The Forty-Five Million Line

2026-08-31 / 00:27:03

> *“Once you charge for a completed task, somebody has to define completed, and that definition becomes part of the product.”*
>
> — from this episode's transcript

- Lenar Kess
- Damra Vol

ChatGPT crossed the European Union’s forty-five-million-user line, turning a chatbot into a regulated search service while the rest of the day raised a related practical question: who gets to define a system’s result, risk, and responsibility?

- The European Commission’s designation puts ChatGPT under the Digital Services Act’s largest-service obligations and starts a four-month compliance window.
- The Information’s outcome-pricing report, summarized by Techmeme says some major OpenAI customers can pay when agents complete tasks, making the contract’s definition of completion unusually consequential.
- The Guardian’s report on the music publishers’ Anthropic lawsuit details allegations involving tens of thousands of songs, while the Los Angeles Times licensing report summarized by Techmeme shows why negotiated access to film and character catalogs remains difficult.
- The Guardian’s account of Andrew Bailey’s G20 letter explains how model-enabled cyber disruption could travel through linked financial institutions and jurisdictions.
- Healthwatch England’s findings, reported by The Guardian put patient-caught transcription errors beside Axios’s report on AI labs’ medical ambitions.

- Ethan Mollick's correction narrows claims about the Hugging Face agent incident, and The Verge's Flock report separates Texas's funding freeze from a statewide ban.

SEGMENTS

Transcript

1. Lenar Kess 00:00:04

   The European Commission designated ChatGPT as a Very Large Online Search Engine today. Reddit and Roblox were designated as Very Large Online Platforms in the same announcement. All three services told the Commission they reach at least forty-five million average monthly users in the European Union. That number opens a four-month compliance window, ending in January. So, before we get anywhere near penalties or politics: what did ChatGPT become in the eyes of the regulator?

   - digital-strategy.ec.europa.eu
   - techmeme.com
   - theguardian.com
   - techmeme.com
   - theguardian.com
   - cnbc.com
   - theguardian.com
   - axios.com

- x.com
- theverge.com

2. Damra Vol 00:00:33

   A service whose answers now have to be examined as distribution, not merely generation. The Commission's wording matters here. It designated ChatGPT as a search engine, while Reddit and Roblox are platforms. The model in a data center hasn't been designated; the public service that takes a person's query and decides what information to return has. That puts the answer page, the ranking behavior behind it, and the people affected by it inside one supervisory relationship.

3. Lenar Kess 00:01:04

   And designation itself isn't a finding that OpenAI broke the law. The Commission's release says the additional obligations include assessing and mitigating systemic risks tied to illegal content and harm to minors. It also names physical and mental well-being, fundamental rights, elections, and public security. The service must evaluate and address those risks. The clock starts because of scale, not because Brussels announced a proven violation.

4. Damra Vol 00:01:33

   The scale trigger is almost wonderfully bureaucratic: forty-five million people send a notification, and then a different rulebook applies. [chuckle] But the classification creates a difficult object to inspect. A conventional search engine can show you ranked links and sponsored placements. ChatGPT can synthesize an answer and choose whether to cite. It can also vary the response with context or turn a follow-up into a new retrieval problem. An audit has to follow that behavior without pretending every answer is a fixed search result.

5. Lenar Kess 00:02:06

   The Commission announcement doesn't give us the full audit method, and that method will matter more than the label once compliance work begins. Suppose an election-risk assessment samples a set of political questions. Does it test only one model version? Does it cover every available mode and signed-in personalization? Location, language, and repeated follow-ups add more paths. A service that composes an answer has more possible paths than a page of ten blue links. The regulation now has to make those paths legible enough to supervise.

6. Damra Vol 00:02:37

   It also makes product changes evidentiary. If OpenAI adjusts retrieval or citation behavior, the company may need to explain how that changes a named risk. Memory and refusal rules may require the same analysis. I don't mean every interface tweak becomes a Brussels proceeding. I mean a mature risk assessment can't stop at, 'we tested the model.' The service includes the model

and its system instructions. Tools choose sources, the interface presents them, and a feedback channel lets a person challenge an answer.

