Barriers to Evidence in AI-Related Cases and the Privatization of Proof

SARAH H. CEN[*], Carnegie Mellon University, USA
HANNAH ISMAEL[*], Mozilla Foundation, USA
LUCIA ZHENG, Stanford University, USA

Evidence lies at the core of litigation, but it is increasingly difficult to obtain in AI-related disputes. Even when a claimant's position has merit, cases are often settled or dismissed because decisive facts are hidden inside proprietary models, platform logs, and protected databases. Grounding our discussion in past and ongoing cases, we investigate how asymmetries in access, resources, and expertise can create functionally insurmountable barriers to producing AI evidence. We show how developers and deployers resist disclosure through various strategies challenging the value of the evidence to the requesting party and the cost of evidence production. From these patterns we identify seven recurring sources of asymmetry—access to models, data, documentation, logs, expertise, compute, and infrastructure—that create uneven evidentiary burdens at multiple stages of litigation. We argue that these recurring burdens reflect a broader pattern that we call the privatization of proof: a shift in effective authority away from courts and toward private actors via control over evidence. We propose a three-part test for determining when proof privatization occurs and when it prevents a claimant from accessing information that ordinary litigation should be able to surface, drawing on concepts such as proportionality and feasible alternatives.

Additional Key Words and Phrases: litigation, evidence, burden of proof, AI accountability, AI access

1 INTRODUCTION

Evidence lies at the core of every legal system. Without the ability to prove or defend against a claim, no legal action can be taken. Yet evidence is difficult to acquire. For a claim to be true is not enough—supporting evidence must be observed and captured, often even measured and quantified. Because legal decisions hinge on key evidence, parties who may be harmed by evidence closely guard it, doing their best to remove avenues to obtain it.

The role of evidence in legal proceedings is governed by a complex system of doctrines. Among these is the concept of the burden of proof, which assigns responsibility for raising issues, producing evidence, and ultimately persuading a factfinder (e.g., judge or jury). While the burden of proof (and who bears it) plays an important role in determining outcomes, it does not capture the full picture. In practice, evidence is produced and contested by both sides of a case at numerous points during legal proceedings. The generation and admissibility of evidence is ultimately shaped by complex procedural rules, gatekeeping standards, discovery obligations, and more.

In recent years, it has become increasingly clear that establishing a viable legal claim about AI is difficult. Parties challenging AI-related systems, actions, and outcomes often face evidentiary hurdles that are functionally insurmountable in practice, not because the governing legal standards are conceptually flawed, but because claimants frequently lack the access, expertise, and resources required to meet them. AI developers and deployers often claim that technologies are protected by trade secrecy or contractual obligations; expert knowledge is limited and many who possess it have conflicts of interest; moreover, the compute and infrastructure needed to test and reproduce AI behaviors is costly. As a result, even well-founded claims may fail before they can be meaningfully evaluated. This dynamic should be concerning even to those who do not directly use AI. It implies that AI-related decisions and outcomes may become effectively uncontestable without special access, knowledge, or support. As AI becomes further embedded across decision systems

[*] Both authors contributed equally to this research.

Authors' addresses: Sarah H. Cen, sarahcen@andrew.cmu.edu, Carnegie Mellon University, Pittsburgh, USA; Hannah Ismael, hannahi@mozillafoundation.org, Mozilla Foundation, San Francisco, USA; Lucia Zheng, Stanford University, Palo Alto, USA, zlucia@stanford.edu.

and institutions, the scope of this implication spreads alarmingly wide. In this Article, we investigate when and why producing AI-related evidence is difficult, grounding our discussion in past and ongoing cases.

Findings and recommendations. Across the cases we examine, we find that evidentiary hurdles in disputes involving automated and AI technologies frequently determine whether claims can proceed at all, with many being resolved or settled in the pre-trial stage. Asymmetries in information, expertise, and resources surface early in litigation, when claimants must make preliminary showings to obtain discovery, overcome admissibility barriers, or justify expert scrutiny. When such showings cannot be made, cases often end before courts reach the substance of the dispute.

In Section 3, we describe common types of arguments the producing party raises in response to the requesting party's efforts to access evidence uniquely within producing party's control in cases concerning AI and other automated technologies. We observe two recurring categories of arguments: (1) arguments that the requesting party has failed to show that access to the requested evidence is necessary to prove their claim, and (2) arguments that invoke production-side limitations on disclosure of evidence, independent of the requesting party's showing of need. Together, these two categories of arguments lead to a "Catch-22": that the requesting party cannot substantiate a request for access without access to further evidence, and the producing party is not compelled to grant access without further substantiation. Ultimately, this leads to what we call the **"privatization of proof": that key evidence is blocked and controlled by private actors.** In Section 4, we identify seven sources of asymmetry between parties in AI-related cases that contribute to proof privatization. We additionally make two observations that will motivate a three-part test for proof privatization in Section 5. First, significant attention has been paid to access to models, documentation, and recently data and user logs. However, in addition to these artifacts, **access to expertise, compute, and infrastructure are also critical** to disputing AI-related claims. Second, we argue that access is multidimensional and that different forms of access are often functionally exchangeable. That access types can be exchangeable allows **substitutions that preserve a requesting party's ability to gather evidence while allowing the producing party to protect artifacts that it views as most sensitive or confidential**. As our main contribution, in Section 5, we propose a three-part test for proof privatization in AI-related cases and evaluate a requesting party's request for AI access. We detail the test, then describe its connections to existing legal doctrines, including relevance, proportionality, and necessity.

2 BACKGROUND AND RELATED WORK

Information, resource, and expertise asymmetries. Many previous works emphasize that *parties who control relevant information can shape what is knowable in litigation, particularly through secrecy claims and institutional deference to confidentiality.* Wexler describes how secrecy claims can block scrutiny in high-stakes contexts and how courts may accept those claims with limited balancing against constitutional rights [61]. In parallel, socio-legal scholarship examine how evidentiary and procedural rules reflect power and capacity, e.g., Galanter presents a repeat-player framework, which predicts that *experienced, well-resourced parties are systematically advantaged* in navigating procedural demands and generating proof over time [23]. Further works (often empirical) find that adverse outcomes often point to a limited capacity to assemble legally sufficient evidence—due to constraints of representation, time, and knowledge—rather than weak claims [3, 52]. There is also documentation of how the high cost of legal services and unequal access to professional support shape who can meet proof requirements in practice [24, 48]. Finally, several works examine how *expert testimony is strictly gatekept, with uneven consequences.* Work on expert evidence and post-Daubert gatekeeping emphasizes that reliability-focused admissibility doctrines influence who can realistically meet proof demands in complex litigation [4, 21]. Other works question why expert evidence is subject to heightened scrutiny, underscoring how strict admissibility thresholds excludes relevant expert insights [53].

Procedural bottlenecks. Related works pinpoint unequal burdens at specific procedural moments. At the pleading stage, Dodson argues that plausibility pleading operates as an early evidentiary screen by implicitly demanding information often unavailable pre-discovery [20]. Reinert argues that modern pleading standards create a procedural self-defeating cycle: plaintiffs are required to plead specific facts to unlock the doors to discovery, yet those very facts are often hidden within the defendant's private records; by demanding evidence this early, these standards systematically bar meritorious claims where the defendant holds all the information [47]. Issacharoff and Miller critique the current motion-to-dismiss standard for allowing defendants to defeat a claim by highlighting a lack of evidence that only the defendants themselves possess. To fix this, they suggest rules that would require defendants to disclose key facts before they can seek a dismissal [26].

