SecurIT360©

WHITEPAPER

AI Security Isn't New. Your Gaps Are.

A Practitioner's Framework for Law Firms:
Governance, Adversarial Testing, and Continuous Monitoring

CONTRIBUTING EXPERTS

Robert McElroy · AI Governance & Risk Frameworks
Brad Causey · Adversarial Testing & Offensive Security
Joey Vandegrift · SOC Operations & Continuous Monitoring

securit360.com | 2026

Executive Summary

Law firms are deploying AI faster than they are securing it. Individual AI use among legal professionals has more than doubled in a single year, reaching 69% in 2026 according to the 8am Legal Industry Report. Firm-level governance has not kept pace: 43% of firms have no formal AI policy and no plans to create one. Only 9% have a written, actively enforced policy. Research copilots, document drafting tools, and agentic workflows are being introduced into environments where a single data exposure can destroy attorney-client privilege, trigger bar complaints, and cost the firm the client relationship that took a decade to build.

The firms that will navigate this well are not the ones moving slowest. They are the ones moving with three capabilities in place: governance that defines what AI systems are allowed to do, adversarial testing that validates whether those definitions hold up, and continuous monitoring that ensures the answer stays accurate over time. Most firms have one of these, at most. Almost none have all three working together.

This paper makes a pointed argument: AI security without governance is guesswork. The only defensible model is a continuous cycle of governance, testing, and monitoring, and the only firms that will be able to answer their clients' growing AI security questions with confidence are the ones that have built it.

The perspective here comes from three SecurIT360 practitioners who work this problem from different angles every day: Robert McElroy on AI governance and risk frameworks, Brad Causey on adversarial testing and offensive security, and Joey Vandegrift on SOC operations and managed detection and response.

The Real Risk Is Not What You Think

Most conversations about AI security in the legal industry focus on the wrong threat. The scenario that dominates boardroom discussion is dramatic: a rogue model, a catastrophic breach, a headline-generating attack. That is not the scenario that security practitioners are losing sleep over.

The realistic AI security incident is quiet. An AI system consumes untrusted content. Subtle manipulation influences its behavior or output. A legal research tool produces a plausible but fabricated case citation. A contract review tool misses an injected clause in a document it was fed. An AI drafting assistant surfaces confidential information from a prior matter because no data boundary was ever defined. No malware runs. No credentials are stolen. No alerts fire. The damage is discovered weeks later, if it is discovered at all.

Joey Vandegrift, who runs SecurIT360's SOC operations, frames the consequence plainly: the result is loss of confidentiality, flawed decision-making, and erosion of trust, often discovered only after the impact has already spread. For a law firm, that impact is not just technical. It is professional. Attorney-client privilege, once destroyed, cannot be restored by a notification letter or a credit monitoring service.

How a Law Firm Loses Privilege Through AI in Five Steps

Step one: an associate uses an AI research tool to pull background on a matter. The tool has access to the firm's internal document repository, which was connected six months ago during a productivity initiative no one fully scoped.

Step two: the AI surfaces a document from a prior, unrelated matter involving a different client. The associate does not recognize the source. The AI presented it with the same confidence it presents everything.

Step three: a summary referencing that prior matter is incorporated into a memo. The memo goes to the client.

Step four: opposing counsel, during discovery, requests all materials that informed the memo. The prior matter document surfaces. Privilege is challenged. The firm has no clean answer for how an unrelated client's confidential information ended up in another client's work product.

Step five: the firm learns that the AI tool has been pulling from the shared repository for months. No one had defined what it was allowed to access. There was no policy. There was no audit trail. There was no monitoring.

No attacker was involved. No system was breached. The firm simply deployed an AI tool without governance, and the tool did exactly what it was configured to do.

> The biggest AI risk is not adoption itself. It is adoption without boundaries. — Joey Vandegrift

Law firms are uniquely exposed to this risk because of where AI is being deployed. It is not in back-office workflows. It is in legal research, document review, client communications, and due diligence, operating directly on the firm's most sensitive and legally protected data. The attack surface is not theoretical. It is the matters currently in progress.

What Is Actually New About AI Security

Here is the part that surprises most practitioners: the security fundamentals have not changed. Identity management, access control, endpoint protection, logging, continuous monitoring, and user training remain the most important controls in any AI security program. Robert McElroy has been making this argument for years: "Security is security is security. Nothing is fundamentally changed. It is logging, it is backups, it is access control."

What has changed is this: in AI environments, risk is defined by policy and context, not just technical configuration. An AI agent that can push documents to external storage is either a severe data exfiltration risk or an approved productivity workflow, depending entirely on what the organization has decided it should be allowed to do. The same behavior. Two completely different findings. Without a policy that defines intent, there is no way to tell them apart.

