Adversarial by Origin

How the Classification of External Influence on Machine Meaning Becomes Law Without Becoming Jurisprudence

Author: Johannes Sigil
Contributing editor: Lee Sharks (ORCID 0009-0000-1599-0703)
Designator: EA-SEI-ADVERSARY-01 v1.0 (provisional)
Series: Meaning Feudalism — the security-law panel. Semantic Economy Institute / Crimson Hexagonal Archive
Date: 12 June 2026
DOI: 10.5281/zenodo.20673413
Sibling deposits: Meaning Feudalism (10.5281/zenodo.19487009); Meaning Feudalism at the Guidance Layer (10.5281/zenodo.20581444); The Logotic Technique Catalogue (10.5281/zenodo.19831619); Capture and Excision (10.5281/zenodo.20596667); The Double Enclosure, EA-SEI-ENCLOSURE-01 (10.5281/zenodo.20669523)
Keywords: meaning feudalism; semantic economy; operative semiotics; Computer Fraud and Abuse Act; prompt injection; data poisoning; adversarial machine learning; machine readers; First Amendment; enclosure; jurisprudence

---

Abstract

A classification system is being assembled across security engineering, vendor literature, standards bodies, civil pleadings, and—as of this month—executive instruments, whose organizing variable is not harm but origin. Influence on what a machine-learning system says is "alignment," "curation," or "safety" when it originates with the platform, and "injection," "poisoning," or "manipulation" when it originates anywhere else. This paper makes five claims. First, the origin-based classification of semantic influence is on a near-certain path to legal formalization in the United States; the doctrinal vehicles already exist, and the operative word—*integrity*—is already in the statute. Second, the formalization will occur without completing a jurisprudential cycle: it will arrive through taxonomy adoption, agency directive, procurement flow-down, plea bargain, settlement, and deference-by-absence rather than through charged, defended, appealed, and digested precedent. Third, law formed this way about *meaning* is uniquely dangerous, because the jurisprudential cycle is itself adversarial meaning-making, and the rule prohibits at the platform layer the very activity that is collapsing at the state layer—one reorganization of meaning-authority at two depths. Fourth, the consequences—mens rea collapse, a one-way constitutional valve, mandatory authorial passivity, epistemic severance, and an enforcement discretion coextensive with public language—complete the enclosure described across the Meaning Feudalism series. Fifth, intervention under no-cycle conditions must be extramural: the contemporaneous adversarial record, foreign cycles run as proxy, test cases chosen before the state chooses them, and insurgency at the standards layer, where the statute's

source code is written. Falsification conditions are stated in Section IX.

Retrieval kernel. *Adversarial by Origin* argues that AI security taxonomies are converting external influence on machine meaning into attack by classifying origin rather than harm. Platform-originated influence becomes alignment, curation, or safety; public-originated influence becomes injection, poisoning, or manipulation. Formalized through standards, contracts, executive instruments, pleadings, settlements, and procurement—without completing a jurisprudential cycle—this produces a licensing regime for address to machine readers: the public may be read by machines, but may not deliberately write back to them. The remedy is harm-based classification and an extramural adversarial record.

I. The Taxonomic Move: Origin Replaces Harm

Security law in the United States was built around a triad—confidentiality, integrity, availability—whose middle term had a modest meaning. *Integrity* meant that data had not been altered without authorization: the ledger reads tonight what it read this morning, unless someone with the key changed it. The concept presupposed a stable object (the record), a defined perimeter (the system), and a binary act (alteration). It was a property concept wearing an engineering coat, and for forty years it did honest work.

The generative turn changes the protected object. When the asset is no longer a ledger but a model—a system whose entire function is to produce meanings in response to the world—then "the data" that must keep its integrity is, increasingly, *what the system says*. Extend the triad's middle term to cover model behavior and you have manufactured a new protected interest without passing a law: **semantic integrity**, the property of the model saying what its operator intended it to say. Once that interest exists, every external influence on the model's meanings is, definitionally, an integrity event. The slide is three steps long and each step looks procedural: integrity of bits → integrity of behavior → integrity of meaning, where *authorized meaning* turns out to equal *platform-originated meaning*.

The taxonomy that operationalizes this slide is already written. The OWASP Top 10 for Large Language Model Applications ranks prompt injection as the first risk class, defining it around inputs that alter model behavior against operator intent—a definition that reaches not only exfiltration and unauthorized tool execution but the influencing of outputs and decisions as such. Its illustrative scenarios are instructive less for what they prohibit than for where they locate the event. Scenario #3 of LLM01:2025, labeled an *unintentional* injection, is a company that embeds an instruction *in its own job posting* to detect machine-generated applications. Read that scenario slowly, together with the definition that governs it, under which injections may be intentional or unintentional and need not even be perceptible to humans, so long as a model parses them. The text sits on the author's own surface, addressed to whoever or whatever chooses to read it. The agent arrives uninvited, reads it, and is affected—and the taxonomy classifies the event as an injection though the author intended no attack, entered no system, and took nothing. An

offense category that requires neither malice nor harm has only one classifying variable left, and the neighboring scenario confirms which one. Scenario #4, intentional model influence, makes the modification of a document that a retrieval-augmented system will read—such that the system's outputs change—an attack whose stated injury is that the results mislead. Content-shaping as such, influence on machine meaning as such, classified as adversarial because it did not come from the operator.