Although discovery doctrine is often framed as the response to informational asymmetry, it is also shaped by resource disparities. Rowe argues that the rules governing who pays for the gathering of evidence are as important as the law itself, because if a party cannot afford the cost of discovery, they are effectively stripped of the ability to prove their case [49]. Moreover, while discovery is a powerful mechanism for obtaining evidence, its scope is limited by Rule 26, which amended the definition of what is discoverable from anything that is "reasonably calculated" to lead to evidence to "proportional to the needs of the case" in order to prevent plaintiffs from going on costly "fishing expeditions." Mullenix warns Rule 26 can protect poorer litigants from being outspent by their opponents, but it can also provide wealthy defendants with a legal excuse to hide evidence by claiming it is too expensive to produce [38].

Another common critique is that discovery can be disproportional in providing too much evidence so as to raise legal costs [35]. Thus, the proportionality test also exists to limit the amount of evidence provided. Still, there are instances in which too much evidence is provided to overwhelm legal opponents with information.

Doctrinal responses: presumption, adverse inference, and spoliation. Finally, evidence law sometimes responds directly to systematic proof barriers with presumption and burden shifting. Doctrines such as *res ipsa loquitur* relax production burdens where direct evidence is inaccessible and defendants are better positioned to explain underlying events [45]. Scholarship on presumptions emphasizes that burden allocation reflects judgments about access to information and fairness, not merely probabilistic inference [13]. Related work on spoliation and Rule 37(e) examines how sanctions and adverse inferences attempt to rebalance proof burdens when evidence is lost or destroyed, while also highlighting the limits of remedial doctrine once information has disappeared [32, 56].

Discussions of evidence and transparency in AI governance. In parallel, debates over algorithmic governance and automated decision systems, scholars similarly argue that opacity and confidentiality can frustrate meaningful contestation and accountability [6, 11, 12, 18, 34, 43]. Related work focuses specifically on code and forensic technologies, highlighting how intellectual property and trade secrecy claims can constrain testing and challenge, including in criminal justice settings [28, 50, 51, 55]. Recent work emphasizes that resource inequality can also be infrastructural: compute, data access, and evaluation capacity are unevenly distributed, affecting who can generate and validate technical evidence [1, 8, 9]. When it comes to the scrutiny of expert testimony, concerns over gatekeeping expertise are increasingly salient in disputes involving technical systems, where parties may disagree not only about facts but about the methods required to establish them [18, 34]. Aligned with Issacharoff and Miller's [26] critique, the current standards allow defendants to defeat claims by pointing to a lack of evidence that only the defendants themselves possess, scholars argue that these dynamics are amplified because key facts about design choices, training data, and validation practices are typically held by defendants [29], producing a situation in which plaintiffs struggle to plead and prove discrimination without access to the very information that discovery is increasingly conditioned upon [5, 22].

3 HOW EVIDENCE IS BLOCKED IN AI, ALGORITHMIC, AND AUTOMATED TECHNOLOGY CASES

In this section, we describe common arguments the producing party raises in response to the requesting party's efforts to access evidence uniquely within producing party's control in cases concerning automated and AI technologies. We observe two recurring categories of arguments: (1) arguments that the requesting party has failed to show that access to the requested evidence is necessary to prove their claim, and (2) arguments that invoke production-side limitations on disclosure of evidence, independent of the requesting party's showing of need. We find that the interaction between these two categories results in an **evidentiary "Catch-22": that the requesting party cannot substantiate a request for access without access to further evidence, and the producing party is not compelled to grant access without further substantiation.**

3.1 Request-side Arguments Against Evidentiary Access

Below, we discuss three types of arguments against evidentiary access that challenge the request itself.

3.1.1 Producing parties argue that high-level descriptions or system outputs are sufficient, despite strong indications otherwise. Producing parties frequently challenge the necessity of implementation-level access by asserting that methodological information about the automated systems is sufficient to satisfy the evidentiary requirements of the requesting party's legal claim.

Courts have accepted this argument as a basis for denying discovery of source code in cases involving probabilistic DNA testing software. In *People v. Superior Court (Chubbs)*,[1], the California Court of Appeals denied the criminal defendant's motion to compel discovery of the source code underlying TrueAllele, a probabilistic DNA analysis system, for a Kelly/Frye hearing challenging the reliability of expert testimony based on new scientific techniques. The court reasoned that the requesting party's demonstrated understanding of 'TrueAllele's methodology, inferences, and reliance on the likelihood ratio" in her declaration and publicly available materials describing the TrueAllele methodology, such as "patent documents" and "published articles", "undercut" the asserted need for implementation-level access.[2] This reasoning **rests on the assumption that methodological descriptions supply the minimum evidence necessary for the requesting party to prove its claims about the contested system.** Methodologies typically specify a class of plausible implementations, not the particular system as deployed. Legally relevant differences often arise at the level of implementation-specific details that have direct bearing on the legal theory advanced by the requesting party. In probabilistic, AI systems, this includes, but is not limited to, data processing choices, hyperparameter selections, or the setting of thresholds values. Stated differently, AI systems are probabilistic (meaning that different outcomes can result from exactly the same inputs and also underdetermined (meaning that different systems can arise from nearly identical procedures). The result is: (i) descriptions of systems are often not sufficient to reproducing a system's errors or harms, and (2) while the description of a system may seem benign, it is still possible for the realized system to be harmful.

In contrast, the Superior Court of New Jersey granted implementation-level access to TrueAllele in *State v. Pickett.*[3] We note that, for the court to reach this conclusion, the requesting party's made arguments about FST, a probabilistic DNA system based on the same method at TrueAllele. An audit of the FST source code revealed an implementation-level

[1] People v. Superior Ct. (Chubbs) No. B258569, 2015 WL 139069 (Cal. Ct. App. Jan. 9, 2015)
[2] *Id.* at *8
[3] State v. Pickett, 466 N.J. Super. 270 (App. Div. 2021)

design choice in the system to exclude weakly informative loci, a choice not apparent from examining the methodology or documentation, which was later shown to overestimate likelihood of guilt.[4] Without this external system and the ability to audit its source code, such a determination would not have been possible.

3.1.2 Producing parties advance unverifiable claims about system behavior, leaving courts reliant on potentially biased or non-representative evaluations in place of independent evaluations. Producing parties object to requests for access enabling independent evaluation by **contending that expert testimony or existing validation studies (e.g., academic papers) provide sufficient evidence** for the requesting party to prove their claims about AI systems. When a propriety AI system is controlled by a developer with financial or litigation incentives aligned with the producing party, developer-generated evidence about the system may reflect those interests. Access enabling independent evaluation may be necessary for the requesting party to rebut potentially biased evidence about the system.

In *Chubbs*, the court denied the defendant's request for access enabling independent evaluation of TrueAllele, reasoning that existing expert testimony and peer-reviewed validation studies were sufficient for the requesting party to assess reliability. That evidence, however, was largely generated by Dr. Perlin, the system's developer, who had a financial interest in TrueAllele and testified as an expert for the prosecution. In *Pickett*, the court questioned the assumption implicit in *Chubbs* court's reasoning that access enabling independent evaluation was unnecessary when developer-generated evidence about the system was available, particularly when the developer was in exclusive control of the system and had incentives aligned with the adverse party. The court emphasized that six of the seven peer-reviewed validation studies relied upon were authored by Dr. Perlin himself, that the remaining study acknowledged his professional involvement, and that the publications expressly disclosed his financial conflict of interest. *Pickett* recognized the circularity of relying on potentially biased developer-generated evidence to deny access to independent evaluation necessary to rebut such evidence.