This is the structural reality that most AI security programs are not built around. It is why governance is not a parallel track to security work in AI environments. It is a prerequisite for security work to have any meaning at all.

> Most AI governance policies in existence today are performative. They state intent. They do not reflect how AI is actually being used inside the organization, and they will not survive contact with a client questionnaire, a regulator, or an incident.

A Three-Pillar Model for AI Security

Effective AI security requires three interconnected capabilities working in a continuous cycle. Remove any one of them and the model fails, not gradually, but structurally.

PILLAR ONE **Governance** Define What Should Happen	PILLAR TWO **Validation** Test Whether It Works	PILLAR THREE **Monitoring** See What's Actually Happening
— Approved AI tools & use cases — Data boundaries & handling rules — AI inventory & impact assessments — NIST AI RMF / ISO 42001 alignment — User training & acceptable use — Ongoing governance & audit cycles	— Test governance assumptions — Prompt injection & jailbreak probing — Data leakage vector assessment — Agent & RAG pipeline testing — OWASP LLM Top 10 coverage — MITRE ATLAS tactics mapping	— Shadow AI detection & control — MDR/EDR coverage for AI activity — SSE, CASB, and DLP integration — AI systems as privileged actors — Prompt injection monitoring — Audit trails for AI interactions

The relationship is cyclical, not sequential. Governance defines intent. Testing validates whether that intent holds under adversarial conditions. Monitoring confirms whether it holds over time. What monitoring surfaces feeds back into governance and triggers new rounds of testing. The cycle does not end because the AI environment does not stop changing.

Pillar One: Governance

Governance is where AI security begins, and it is where most organizations are weakest. The core question governance must answer is simple: what are our AI systems supposed to do? Most firms have not answered it in any structured way. AI tools have proliferated through procurement, individual adoption, and vendor integrations, without the policy framework needed to manage them as part of a coherent security posture.

The most immediate consequence of that gap is shadow AI. Research consistently finds that 59% of employees use unapproved AI tools, and 75% share sensitive data with them. At law firms,

employees are uploading client documents to consumer AI platforms, using personal subscriptions for work tasks, and installing open-source models locally, often without any awareness that the data they are uploading is no longer under the firm's control. As Joey Vandegrift puts it: "If you haven't written a policy, or even if you have but you're not doing good education, employees are uploading documents into AI tools — and once that data is uploaded, you cannot control what happens to it." For a law firm, that is not a policy violation. It is a potential privilege waiver.

What Governance Actually Requires

Robert McElroy's governance framework draws on the NIST AI Risk Management Framework and ISO 42001, not because frameworks are the point, but because they provide a repeatable structure for answering questions that organizations consistently avoid: which AI tools are approved, what data can they access, who owns decisions when something goes wrong, and how do we know our policies are actually being followed.

The work is less glamorous than most firms expect. It is building an AI inventory, cataloging every system, integration, and RAG pipeline in active use. It is conducting impact assessments before deployment, not after an incident. It is writing a policy that reflects the firm's actual tools and workflows, not a downloaded template with the firm's name pasted in. Robert McElroy is direct about the template problem: a policy that only states intent will not hold up under scrutiny from clients, regulators, or in the aftermath of an incident.

The governance cycle connects directly to monitoring. DNS filtering, CASB tools, and SSE platforms are not just IT controls. They are the enforcement layer that allows governance to move from a document to an operational reality, surfacing shadow AI usage, enforcing data handling rules, and creating the audit trails that clients are increasingly asking to see.

> **The governance-before-testing rule**
>
> AI pen testing without governance in place is largely a waste of money. Brad Causey encounters this regularly: clients who want an AI penetration test before they have defined what their AI systems are supposed to do. The problem is structural. Without governance, there is no baseline to test against. A pen test produces a list of AI behaviors. Governance determines which of those behaviors are violations and which are features. Without that distinction, the report is a collection of observations, not findings. Governance first. Then test.

Pillar Two: Adversarial Validation

Policy expresses intent. Testing determines whether intent survives contact with a motivated attacker. These are not the same thing, and in AI environments the gap between them is often larger than organizations expect.

Why AI Pen Testing Is Different

Traditional penetration testing produces largely universal findings. A critical vulnerability in a public-facing server is a critical vulnerability regardless of organizational context. The remediation path is clear.

AI pen testing does not work that way. The same behavior, an AI agent capable of pushing documents to external storage, might be a severe data exfiltration risk in one firm and an approved workflow in another. Context defined by governance determines which it is. This is why Brad Causey refuses to start an AI engagement without first understanding the client's policy framework: "A finding is not a finding. A finding is a piece of information that we bump up against policy — and then it becomes a violation. Without your policy, I don't even know what to test."