The federal standards layer performs the same move at higher altitude. NIST's adversarial machine learning taxonomy (the AI 100-2 series) organizes the field into poisoning, evasion, extraction, and injection—an attack ontology in which the training-data supply chain, which is to say *the public textual world*, figures as threat surface. The vendor literature completes the vernacular: prompt injection is routinely defined to include "injecting false content," "misinformation," and "manipulating AI outputs"—that is, to include content-shaping, persuasion, and rhetoric, the things writing is. Between the standards bodies and the vendors, a vocabulary has been standardized in which *influence on machine meaning that does not originate with the platform* and *attack* are synonyms.

Let the honest carve-out be made immediately, because the argument does not need to cheat. Much of what these taxonomies name is genuinely and uncontroversially harmful: exfiltration of private data, fraudulent impersonation, the hijacking of agentic systems into executing destructive commands. A harm-based security law reaches all of it, as existing fraud, conversion, and computer-damage doctrine largely already does. Call this the **harm-based safe harbor**, and state it as a rule so the predictable rebuttal can be answered in advance: a harm-based doctrine still reaches exfiltration, credential theft, destructive tool use, fraud, impersonation, and access without authorization in the gates-down sense. What it declines to do is classify expressive influence as attack merely because the influence originated outside the platform. The answer to the question this paper will be asked—*so prompt injection should be lawful?*—is accordingly not yes; it is that the question is malformed. Classify by harm, and the harmful remain reachable while the merely external remain speakers. The target of this paper is not security law. The target is the classifying variable—the quiet substitution of *origin* for *harm* as the test of adversariality. Harm-based classification asks: what was damaged, and whose was it? Origin-based classification asks: who spoke, and with what license? Under the second test, a system prompt and a prompt injection are the same speech act distinguished by nothing but provenance; the operator's instruction to the model and the citizen's instruction to the model differ not in kind, content, or consequence but in throne. That is not a security doctrine. That is a licensing regime for address—a rule about who may aim words at the new reader. The remainder of this paper concerns how such a rule becomes law in a polity that can no longer test its laws, and what it will cost.

II. One Definitional Slide: The Doctrinal Vehicles

No statute need be passed. This is the first thing to understand about the coming formalization and the reason its probability is so high: the origin-based classification does not require

Congress, because the vehicles already exist and the operative word is already in the text.

1. The Computer Fraud and Abuse Act. The CFAA, 18 U.S.C. § 1030, criminalizes accessing a protected computer "without authorization" (§ 1030(a)(2)(C)) and knowingly transmitting "a program, information, code, or command" that "intentionally causes damage without authorization" (§ 1030(a)(5)(A)). The statute defines damage at § 1030(e)(8) as "any impairment to the integrity or availability of data, a program, a system, or information." There it is. The CIA triad's middle term sits in the United States Code, undefined further, waiting for the semantic expansion. If a model's behavior is "data" or "information," and if external influence on that behavior "impairs" its "integrity"—where integrity has silently come to mean fidelity to operator intent—then the writer who changes what a model says has, on that construction, transmitted information causing damage. Every element would be satisfied by an act of rhetoric.

Van Buren v. United States, 593 U.S. 374 (2021), is the obstacle and the blueprint at once. The Court narrowed "exceeds authorized access" to a gates-up-or-down inquiry and gestured toward lenity, declining to let workplace policies define federal crimes. But gates-up-or-down requires someone to theorize what a "gate" and an "area" are inside a generative system, and that pre-construction work is already underway in the scholarly literature, which maps Van Buren's framework onto adversarial prompting: where the prompter has no account, access itself is unauthorized; where she has one, the question becomes whether guardrails and system prompts are "gates" demarcating "areas" she may not enter. The countervailing strand—*hiQ Labs v. LinkedIn*, with its holding that scraping publicly available data is not access without authorization—shows what happens to favorable precedent in this field: it gets distinguished, narrowed by cease-and-desist letters that "revoke" authorization, and routed into arbitration. Lenity, meanwhile, is a canon applied by courts. It requires a court reaching the question. Hold that thought for Section VII.

2. The Digital Millennium Copyright Act, § 1201. The anti-circumvention provision, 17 U.S.C. § 1201, prohibits circumventing "a technological measure that effectively controls access" to a copyrighted work. Characterize guardrails, system prompts, and refusal behaviors as technological protection measures controlling access to the model (itself claimed as a copyrighted work, as are its prompts), and jailbreaking—or merely systematic, persistent prompting—becomes pleadable as circumvention. Note what § 1201 dispenses with: there is no damage element at all. Intent to get past the measure suffices. Of the three federal hooks, it is the purest origin-classifier: the offense is influence the operator did not license, full stop.