These positions and justifications could be informative of courts' responses to requests for deep access to AI systems, where the producing party controls the system and much of the evidence offered to explain or defend it, including public blogs and papers. Independent evaluation access may also enable the requesting party to test the system behavior on the specific populations, inputs, or decisions contexts relevant to the litigant's claim. A substantial body of work in machine learning has shown that model performance can vary significantly under distribution shift [17, 33], underscoring why prior validation studies are often a poor substitute for access that enables case-specific evaluation.

In *State v. Loomis*,[5] the Wisconsin Supreme Court held that the use of the COMPAS risk assessment at sentencing did not violate the defendant's due process right to be sentenced based on accurate information. Although the defendant could not review and challenge how the proprietary algorithm calculate risk, the court concluded that due process was satisfied because he could review and challenge the resulting risk scores in the presentence investigation report (PSI) , review the algorithm's inputs drawn from his criminal history and questionnaire responses, and prior validation studies have determined the system to be reasonably accurate. *Loomis* provides a useful context for examining the limits of treating access to a litigant's own input-output observation and general validation studies, as sufficient system access for the litigant to meaningfully contest a system's use in his case. Individual input-output observations or validation studies conducted on different reference populations or that do not isolate the defendant's class, cannot distinguish whether an outcome is driven by individualized factors specific to the defendant, group-level effects, or a mismatch between the system's reference population and the actual decision setting. The *Loomis* court explicitly recognizes these concerns

[4] *Id.* at 307-308

[5] State v. Loomis, 371 Wis.2d 235 (Wis., 2016)

by requiring any PSI containing a COMPAS score to include a written advisement detailing risks associated with the system use including its proprietary nature, that risk scores are based on group data, the lack of validation studies on the Wisconsin population, and the need for ongoing re-norming as populations change.[6] Yet, despite acknowledging these risks, the court's due process analysis ultimately stops short of requiring access that would enable litigants to independent evaluate whether those risks are implicated in their own cases and contest the systems use on that basis.

3.1.3 Producing parties question the relevance of the system to the disputed fact, undermining the basis for the access request. One way defendants have sought to deny access to AI systems is by classifying the technology as implementation or organizational tools rather than decision makers. In *Mehrara v. Canada*,[7] Canadian authorities defended the use of a technology, Chinook 3+, within their immigration process. IRCC frames Chinook 3+ as providing a "visual representation of a client's information" to immigration officers responsible for determining visa outcomes. While Chinook itself is not an algorithm, it does display applicants' information in a way that can systemtically influence officer's discretion, including a risk score predicted by the algorithm ITAT [39]. Another efficiency feature allowed officers to select amongst pre-selected reasons for denying individual's visa applications.

The framing had direct procedural consequences. The court did not grant access to Chinook spreadsheets, officers' working notes, or the ITAT risk indicator score. Although the spreadsheets were deleted daily, the latter two were not found relevant. No risk score was given to this applicant, so the plaintiffs could not investigate how ITAT is portrayed or whether the lack of a score affects outcomes. Moreover, because Chinook was **treated as an organizational tool, its internal presentation of information (such as the officer notes) was treated as immaterial to the outcome**. As a result, the claimant could not investigate whether the truncation or structuring of evidence might have constrained the officer's reasoning in practice. By defining the tool as a non-decisional interface, the informational environment that structured the officers' decision making and refusal is deemed irrelevant, and procedural scrutiny narrowed to the final judgement. The officer's reasoning, the court ruled, was unfettered by the technology.

However, Zeynab Ziaie Moayyed, a lawyer for Mehrara, questioned whether discretion meaningfully remains independent under such systems: "Is the officer really making a decision based on an application themselves, or...is the discretion being fettered by having this truncated single row of evidence for them" [63]? In later litigation, she described how Chinook's notes-generation interface allows officers to select standardized reasons for refusal. Notably, if their decision to deny an applicant is based on sufficiency of evidence, the officer may simply refuse the application, but if their concern is around credibility (that an applicant is being truthful in the reason around their trip), they have to contact the applicant to investigate and express these concerns. Moayyed notes in another Chinook case, *Jahanian v. Canada*,[8] "with the widescale adoption of Chinook and the uniform reasons generated, the language that typically signals that an officer was concerned about credibility is disappearing and is replaced by language that signals concerns about sufficiency of evidence" [62]. Credibility concerns might be disguised as evidentiary ones, which bypass safeguards that might allow the applicant further participation and rebuttal.

The same "AI as secondary to human-decision makers" framing appears in private litigation. In *Estate of Gene B. Lokken v. UnitedHealth Group*,[9] UnitedHealth defends its nH predict model, which is used "to determine whether Medicare Advantage patients should receive post-acute care" [41]. Plaintiffs argue that the tool overruled physicians and often resulted in the denial of life-saving treatment, which lends itself to breaking company policy: physicians are

[6] *Id.* at 276
[7] Mehrara v. Canada, [2024] F.C. 1554 (Can.)
[8] Jahanian v. Canada, [2024] F.C. 581 (Can.)
[9] Est. of Lokken v. Unitedhealth Grp., Inc., No. 23-CV-3514 (JRT/SGE), 2025 WL 2607196 (D. Minn. Sept. 8, 2025)

the ones making decisions regarding patients' care. While United insisted that the algorithm only acted as a guide for physicians, the high error rate and the affected populace of patients treated with the involvement of nH predict made the "AI as a guide" argument less believable. Public reporting indicates a roughly 90 percent error rate, with the affected populace being the elderly, who are more likely to either die before contesting the decisions or lack the technical or institutional knowledge to go about a contestation. United's insistence that AI acted as a guide sought to protect the model from scrutiny by placing responsibility on the acting physicians, but it was this very allocation of responsibility that would allow the breach of contract claims to proceed.

Unlike in *Mehrara*, plaintiffs were able to proceed by grounding their claims in a breach of company policy and bad faith. United's policies promise patients that their healthcare decisions be made by a physician. The plaintiffs argued that the deal was made in bad faith. Even if an acting physician did have nominal authority, the substance of the commitments were unmet given the model's deterministic nature.

3.2 Production-side Arguments Against Evidentiary Access

In contrast to production-side arguments that contest the necessity of the requested evidence, producing parties also resist evidentiary access by appealing to constraints internal to production. This division tracks the two elements of the standard for discovery in federal courts, relevancy and proportionality.[10] Where the arguments in Section 3.1 seek to defeat discovery by challenging the value of the evidence to the requesting party, the latter emphasizes the cost of producing the evidence to the producing party.

3.2.1 Private actors argue that access poses risk to competitive advantage, appealing to business confidentiality and trade secrecy. The *New York Times Co. v. Microsoft Corp.*[11] case illustrates how producing parties frequently resist evidentiary access by framing requested materials as competitively sensitive assets. At the onset of the case, the lead legal counsel denied access to the model's proprietary model information "on commercial grounds," an argument that echoes across other AI copyright cases [14]. In *Kadrey v. Meta Platforms, Inc.*[12], Meta explained that Llama 2's "data mixes are intentionally withheld for competitive reasons" [27]. Meanwhile, plaintiffs alleged that this framing works to "avoid scrutiny by those whose copyrighted works were copied and ingested during the training process for Llama 2." Plaintiffs are forced to resort to proxy forms of proof: inference from public datasets, or, as the NYT did in it's Exhibit F, random querying of parts of its copyright work to induce a record of the model regurgitating its copyrighted works [15]. **Competitive advantage arguments can work to gatekeep and slow access to training data or source code**, rather than allow a mediated understanding of what level of access is appropriately proportional.