The testing methodology covers the full OWASP Top 10 for LLM Applications 2025 and MITRE ATLAS threat taxonomy: prompt injection and input manipulation, data exposure and leakage, model and pipeline security, supply chain vulnerabilities, shadow AI and access control failures. Critically, it blends AI-specific attack techniques with traditional security testing, because the most significant real-world attack paths combine both. In a law firm environment this means testing whether a RAG pipeline connected to a matter management system can be manipulated to surface documents from unrelated client matters, whether a contract review tool can be fed injected instructions through the documents it processes, and whether an AI research assistant will hallucinate case law in ways that are plausible enough to pass initial review.

The output is structured for multiple audiences: an executive summary calibrated for senior leadership, detailed findings with OWASP and ATLAS mappings for security teams, and a defensive playbook that closes the loop back to governance with specific policy updates and a schedule for the next round of testing.

Testing as a Feedback Loop

Adversarial testing is not a one-time event that produces a report and gets filed. It is a recurring input to the governance cycle. When testing reveals that an AI agent can exfiltrate documents in a way the policy did not anticipate, the response is a policy update, followed by retesting to confirm the fix held, followed by monitoring to confirm it continues to hold. Brad Causey

describes the dynamic: "We loop through that triangle over and over until the pen test comes back clean — and then we schedule the next one."

The firms that treat AI pen testing as a recurring discipline rather than a compliance checkbox are the ones whose governance frameworks actually mature over time. The ones that treat it as a one-time event are the ones who discover their policy gaps the hard way.

Pillar Three: Continuous Monitoring

Governance and testing describe what should be happening and whether the current state matches that description. Monitoring answers a different question: what is actually happening right now? It is the pillar most organizations underinvest in, and the gap between the security posture documented on paper and the one that exists in practice is widest here.

The SOC Reality: You Are Already Ahead of Where You Think

Here is a contrarian claim that the data supports: for the majority of AI-related threats, the SOC already has what it needs. This is not a reason to relax. It is a reason to stop misdirecting security investment.

When an AI-assisted attack reaches the endpoint, identity layer, or network, it generates behavioral signals that are structurally identical to what a human attacker generates. Account compromise looks like account compromise. Lateral movement looks like lateral movement. Modern MDR and EDR tools detect these behaviors well regardless of whether the attacker is human or AI-assisted, and the response is the same: disable the account, revoke sessions, investigate. Firms spending heavily on AI-specific monitoring tools before they have continuous SOC coverage are solving the wrong problem in the wrong order.

The monitoring gaps that actually matter are upstream, before the threat reaches the endpoint. Three areas stand out: prompt injection attacks that manipulate AI system behavior without generating any system-level events that traditional security tools can detect; shadow AI usage that is invisible unless specific controls like DNS filtering and CASB solutions are in place to surface it; and AI agents operating as non-human identities, taking actions that look like legitimate automation until they do not. In a law firm, these gaps are not abstract. They are the associate running a client matter through a personal AI subscription, the research tool pulling from a repository it was never authorized to access, and the agentic workflow drafting correspondence

on matters it was given no data boundary for. None of these generate alerts. All of them generate exposure.

These gaps are not solved by AI-specific monitoring platforms, most of which are still maturing. They are solved by applying security fundamentals correctly, extending SOC coverage to AI systems as privileged actors, and adding targeted tooling for the specific blind spots that fundamentals alone cannot close.

The Six Controls Most Firms Are Not Applying Correctly

The conversation about AI security fundamentals usually goes one of two ways: either the controls get dismissed as too basic to bother with, or they get checked off a list without being implemented with the depth they require. Neither produces security. Here is what actually matters and why most organizations are getting it wrong:

SecurIT360's Six Security Priorities

1. **Regular Backups**
   Isolated from the main network, tested through actual restores, following the 3-2-1 strategy (disk, cloud, tape). AI systems increasingly generate business-critical data that must be captured in backup policies.
2. **Multi-Factor Authentication**
   Across all systems, including service accounts used by AI agents. Compromised credentials used by AI-assisted attackers stop at the MFA layer before lateral movement begins.
3. **Restrict Administrative Privileges**
   The blast radius of any compromise is directly proportional to the privileges of the compromised account. AI-assisted lateral movement is faster; containing it starts here.
4. **Patch Applications and Operating Systems**
   Unpatched systems are the most reliably exploited targets. AI-assisted attackers scan for them at a scale and speed that makes delayed patching increasingly indefensible.
5. **User Training**
   54% of law firms provide zero AI training, and only 11% mandate it for all staff (8am 2026 Legal Industry Report). AI-generated phishing is more convincing than anything a human attacker writes. For law firms, training must explicitly address the risks of uploading client documents to unapproved AI platforms.
6. **Application Control**
   Blocklisting known malicious tools first, then progressing to allowlisting. Application control is also the primary mechanism for preventing unauthorized AI tool installations.