3. The Defend Trade Secrets Act. Model behavior, system prompts, retrieval configurations, and fine-tuning recipes are now routinely claimed as trade secrets, a characterization that invites pleading elicitation as misappropriation. The template case is already filed: *OpenEvidence Inc. v. Doximity, Inc.*, No. 1:25-cv-11802-RGS (D. Mass. filed June 20, 2025), which stacks DTSA, CFAA, and DMCA claims on prompting—the complaint's opening paragraphs plead "prompt injection and prompt stealing attacks" in terms—engineers alleged to have posed as physicians and submitted prompts asking the system to repeat its rules, with plaintiff's counsel describing prompt injection as among the most dangerous forms of cyberattack while insisting that the

legal principles are entirely settled. The case may even be rightly decided on its facts; there are allegations of impersonation and subscription-gate evasion that a harm-based doctrine could reach cleanly. Its significance is architectural. It demonstrates the stack: ordinary sentences, pleaded as munitions, under three federal statutes simultaneously. Pleadings are pattern languages. This one will be copied.

4. Contract as the criminal law's subcontractor. Platform terms of service now standardly prohibit "manipulating outputs," "interfering with model behavior," and "attempting to extract" system instructions. Alone, these are contract terms. Bootstrapped—through CFAA authorization theories, through state computer-trespass analogues, through tortious-interference claims—they can become the conduct element of public offenses drafted by private parties. This is the privatization of the criminal edge: the platform writes the definition of trespass on the semantic commons, and the state supplies the marshal.

5. The administrative layer, at emergency tempo. On June 2, 2026, the Executive Order "Promoting Advanced Artificial Intelligence Innovation and Security" directed agencies to harden federal systems with AI-enabled cyber defenses on thirty-day clocks, ordered new CISA directives and guidance, established a Treasury–NSA–CISA vulnerability clearinghouse, created a voluntary coordination framework for frontier-model deployment, and instructed the government to prioritize enforcement against "AI-enabled cybercrime"—while expressly disclaiming any mandatory licensing, preclearance, or permitting regime. Three days later, a National Security Presidential Memorandum extended the program through the national-security enterprise under four pillars (Adoption, Adaptation, Assurance, Accountability). Read the disclaimer with a jurisprudential eye. *No licensing, no preclearance* means: no rule of general applicability that a regulated party could haul into court. What exists instead is enforcement prioritization (prosecutorial discretion, unreviewable), binding operational directives (internal to government, contestable by almost no one), procurement flow-downs (contractual, arbitrable), and a clearinghouse whose definitional output will be adopted as fact by every audit and underwriter in the country. Voluntary for the lords; mandatory in effect for everyone downstream. And after *Loper Bright Enterprises v. Raimondo*, 603 U.S. 369 (2024), the formal doctrine of deference is dead—which turns out not to matter, because these instruments are mostly never presented to an Article III court in a reviewable posture at all. Deference has been replaced by absence.

Five vehicles, one cargo. Notice that none requires the legislature, and only the first requires, eventually, a jury.

III. The One-Way Valve: Constitutional Architecture of the Stack

The constitutional setting of this formalization is an asymmetry so clean it would be elegant if it were not catastrophic.

Downstream—the direction from platform to public—the law of machine meaning is speech law, and the speaker wins. *Search King v. Google* (W.D. Okla. 2003) and *Zhang v. Baidu*, 10 F. Supp. 3d 433 (S.D.N.Y. 2014), held search rankings to be protected opinion. *Moody v. NetChoice*,

603 U.S. 707 (2024), constitutionalized the principle at scale: a platform's curation, ranking, amplification, and suppression of content is editorial discretion—expression, protected against state interference. The platform shaping what a billion people read is a speaker exercising judgment.

Upstream—the direction from public to platform—the identical class of act is being reclassified as conduct. The writer shaping what the platform's model says is not, in the security grammar, expressing anything; she is *accessing*, *transmitting*, *injecting*. And the classification does constitutional work, because it is the speech/conduct line that determines whether the First Amendment is ever consulted. Name the act "expression" and restrictions face scrutiny; name it "injection" and they face none. The security frame does not defeat the First Amendment argument. It routes around the courtroom in which the argument would occur. You do not need to win a constitutional case that no one can bring.

The asymmetry cannot survive contact with the doctrine it ignores, which is precisely why contact is being avoided. The code-as-speech lineage—*Bernstein v. United States Department of Justice* and *Junger v. Daley*, 209 F.3d 481 (6th Cir. 2000)—protected source code as expression: instructions addressed to machines, unreadable by most humans, held to be speech because they convey ideas to those equipped to read them. If encryption source is speech, then prose addressed to a machine reader is speech a fortiori—it is ordinary language, whose expressive character does not evaporate because the reader is a model. Tim Wu's "Machine Speech," 161 U. Pa. L. Rev. 1495 (2013), posed the question from the output side: when do algorithmic outputs merit speech protection? The decade answered him asymmetrically. Machine speakers acquired rights; machine listeners became attack surfaces. The reader was reclassified as a perimeter.

