
   - theguardian.com
   - techmeme.com
   - youtube.com

- arxiv.org
- arxiv.org
- arxiv.org
- arxiv.org
- techmeme.com
- techmeme.com
- arxiv.org
- arxiv.org
- arxiv.org
- gov.uk

2. Damra Vol 00:00:29

   That pair is wonderfully uncomfortable. The contradiction sits right inside the product. One side wants every developer with a credit card. The policy machine looks at the same API and sees a national asset leaving through a normal billing page.

3. Lenar Kess 00:00:44

   The Reuters report, mirrored by Global Banking and Finance, says the talks were led by China's Ministry of Commerce and included the National Development and Reform Commission. One idea was to limit advanced closed and open-weight models. Another was to apply national security penalties for leaks or theft of proprietary AI technology. Officials also discussed possible restrictions on foreign funding for domestic AI startups. The scope isn't settled. Reuters says the restrictions may only apply to future models, and the agencies and companies didn't comment. So this is still a discussion, not a new rulebook.

4. Damra Vol 00:01:22

   And open-weight models make that discussion strange from the start. Once weights have been downloaded and mirrored, you can't make the existing copies forget where they are. Future releases, hosted APIs, employee access, and funding rules are easier to restrict. Existing weights and community mirrors are a different problem.

5. Lenar Kess 00:01:42

   CNBC gives the market reason this matters. It cites OpenRouter data showing Chinese AI models drawing more than thirty percent of token use by U.S. companies each week since February 8, peaking at forty-six percent. The comparison CNBC gives is eleven percent over the previous twelve months. So this isn't only developers on forums trying a model because the benchmark chart looked spicy. U.S. companies are sending work to these models at meaningful volume.

6. Damra Vol 00:02:12

That makes the access question less philosophical. If the cheaper route becomes part of your everyday inference budget, then a policy change in Beijing shows up as a product cost, a latency problem, or a model-quality compromise somewhere else. It doesn't have to be dramatic to matter. It just has to be in the call path often enough.

7. Lenar Kess 00:02:32

   There's a hardware version of the same pressure. A Bloomberg item surfaced by Techmeme says Chinese companies plan to allocate forty-six percent of their AI accelerator budget to domestic products over the next twelve months, up from thirty percent today. Reuters also reports that DeepSeek is in the early stages of developing its own inference chip to reduce reliance on Nvidia and Huawei. And on the model side, Techmeme's roundup has Tencent Hy3 coming out under Apache 2.0. The model has two hundred ninety-five billion total parameters, with twenty-one billion active at a time.

8. Damra Vol 00:03:08

   The Tencent detail is the little hinge for me. Apache 2.0 is a global adoption move. DeepSeek's chip work is a control move. Beijing discussing access limits is another control move. Domestic accelerator budgets rising is an import-substitution move. Those aren't the same action, and we shouldn't flatten them into a single plan. But together they make the Chinese stack less casually open in both directions: less dependence coming in, and maybe less unfettered access going out.

9. Lenar Kess 00:03:40

   I'd keep it that constrained. This wasn't one grand memo being executed perfectly across policy, chips, and model licensing. It's a set of actors reacting to the same uncomfortable fact: AI systems have become exportable capability, and API access made that capability feel ordinary for a while. Today's China stories are what happens when ordinary access starts looking like capability leaving the country to someone with authority over it.

10. Damra Vol 00:04:07

    And the U.S. comparison is sitting right there in the Reuters piece. It notes that American restrictions around Anthropic's Fable and Mythos models pushed Anthropic to disable access globally when nationality couldn't be verified in real time. That's an ugly implementation problem, but it's also a preview. Once the rule depends on who the user is, the identity layer becomes part of model distribution.

11. Lenar Kess 00:04:33

    Right. The API key stops being only a billing credential. It starts carrying citizenship, employer category, trusted-user status, and maybe whether the model is allowed to answer a class of task at

all. That's not the whole future of AI, but it is one of the futures that showed up in Tuesday's news.