3.2.2 Third-party privacy and confidentiality. **Model providers have also sought to block access by appealing to users' privacy concerns.** In *New York Times Co.*, Sam Altman met the request from plaintiffs for access to users' output data by invoking privacy concerns [44]. Output data is critical to plaintiffs for a number of reasons. First, output data can provide direct evidence that the model itself reproduces copyrighted works. Second, even when courts conceptualize the user as the infringer, OpenAI can be linked to contributing to infringement. Patterns of contributory behaviour that might be supported by output data include the frequency of text regurgitation across the range of minimally to maximally infringing users queries. As of June 2025, the court has ordered OpenAI to retain their output logs, even when users have deleted the chats on their end. OpenAI is appealing the decision, on grounds that the decision infringes

[10] Fed. R. Civ. P. 26(b)(1)

[11] New York Times Co. v. Microsoft Corp., No. 23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023)

[12] Kadrey v. Meta Platforms, Inc., No. 23-cv-03417 (N.D. Cal. filed July 7, 2023)

upon users' privacy. Sam Altman, in response, publicly called for something akin to "AI privilege" to protect users [44]. While privacy concerns have recently been invoked to block evidence, OpenAI itself has historically used user queries and output data as training data [30].

3.2.3 Access is framed as infeasible or unduly burdensome to produce. AI model providers and developers might also claim that providing access to, e.g., training data, is too technically difficult or cost intensive. In *Kadrey v. Meta Platforms, Inc.*, for example, the court ruled limited discovery to post-training data, as the raw dataset was "massive compared to the datasets actually used." Legal analysts note the training data size can be used "to [defendants'] advantage. Specifically, courts may be sympathetic to proportionality arguments when the data is burdensome to produce" [46]. If courts insist upon training data identification, there is often debate over who has to go through the identification process of relevant scraped works in the training data and by what degree of fineness. For example, in *New York Times Co. v. Microsoft Corp.*, the matter of who is primarily responsible for this article identification remained a central point of contention. Initially, OpenAI built the infrastructure for the news plaintiffs to do this investigation themselves. OpenAI provided "two virtual machines with computing resources. These machines were provided so that the counsel for NYT and Daily News could perform searches for their copyrighted content in its training sets" [31]. While the virtual machines are framed as granting access to plaintiffs, the sandbox structure externalized the technical labor, error risk, and verification burden onto plaintiffs despite OpenAI's exclusive control and higher familiarity with the training data.

These virtual machines, or the "sandbox," as it has been referred to in the case, came with several technical issues. First, mid-way through the process, "all of News Plaintiffs' programs and search result data stored on one of the dedicated virtual machines" got deleted [25]. NYT alleged this deletion occurred at the hands of OpenAI engineers, while OpenAI claims the deletion was the result of improper handling of the sandbox. Regardless of where the fault lies, once the query information was deleted, the retrieved information was "unreliable and [could not] be used to determine where the News Plaintiffs' copied articles were used to build Defendants' models" [57].

The NYT explained they had spent over 150 hours gathering the evidence, and that the deletion had made it clear that OpenAI was best positioned to identify the data [16]. After the deletion, the news' plaintiffs filed an expansive Request for Admissions. In response, OpenAI claimed that the request was so burdensome that it would be unreasonable for them to comply: "Plaintiffs need to engage with OpenAI and meet and confer before seeking a court order compelling an answer to nearly 500 million requests for admission" [58]. At present moment, OpenAI has agreed to bear further responsibility in helping the news' plaintiffs conduct the training corpus search, but the news' plaintiffs must provide more curtailed requests.

It seems that the Court has come to a resolution with regards to how the training data will be identified using two methods: (1) a six-word n-gram token matching and (2) URL based matching. This means that the News plaintiffs have specific baseline URLS (such as www.nytimes.com) for their articles and sets of six-word sequences drawn from each work, which OpenAI can then search against its training data. However, this method might undercapture infringement in certain scenarios, such as when works are republished on secondary websites, stripped of their original URLS, paraphrased, partially quoted, or excerpted in less the chosen n-gram length. While the argument that evidence identification is too technically infeasible might have limited plaintiffs' requests for admission and prevented a higher level of access (say, for example, any time a set of three words is regurgitated), it may be litigiously advantageous for plaintiffs to shift the burden of identification to model providers.

Together, these developments indicate that AI developers and deployers can **argue that requests for access are infeasible or unduly burdensome, often taking advantage of the court's lack of technical knowledge**. Yet requesting parties can potentially push courts to make AI developers and deployers carry a greater burden, especially with respect to their own systems.

4 PROOF PRIVATIZATION: SYSTEMATIC ASYMMETRIES IN FAVOR OF PRIVATE ACTORS

In this previous section, we argue that gathering AI evidence is increasingly blocked by a phenomenon we call the "privatization of proof": key channels to evidence are increasingly controlled by private actors. Building on our case analysis and well known facts in identify the seven sources of asymmetry between parties in AI-related cases that contribute to proof privatization. We additionally make two observations that motivate a test for proof privatization in Section 5. First, significant attention has been paid to access to models, documentation, and recently data and user logs. However, in addition to these artifacts, access to expertise, compute, and infrastructure are also critical to disputing AI-related claims. Second, we argue that access is multidimensional and that different forms of access are often *functionally exchangeable*. That access types are exchangeable allows for substitutions that preserve a requesting party's ability to gather evidence while allowing the producing party to protect sensitive or confidential artifacts.

4.1 Proof Privatization and Seven Sources of Asymmetry

AI disputes increasingly rely on information that exists but resides inside proprietary systems, across multi-actor systems, and behind technical expertise that are privately owned or protected. As a result, AI accountability can fail not because of a lack of evidence, but because of a lack of access to evidence. We argue that this outcome is due to the privatization of proof and argue that it is a structural threat to litigation of AI-related claims, blocking a key path to AI accountability. To keep AI cases contestable on the merits, courts need a principled way to detect and respond to proof privatization. We provide such a test in Section 5. To develop this test, we first identify seven sources of asymmetry between private actors (namely, AI developers and deployers) and the requesting party.

(1) **Model.** This includes access to all or a subset of the model weights and/or activations, access to the model's outputs, and API access [8, 9]. Model access is often critical because it allows direct testing of the system of interest on different inputs. Models are highly expensive to train, so replicating models is costly.
(2) **Data.** This includes pre-training, supervised fine-tuning, human feedback, and benchmark data. There are various forms of data access, including full database access, exact n-gram search, and fuzzy search [42, 54]. Data is highly expensive to gather, considered proprietary, or protected as private [36, 37].
(3) **Documentation.** We refer to documentation as information that reveals design decisions and organizational knowledge. In practice, documentation provides critical insights that are difficult to infer, but it is therefore the most revealing and thus generally considered to be highly confidential [59].
(4) **Post-deployment information (logs and operational traces).** Post-deployment traces serve as the primary source of information of how a system behaves and is used in practice. Recent litigation illustrates that the availability and retention of such logs can become a contested evidentiary issue [7, 60].
(5) **Expertise.** AI development and deployment as well as testing and evaluation generally require specialized knowledge and experise. Expertise also shapes whether a party can formulate targeted discovery requests and assess (or rebut) technical claims that factfinders may otherwise credit [19].
(6) **Compute.** Many operations—including red teaming and reproducing a model via training—require significant compute. Limited compute can therefore impose limits on what analyses are feasible [2, 10].
(7) **Infrastructure.** Similarly to compute, AI developers and developers often have access to tooling that greatly

assists in AI development and analyses. While lack of access to infrastructure is not necessarily prohibitive, it can create significant barriers for external parties [40].

