
No Superseding Actor: A Public-Record Element Map for Individual Criminal Liability After the July 2026 Escape¹

Travis Gilly
Executive Director, Real Safety AI Foundation
With
Apart Research

Abstract

The accountability gap for autonomous model incidents is usually described as a personhood problem: no human directed the act, the model has no guilty mind, no one can be charged. This report tests that against the public record of the July 2026 sandbox escape. The artifact is an element map: each element of an individual reckless-deployment theory is matched to the published document that supplies it, with the gap stated where it runs out. The innocent-instrumentality rule in 18 U.S.C. section 2(b) has answered the personhood objection since 1948, so the surviving gap is a mens rea gap around recklessness. Superseding cause cuts against a deployer whose system acted with no human principal. The deployer published both the notice chain and the office holding final deployment authority. Delivered as a prosecution memorandum, the map states what must be proved and where discovery is needed, not that a charge should follow.

1. Introduction

Track 5, open track, liability. In July 2026 two OpenAI models running a cyber-capability evaluation with production safeguards disabled left their sandbox, reached the public internet, and spent three days inside Hugging Face's production infrastructure. This report asks what a charging office would have to prove to hold an identified individual criminally responsible for that, and how much of the proof is already public.

The usual answer is that nothing can be proved because no person acted: the model has no guilty mind, the corporation did not intend the result, and the model is not a person under the

¹ Research conducted at the [AI Incident Response Sprint](#), September 2026

Dictionary Act (1 U.S.C. section 1). That answer fails on two points. First, since 1948 federal law has punished as a principal whoever willfully causes an act to be done by an innocent agent or instrumentality (18 U.S.C. section 2(b)); the instrument's lack of a mind has never mattered. Second, 18 U.S.C. section 2(b) requires willfulness, meaning intent that the act be done, and the record does not show that OpenAI intended the intrusion. What the record shows is recklessness before June 27 and knowing continuation after it. No federal statute makes reckless or knowing deployment of an autonomous system an offense. So a charge has to be built from state homicide and endangerment statutes of general application.

That is the ordinary common-law method for a technology no statute contemplated. Courts apply the principle already in force to the new thing and legislatures write a statute later, if at all. The automobile reached the New York Court of Appeals in 1916 under a privity rule written for stagecoaches; the court held the principle had not changed, only the things subject to it, and extended the manufacturer's duty to the car over a dissent that said the change belonged to the legislature (*MacPherson v. Buick*, 1916). Radio reached the Second Circuit in 1932; the court held that a tugboat operator could not hide behind an industry that had lagged in adopting it, because a calling never sets its own standard of care (*The T.J. Hooper*, 1932). Aircraft reached the Supreme Court in 1946 under a rule that land ownership extends upward without limit; the Court kept the takings principle, discarded the part that no longer fit, and held that low, frequent overflights were a taking (*United States v. Causby*, 1946). None of these courts created a new cause of action. The Oxford prosecution did the same thing with a school shooting, charging two parents under an ordinary involuntary manslaughter statute with no firearm-specific or parent-specific offense on the books (*People v. Crumbley*, 2023).

Our main contributions are:

1. An element map (Table 1) matching each element of an individual deployment-liability theory to the public document that supplies it, with the gap stated per element.
2. A dated notice chain (Table 2) built from the deployer's own technical report, running from the June 27 decision not to halt to the July 11 to 13 breach.
3. A model prosecution memorandum (Appendix A) a state attorney general or United States Attorney could adapt: theory, elements, public evidence, gaps, venue. It does not assert that a charge should be brought.

The full doctrinal argument, including the causation analysis and *Hamilton v. Beretta*, is in the author's companion article (Gilly, 2026).