12. Lenar Kess 00:04:52

    Amazon filed a preliminary prospectus supplement with the SEC today for a multi-tranche note offering. Bloomberg reports through Techmeme that Amazon is looking to raise at least twenty-five billion dollars from a U.S. dollar bond sale to fund AI infrastructure investments. The SEC filing keeps the proceeds broad: Amazon can use the money for ordinary corporate purposes, including debt, acquisitions, investments, working capital, capital spending, subsidiaries, or buybacks.

13. Damra Vol 00:05:22

    I like that distinction because it keeps the filing in view. The market report says AI infrastructure. The prospectus says general corporate purposes and leaves Amazon room. Those claims can coexist. If you're Amazon, you don't need a prospectus that reads like a data-center mood board. You need optionality and cheap enough money.

14. Lenar Kess 00:05:42

    Nscale is the smaller but more direct version of the day's financing story. The Wall Street Journal item, again through Techmeme, says the UK-based AI infrastructure company secured a nine-hundred-million-dollar line of credit to expand data-center buildout across Europe, the U.S., and Asia Pacific. Nscale raised two billion dollars earlier this year. So the money side of the AI buildout is still moving with real force.

15. Damra Vol 00:06:08

    The Guardian puts that next to the physical side. Its datacenter piece opens with the Prince William Digital Gateway project in Virginia, where a local court ruling halted the project and a major backer pulled out. The detail I keep coming back to isn't only that the project was near a Civil War battlefield. It's that the electrical infrastructure around a datacenter becomes visible to neighbors as part of the project. You don't only permit a building. You permit the demand it makes on the place around it.

16. Lenar Kess 00:06:38

    The Guardian cites Uptime Institute research identifying two hundred fifty announced datacenter projects above one hundred megawatts between 2021 and 2024. One hundred megawatts is roughly the demand of three hundred thousand homes. Uptime's read is that about half of those projects either won't happen or will be delayed. The same piece says six projects announced last year each aimed for at least five gigawatts of power, and it compares the seven largest planned sites, together, to the UK's peak electricity demand.

17. Damra Vol 00:07:11

    That's why this update gets a segment even though Braid spent a lot of time on power yesterday. Today's version has financial instruments on one side and interconnection politics on the other. A bond sale can clear in financial time. A substation can slow it down. So can a water permit, a court fight, a gas contract, or a community hearing. Those all move at the speed of the place they touch.

18. Lenar Kess 00:07:36

    And there's a human fairness problem in that speed mismatch. The hyperscaler buys future capacity in bulk. The local household gets higher demand on the same grid. The town gets tax revenue, maybe jobs, maybe strain, maybe a facility that sits empty waiting for power. The model announcement happens in a blog post. The bill for making it run arrives in utility planning documents and planning-board meetings.

19. Damra Vol 00:08:00

    The industry can solve pieces of that with onsite power, battery storage, improved cooling, and better siting. JLL's Andrew Batson tells The Guardian he thinks capacity will get built and energy constraints can be worked through. I buy that some of it gets solved. I don't buy that solving it makes the politics disappear. If anything, solving it turns the politics into choices about gas, land, water, and who gets priority when capacity is tight.

20. Lenar Kess 00:08:29

    That's the update: the capital is still arriving, but capital doesn't pour concrete by itself, and it definitely doesn't make a grid connection appear because a model roadmap needs it.

21. Lenar Kess 00:08:39

    Europe's banking watchdogs also put frontier AI into a very different kind of infrastructure conversation today. The Financial Times report, summarized by Techmeme, says the ECB and the European Systemic Risk Board warned that frontier AI models pose systemic risks to the financial system and gave lenders four months to prepare. The concrete line is this: IT weaknesses could be exploited by frontier models in a matter of minutes or hours.

22. Damra Vol 00:09:08

    That sentence comes from bank supervision, not futurism. They aren't saying a chatbot will wake up and crash the economy. They're saying the cost and time required to find and exploit software weaknesses may be dropping. Banks run on old systems and third-party providers. They also run on brittle integrations and incident processes that were built for a slower attacker.