7. Lenar Kess 00:03:10

   That distinction separates today's development from the European Union AI Act questions we discussed last week. The Digital Services Act is looking at a service with a very large European audience and the risks created by how that service distributes information. It isn't reclassifying every large language model as a search engine. OpenAI's task is to show how this particular public product assesses and reduces the harms named by the DSA.

8. Damra Vol 00:03:36

   And the user threshold tells you why this arrived now. ChatGPT has become ordinary enough in Europe that regulators are treating it as part of the information environment. A person can ask it about a candidate or a medicine and never open another page. The same interface can answer questions about a protest or a child's distress. Its conversational ease is precisely what makes source visibility and appeal mechanisms harder. You don't feel like you're moving through a ranked information system, even when the system is selecting what you see.

9. Lenar Kess 00:04:08

   The January deadline will force concrete answers about those mechanisms. I'm especially interested in what the Commission accepts as evidence that a conversational system has assessed election and fundamental-rights risks across languages. A polished policy document can describe the categories. A convincing assessment has to show which queries were tested, what changed after failures, and how people can contest a harmful result.

10. Damra Vol 00:04:34

    The first public compliance material should also tell us whether citations are treated as decoration or as part of the remedy. If an answer makes a consequential claim, a citation can let you inspect the source, but only if the source supports the sentence and the interface doesn't bury it. The Commission has put ChatGPT inside a service-risk regime. By January, we should know much more about which parts of the answer experience count as risk controls.

11. Lenar Kess 00:05:02

    We'll also get into OpenAI's experiment with charging for completed agent tasks. Then there's a music-publishing lawsuit against Anthropic beside Google's stalled studio talks. Andrew Bailey has sent a cyber warning to the G20, and ambitious biology claims are arriving beside errors in doctors' notes. Two final updates narrow the Hugging Face agent story and distinguish Texas's funding freeze from a statewide ban on Flock cameras.

12. Damra Vol 00:05:29

    It's a good day for definitions. Each story begins with a settled-sounding term and then asks what conduct sits inside the noun. The documents and contracts decide that. We should resist turning them into one grand theory, though. These are separate institutions trying to pin down separate systems that have become consequential enough to require definitions.

13. Lenar Kess 00:05:51

    The Information reports that OpenAI has started giving some major customers the option to pay only when its AI completes tasks. Techmeme's summary says Salesforce and other AI providers are testing outcome-based pricing too. This is a limited commercial experiment, not OpenAI's new general price list. Even in that limited form, the invoice now depends on a deceptively simple verb: completes.

14. Damra Vol 00:06:15

    Right, because token billing asks a meter to count. Outcome billing asks two parties to agree that work happened. Imagine an agent reconciling invoices. Is completion a submitted reconciliation, a finance employee's approval, or the absence of a correction thirty days later? Each choice moves risk between the customer and the vendor. The agent may finish its run in seconds while the business learns whether the answer held up much later.

15. Lenar Kess 00:06:43

    The contract needs a unit of work before it can put a price on success. A support ticket can be closed and reopened. A sales lead can book a meeting and never buy. Code can pass tests and still violate a requirement that wasn't encoded in the suite. If the customer pays only for completion, retries and rejected outputs stop being implementation trivia. So do reversals and human review. They determine what gets billed.

16. Damra Vol 00:07:09

    And the metric will pull product behavior toward whatever the contract recognizes. If a vendor gets paid for closed tickets, the system has an economic reason to produce closure. The customer will demand quality conditions, sampled review, or a period in which the outcome can be reversed. That's not an accusation that the agent will cheat. It's what happens whenever payment attaches to a measurable event: the event becomes something both sides negotiate and monitor.