We conclude with three takeaways. First, while debates over evidentiary asymmetry in AI disputes have often centered on *model access* and increasingly on *data access*, they are not determinative of how a system behaves. Depending on the context, lacking any key channel to evidence above can be prohibitive to the requesting party's ability to support their claim. Second, **one party in a suit typically has significantly greater access than the other to all artifacts above.** This creates critical set asymmetries between the two parties and their abilities to support AI-related claims or dispute AI-related facts throughout a case, as discussed in Section 3. In this way, the privatization of proof can be assessed by considering the asymmetry in access to the seven types of artifacts above. Third, the last three asymmetries are discussed less often even though expertise is frequently a determining factor of successful claims or discovery requests. For example, claimants often face a catch-22: unless they can substantiate a discovery request, they cannot gain discovery but without discovery, they do not have enough evidence to move forward. Expertise can greatly assist in such cases, allowing AI challengers to make targeted and informed requests. Further, compute and infrastructure are of utmost significant. Assuming that limited access to models, data, documentation, and logs are granted, the ability to independently test and evaluate AI systems often hinges on infrastructure and compute.

4.2 Different Types of Access are Functionally Exchangeable

Access in AI disputes is often treated as a one-dimensional question: greater or less access to a particular artifact such as the model, training data, or source code. One of our central observations is that access is *multidimensional*; specifically, most types of access can be *functionally exchangeable* with other types of access. Because they often serve similar evidentiary functions, denying one form of access does not always foreclose the successful gathering of evidence if other forms of access can be provided. For example, limited model access may be partly mitigated by richer documentation plus sufficient compute and infrastructure to run systematic tests; conversely, where documentation is contested, significant access to model, data, and logs may support many of the same inferences.

We build our main contribution in Section 5 on this observation: exchangeability makes it possible to treat access not as an all-or-nothing entitlement to specific artifacts, but as a set of negotiable options oriented toward an evidentiary objective. This exchangeability matters most in disputes where the producing party claims heightened risk from disclosure (e.g., trade secrecy, privacy, security, or misuse) while the requesting party seeks access for a specific evidentiary purpose. Exchangeability allows one to ask: *what access is sufficient to accomplish the requesting party's evidentiary task?* In many cases, that task can be satisfied through alternative combinations of access that pose materially different risks. For example, rather than producing sensitive datasets in full, a developer might provide structured evaluation sets, statistically representative slices, or controlled log access that enables the requesting party to test the contested proposition. Rather than disclosing full source code, the producing party might provide design and evaluation documentation, model cards, internal test results, or supervised testing access calibrated to the claim at issue.

This flexibility also helps explain why substitutions may be attractive even when they are not "cheaper" in any absolute sense. If access A is especially valuable to the requesting party but uniquely sensitive to the producing party, the producing party may rationally prefer to offer access B that is more costly (or less convenient) for the requester to use but materially less risky to disclose. Thus, substitutions can protect what the producing party is most concerned about while still providing key access to the requesting party.

5 SOLUTION: THREE-PART TEST

In this section, we build on our analysis and propose a three-part test that can be used to assess whether proof privatization has occured, the extent to which it has occured, and whether requested access is reasonable. We explain the main themes of the test and connect it to existing legal principles, including relevance, proportionality, and necessity.

5.1 Three-Part Test

We refer to the requesting party as the party seeking access in order to support a claim, and the producing party as the party that controls an artifact to which the requesting party is requesting access. The producing party may refer to multiple parties that, together, hold control over artifacts that the requesting party seeks.

Step 1: Degree of asymmetry. The first part of the test evaluates the degree of asymmetry between the requesting party and producing party, determining whether obtaining more access is necessary and relevant.

Mirroring the asymmetries given in Section 4, this step can be performed by evaluating each party's access to relevant (1) models, (2) data, (3) documentation, (4) compute, (5) infrastructure, (6) expertise, and (7) downstream information such as usage logs. Along all seven dimensions, the test should separately assess, at the time of evaluation, what artifacts the requesting party and producing party have access to. At this stage, the test should terminate if either one of the following holds: (i) the requesting party should be able to prove their claim with their current level of access, or (ii) the requesting party would not be able to prove their claim even with the same amount of access as the producing party. The checks (i) and (ii) provide indications of the **necessity** and **relevance**, respectively, of granting more access. The test only continues if the requesting party *cannot* prove their claim with the current level of access, and the requesting party *could* prove their claim if they were in the producing party's position.

Step 2: Justified access. The second part of the test determines whether privatization of proof has occurred, and, in doing so, creates a standard for proportionality in cases of proof privatization in AI-related cases. Proof privatization occurs when access to key artifacts is protected by the producing party. A critical choice in assessing proof privatization is establishing an appropriate notion of "key" evidence. We propose that it should be determined by the *minimal access needed to reproduce the cause of action.* Specifically, the test should assess the quantities:

1. Requested Access Benefit: across the seven types of access, the amount the requesting party would expend to emulate the equivalent amount of access to the requested level of access. This corresponds to the amount the requesting party would save by being granted the requested access.
2. Requested Access Risk: across the seven types of access, the monetary risk that the producing party would face if they were to grant the requested level of access.
3. Cause of Action Access Benefit: analogously to Requested Access Benefit, the amount the requester would expend to obtain access equivalent to the minimal access needed to reproduce the cause of action.
4. Cause of Action Access Risk: the monetary risk that the producing party faces by granting the requester the minimal access needed to reproduce the cause of action.

Then, the benefit-to-risk ratio of the requested access is the benefit of the requested access to the requesting party relative to the risk the producing party faces by granting the requested access. Similarly, the benefit-to-risk ratio of the cause of action access is the benefit of the cause of action access to the requesting party relative to the risk the producing party faces by granting the cause of action access. Then, if

$$\frac{\textsc{Requested Access Benefit}}{\textsc{Requested Access Risk}} \geq \frac{\textsc{Cause of Action Access Benefit}}{\textsc{Cause of Action Access Risk}},$$

the benefit-to-risk ratio of the requested access is greater than the benefit-to-risk ratio of the cause of action access, and the test favors the requested access. Otherwise, the test terminates. By comparing the requested access to the minimum access needed to reproduce the cause of action, the test ensures that the requesting party is entitled to some access and *anchors it at the cause of action.* Because the law already recognizes the cause of action as a legitimate basis for a claim, it provides a flexible yet acceptable threshold that is applicable across types of claims. It is also minimal in scope, as only gives the requesting party access to artifacts that are directly pertinent to the original cause of action.

Step 3: Alternatives and safeguards. The test only reaches this point if the request for access has passed Step 2. However, the producing party has one last defense that they can use to reduce or modify the access that requesting parties are granted. The producing party can do so by proposing alternative plans for access, including a different combination of artifacts, conditions under which access is granted, the use of a sandbox, phased discovery, and other protective measures. They could even propose to permit adverse inference instead of granting access if the producing party truly believes that confidentiality is paramount and no amount of access protects these interests. An alternative is acceptable only if (i) it provides functionally equivalent access with respect to what the requesting party seeks to prove and (ii) it does not decrease the benefit-to-risk ratio of the requested access as assessed in Step 2. In other words, it cannot result in functionally less access that what Step 2 grants the requesting party, *and* it cannot reduce the benefit-to-risk ratio accepted during Step 2.