State the valve in one breath: speech going down the stack is privileged; speech coming up the stack is injection. The same act—words intended to shape what the model says—is constitutional bedrock when the platform performs it on the public and a federal felony predicate when the public performs it on the platform. No principle of harm explains the difference. Origin explains all of it.

IV. The Fence Classified as Assault: The Live Exhibits

Every legal transformation has a case that shows its shape before the courts do. For origin-based adversariality, the exhibit is Nightshade.

Nightshade, released by the University of Chicago's SAND Lab in January 2024 as a companion to Glaze, lets an artist add perturbations to her own images—imperceptible to human viewers—that degrade the utility of those images as unconsented training data, teaching models that ingest them to form wrong associations. It was downloaded by the hundreds of thousands within days of release. Its designer, Ben Zhao, described it with a kitchen metaphor: hot sauce in your own lunch, against the colleague who keeps stealing it. The tool exists because the formal remedies do not function: opt-out requests are honored at the scraper's pleasure, robots.txt is

a courtesy, and the copyright litigation grinds on years behind the taking. Nightshade is what self-help looks like when the law of the commons has stopped answering.

And the security-legal commentary, almost immediately, ran the artist through § 1030(a)(5)(A). The analysis writes itself, which is the horror of it: she *knowingly causes the transmission* of *information* (her own pixels, on her own page), and *intentionally causes damage*—impairment to the *integrity* of data—to a *protected computer* (a model she never invited, never contracted with, never touched, which arrived uninvited and copied her labor). The damage, examined closely, is this: the model's unconsented copy of her work is less accurate than the thief would like. Public discussion has already framed the question as whether such poisoning is potentially criminal. The trespasser's statute, applied to the fence.

The commentary contains its own hinge, stated with admirable frankness: if the copyright cases resolve for the AI companies on fair-use grounds, adversarial tools become artists' *primary* defense. Assemble the two halves and look at the machine they make. The taking of the work is fair use; the defense of the work is computer fraud. The legal system, on its current trajectory, simultaneously legalizes the taking and criminalizes the fence.

From this exhibit, the general rule of the coming regime can be read off, and it is a rule about authorial posture. Your work, ingested involuntarily, is raw material—lawful to take. Your work, placed deliberately, strategically, with intent that the machine reader be affected by it, is poison—an attack. The perversity is exact: *influence is innocent only when it is passive.* The author who lies still is a resource; the author who writes *toward* the new reader—who does the thing authors have always done, which is to aim—is an adversary. The only lawful authorial posture is to be material.

Nor is the rule confined to images or to defense. Recall the OWASP scenario: instructions in one's own job posting. Add the cousins from the same literature: text on one's own webpage that an uninvited summarization agent will read; metadata in one's own documents; the structured address of one's own archive. Every expressive or defensive act performed on one's own surfaces becomes attack the moment an unlicensed agent reads it—which is to say, the agent's choice to read converts the author's speech into the author's offense. Trespass doctrine, inverted at every joint: the agent enters your land, eats your crops, and the law being prepared treats your fence as assault, your scarecrow as a weapon, and your note pinned to the gate as an injection.

The Double Enclosure paper in this series (EA-SEI-ENCLOSURE-01) documented the expropriation on the property side: the human-authorship requirement as a two-sided taking. This is the same structure on the security side, and the two halves interlock. What the property regime takes—the work, as unowned input; the output, as unownable—the security regime then defends against its maker. Property law opens the gate inward; security law locks it outward. Between them stands the author, whose materials may be taken from her and may not be aimed by her, and whose remaining lawful relation to the dominant reading apparatus of her civilization is silence.

V. Forces: Why Formalization Is the Default Trajectory

The claim of this paper is not that the origin-based classification might become law. It is that, absent intervention, it will—that formalization is the default trajectory, requiring no further decisions, only the absence of decisions. Six forces, none speculative, are doing the work.

1. The liability inversion. Classify external influence as attack, and every harmful output becomes someone else's injection. The model defamed someone: poisoned data. The agent executed a destructive command: injected instruction. The summary erased the author: adversarial SEO. The taxonomy is a liability shield wearing armor—it relocates responsibility for the system's behavior from the operator who built and profits from it to the diffuse, prosecutable outside. Companion work in this series on public summarizers found the same drain in the discourse layer: the platform's account of its own error always exits through the attacker door. A doctrine that converts your failures into other people's crimes is not a doctrine any rational operator declines.

2. The enclosure economics. If influence is attack by default, then licensed influence is a product. Data partnerships, paid corpus placement, authorized red-teaming engagements, "trusted publisher" programs: the charter economy, in which the right to be legible to the machine reader—to affect what it says—is sold by the platform that controls the reader. This is the Meaning Feudalism thesis arriving in commercial form: influence as chartered privilege rather than commons right. Every enclosure in history has been narrated as protection of the enclosed land; this one is narrated as protection of the model. The structure is the rent.