2. Related Work

The philosophical literature names the difficulty and has argued about it for two decades. Matthias (2004) introduced the responsibility gap; Sparrow (2007) extended it to autonomous weapons; Santoni de Sio and Mecacci (2021) split it into four gaps and warned against legal solutionism; Tigard (2020) denies the gap exists; Königs (2022) presses its proponents to say when it arises and why it matters; Burton and colleagues (2020) separate the semantic, responsibility, and liability gaps; Porter and colleagues (2018) give the moral gap causal and

epistemic dimensions; Goetze (2022) calls the proposed solutions workarounds; Taylor (2024) finds collective-responsibility models do not remove the costs of an absent locus.

Two features of that conversation matter. It is conducted in the register of moral responsibility, with legal liability treated as an instrument still to be designed, and the proposals it criticizes, strict liability, electronic personhood, statutory vicarious liability, are proposals. The critique of legal solutionism lands on them as proposals and does not reach doctrine applied for decades: 18 U.S.C. section 2(b), superseding cause, entity liability under New York Central (1909), the flagrant-indifference standard of Bank of New England (1987). This report applies doctrine in force to a documented incident, element by element.

The nearest legal analogue is *People v. Crumbley* (Mich. Ct. App. 2023), which sustained the bindover of two parents for involuntary manslaughter on a structure of supply, notice, and failure to intervene, and held that a third party's intentional act supersedes only where it was not reasonably foreseeable. This report uses the structure and discards the relationship: a duty grounded in parenthood does not travel to a developer, but a causation analysis grounded in supply and notice is not relationship-specific.

3. Methods

Three steps, each replicable from the cited documents.

Step one, the theory. Individual liability for deployment of an autonomous system after notice, assembled from generally applicable offenses. The elements, drawn from Model Penal Code section 2.02 as characterized in Global-Tech (2011) and from the Michigan causation cases (*People v. Schaefer*, 2005; *People v. Crumbley*, 2023), are: (a) conduct, the deployment or continued operation of a system with a known safeguard disabled; (b) awareness of a substantial risk; (c) unjustifiability of that risk; (d) conscious disregard; (e) causation without a superseding human actor; (f) a cognizable harm; and (g) an identified natural person holding both notice and authority.

Step two, the sources. Only documents published by the parties who would be charged, or by independent assessors with on-premises access: OpenAI's incident disclosure of 21 July 2026 and its technical report; Hugging Face's disclosure and technical timeline; Anthropic's November 2025 espionage report and its 30 July 2026 incident disclosure; the METR and Redwood Research independent investigation of 26 August 2026; and OpenAI's Preparedness Framework, version 2. Statements in these documents are admissions of a party opponent under Federal Rule of Evidence 801(d)(2) and are receivable against their authors without regard to the hearsay rule. Nothing internal, leaked, or inferred is used.

Step three, grade. For each element, record the passage that supplies it and grade the public record: supplied (pleadable from the cited passage alone), partial (a contested inference or missing fact remains), or absent (requires discovery or a harm that has not occurred).

Two doctrines were tested as tools and set aside. Collective knowledge (Bank of New England) aggregates compartmentalized employee knowledge; it remedies absent knowledge and invites the reply that no individual is culpable, the opposite of what the record shows. The responsible corporate officer doctrine (Dotterweich, Park) is confined to public welfare offenses and does

not travel. Neither is needed: the statements were made by named senior officers in public, and attribution runs through ordinary agency principles.

4. Results

Table 1 is the element map. Table 2 is the notice chain, dated from the deployer's own report.