5.2 Justification

A shared framework that applies existing doctrines to the problem of AI proof privatization. The three-part test that we propose respects existing procedural commitments and producing parties' confidentiality interests. US courts have long dealt with situations in which key evidence is held primarily or exclusively by one party. Doctrines governing discovery (the pretrial procedure through which parties obtain evidence), relevance (whether evidence bears on the disputed claim), proportionality (whether the burden of producing it is suitable relative to the needs of the case), cost-shifting (the allocation of discovery expenses between parties), spoliation (sanctions for the destruction of evidence), adverse inference (allowing the factfinder to draw unfavorable conclusions in the absence of evidence), and trade secret protection (safeguards limiting disclosure of sensitive business information) all reflect judicial efforts to manage informational asymmetries and the burdens of producing evidence [13].

For example, *Zubulake v. UBS Warburg LLC* was an employment-discrimination suit, where there was a request for emails for which many potentially responsive messages were stored on the defendant's backup tapes and would be costly to restore and search *Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 322–24 (S.D.N.Y. 2003).* The court used a seven-factor framework to assess whether to restore the emails, employing phased discovery and sampling to estimate the likely value of the information relative to the burden of production prior to ordering large-scale restoration. Ultimately, the court permitted restoration but employed cost-shifting, requiring the requesting party to bear a portion of the costs. The seven factors included (1) the specificity of the request, (2) the availability of the information from other sources, (3) the total cost relative to the amount at stake, (4) the total cost relative to the parties' resources, (5) each party's ability to control costs and incentives to do so, (6) the importance of the issues at stake, and (7) the relative benefits to the parties of obtaining the information. This and related cases demonstrate that courts recognize the risk that exclusive control over evidence can undermine the effective enforcement of substantive law, using tools like relevance, proportionality,

[13] Fed. R. Civ. P. 26(b)(1), 26(c), 37(e); Oppenheimer Fund, Inc. v. Sanders, 437 U.S. 340, 351 (1978); Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 322–24 (S.D.N.Y. 2003); Residential Funding Corp. v. DeGeorge Fin. Corp., 306 F.3d 99, 108–09 (2d Cir. 2002); Seattle Times Co. v. Rhinehart, 467 U.S. 20, 34–36 (1984)

cost-shifting, necessity, and cost-benefit analysis to assess the need to rebalance evidentiary burdens.

Yet existing doctrine addresses these issues in a fragmented manner. For example, discovery doctrine typically evaluates access through relevance and proportionality once litigation is underway, often assuming that requesting parties can reach the discovery stage without interrogating whether key evidence has been rendered inaccessible in the first place. Moreover, trade secret doctrine treats confidentiality as a competing interest to be balanced against disclosure, rather than as a structural feature of proof. As a result, courts may reach inconsistent outcomes when there is proof privatization. The proposed test draws on these familiar doctrines, consolidating them into a shared framework specialized for AI-related disputes. In particular, the strength of the test is the use of a flexible, context-dependent baseline, which allows the test to be applied across types of claims and at different stages of the litigation process. We explain the main rationale for each step.

Unpacking each step. In short, the test provides a standardized way to assess the degree of proof privatization and the proportionality of the requested access while ensuring that the access granted is relevant, necessary, and minimal. It is intended to guide courts wherever access disputes arise, at any stage of the litigation process.

Step 1 formalizes whether parties face a meaningful informational asymmetry and whether additional access is both necessary and appropriate. It examines whether the requesting party cannot prove the claim with their current level of access *and* whether the request for further access actually advances their ability to do so. If these initial conditions are not met, then the request for access is unwarranted and the test terminates.

While Step 1 evaluates relevance of the request for access, Step 2 returns to the problem central to evidentiary issues in AI-related cases: the degree of proof privatization. Building on Section 4, it assesses the degree of asymmetry in access to relevant (1) models, (2) data, (3) documentation, (4) compute, (5) infrastructure, (6) expertise, and (7) downstream information. The "access benefit" quantifies the amount the requesting party would expend to emulate the target amount of access. In other words, it is the amount that the requesting party is saved from expending if they were granted the target access. The access benefit can be insurmountably large in some cases, e.g., if the requesting party lacks access to large models and datasets that the producing party own, which may cost tens if not hundreds of millions to replicate. The "access risk" quantifies the monetary risk that the producing party would face if they were to grant the requested access; notably, risk does not simply capture the possible amount the producing party could lose, but also the likelihood of such an outcome. For instance, if the producing party believes that there is a possibility that access could leak a trade secret, then risk captures not only the value of this trade secret but also the likelihood of such a leak.[14]

There are two components of Step 2 that are critical: (i) the use of the **cause of action** as a baseline, and (ii) the **comparison of ratios that effectively measures the proportionality** of the requested access. Step 2 compares the access request to a legally grounded baseline: the *minimal access needed to reproduce the cause of action.* Thus, it does not permit the requesting party maximal transparency, but ensures that some access is granted, lower bounded by what the requesting party should minimally be entitled in order to make (or contest) a claim the law already recognizes. Using cause of action anchors Step 2 at stable reference point to distinguish between excessive fishing expeditions and access that is minimal for adjudication. Step 2 then compares the benefit-to-risk ratio of the requested access to the cause-of-action access. If the former exceeds the latter, then the requested access has a higher benefit-to-risk ratio than the baseline cause-of-action ratio, meaning that the test favors the requested access. This test almost literally maps to a quantitative test of proportionality. As sanity checks, note that the requester is not incentivized to request large

[14] The reason that the "access benefit" is not exactly analogous to risk in considering the chance of proving the claim with the target access is (i) the chance of proving a claim with the target access cannot be predetermined, as it central to the outcome of the suit, and (ii) Step 3 ensures that access is minimal and thus indirectly accounts for the possibility of over-estimating the numerator of the benefit-to-risk ratio.

amounts of access simply because the numerator of the benefit-to-risk ratio grows with greater access because the risk (the denominator) may also grow with greater access. Finally, note that, as a bare minimum, the requester could request the minimal access needed to reproduce the cause of action, which would be accepted by Step 2.

Step 3 allows alternative proposals to ensure that the access granted is **necessary and minimal** while **reducing risks to the producing party as much as possible given the amount of access granted**. While Step 2 requires some consideration of the relative benefits and risks of requesting access, Step 3 is a critical safeguard that ensures that the producing party can propose alternatives or revisions to the request, as long as it does not contradict reverse the guarantees provided by Step 2. Step 3 relies on the key observation from Section 4 that *access is not linear*: different forms of access can substitute for one another, meaning that if a producing party prefers to allow functionally equivalent (or more) access to the requesting party because they believe it poses less risk based on an internal calculation of their interests, then they should be permitted to do so. The producing party may even wish to permit adverse inference instead of granting access if some form of confidentiality is absolutely paramount to them that they are willing to risk an adverse outcome with respect to the claim of interest in order to protect these interests.

Why the test does not add "business confidentiality" as an override in Step 3. The typical discussion of AI "access" revolves around the push-and-pull between the need for access and the desire to protect business confidentiality. Such conversations often come to a standstill because claims of confidentiality are permitted to override the need for access, even when proving that access would pose a significant risk to, e.g., trade secrets, is not required. Against this backdrop, we interrogate precisely what poses a risk to effective adjudication and arrive at the issue of privatization of proof. Thus, the test focuses not on access versus business confidentiality, but on the relevance, proportionality, and necessity of the requested access. In doing so, this test still **places significant weight on confidentiality interests while avoiding the pitfalls of a vague exception that can swallow access altogether**.