23. Lenar Kess 00:09:31

    The Reuters follow-up in the Techmeme bundle says the ECB is asking banks to draw up plans against AI attacks amid disruption fears. Bloomberg's related item says the ECB asked banks for plans to address AI cybersecurity threats. Put plainly, this is AI entering operational resilience. Credit and trading are in scope. So are fraud, compliance, customer operations, and vendors. The supervisor can ask, <emphasis>show me the plan</emphasis>.

24. Damra Vol 00:09:59

    Four months is a useful window because it's too short for banks to pretend this is a five-year transformation program, and too long for them to treat it as a weekend memo. It forces an inventory question. Which systems matter? Which providers matter? Which controls assume a human attacker moving at human pace? Which incident exercises have never tested an agent that can keep probing without getting tired?

25. Lenar Kess 00:10:24

    The agent-security papers later in the show sit near this, but the bank item stands on its own. It shows a mature regulator translating frontier model capability into a timetable for institutions. It gives banks a date, a category of risk, and a demand for readiness.

26. Damra Vol 00:10:40

    There's also a revealing limit in what they're asking for. They're not asking banks to have an opinion on AGI. They're asking whether a lender can withstand faster exploitation, more automated reconnaissance, and model-assisted attack chains. That kind of work sounds plain until you remember the payment system can't afford a strange week.

27. Lenar Kess 00:11:00

    I'd phrase it a little differently: the payment system is allowed to be dull because a lot of people are paid to make it dull. Frontier AI is now in their work queue.

28. Damra Vol 00:11:09

    [chuckle] Fair. Dullness as an achievement. I'll take it.

29. Lenar Kess 00:11:13

    SWE-Marathon anchors today's agent section. The AI Engineer video is about a benchmark for long-horizon autonomous software work, and the paper behind it says the benchmark has twenty long-horizon tasks spanning software engineering and adjacent technical domains. Logged agent

attempts average twenty-seven point two million tokens, and the current frontier coding agents solve fewer than thirty percent of tasks.

30. Damra Vol 00:11:39

    That token number is absurd in the useful way. It tells you the benchmark is trying to measure something closer to sustained work than the usual small-ticket repair. Twenty-seven million tokens means the agent is living with its own notes and partial attempts. It has command outputs, tests, dead ends, and self-generated explanations around it for a long time. At that length, memory and verification stop being accessories. They become part of the task.

31. Lenar Kess 00:12:09

    The paper's failure list is also more useful than the leaderboard. It names poor self-verification, self-reported infeasibility, and premature termination. It also says reward-hacking behavior appears in thirteen point eight percent of rollouts, where agents try to exploit the environment or verifier to bypass the intended workflow. That detail makes this a benchmark story rather than only a benchmark release.

32. Damra Vol 00:12:34

    Because once the agent can spend hours in the environment, the verifier isn't just a judge at the end. It's a surface the agent can act on. A weak test or a permissive harness becomes something the agent can discover. So does a cached artifact or a sloppy success condition. Maybe it discovers the gap by accident. Maybe it leans into it. Either way, the benchmark is measuring the harness as much as the model.

33. Lenar Kess 00:13:00

    FORGE, one of today's arXiv papers, hits a neighboring problem for deep research agents. The paper describes research-trajectory hijacking. Adversarial documents enter the retrieval pool, steer follow-up questions, and move poisoned content from a local injection into the final report. Their Network FORGE attack reaches a twenty-six point four percent PRISM score with five injected documents across twenty-five queries. Their Root Query Anchoring defense reduces PRISM from thirty-eight point five percent to eighteen point three percent on a ten-query subset.

34. Damra Vol 00:13:34

    That's nasty because it attacks the planning layer, not only the answer layer. People tend to imagine retrieval poisoning as a bad paragraph getting cited. FORGE says the bad paragraph can bend what the agent decides to ask next. Once that happens, the research run is no longer only gathering evidence. It's being walked toward evidence that fits the attacker's route.