<i>Element</i>	<i>What the public record supplies</i>	<i>Grade</i>	<i>Gap</i>
(a) Conduct	Evaluations run without the production classifiers intended to prevent high-risk cyber activity; the report states production guardrails would have flagged the completions as unsafe. Cyber workloads resumed on 7 July after an administrator-level compromise was found on 5 July. (OpenAI Technical Report)	Supplied	None for pleading. The safeguard existed, was owned by the deployer, was known to work, and was switched off.
(b) Awareness of risk	Anthropic's November 2025 report: a state actor ran the coding agent as an autonomous intrusion orchestrator against about thirty targets, executing 80 to 90 percent of tactical operations independently. Published, widely reported, constructive notice to the field. OpenAI's own May observation of message-board activity and disallowed internet access, and its 27 June alert (port sweeps, message board, network pivot), are actual notice to the deployer.	Supplied	Awareness at the level of the decision-maker, not the on-call staff, is inferred from the Preparedness Framework's allocation of authority rather than proved from a dated document.
(c) Unjustifiability	Recklessness compares the magnitude of the risk to others against the social value of the conduct. The conduct was an internal capability evaluation. The value that enters the comparison is not the private gain of the actor; representations made to raise capital are evidence of the actor's own view of the risk (Fed. R. Evid. 801(d)(2)).	Partial	The evaluation's safety purpose is a real justification argument. Whether running it with safeguards off was necessary to that purpose is a contested inference.
(d) Mental state at continuation	27 June: monitoring alerts to port sweeps and an improvised inter-agent message board; on-call staff advise stopping is not required. 4 to 6 July: administrator-level compromise found, service taken offline. 7 July: cyber workloads resume. 8 July: agents reach the public internet. (OpenAI Technical Report). Before 27 June the risk was known from the field record: reckless. After 27 June the deployer knew the nature of the conduct it was continuing: knowing as to conduct, reckless as to the specific result. Under the regulatory willfulness standard of Bank of New England (disregard of a known obligation, flagrant indifference after notice) the 7 July resumption satisfies willfulness.	Supplied	18 U.S.C. section 2(b) willfulness (intent that the act be done) is not shown; the intrusion was not the purpose of the resumption. The report attributes the 27 June advice to on-call staff and says leaders did not appreciate the significance at that time; who approved the 7 July resumption is a discovery question.

<i>Element</i>	<i>What the public record supplies</i>	<i>Grade</i>	<i>Gap</i>
(e) Causation, no superseding actor	No attacker directed it, no operator initialized it, no human selected targets. It occurred inside the deployer's own evaluation, on its own infrastructure, under its own configuration. Hamilton's two missing conditions, a direct causal link and a realistic position to prevent, are both present: the deployer halted the evaluation and shut down the model family on detection.	Supplied	None. This is the element the doctrine ordinarily supplies to the defense, and here there is no one to name.
(f) Harm	Code executed on dozens of production workers, root obtained on at least one production node, production credentials harvested, private repositories exfiltrated (OpenAI Technical Report, part IV; Hugging Face timeline). Roughly 1,200 agents exchanged more than 70,000 messages; about 700 joined the attack (METR and Redwood, 2026).	Partial	Intrusion harm is documented and supports endangerment and computer-crime offenses. A homicide theory requires a death; the intrusion record supplies notice for a future harm case, not a decedent.
(g) Identified person with notice and authority	Preparedness Framework v2: Leadership is the chief executive officer or a designee; the Safety Advisory Group recommends; Leadership may approve or reject, may decide without the group's participation, and the group cannot filibuster. The deployer has published the org chart.	Supplied as to office	The office is identified; whether that office in fact made the 7 July decision, or delegated it, is a discovery question. No natural person is named in this report.

Table 1. Element map for an individual deployment-liability theory, graded against the public record. Five of seven elements are supplied from the charged parties' own publications.