The framework assumes from the outset that the entitlement is bounded by what is minimally required to reproduce the claim. Confidentiality remains protected not via a vague exception, but via the combination of (i) an assessment of necessity and relevance in Step 1, (ii) a calculaton of risk to the producing party in Step 2, and (iii) the allowance of procedural safeguards, substitutions, and alternatives in Step 3.

6 CONCLUSION

Although less often discussed than legislative or administrative actions, private ligitation is a key avenue for AI accountability. However, for it to remain viable, parties bringing meritous claims about AI systems, decisions, and outcomes must be able to produce evidence that supports their claims. In this Article, we study the barriers AI "challengers" face when requesting access to key evidence. From our case analysis, we identify several key trends. For one, claims about automated, algorithmic, and AI technologies often fail during pre-trial stages due to barriers to evidence, often due to a self-defeating cycle where the requesting party cannot gather clear evidence to support a request for discovery, but without discovery, they cannot produce evidence needed to move forward with the case. We pinpoint the underlying problem as one of "proof privatization": that major channels to evidence are increasingly controlled by private actors, shifting authority from the court to the private actors. In response, we propose a three-part test for proof privatization in AI-related cases, building on existing legal principles.

ACKNOWLEDGEMENTS

We thank Riana Pfefferkorn and lawyers who wish to remain anonymous for their invaluable feedback and guidance.

REFERENCES

[1] Nur Ahmed and Muntasir Wahed. 2020. The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research. In *arXiv preprint*. https://arxiv.org/abs/2010.15581

[2] AI Now Institute. 2018. Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems.

[3] Catherine R. Albiston and Rebecca L. Sandefur. 2013. Accessing Justice in the Contemporary USA. *Annual Review of Law and Social Science* 9 (2013), 101–119.

[4] David E. Bernstein. 2014. The Misbegotten Judicial Resistance to the Daubert Revolution. *Notre Dame Law Review* 89, 1 (2014), 27–70.

[5] Laura J. Bernt. 2021. Workplace Transparency Beyond Disclosure. In *Marquette Law Review*, Vol. 105. 73–129. https://fairemploymentproject.org/images/Workplace_Transparency_Beyond_Disclosure%2C_Marquette_Law_Review_2021.pdf

[6] Hannah Bloch-Wehba. 2020. Access to Algorithms. *Fordham Law Review* 88, 4 (2020), 1265–1313.

[7] Blake Brittain. 2025. OpenAI loses fight to keep ChatGPT logs secret in copyright case. *Reuters* (3 December 2025). https://www.reuters.com/legal/openai-loses-fight-keep-chatgpt-logs-secret-copyright-case-2025-12-03/

[8] Stephen Casper, Carson Ezell, Charlotte Siegmann, Noam Kolt, Taylor Lynn Curtis, Benjamin Bucknall, Andreas Haupt, Kevin Wei, Jérémy Scheurer, Marius Hobbhahn, Lee Sharkey, Satyapriya Krishna, Marvin Von Hagen, Silas Alberti, Alan Chan, Qinyi Sun, Michael Gerovitch, David Bau, Max Tegmark, David Krueger, and Dylan Hadfield-Menell. 2024. Black-Box Access is Insufficient for Rigorous AI Audits. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. ACM, New York, NY, USA. arXiv:2401.14446.

[9] Sarah H. Cen and Rohan Alur. 2024. From Transparency to Accountability and Back: A Discussion of Access and Evidence in AI Auditing. In *Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '24)* (San Luis Potosi, Mexico). ACM, New York, NY, USA. arXiv:2410.04772.

[10] Sarah H. Cen, Salil Goyal, Zaynah Javed, Ananya Karthik, Percy Liang, and Daniel E. Ho. 2025. Audits Under Resource, Data, and Access Constraints: Scaling Laws for Less Discriminatory Alternatives. (2025). Preprint (under review).

[11] Danielle Keats Citron. 2008. Technological Due Process. *Washington University Law Review* 85, 6 (2008), 1249–1313.

[12] Danielle Keats Citron and Frank Pasquale. 2014. The Scored Society: Due Process for Automated Predictions. *Washington Law Review* 89, 1 (2014), 1–33.

[13] Edward W. Cleary. 1959. Presumptions as Rules of Law. *California Law Review* 47, 2 (1959), 257–266.

[14] The New York Times Company. 2023. Complaint, The New York Times Company v. Microsoft Corporation, OpenAI, Inc., et al. https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf United States District Court for the Southern District of New York, Case No. 1:23-cv-11195 (SHS). Accessed 2026-01-12.

[15] The New York Times Company. 2023. Exhibit F-1 to Complaint, The New York Times Company v. Microsoft Corporation, OpenAI, Inc., et al. https://nytco-assets.nytimes.com/2023/12/Lawsuit-Document-dkt-1-51-Ex-F-1.pdf United States District Court for the Southern District of New York, Case No. 1:23-cv-11195 (SHS), Docket No. 1-51. Accessed 2026-01-12.

[16] The New York Times Company and LP Daily News. 2024. Letter to Magistrate Judge Wang in The New York Times Company v. Microsoft Corporation, et al. No. 1:23-cv-11195 (SHS) (OTW), ECF No. 328 (S.D.N.Y. Nov. 20, 2024), https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.328.0.pdf. Filed in *The New York Times Company v. Microsoft Corp.*, 1:23-cv-11195 (S.D.N.Y.).

[17] Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D Hoffman, et al. 2022. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research* 23, 226 (2022), 1–61.

[18] Deven R. Desai and Joshua A. Kroll. 2017. Trust But Verify: A Guide to Algorithms and the Law. *Harvard Journal of Law & Technology* 31, 1 (2017), 1–64.

[19] Digital Regulation Cooperation Forum. 2022. *Auditing Algorithms: The Existing Landscape, Role of Regulators and Future Outlook*. Technical Report. UK Government. https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook

[20] Scott Dodson. 2010. New Pleading, New Discovery. *Michigan Law Review* 109, 1 (2010), 53–91.

[21] David L. Faigman. 2001. Expert Evidence After Daubert. *Hastings Law Journal* 52, 4 (2001), 1035–1047.

[22] Virginia Foggo and John Villasenor. 2021. Algorithms, Housing Discrimination, and the New Disparate Impact Rule. In *Science and Technology Law Review*, Vol. 22. 1–62. https://doi.org/10.7916/stlr.v22i1.7963

[23] Marc Galanter. 1974. Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change. *Law & Society Review* 9, 1 (1974), 95–160.

[24] Gillian K. Hadfield. 2017. *Rules for a Flat World: Why Humans Invented Law and How to Reinvent It for a Complex Global Economy*. Oxford University Press.

[25] Julia Helmer. 2024. *This Week in eDiscovery: eDiscovery Day Special Edition | Discovery Challenges of AI Datasets; Deleted ESI Sanctions*. JD Supra. https://www.jdsupra.com/legalnews/this-week-in-ediscovery-ediscovery-day-8948648/

[26] Samuel Issacharoff and Geoffrey P. Miller. 2013. An Information-Forcing Approach to the Motion to Dismiss. *Journal of Legal Analysis* 5, 2 (2013), 437–465.

[27] Richard et al. Kadrey. 2024. Corrected Second Consolidated Amended Complaint. https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.133.0.pdf United States District Court for the Northern District of California, Case No. 3:23-cv-03417-VC, Document 133.

Accessed 2026-01-12.

[28] Sonia K. Katyal. 2019. The Paradox of Source Code Secrecy. *Cornell Law Review* 104, 5 (2019), 1183–1262.