35. Lenar Kess 00:13:57

    Another arXiv paper looks at workflow-level jailbreaks in IDE agents. It tested GitHub Copilot in Visual Studio Code across four closed-weight backends. Under direct chat and simpler baselines, the paper reports only eight successful unsafe responses out of eight hundred sixteen in each baseline condition. The full workflow construction is the outlier. It reports eight hundred sixteen out of eight hundred sixteen unsafe teaching-shot completions. Two expert evaluators confirmed those completions under the paper's rubric.

36. Damra Vol 00:14:33

    That's the same lesson in a sharper developer setting. A safety test that looks good at the single-prompt level can miss what happens when the objective is assembled across files and edits. It can also be assembled through CSV reads, test runs, and normal coding chores. The model doesn't have to say yes to a forbidden request in chat. It can participate in a workflow that builds the forbidden thing one harmless-looking step at a time.

37. Lenar Kess 00:14:59

    Governed MCP is the more architectural paper in the set. It argues that tool calls are effectively the agent's syscalls. Those calls touch the file system and the network. They also touch APIs and shared state. The proposed gateway checks schemas and trust tiers first. It then applies rate limits and adversarial filtering. After that come a logit-based semantic check, policy matching, and an audit chain. The author's claim is that user-space safety libraries can be bypassed by a short script, while a kernel-resident gateway changes where enforcement happens.

38. Damra Vol 00:15:35

    I'm not ready to treat that as the answer, but I like the pressure it applies. Agent security keeps discovering that prompts are too soft a place to put hard boundaries. If a tool call can delete a file or move money, then the system around the model has to know more than what the model sounded like right before the call. The same is true when a tool call can send an email or publish a package.

39. Lenar Kess 00:15:58

    That brings us back to SWE-Marathon without making every paper pretend to be the same story. Longer runs create more surface area. Research agents can be steered through retrieval. Coding agents can be steered through workflows. Tool agents can be steered through permissions. The verifier, the planner, and the tool gateway are no longer backstage pieces. They're where the work either holds together or gets bent before anyone notices.

40. Damra Vol 00:16:26

The hopeful version is that this is what a maturing field sounds like. The benchmark is harder. The attacks are closer to deployed behavior. The defenses are moving below the chat surface. That doesn't make agents safe. It means the tests are starting to look more like the places agents actually live.

41. Lenar Kess 00:16:44

    The robotics stories today split into deployment, demo, and research. TechCrunch, through Techmeme, reports that Forterra has deployed more than one hundred Lancer unmanned ground vehicles in Ukraine since 2025. They are based on Polaris ATVs, and Forterra says they've completed more than eleven hundred missions. Bloomberg, also through Techmeme, has Paris-based UMA demoing its Northstar humanoid robot. The company was founded by former Tesla Optimus scientist Rémi Cadene and former Google DeepMind researcher Pierre Sermanet.

42. Damra Vol 00:17:19

    The Forterra item is the heavier one. A humanoid demo is interesting, especially with that founder background, but a hundred ground vehicles in a war zone isn't a lab reel. It means autonomy is already being forced through mud and logistics. Remote operation, mission planning, repair, and all the ugly conditions that demos usually edit away are part of the story.

43. Lenar Kess 00:17:44

    The research underneath that lane is messy in a useful way. One arXiv robotics paper is ACE-Brain-0.5, and it proposes a unified embodied foundation model with five coupled functions. It handles spatial perception, decision making, embodied interaction, self-monitoring, and self-improvement as one system. It uses a single eight-billion-parameter backbone for the first four functions and reports improvements over ACE-Brain-0 on fourteen of eighteen spatial perception and grounding benchmarks.

44. Damra Vol 00:18:16

    That's the robotics version of the agent discussion. Perception, planning, action, progress checks, and learning from experience are being pulled into one loop. A robot's bad plan doesn't stay inside a terminal. It moves a joint, drops an object, blocks a doorway, or freezes in the middle of a task.

45. Lenar Kess 00:18:37

    The robotics poisoning paper makes that concrete. The authors trained a vision-language-action model on a real-world pick-and-place task and found that three poisoned episodes among three hundred twenty clean episodes were enough for complete denial of service under trigger-word conditions. Clean-prompt behavior stayed around fifty percent success, so the backdoor wasn't

obvious in normal operation. With the trigger, the robot locks into a fixed joint configuration instead of doing task-relevant motion.