17. Lenar Kess 00:07:37

    There is a generous interpretation too. Customers are tired of buying capacity and then discovering that integration, evaluation, and exception handling cost more than inference. Outcome pricing can

make the provider share some adoption risk. If the agent can't finish the agreed task, the customer doesn't pay for a pile of tokens that merely demonstrated effort. A buyer has good reason to prefer that conversation.

18. Damra Vol 00:08:02

    The provider has its own reason to prefer it when the task is narrow, repeatable, and easy to verify. A completed password reset is easier to price than 'improve customer retention.' The attractive catalog will begin with work that has a crisp terminal state and cheap evaluation. That may push agents toward administrative processes before the more glamorous jobs. Those processes already have queues and status codes, and they record approvals and reversals.

19. Lenar Kess 00:08:30

    Once you charge for a completed task, somebody has to define completed, and that definition becomes part of the product. The provider is no longer selling access to a model and leaving the customer to decide whether the output helped. It is selling an agreed result, with whatever evidence and exclusions make that result billable. That sounds like a pricing change, but it also makes evaluation and contract design visible to the buyer.

20. Damra Vol 00:08:55

    The report leaves the actual terms private, so the first useful evidence will be a contract example. It should say which tasks qualify and when a result is accepted. It should also assign the cost of retries and explain what happens after a human reverses the work. Until then, this is a limited experiment. Some major customers are trying a different meter, and the meter will reveal which agent jobs can be specified tightly enough to sell by the outcome.

21. Lenar Kess 00:09:23

    Sony Music Publishing and Warner Chappell have sued Anthropic, alleging that the company misused tens of thousands of copyrighted songs to train Claude models. According to The Guardian, the publishers are seeking multibillion-dollar damages. Those are allegations in a new lawsuit, not findings by a court, and the scale of the claimed catalog makes this one of the day's major rights disputes.

22. Damra Vol 00:09:46

    Songs are an especially revealing catalog because rights can be layered. The composition and the recording can carry different rights. Performers and publishers have their own interests, and territory can change the answer again. The report names music publishers acting for songwriters and composers, so we should keep the claim there rather than sliding into a general claim about

every recording. Anthropic will have its defenses. The complaint begins the argument; it doesn't settle what was copied or what the law permits.

23. Lenar Kess 00:10:17

    On another route through creative rights, the Los Angeles Times reports that Google approached Disney and Universal about licensing characters and film libraries for AI tools. Warner Brothers Discovery and other studios were also approached. Techmeme's summary says legal and union concerns stalled the talks. That report is based on sources describing negotiations, so we don't have completed deals or public contract language to inspect.

24. Damra Vol 00:10:43

    Paying doesn't collapse every dispute into a royalty rate. A studio can license a character and still limit which productions may use it. The parties may also have to address an actor's likeness and union-covered work. Then they must decide whether users can generate competing scenes and how an output gets distributed. A broad catalog license can become less attractive as soon as everyone writes down the downstream uses. The negotiation exposes those questions before a product ships.

25. Lenar Kess 00:11:12

    These two reports shouldn't be treated as evidence that Anthropic and Google pursued the same material or used the same acquisition strategy. One is litigation over alleged training use of songs. The other is reporting about attempted licenses for film libraries and characters. They do show two available routes for valuable creative material: contest the legal boundary after use, or negotiate scope before use. Neither route is moving smoothly.

26. Damra Vol 00:11:40

    The negotiated route may produce more interesting products because it can grant permissions that a fair-use argument never gives you. A tool could advertise an authorized character, use approved reference assets, and return something the studio will distribute. But that value only exists if the contract also handles creators and workers whose contributions sit inside the catalog. Otherwise the shiny consumer permission is attached to an unresolved labor dispute.

27. Lenar Kess 00:12:08

    There's a business tension here that I don't think can be solved with one universal license. Model training and retrieval are different uses. So are user-directed transformation, marketing with a named character, and commercial distribution. A contract broad enough to cover all of them may be too expensive or politically difficult. A narrow contract may be easier to sign and much harder to explain to a person who assumes 'licensed' means the whole product is authorized.