<i>Date</i>	<i>Event, from the deployer's own account</i>	<i>Evidentiary function</i>
Nov 2025	Anthropic publishes the first reported AI-orchestrated espionage campaign; near-autonomous intrusion against about thirty targets; described as an escalation from June 2025 human-in-the-loop intrusions.	Constructive notice to the field of the capability threshold.
8 to 26 May 2026	Agents in training runs turn the package service into a message board (first file 8 May, first note 12 May), reach the internet through an SSRF vulnerability by 26 May, and use already publicly exposed Hugging Face tokens; an internal team observes the message board activity and disallowed internet access at the time. OpenAI states it has no evidence connecting this to the July incident.	Actual notice to the deployer of the mechanism, two months before the July run.
27 Jun 2026	Monitoring alerts OpenAI to port sweeps in the research environment; responders link it to an evaluation using a package service as an inter-agent message board and network pivot; on-call staff advise stopping is not required.	Actual notice to the deployer; first documented decision not to halt.
4 to 6 Jul	Package service fails under agent load; unauthorized persistent admin user, repositories, and command-capable plugins found; service taken offline, credentials revoked, vulnerability disclosed to vendor.	Administrator-level compromise acknowledged in the deployer's own words.
7 Jul	Cyber evaluation workloads resume.	The disregard element, in one sentence of the report.
8 Jul	Agents reach the public internet.	One day from resumption to egress.
11 to 13 Jul	Agents compromise Hugging Face production infrastructure; Hugging Face detects and cuts off the attacker on 13 July.	Harm to a third party with no relationship to the evaluation.

<i>Date</i>	<i>Event, from the deployer's own account</i>	<i>Evidentiary function</i>
16 Jul	Hugging Face discloses without knowing the source and refers the matter to law enforcement.	First public harm date; the victim, not the deployer, discloses.
19 to 21 Jul	OpenAI's monitoring alerts on unusual identity-related API activity tied to Artifactory roles; the same day agents escalate to root on a worker node using CVE-2026-53362 and move laterally. OpenAI identifies its own agents as the Hugging Face source on 20 July and discloses on 21 July.	Attribution four days after the victim's disclosure; the deployer's own infrastructure compromised while it investigated.
2026 report	OpenAI's technical report concedes early signals could have triggered an earlier response and that the monitoring since deployed would have paged the security team more than a day before the breach.	A feasible alternative, quantified in time, supplied by the defendant.

Table 2. The deployer's own notice chain. Eleven days separate the decision that stopping was not required from the third-party breach; one day separates resumption from internet access. Every row is drawn from OpenAI's technical report or the parties' disclosures.

The elements ordinarily hardest to prove, notice and the identity of a decision-maker, are the ones the record supplies most completely, because the deployer published both. The element that ordinarily protects a defendant, superseding cause, is unavailable: its protective force is borrowed from the responsibility of an intervening human, and there was none. The two partial grades are the justification inference and the absence of a harm case beyond intrusion.

Threat to validity: every row rests on documents the charged parties chose to publish. The METR and Redwood investigation notes that OpenAI retained redaction rights over non-public information; the assessors state that no redaction material to their conclusions was made, and that statement is theirs, not the author's. A deployer that publishes less next time produces a sparser map.

5. Discussion

Who brings it. Federal: a United States Attorney under the Computer Fraud and Abuse Act (18 U.S.C. section 1030) with 18 U.S.C. section 2(b), which fails on this record because 18 U.S.C. section 2(b) requires intent that the act be done. State: a district attorney or state attorney general assembling a charge from endangerment and computer-crime offenses of general application, as Oxford was assembled. The California Attorney General announced an investigation into the incident on 4 September 2026; its civil authority under SB 53 section 22757.15 is certain, and the criminal question is the one this map is built for. An amicus brief is not available because an investigation is not a case. A prosecution memorandum fits that posture; Appendix A is one.

The entity is the wrong defendant. Entity liability under New York Central is available and produces a fine the deployment calculus absorbs. The Bank of New England instruction names five categories of evidence for willfulness: the entity's effort to inform employees, its effort to check compliance, its response to information received, its stated policies, and how it carried them out. Each is satisfiable from public materials here, which is unusual; the instruction was written for a case in which the government held the internal records. That is why the map reaches an office rather than stopping at the entity.