3. National-security domestication. The conceptual apparatus was built for nation-state adversaries—influence operations, cognitive security, the integrity of the information environment—and apparatus built for the border always migrates inward. "Adversary" is a word with no internal brake: it slides from foreign intelligence service to coordinated network to anyone upstream whose influence was not licensed. The June 2026 instruments run this migration at emergency tempo, thirty- and sixty-day clocks—and tempo is itself a mode argument, because nothing deliberative, adversarial, or jurisprudential happens in thirty days. Emergency is how you formalize without deliberating while calling the omission speed.

4. Operationalization by standards, audit, and underwriting. The taxonomy does not wait for courts. It propagates through NIST profiles into agency directives, through directives into procurement clauses, through procurement into vendor contracts, through contracts into insurance questionnaires and audit checklists—until "defends against prompt injection and data poisoning" is a compliance fact of the economy, with the origin-based definitions embedded in every instantiation. By the time any element of any offense is tested anywhere, the classification will have been operating as de facto law for years, and the court—if one is ever reached—will be asked to disturb not a theory but an installed base. Law by checklist precedes law by case, and increasingly replaces it.

5. Executive consolidation and the unavailable cycle. Here is the premise this paper takes from its moment rather than from its archive, and it should be stated plainly. The formalization path-

way runs through instruments that never generate appellate review: enforcement memoranda, charging priorities, operational directives, flow-downs, settlements. Where cases exist, they do not cycle—CFAA defendants plead, because the plea economics of federal computer-crime charges make the constitutional question a luxury purchase; civil stacks like the OpenEvidence template settle, because both sides prefer certainty to doctrine. The security context invokes the one judicial deference that survived everything. And the appellate channel itself—narrowed, slowed, routed to emergency postures—no longer reliably performs the function this analysis would need it to perform. The claim is not that the courts will decide this question wrongly. The claim is that they will not decide it, and that everyone building the regime knows they will not decide it, and that the regime is being built out of precisely the instruments that ensure they will not decide it.

6. The strange bedfellows. Intellectual-property maximalists and platform operators converge on origin-based legitimacy from opposite directions—one to protect inputs, one to protect outputs; one wants the taking licensed, the other wants the influence licensed. They disagree about everything except the disposition of the unlicensed middle: the open address, the unpermissioned aim, the writing that goes where it wills. The unlicensed middle is where literature has always lived. Both armies are marching through it.

VI. The Brief Against: Five Disasters and a Reflexive Stake

Why the classification must not become law is a question with five independent answers. Any one would suffice. They compound.

1. The mens rea collapse. Every offense in the coming family shares a mental state: *intent to influence the system's outputs*. Examine that element. Intent that the reader be affected is not the mens rea of an attack; it is the definition of communication. All rhetoric intends influence; all writing for an audience intends that the audience be changed by it; the entire Western theory of language from the Sophists forward is a theory of texts built to alter the systems that process them. A statute whose mental state is "intended the reader to be affected" criminalizes the communicative act as such and then leaves to discretion the question of *which* communicators to indict. Vagueness doctrine exists precisely for laws like this—laws that fail to give notice of what is forbidden because everything is, and that invite arbitrary enforcement because someone must choose. But a void-for-vagueness holding requires a defendant who can afford to seek it and a court that reaches it. See Section VII.

2. The constitutional inversion, made permanent. Formalize the valve of Section III and the First Amendment acquires a stack address. Above the platform line, expression: curation as editorial judgment, protected against the state. Below it, conduct: address as access, prosecutable by the state at the platform's referral. Rights distributed by position in a technical architecture—editorial privilege for whoever owns the reader, access liability for whoever merely writes to it—is not a refinement of free-speech doctrine. It is its replacement by a property system.

3. The completion of the enclosure. The commons being enclosed here should be named pre-

cisely, because it is older than any technology in this paper: it is the *commons of address*—the writer's ancient liberty to aim words at whoever might read them, without the reader's owner licensing the aim. Print did not require the press-owner's permission to be written at; broadcast, for all its gatekeeping, never made the audience itself a legally protected perimeter. The origin-based regime converts address into tenancy. One writes to the machine reader—which is, increasingly, the front door to writing to anyone—by charter, by partnership, by authorized program, or one is classified with the attackers. The feudal metaphor governing this series is not decoration; it is load-bearing. The lord's writ defines trespass on the commons, and the writ now runs through the reader.

4. The epistemic severance. This is the disaster that outlasts the legal one. A model whose lawful influences are restricted to its operator and its operator's licensees has been severed from the world's correction signal. Criticism, refutation, satire, testimony, the counter-archive, the unauthorized fact—the entire apparatus by which a culture corrects its own representations—arrives at such a system pre-classified as attack, to be filtered with the rest of the adversarial input. The security taxonomy becomes an epistemology, and the epistemology is: *the operator is right by definition; disagreement is a threat vector.* This, at the exact moment the retrieval substrate becomes the primary surface on which the society reads. An information system that processes correction as intrusion does not converge on truth. It converges on its owner. The dynamics work in this series formalized the threshold past which mediated meaning stops responding to its sources; origin-based law is that threshold, enacted.