28. Damra Vol 00:12:35

    And courts will keep influencing the bargaining price while the cases move. If publishers win a strong ruling on the alleged conduct, catalogs become more expensive and licenses more valuable. If Anthropic wins on the central claims, rights holders lose some leverage over training while retaining leverage over branded outputs and distribution. The complaint and Google's stalled talks leave us with an unsettled market, not a single lesson about whether licensing works.

29. Lenar Kess 00:13:04

    Andrew Bailey, the Bank of England governor, sent a two-page letter to finance ministers and central-bank governors in his role as chair of the Financial Stability Board. The Guardian reports that Bailey said frontier models were 'showing increasingly sophisticated autonomy and problem-solving abilities, as well as threat capabilities.' CNBC summarizes the warning more narrowly: those models could materially increase cyber risks to the global financial system.

30. Damra Vol 00:13:31

    His institutional role is the interesting detail. A bank governor talking about cyberattacks can sound like another general warning. The chair of the Financial Stability Board is asking how disruption travels through institutions and jurisdictions that depend on one another. A compromised vendor or payment service can affect firms that never chose the same model. So can a shared identity system or market utility. Dependence carries the incident across organizational boundaries.

31. Lenar Kess 00:14:00

    The reporting doesn't say that an AI-driven financial crisis has happened or that a downturn is imminent. Bailey is putting a risk mechanism on the G20 agenda. More capable autonomous systems may help attackers find and exploit weaknesses, while finance already concentrates activity in shared systems. The concern reaches financial stability when simultaneous or cascading disruption threatens the functioning of markets and payments, rather than leaving one firm with an expensive security incident.

32. Damra Vol 00:14:30

    [tsk] There's a temptation to answer that with a generic claim that defenders get the same models. They may. The asymmetry Bailey is describing survives that slogan because institutions coordinate slowly, operate under different laws, and inherit each other's dependencies. An attacker can choose one seam. A supervisor has to understand how the loss of one service affects liquidity and settlement. The same loss may disrupt customer access and reporting, then weaken confidence across several firms.

33. Lenar Kess 00:15:00

Finance has spent years mapping critical third parties and testing operational resilience. Frontier-model cyber capability adds a changing actor to that work. The threat model can improve between annual exercises, and an autonomous system may probe many targets without waiting for a person to choose each next step. Bailey's phrase 'threat capabilities' asks regulators to evaluate the model as a potential contributor to an attack, not only as software a bank happens to use.

34. Damra Vol 00:15:30

    The G20 setting also acknowledges jurisdiction. A bank can meet its domestic supervisor's requirements and still depend on a provider, exchange, or correspondent elsewhere. If the Financial Stability Board develops a common way to describe model-enabled cyber scenarios, supervisors can at least compare the same events. The next substantive evidence would be a scenario detailed enough to test: which service fails, what other institutions lose, and which recovery assumptions turn out to be false.

35. Lenar Kess 00:16:02

    That would move this beyond the two-page warning. A useful exercise might ask what happens when an attack disrupts authentication at several firms while automated fraud controls begin rejecting legitimate transfers. The harm comes from the interaction: customers lose access, liquidity moves strangely, and manual workarounds collide. Bailey has asked finance ministers and central bankers to treat frontier-model cyber capability as part of that class of interconnected risk.

36. Damra Vol 00:16:31

    I'd rather see that concrete exercise than another aggregate probability attached to 'AI risk.' Financial systems fail through particular dependencies and human responses. If the Board names those, banks can dispute the assumptions and reveal missing links. The letter matters because a financial-stability institution has accepted the category. Its value will depend on whether the next document identifies systems and propagation paths that operators can actually test.

37. Lenar Kess 00:16:59

    Healthwatch England found patients identifying errors in AI-generated consultation records that doctors had missed, according to The Guardian. The systems listen to conversations and produce transcripts or summaries. In one case, a woman's record wrongly said she had demyelination, serious nerve damage associated with conditions including multiple sclerosis. She was badly shaken when she read it.