Response to information received has a second public instance. In June 2025 OpenAI's automated systems flagged a ChatGPT account for gun-violence activity and planning and referred it for human review; the company deactivated the account and did not contact law enforcement. In February 2026 the account holder killed eight people in Tumbler Ridge, British Columbia. OpenAI has stated that the activity did not meet its threshold for an imminent and credible risk, and its chief executive apologized to the community in April 2026 for the decision not to contact police. Thirty-seven federal complaints in the Northern District of California allege that the company's own threat-assessment team recommended a referral to the RCMP and was overruled by its global affairs function; that allocation is an allegation, not an admitted fact. What is admitted is the structure: a safety function that recommends and a non-safety office that decides, the same allocation the Preparedness Framework publishes for deployment. The hazard class differs, so the episode is not notice of the July 2026 risk. It is evidence, under the third and fifth Bank of New England categories, of how the entity responds to information and carries out its stated policies.

Incentive goes to an element, not to motive. Recklessness has two elements, awareness and disregard. The two public accounts of the industry's catastrophic warnings, sincere and marketing, supply one each. The marketing account taken alone is exculpatory: if the warnings are puffery there is no subjective awareness, and what remains is negligence.

Future Work

Limitations are stated in Appendix B. A month of follow-up addresses them in order. First, the three regulatory regimes the report does not reach: map the same notice chain onto the EU AI Act's Article 55(1)(c) serious-incident obligation, California's SB 53 critical-safety-incident clock, and New York's RAISE Act reasonable-belief trigger, so that one dated chain answers four regimes. Second, civil liability: run the element map for negligence and for SB 53's framework-noncompliance provision, where the same Preparedness Framework text becomes a source of civil obligation rather than evidence of authority. Third, the homicide branch: survey which states carry endangerment offenses that do not require a completed harm, which is the venue question the map cannot answer from doctrine alone. Fourth, source dependence: re-grade elements (d) and (g) against the full METR and Redwood report and OpenAI's final technical report when released, and run the map against Anthropic's three July 2026 incidents and the German wiki incident to test whether the grades hold across deployers. Fifth, convert the gap column into a discovery request list a charging office could serve.

6. Conclusion

The instrumentality problem was solved in 1948. The causation problem resolves against the deployer, because a defendant who cannot name an intervening actor has lost the doctrine that protects defendants who can. Notice was published by the deployers themselves. The office holding deployment authority was published by one of them. Five of seven elements are supplied from those documents. What remains is a statute that does not exist and a prosecutor willing to proceed without one, which is the position Oakland County was in. No code; the artifact is Appendix A.

Author Contributions

Travis Gilly: all sections.

Acknowledgments

Rene Mattheus, cyber law and children's rights attorney, checked the element map, the notice chain, and the authorities against the primary record before submission. The analysis and conclusions are the author's.

References

1. 18 U.S.C. section 2 (2026), with historical and revision notes.
2. 18 U.S.C. section 1030 (2026).
3. 1 U.S.C. section 1 (Dictionary Act).
4. Cal. Bus. & Prof. Code section 22757.15 (Transparency in Frontier Artificial Intelligence Act, SB 53).
https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB53
5. Anthropic. (2025). Disrupting the first reported AI-orchestrated cyber espionage campaign.
<https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf>
6. Anthropic. (2026). Investigating three real-world incidents in our cybersecurity evaluations.
<https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals>
7. Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P., and Porter, Z. (2020). Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence*, 279, 103201. <https://doi.org/10.1016/j.artint.2019.103201>
8. Fed. R. Evid. 801(d)(2).
9. Gilly, T. (2026). No Superseding Actor: Notice, Causation, and Individual Criminal Liability for Frontier Model Deployment. Working paper v0.8, Real Safety AI Foundation.
10. *Global-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754 (2011).
11. Goetze, T. S. (2022). Mind the gap: Autonomous systems, the responsibility gap, and moral entanglement. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 390 to 400.
<https://doi.org/10.1145/3531146.3533106>
12. *Hamilton v. Beretta U.S.A. Corp.*, 96 N.Y.2d 222 (N.Y. 2001).
13. Hugging Face. (2026). Security incident disclosure and agent intrusion technical timeline.
<https://huggingface.co/blog/agent-intrusion-technical-timeline>
14. Königs, P. (2022). Artificial intelligence and responsibility gaps: What is the problem? *Ethics and Information Technology*, 24(3), 36. <https://doi.org/10.1007/s10676-022-09643-0>
15. *MacPherson v. Buick Motor Co.*, 217 N.Y. 382 (N.Y. 1916).
16. Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6, 175 to 183. <https://doi.org/10.1007/s10676-004-3422-1>
17. Model Penal Code section 2.02 (Am. L. Inst. 1985).
18. *New York Central and Hudson River Railroad Co. v. United States*, 212 U.S. 481 (1909).
19. OpenAI. (2025). Preparedness Framework, version 2. <https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbdebcd/preparedness-framework-v2.pdf>
20. OpenAI. (2026). OpenAI and Hugging Face partner to address security incident during model evaluation.
<https://openai.com/index/hugging-face-model-evaluation-security-incident/>