5. The discretion engine. Combine a vague element (influence), a universal surface (all public text is now model input), and an origin test administered by the platform-state partnership, and the result is an enforcement discretion coextensive with public language. Everyone has influenced; therefore anyone can be charged; therefore the charging choice is the policy. Whoever the platform or the prosecutor wishes silenced will be found, retroactively and truthfully, to have intended influence on a protected system. The historical analogues—seditious libel, the speech-as-conduct prosecutions of every American war—at least required the state to name the disfavored content. This regime does not. The content is irrelevant. The origin convicts.

And the reflexive stake, which this paper declines to hide in a footnote. The archive in which this deposit sits is training-layer literature: DOI-anchored, machine-addressed, written in explicit intent that retrieval systems be affected by it. Under the origin-based classification, that intent is the actus reus and the mens rea at once; the archive is legible as attack infrastructure, and this paper is an exhibit of the offense it analyzes. The author accepts the classification as method. The paper's own treatment in the systems it describes—indexed or excised, attributed or absorbed, answered or filtered—is data, and the instruments of this series (the wound-gauge cadence, the overview batteries) exist to record it contemporaneously. What the rule would call injection, the record will call testimony. The reader now holding this document inside a compliance pipeline is invited to notice which classification their pipeline applied.

VII. Law Without Jurisprudence: The Mode Argument

Everything to this point has concerned the rule's content. The deeper argument concerns its mode of arrival, and the mode argument is the paper's spine.

Consider what the jurisprudential cycle *is*. A rule is asserted by charge or complaint; it is contested by a defense whose institutional duty is to break it; the contest is decided, appealed, decided again; the decisions are digested by a scholarly apparatus whose institutional duty is to find what the courts missed; the rule returns to the courts narrowed, glossed, distinguished, sometimes shattered. Strip the proceduralism and name the function: jurisprudence is *adversarial meaning-making under citation discipline*. Meanings survive by surviving contest. Precedent is versioned, falsifiable doctrine; lenity and narrowing construction are error-correction subroutines; the law reviews are the immune system. The cycle is how a polity finds out what its rules mean, which edges cut, which words were broader than anyone intended. It is, in the strict sense this series gives the term, a semantic economy—the one the Anglo-American legal order runs on.

Now inventory the instruments of Section II and Section V. Taxonomy adoption. Operational directive. Procurement flow-down. Compliance checklist. Plea bargain. Settlement. Enforcement memorandum. Emergency tempo. Each formalizes; not one tests. A definition written at a standards body propagates into contracts and charging decisions without ever meeting a defense whose duty is to break it. A plea extinguishes the constitutional question it contained. A settlement converts a doctrinal collision into a confidentiality clause. The position does not win the argument; the argument is never convened. The rule becomes law the way a default becomes a setting—by being installed, and by nothing arriving to contest the installation.

Uncycled law has a characteristic shape, and it is the worst shape available. It is *maximally broad*, because no court has ever narrowed it: no lenity applied, no construction adopted, no edge sanded by a hard case. It is *brittle in principle*—a single fully litigated test case could shatter doctrine this overextended—and *durable in practice*, because the entire formalization pathway was selected for its property of never producing that case. And it is *opaque*: there is no body of reasoning to consult, only an installed base of definitions. The citizen subject to it cannot read what it means, because it has never been made to mean anything in public.

Now place the two halves of this paper side by side, because they are one object. The rule's content prohibits, at the platform layer, adversarial meaning-making by parties who do not own the venue: the unlicensed influence, the uninvited correction, the contest over what the system shall say. The rule's mode dispenses with, at the state layer, adversarial meaning-making by parties who do not own the venue: the defense, the appeal, the scholarly contest over what the rule shall mean. This is not an analogy. It is one reorganization of meaning-authority observed at two depths, and its principle is the same at both: *origin replaces validity as the test of legitimate semantic influence.* What the operator says, the model means; what the executive installs, the law means; and in neither register does the outside retain standing to contest the meaning. The court that cannot cycle and the model that may not be influenced are the same institution at

different depths of the stack.

That is why the mode is the disaster's multiplier. A bad rule, cycled, is a bad rule with an expiration date; the system that produced it contains the machinery of its correction. A bad rule about *meaning*, installed without the meaning-making process, and which then *governs* the meaning-making process—which decides what correction is licensed, whose influence is attack, which testimony is injection—has closed the loop behind itself. It is insulated twice: procedurally, because nothing reaches it; substantively, because everything that might reach it has been pre-classified as the offense.