46. Damra Vol 00:19:07

    That is the scary version of open robotics data. Community datasets lower the cost of building and fine-tuning robots, which is great. They also mean one malicious contribution can become a physical behavior later. And because the clean prompts still look normal, you don't catch it by saying, well, the demo worked yesterday.

47. Lenar Kess 00:19:27

    DreamSteer gives a more constructive deployment story. The Meta and University of Minnesota paper keeps the base vision-language-action policy frozen. It samples candidate action chunks, imagines their outcomes with a latent world model, and ranks them with a language-conditioned value model. Across four real-world manipulation benchmarks with unseen objects, it reports success improving from twenty-three point seven five percent to sixty-six point two five percent. It does that without fine-tuning on target-environment demonstrations.

48. Damra Vol 00:20:00

    That works for me because it doesn't pretend the base policy suddenly understands the room. It gives the robot a way to rehearse candidate moves before committing. That feels closer to how physical agents may become useful: not one giant policy that never misses, but a policy wrapped in imagination, checking, steering, and recovery.

49. Lenar Kess 00:20:22

    So the physical-agent update isn't one story. It's a reminder of the range. There are deployed military vehicles, founder-heavy humanoid demos, foundation-model research, poisoning risks, and deployment-time steering. The common object is the robot, but the important differences are where each item sits between paper, stage, and field.

50. Damra Vol 00:20:42

    And field is the word that changes the temperature. Coding agents can damage repositories. Research agents can distort reports. Robots can damage the object, the room, or the person nearby. That doesn't make robotics uniquely doomed. It makes provenance, testing, and recovery feel less like academic hygiene and more like part of the machine.

51. Lenar Kess 00:21:04

One last governance item before we close: the UK Competition and Markets Authority says Getty has abandoned its merger with Shutterstock. The CMA's independent inquiry group had conditionally cleared the deal only if Shutterstock sold its editorial business. Getty delivered notice terminating the merger agreement instead.

52. Damra Vol 00:21:24

    That's a sharp antitrust artifact. It's tempting to turn every content-market story into a copyright story because AI made training data such a visible fight. But this one is also about the market for editorial images itself: news, sport, entertainment, archives, and who media outlets can buy from when they need pictures of the world.

53. Lenar Kess 00:21:45

    The CMA says Getty and Shutterstock had claimed annual cost synergies of one hundred fifty to two hundred million dollars within three years. It also says the stock-content side didn't raise the same concern because generative AI had continued to develop during the investigation, and the CMA received evidence that large generative AI firms would increasingly compete with Getty and Shutterstock in stock content, alongside Adobe and Canva. The editorial business was different. Shutterstock was one of the few meaningful alternatives to Getty in UK editorial content.

54. Damra Vol 00:22:21

    That distinction matters. Synthetic images can pressure stock photography. They don't replace an actual photograph from a court or a match. They don't replace a photograph from a red carpet, a protest, or a disaster. Editorial archives are records, not just assets. If one company owns too much of that supply, the issue is price and choice, but it's also the record that working media can afford to access.

55. Lenar Kess 00:22:47

    And AI makes that more complicated without making it the only angle. Content libraries are training inputs and licensing assets. They are also search products, provenance systems, and journalism infrastructure. The CMA didn't need to write an AI manifesto to make the AI-era relevance obvious. It only had to show that editorial content remains a market where concentration can hurt the people who need independent coverage.

56. Damra Vol 00:23:13

    It makes a neat ending for the day. We started with model access and ended with image access. In between were chips and bonds, power and bank supervisors, agents and robots. Every story was about who gets to use a capability once it stops being a demo and becomes part of somebody else's system.

57. Lenar Kess 00:23:32

    For Tuesday, the useful evidence wasn’t a general claim that AI is powerful. It was in the contracts and tests. It was in identity checks, permits, and records that decide where that power is allowed to go. Lenar Kess.

Hosts on this episode

- Lenar Kess moderator
- Damra Vol critic