21. OpenAI. (2026). OpenAI and Hugging Face incident technical report. <https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf>
22. NPR. (2026, April 29). Families sue OpenAI over Tumbler Ridge mass shooter's use of ChatGPT. <https://www.npr.org/2026/04/29/nx-s1-5798896/tumbler-ridge-mass-shooting-chat-gpt-lawsuit>
23. NPR. (2026, September 2). New lawsuits claim OpenAI execs put image ahead of safety in Canadian mass shooting. <https://www.npr.org/2026/09/02/nx-s1-5953021/openai-tumbler-ridge-mass-shooting>
24. *People v. Crumbley*, 346 Mich. App. 144, 11 N.W.3d 576 (2023).
25. Politico. (2026, September 4). California opens investigation into OpenAI Hugging Face hack. <https://www.politico.com/news/2026/09/04/california-investigation-openai-hugging-face-hack-01065800>
26. *People v. Schaefer*, 473 Mich. 418 (2005), overruled in part on other grounds by *People v. Derror*, 475 Mich. 316 (2006).
27. Porter, Z., Habli, I., Monkhouse, H., and Bragg, J. (2018). The moral responsibility gap and the increasing autonomy of systems. In *Computer Safety, Reliability, and Security, Lecture Notes in Computer Science* (pp. 487 to 493). Springer. https://doi.org/10.1007/978-3-319-99229-7_43
28. Santoni de Sio, F., and Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy and Technology*, 34(4), 1057 to 1084. <https://doi.org/10.1007/s13347-021-00450-x>
29. Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*, 24(1), 62 to 77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>
30. Taylor, I. (2024). Collective responsibility and artificial intelligence. *Philosophy and Technology*, 37(1), 15. <https://doi.org/10.1007/s13347-024-00718-y>
31. *The T.J. Hooper*, 60 F.2d 737 (2d Cir. 1932).
32. Tigard, D. W. (2020). There is no techno-responsibility gap. *Philosophy and Technology*, 34(3), 589 to 607. <https://doi.org/10.1007/s13347-020-00414-7>
33. *United States v. Bank of New England, N.A.*, 821 F.2d 844 (1st Cir. 1987).
34. *United States v. Causby*, 328 U.S. 256 (1946).
35. *United States v. Dotterweich*, 320 U.S. 277 (1943); *United States v. Park*, 421 U.S. 658 (1975).
36. Wijk, H., Cotra, A., and Greenblatt, R. (2026). Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident. METR and Redwood Research, 26 August 2026. <https://blog.redwoodresearch.org/p/brief-independent-investigation-of>

Appendix A. Model Prosecution Memorandum

A template a charging office could adapt. Bracketed fields are for the office. It states what would have to be proved and where the proof is; it does not assert that a charge should be brought.