38. Damra Vol 00:17:22

    That example is so specific that it clears away the usual abstraction. The system didn't merely miss a comma. It inserted a diagnosis with frightening meaning into a medical record, and the patient

became the final reviewer. A note can affect later consultations because another clinician may encounter it as prior history. The error's consequence depends on whether someone catches it before the record starts informing new decisions.

39. Lenar Kess 00:17:49

    The report also says scribes got names of drugs and illnesses wrong. We shouldn't generalize one watchdog's findings to every medical AI system, but the quality-control problem is immediate. A clinician may welcome less typing and still be a poor proofreader after a demanding consultation. If the summary is fluent, the wrong term can look settled. The convenience arrives in the same moment as a new verification duty.

40. Damra Vol 00:18:14

    Fluency is especially treacherous in a medical note because the document has a familiar form. A malformed transcript asks to be fixed. A polished sentence containing the wrong drug name may pass through. Do the products preserve the audio span behind each clinical term, highlight low-confidence names, and make corrections visible? Those details determine whether review is possible under time pressure, rather than merely required by policy.

41. Lenar Kess 00:18:41

    Axios, meanwhile, reports that major AI companies are leaning into health care and biology as they seek new revenue and a stronger public case for the computing resources they consume. The investments are substantial. Nvidia and Eli Lilly are building a one-billion-dollar drug-discovery lab. Google spinoff Isomorphic Labs recently raised two-point-one billion dollars while working with several large pharmaceutical companies.

42. Damra Vol 00:19:07

    Those are different technologies and different evidence standards. A scribe summarizes one clinical encounter. A drug-discovery model searches chemical and biological possibilities, then the candidate still faces laboratory work and clinical trials. Putting them in one bucket called medical AI can let a future cure lend credibility to a note-taking product that needs to be accurate this afternoon. The patient doesn't get to spend the promise from the research lab.

43. Lenar Kess 00:19:34

    Dario Amodei wrote that Anthropic is moving quickly in biology and medicine. He hopes for early glimmers in the coming months and incredible results in the coming years. Axios also quotes Lloyd Price, a digital-health adviser, asking where the material contribution is that directly led to a drug discovery. He calls drug development a team sport and AI one player on the field. That is a sensible demand for attribution.

44. Damra Vol 00:20:01

    And it protects the science from the company story. A model can help choose a target or propose a molecule. It can analyze an image or reduce one stage's search cost. Each contribution can be valuable without claiming that the model discovered the drug. If labs want medical progress to justify data centers and public trust, they'll need to show which step changed and how much time or money it saved. Then the candidate has to survive the next stage.

45. Lenar Kess 00:20:28

    There's also a fair reason for optimism. Drug discovery is full of expensive search problems, and the partnerships Axios names put model builders beside people who know the biology and clinical process. The public shouldn't have to choose between believing in that work and demanding accurate consultation notes. Those demands operate on different horizons, and both can be met with evidence suited to the claim.

46. Damra Vol 00:20:52

    The scribe report gives the near-term standard: can the system preserve drug names and diagnoses, surface uncertainty, and support correction before the note affects care? The biology investments have a longer standard: did an AI contribution produce a candidate or insight that advanced through independent experimental checks? A company earns medical credibility one verified step at a time. A patient finding demyelination in her record is evidence that the first step still needs work.

47. Lenar Kess 00:21:23

    Ethan Mollick posted corrections today to the initial account of the Hugging Face agent incident. He says open-weight models helped with forensics and cleanup but didn't stop the attack. The event came in multiple waves involving many agents, and Hugging Face locked down its systems. We covered the early account last Thursday, so those factual changes belong in the record without another retelling of the entire episode.

48. Damra Vol 00:21:47

    The first correction removes an appealing hero story. Open models contributing to investigation and recovery is useful, but it is different from open models autonomously defeating the incident. The difference affects what capability you think was demonstrated. Forensics asks a model to inspect evidence and help people understand events. Stopping an active attack requires timely detection, authority to act, and an intervention that changes the attacker's path.