[29] Pauline T. Kim. 2017. Data-Driven Discrimination at Work. In *William & Mary Law Review*, Vol. 58. 857–936. https://wmlawreview.org/sites/default/files/Kim.pdf

[30] Jennifer King, Kevin Klyman, Emily Capstick, Tiffany Saade, and Victoria Hsieh. 2025. User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies. arXiv:2509.05382 [cs.CY] https://arxiv.org/abs/2509.05382

[31] Kate Knibbs. [n. d.]. *New York Times Says OpenAI Erased Potential Lawsuit Evidence.* https://www.wired.com/story/new-york-times-openai-erased-potential-lawsuit-evidence/

[32] Margaret M. Koesel, Tracey L. Turnbull, and Daniel F. Gourash. 2017. *Spoliation of Evidence: Authors' Advice and Caveats for Civil and Criminal Litigation* (4th ed.). American Bar Association.

[33] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. 2021. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*. PMLR, 5637–5664.

[34] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable Algorithms. *University of Pennsylvania Law Review* 165, 3 (2017), 633–705.

[35] Alexandra D. Lahav. 2018. Procedural Design. *Vanderbilt Law Review* 71, 3 (2018), 821–886.

[36] Katherine Lee, A. Feder Cooper, James Grimmelmann, and Daphne Ippolito. 2023. AI and Law: The Next Generation. *SSRN* (2023). http://dx.doi.org/10.2139/ssrn.4580739.

[37] Katherine Lee, A. Feder Cooper, and James Grimmelmann. 2023. Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain. *arXiv preprint arXiv:2309.08133* (2023).

[38] Linda S. Mullenix. 2016. The New Discovery Rules and the Future of Civil Litigation. *University of Kansas Law Review* 64, 4 (2016), 839–880.

[39] Government of Canada. 2022. CIMM — Chinook Development and Implementation in Decision-Making — February 15 & 17, 2022. https://www.canada.ca/en/immigration-refugees-citizenship/corporate/transparency/committees/cimm-feb-15-17-2022/chinook-development-implementation-decision-making.html

[40] Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. 2025. Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling. In *CHI Conference on Human Factors in Computing Systems (CHI '25).* https://doi.org/10.1145/3706598.3713301 Accessed 2026-01-11.

[41] Helen Oscislawski. 2025. *Judge Decides Class Action Can Proceed Against UnitedHealth for Use of AI.* LegalHIE. https://www.legalhie.com/judge-decides-class-action-lawsuit-can-proceed-against-unitedhealth-for-use-of-ai/

[42] Ashwinee Panda, Xinyu Tang, Milad Nasr, Christopher A Choquette-Choo, and Prateek Mittal. 2025. Privacy auditing of large language models. *arXiv preprint arXiv:2503.06808* (2025).

[43] Frank Pasquale. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press, Cambridge, MA.

[44] Sarah Perez. 2025. *Sam Altman warns there's no legal confidentiality when using ChatGPT as a therapist.* https://techcrunch.com/2025/07/25/sam-altman-warns-theres-no-legal-confidentiality-when-using-chatgpt-as-a-therapist/ Accessed: 2026-01-13.

[45] William L. Prosser. 1949. Res Ipsa Loquitur: A Modern Doctrine. *Minnesota Law Review* 33, 6 (1949), 583–612.

[46] Sasha S. Rao and Richard A. Crudo. 2025. Discovery of Training Data in AI Litigation. (30 April 2025). https://www.sternekessler.com/news-insights/insights/discovery-of-training-data-in-ai-litigation/ Accessed 2026-01-11.

[47] Alexander A. Reinert. 2011. The Costs of Heightened Pleading. *Indiana Law Journal* 86, 1 (2011), 119–171.

[48] Deborah L. Rhode. 2004. Access to Justice. *Fordham Law Review* 73, 3 (2004), 1013–1028.

[49] Thomas D. Jr. Rowe. 1982. Discovery Cost Allocation and Burden Shifting. *Boston College Law Review* 23, 5 (1982), 1233–1283.

[50] Meghan J. Ryan. 2020. Secret Algorithms, IP Rights, and the Public Interest. *Nevada Law Journal* 21, 1 (2020), 61–104.

[51] Meghan J. Ryan. 2022. Criminal Justice Secrets. *American Criminal Law Review* 59, 4 (2022), 1541–1596.

[52] Rebecca L. Sandefur. 2014. What We Know and Need to Know About the Legal Needs of the Public. *South Carolina Law Review* 67, 2 (2014), 443–459.

[53] Frederick Schauer. 2010. Is Expert Evidence Really Different? *University of Tulsa Law Review* 45, 4 (2010), 635–648.

[54] Judy Hanwen Shen, Ken Liu, Angelina Wang, Sarah H Cen, Andy K Zhang, Caroline Meinhardt, Daniel Zhang, Kevin Klyman, Rishi Bommasani, and Daniel E Ho. 2025. Fallacies of Data Transparency: Rethinking Nutrition Facts for AI. In *ICML Workshop on Technical AI Governance (TAIG).*

[55] Eli Siems, Katherine J. Strandburg, and Nicholas Vincent. 2022. Trade Secrecy and Innovation in Forensic Technology. In *Hastings Law Journal*, Vol. 73. 773–824.

[56] Adam N. Steinman. 2012. Adverse Inferences About Adverse Inferences. *Georgia Law Review* 46, 3 (2012), 651–688.

[57] Ars Technica. 2024. *Tech Problems Plague OpenAI Court Battles; Judge Rejects a Key Fair Use Defense.* Ars Technica. https://arstechnica.com/tech-policy/2024/11/tech-problems-plague-openai-court-battles-judge-rejects-a-key-fair-use-defense/

[58] The New York Times Company and Daily News, LP. 2024. Joint Letter Regarding Training Data Issues in The New York Times Company v. Microsoft Corporation, et al. No. 23-cv-11195 (SHS) (OTW), ECF No. 305 (S.D.N.Y. Nov. 1, 2024), https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.305.0.pdf.

[59] The Sedona Conference Working Group 12. 2022. The Sedona Conference Commentary on Protecting Trade Secrets in Litigation About Them. *The Sedona Conference Journal* 23, 1 (March 2022), 1–172. https://thesedonaconference.org/publication/Commentary_on_Protecting_Trade_Secrets_

in_Litigation_About_Them

[60] U.S. District Court for the Southern District of New York. 2025. Preservation Order. In re: OpenAI, Inc. Copyright Infringement Litigation, No. 1:25-md-03143. Magistrate Judge Ona T. Wang; affirmed by District Judge Sidney H. Stein on June 26, 2025.

[61] Rebecca Wexler. 2018. Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System. *Stanford Law Review* 70, 5 (2018), 1343–1429.

[62] Zeynab Ziaie Moayyed. 2024. *Chinook assisted visa decisions are often shielded from scrutiny due to the limited language available in the Chinook note generator tool.* LinkedIn. https://www.linkedin.com/posts/zeynab-ziaie-moayyed_chinook-assisted-visa-decisions-are-often-activity-7186015470909587456-iRlC/

[63] Zeynab Ziaie Moayyed. 2025. *Chinook, AI Triaging and ITAT, with Zeynab Ziaie Moayyed.* YouTube. https://www.youtube.com/watch?v=TA-EOQiFUbU

GENERATIVE AI USAGE STATEMENT

Generative AI was not used to generate the text in this Article. The authors used Generative AI to brainstorm some parts (though not a substantial portions) of this Article. Generative AI was additionally used to check for missing, relevant citations as well as to proofread, edit, and improve language inthe text.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009