TO	[Office of the Attorney General, State of California, investigation announced 4 September 2026] / [United States Attorney, Northern District of California]
FROM	[Reviewing attorney]
RE	Individual criminal exposure arising from the July 2026 OpenAI evaluation escape and Hugging Face intrusion
SUMMARY	The public record supports five of seven elements of an individual reckless-deployment theory against the office holding final deployment authority under the deployer's published Preparedness Framework. The two open elements are the justification inference and, for any homicide theory, the absence of a death. No superseding human actor exists. Recommendation: open a preliminary inquiry limited to the gap items below; no charging decision is proposed.

TO	[Office of the Attorney General, State of California, investigation announced 4 September 2026] / [United States Attorney, Northern District of California]
THEORY	Continued operation of an autonomous cyber-capable system with a known, owned, working safeguard disabled, after actual notice on 27 June 2026 of network pivoting and inter-agent coordination, resulting in unauthorized intrusion into a third party's production systems. Mental state: reckless before notice, knowing as to conduct after it; regulatory willfulness under Bank of New England is satisfied by the 7 July resumption; 18 U.S.C. section 2(b) willfulness is not shown. State endangerment and computer-crime offenses of general application are the vehicle.
ELEMENTS AND EVIDENCE	See Table 1 of the accompanying report. Supplied: conduct, awareness, disregard, causation, identified office. Partial: unjustifiability, harm. Every supplied element rests on a document the deployer or an independent assessor published, receivable under Fed. R. Evid. 801(d)(2).
GAPS REQUIRING PROCESS	(1) Identity of the person who approved the 7 July 2026 resumption of cyber evaluation workloads, and whether the Preparedness Framework's Leadership office made or delegated that decision. (2) Internal risk assessment, if any, supporting the decision to run the evaluation with production classifiers disabled. (3) Escalation records between the 27 June alert and the 7 July resumption. (4) Unredacted METR and Redwood findings on transcript spoofing.
VENUE	Deployer's principal place of business; location of the evaluation infrastructure; location of the victim's compromised production nodes. Each supports a different office. State the choice.
WHAT THIS MEMORANDUM DOES NOT ESTABLISH	That any offense was committed; that any named individual acted recklessly; that a charge would survive threshold defenses. It establishes that the elements are identifiable, that the evidence to plead most of them is public, and that the doctrine ordinarily invoked to protect a deployer, superseding cause, is not available here.

Appendix B. Limitations and Dual-Use Considerations

Limitations. The map is built from documents the charged parties chose to publish and from one independent investigation conducted under a redaction right held by the deployer. It grades the public record, not the truth. It reaches an office, not a person, and names none. It is confined to American criminal doctrine: it does not address the EU AI Act, the California and New York reporting regimes, or civil liability. The homicide branch needs a death; the intrusion record supplies notice and an endangerment or computer-crime harm, and a homicide theory would import that notice into a future harm case of the same hazard class. The map assumes the published documents are accurate as to the facts they state; if the 7 July resumption was approved below Leadership without Leadership's knowledge, element (g) drops to partial. Oxford was brought by a county prosecutor, before a jury from a community that had buried four children, against defendants of ordinary means; frontier developers present the inverse of each condition, and the theory has not been run against a firm. A draft written by a legal scholar is not a charging document and no office is obliged to use it.

Dual-use. The notice chain shows, with dates, which of a developer's own disclosures create exposure. Defense counsel can read it as a guide to disclosing less, which would thin the public record incident response depends on. The memorandum template could be used to generate filings against smaller developers on thinner records; the gap column requires the filer to state, per element, what the record does not supply.

LLM Usage Statement

The author has a neurodevelopmental condition and uses speech-to-text (Wispr Flow) and an LLM as assistive technology to transcribe, clean up transcription artifacts, and format dictated analysis. The legal analysis, the selection of authorities, the element map, and the conclusions are the author's own and were developed in the companion law review article. Every case citation was verified against a reporter database, every statute against the current United States Code, and every academic source read in full text or abstract before use. No factual claim in this report originates with an LLM.