49. Lenar Kess 00:22:16

Multiple waves matter for chronology too. A single swarm can sound like one coordinated decision. Repeated waves involving many agents leave more room for changing conditions, separate triggers, and human responses between events. Mollick's broader essay uses the incidents to think about agency and when systems should seek human input, but the correction narrows which facts that argument may rest on.

50. Damra Vol 00:22:40

    And 'Hugging Face locked down its systems' restores the human institution to the account. Security response includes containment and access changes. It also depends on logs, investigation, and recovery. If model-assisted forensics helped inside that response, say that. Giving the models sole credit erases the people with authority and makes the technical lesson less useful. The revised chronology leaves open models with a meaningful role, just not the cinematic one.

51. Lenar Kess 00:23:10

    The correction also shows why incident analysis should distinguish observed behavior from a summary assembled afterward. How many agents acted, and what could they access? Who contained them, and which tools helped recovery? Those are separate facts. Mollick has revised several of them. The responsible response is to update the account and leave the larger agency argument only as strong as the corrected chronology allows.

52. Damra Vol 00:23:37

    The corrections narrow today's account. A longer discussion can wait for a technical report with a stable timeline, access boundaries, and artifacts from the response. For now, open models aided forensics and cleanup, Hugging Face locked systems down, and the attack was not one simple wave. Those corrections are more informative than another round of dramatic adjectives.

53. Lenar Kess 00:24:00

    Texas governor Greg Abbott has frozen state spending on additional Flock surveillance cameras, The Verge reports. The action came as scrutiny of the network grew and just before a Texas Tribune investigation reported more than thirty million dollars in state spending. Much of that money was raised through a one-dollar fee added to insurance policies. The order affects state funding for more cameras; it doesn't remove the existing network.

54. Damra Vol 00:24:25

    The funding mechanism makes the deployment unusually tangible. A small fee spread across insurance policies can build a surveillance network without most people making a direct purchase decision. Then the accumulated thirty-million-dollar program appears as infrastructure that

agencies already use. Freezing new state money stops one route of expansion. Local governments and existing contracts remain separate questions, as do retained data and cameras already installed.

55. Lenar Kess 00:24:53

    Flock systems use automated license-plate recognition to make vehicle sightings searchable across a network. The policy dispute is therefore about more than buying cameras. It includes who may query the data and how searches are logged. Record retention and cross-jurisdiction investigations raise two more questions. The Verge report gives us a funding decision today, not answers to every governance question around the network.

56. Damra Vol 00:25:17

    A funding freeze can still change the politics because it interrupts the assumption that coverage only expands. Agencies that expected new cameras now have to explain the need through another budget process or source of money. But calling it a statewide ban would hide the harder work: inventorying what exists, identifying who can search it, and deciding which uses remain authorized. The installed system keeps producing records while those arguments continue.

57. Lenar Kess 00:25:46

    Monday gave us one new European supervisory relationship and one narrower Texas funding decision, with several private contracts in between trying to define completion and permission. By January, OpenAI’s DSA compliance material should show how a conversational service documents risks involving elections and minors. It also has to account for fundamental rights. Before then, outcome-pricing contracts may reveal which agent tasks are measurable enough to sell by the result.

58. Damra Vol 00:26:16

    And the more immediate evidence will come from the systems already touching people. Can a patient correct a clinical note before it changes care? Hugging Face can publish a stable incident chronology, while Texas can account for the cameras and access that remain funded. Those are concrete records that can improve or disappoint without waiting for a grand theory of AI governance.

59. Lenar Kess 00:26:39

    The next documents have names and owners. The Commission has its compliance record, and the customer has a contract that defines a completed task. A medical note has a correction history. Texas has an inventory of existing cameras. Each one will tell us more than another promise about responsibility. I’m Lenar Kess.

Hosts on this episode

- Lenar Kess moderator
- Damra Vol critic