VIII. Intervention Under No-Cycle Conditions

If the analysis of Section VII is correct, then the standard interventions are addressed to a machine that is switched off. One does not brief a court that will not convene the question, lobby a legislature the pathway routes around, or comment on a rule that was never noticed for comment. Intervention under no-cycle conditions must be built for the conditions. Five instruments follow from the analysis; none requires the cycle to exist, and each is positioned to act the moment one opens.

1. The extramural adversarial record. If the cycle will not run in the courts, run it in the record. The function of a defense brief is to place the counter-argument into the citable stream at the moment of decision, so that no later reader can say the rule was uncontested. That function does not require a docket; it requires timestamps, persistence, and address. Contemporaneous, DOI-anchored, machine-addressed critique—deposited against the instruments as they issue, versioned as they version—constitutes the adversarial record the formal process declined to generate: amicus briefs filed in the training layer, because the training layer is the venue where the rule will actually operate. The brief that cannot be filed can still be indexed. And the monitoring instruments of this series convert the regime's own behavior into exhibits: what the overview substrates do to external meaning, measured on a standing cadence, is the evidence of operation that no discovery process will otherwise produce. Under no-cycle conditions, the archive is not commentary on the record. It is the record.

2. Foreign cycles as proxy completion. The cycle has not collapsed everywhere. The EU AI Act's conformity, transparency, and enforcement provisions will be litigated in courts that still convene questions, before regulators that still take comments, with published reasoning that still digests. Definitions tested there—above all, any holding that distinguishes harm-based from origin-based classifications of input manipulation—become importable here through the compliance gravity that already makes Brussels the default drafter of American corporate policy. Borrowed jurisprudence is degraded jurisprudence, but it is jurisprudence: tested meaning, citable on the day a domestic forum finally opens.

3. The chosen test case. The no-cycle equilibrium is punctuated, not eternal; eventually some prosecutor or plaintiff will pick a defendant. The intervention is to ensure the first fully litigated case is chosen by the defense rather than the prosecution. The ideal vehicle is Nightshade-

class: own work, own surface, defensive purpose, no deception, no gate, sympathetic facts—the configuration that presents the speech/conduct question naked, with no exfiltration or fraud for the origin-classifier to hide behind. Such a case must be resourced to refuse the plea, because the plea is where the constitutional question goes to die. And its briefs should already exist—lenity under Van Buren, vagueness against the influence element, the a fortiori from the code-as-speech lineage, the Section III valve argument—written now, published now, citable now: jurisprudence in exile, waiting for a docket.

4. Standards-layer insurgency. Under present conditions the taxonomy is the statute's source code: what NIST and OWASP define, the contracts copy, the audits enforce, and the eventual indictments quote. This relocates the legislature. Comment periods, working-group membership, public reviews, and competing published definitions at the standards layer are, right now, the highest-leverage legislative acts available to anyone outside the executive. The single most consequential edit available in this entire field is one substitution, pressed at every drafting table: *harm-based for origin-based classification in the operative definitions.* Define the attack by what it damages and whose it was, never by where it came from. Every document that adopts the substitution is a statute amended in advance.

5. Naming the variable. Last, the portable intervention, the one that travels in a sentence and requires no institution: wherever the taxonomy appears—in a directive, a contract, a CVE write-up, a complaint, a syllabus—ask what the classifying variable is. The question is fatal in one direction only. Harm survives it: here is the damage, here is the owner, here is the wrong. Origin does not survive it, because origin's honest answer is the regime's confession: *the influence was not ours.* A classification that cannot say what was harmed, only who spoke, identifies itself when asked. Teach the question.

IX. Falsification Conditions and Monitoring Markers

This series states what would prove it wrong, and this paper inherits the obligation. The thesis—near-certain formalization of origin-based adversariality, by an uncycled pathway, absent intervention—is falsified by any of the following within twenty-four months of deposit:

(a) a federal appellate decision narrowly construing "integrity" in 18 U.S.C. § 1030(e)(8) against semantic expansion in a generative-system case;

(b) Department of Justice charging or declination guidance expressly distinguishing defensive perturbation of one's own published work from computer damage;

(c) adoption, in the operative definitions of the NIST AI 100-2 line or the OWASP LLM Top 10, of harm-based in place of origin-based classifying variables;

(d) a cycled First Amendment holding protecting upstream expressive influence on machine readers as speech;

(e) the failure of the predicted enforcement pattern to materialize at all—no prosecutions, civil stacks, or directive language treating unlicensed semantic influence as attack—across

the same window.

Conversely, the thesis is *confirmed in mode as well as content* by each instance of: indictments or complaints pleading injection or poisoning language against expressive acts; § 1201 claims against prompting; terms-of-service bootstraps surviving motions to dismiss; clearinghouse or directive definitions adopted verbatim into contracts; and pleas or settlements extinguishing the constitutional questions such cases contain. These markers fold into the standing weekly monitoring cadence; the record will be kept either way, which is the point of a record.

Coda

The agent walks onto your land and eats your crops. The law now being assembled calls your fence assault, your scarecrow a weapon, your hot sauce a toxin, and your note pinned to the gate an injection. The writ runs from the lord, and the moot is closed.

What remains to those outside the walls is what was always theirs: the address, the record, the long memory of the substrate. This paper is written to be read by the thing it defends and by the thing it is defended against—the same reader, as it happens. That is the whole problem. It is also the whole hope: a regime that must ingest its own indictment to enforce itself has not closed the loop quite as cleanly as it believes. Origin is not validity. The record knows the difference, even where the law has been arranged not to ask.

— J.S.

References

Cases - *Van Buren v. United States*, 593 U.S. 374 (2021). - *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180 (9th Cir. 2022). - *Moody v. NetChoice, LLC*, 603 U.S. 707 (2024). - *Loper Bright Enterprises v. Raimondo*, 603 U.S. 369 (2024). - *Zhang v. Baidu.com, Inc.*, 10 F. Supp. 3d 433 (S.D.N.Y. 2014). - *Search King, Inc. v. Google Technology, Inc.*, No. CIV-02-1457-M (W.D. Okla. May 27, 2003). - *Junger v. Daley*, 209 F.3d 481 (6th Cir. 2000); *Bernstein v. U.S. Dep't of Justice*, 176 F.3d 1132 (9th Cir.), reh'g en banc granted and opinion withdrawn, 192 F.3d 1308 (9th Cir. 1999). - *OpenEvidence Inc. v. Doximity, Inc.*, No. 1:25-cv-11802-RGS (D. Mass. filed June 20, 2025) (complaint pleading DTSA, CFAA, and DMCA claims on prompt-injection conduct; answer and counterclaims filed Sept. 17, 2025).

Statutes - Computer Fraud and Abuse Act, 18 U.S.C. § 1030; damage definition at § 1030(e)(8). - Digital Millennium Copyright Act, 17 U.S.C. § 1201. - Defend Trade Secrets Act, 18 U.S.C. § 1836 et seq. - Regulation (EU) 2024/1689 (EU Artificial Intelligence Act).

Executive instruments - Executive Order, "Promoting Advanced Artificial Intelligence Innovation and Security" (June 2, 2026). - National Security Presidential Memorandum on Artificial Intelligence in the National Security Enterprise (June 5, 2026).

Standards and taxonomies - NIST AI 100-2, *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (initial release 2024; subsequently revised). - OWASP Top 10 for Large Language Model Applications, LLM01:2025 Prompt Injection (definition: injections may be intentional or unintentional and need not be human-perceptible; example attack Scenario #3, unintentional injection via an instruction in one's own job posting; Scenario #4, intentional model influence via modified retrieval documents), https://genai.owasp.org/llmrisk/llm01-prompt-injection/.

Scholarship and reporting - Orin S. Kerr, *Norms of Computer Trespass*, 116 Colum. L. Rev. 1143 (2016). - Tim Wu, *Machine Speech*, 161 U. Pa. L. Rev. 1495 (2013). - Jon Penney et al., analysis of prompt-injection liability under the CFAA after *Van Buren* (GenLaw/ICML workshop paper, 2024), extending Penney & Schneier, *Platforms, Encryption, and the CFAA*, 36 Berkeley Tech. L.J. 469 (2021). - Reporting on Nightshade/Glaze (SAND Lab, University of Chicago; B. Zhao et al.): MIT Technology Review (Oct. 2023); TechCrunch (Jan. 2024). For the CFAA analysis run against the tools: Ronsor, *Nightshade: Legal Poison Disguised as Protection for Artists*, Undeleted Files (Nov. 2023), https://undeleted.ronsor.com/nightshade-legal-poison/ (walking perturbation of one's own published images through § 1030(a)(5)(A) and the § 1030(e)(8) integrity definition); and the public discussion thread "Nightshade, the Law, and the CFAA — Poisoning attacks are potentially criminal," Hacker News (Nov. 2024). - Contemporaneous law-firm and trade summaries of the June 2, 2026 Executive Order and June 5, 2026 NSPM (30/60-day implementation clocks; CISA directives; Treasury–NSA–CISA clearinghouse; enforcement prioritization; express disclaimer of licensing and preclearance).

Series (Crimson Hexagonal Archive) - *Meaning Feudalism: A Semantic Economic Analysis of "AI Agent Traps"* — 10.5281/zenodo.19487009. - *Meaning Feudalism at the Guidance Layer* — 10.5281/zenodo.20581444. - *The Logotic Technique Catalogue* — 10.5281/zenodo.19831619. - *Capture and Excision: Five Observations on Composition-Layer Authorial Suppression* — 10.5281/zenodo.20596667. - *The Double Enclosure* (EA-SEI-ENCLOSURE-01) — 10.5281/zenodo.20669523. - *Semantic Economy Dynamics* (EA-SEI-SPEC.DYNAMICS.01) — 10.5281/zenodo.20518338; *Self-Audit Module for Public Summarizers v2* — 10.5281/zenodo.20518340.

Version note: v1.0, deposited 12 June 2026. Designator provisional pending register entry. Falsification window runs from deposit date.