Joey Vandegrift's framing of these fundamentals is worth holding onto: "If you couple these with continuous monitoring, you are going to be better positioned than most — because without people watching your environment, who responds when an attack happens at 2 AM on a Saturday?" The

fundamentals limit the blast radius. Continuous monitoring ensures someone is watching when the perimeter is tested.

Why Law Firms Face a Different Version of This Problem

The three-pillar model applies to any organization adopting AI. But the stakes in the legal industry are structurally different, and the margin for error is smaller.

Law firms do not just handle sensitive data. They handle privileged data, and privilege is a legal protection, not a security classification. When a client's confidential communications are exposed through an AI tool, the injury is not a privacy violation that can be remediated through notification and credit monitoring. The privilege may be waived. The work product may be compromised. The professional consequences for the attorneys involved can be career-defining. None of that can be reversed after the fact.

The gap between AI use and AI governance in legal is striking. While 79% of legal professionals report using AI, more than half say their firm has no AI policy or they are unaware of one. Forty-one percent of lawyers cite data privacy as a fundamental concern about AI adoption, yet adoption continues to accelerate regardless. The tools are in the workflow. The governance is not. That gap is where the risk lives.

At the same time, law firms are under growing pressure from their own clients to demonstrate that their AI practices meet a standard of care. Corporate legal departments are building AI-specific questions into vendor security questionnaires: which tools does the firm use, how is client data handled, what governance structures are in place, what testing and monitoring is conducted. These questions are not hypothetical. Firms that cannot answer them confidently are losing pitches to firms that can. Robert McElroy observes: "The firms that can answer those questions confidently and specifically are going to be at a meaningful advantage."

> AI is operating very close to a law firm's most sensitive data, often without the oversight that would apply if a human were performing the same function.

The risk concentrates in the workflows where that proximity is highest: legal research pulling in external content with no prompt injection protection, document review processing privileged communications through AI tools with no data boundary enforcement, agentic workflows executing multi-step tasks across case management and billing systems with minimal human review. These are not edge cases. They are how legal AI is actually being used today.

The regulatory horizon will sharpen this further. There are no AI-specific security standards that directly govern law firms today. But client vendor management requirements are already creating de facto standards, and formal regulatory requirements will follow. Firms building mature AI governance, testing, and monitoring programs now will be ahead of that curve when it arrives.

> Most law firms that believe they have an AI security program have an AI security document. The two are not the same thing.

Where to Start, and Why It Cannot Wait

The honest picture of where the legal industry stands on AI security is sobering. The most advanced law firm AI security program SecurIT360's team has encountered, one with documented governance, active testing, and real monitoring coverage, is still working through significant gaps. That firm represents the top one percent of the industry. The vast majority are starting from a much earlier position.

That is not cause for paralysis. It is an argument for starting now, even imperfectly, because the gap between firms that are building AI security programs and firms that are waiting for the landscape to stabilize is widening every month. The practical path forward is not complicated: build the AI inventory, establish governance, validate with targeted testing, connect AI systems to continuous monitoring, and treat all three as a cycle rather than a project.

The most dangerous position is not moving too fast. It is moving without structure. Deploying AI into legal workflows without defined data boundaries is not a calculated risk. It is an unquantified one, and the consequences in a law firm environment, privilege waiver, client loss, bar exposure, cannot be undone with a patch or a policy update once the damage is done.

Here is the uncomfortable truth that this entire paper points toward: the firms that are going to get hurt are not the ones being targeted by sophisticated attackers. They are the ones that deployed AI tools into privileged workflows, wrote a policy they never operationalized, ran a pen test they were not ready for, and assumed the SOC would catch anything that went wrong. The threat is not coming from outside. It is already inside, in the form of AI systems operating without boundaries in environments that have never defined what those boundaries should be.

> Govern. Validate. Monitor. Improve. This is not a framework to implement once and review annually. It is the operating model for AI security, a continuous practice that evolves with the technology it is designed to protect.

SecurIT360 builds this model with clients across the legal industry. Robert McElroy and the governance team establish the policy foundations that make security work coherent. Brad Causey and the offensive security team stress-test those foundations and close the gaps that governance alone cannot see. Joey Vandegrift and the SOC team provide the continuous monitoring coverage that connects the policy on paper to the reality in the environment. The three perspectives in this paper are not theoretical. They are the three disciplines that have to work together for AI security to actually function.

Sources

8am Legal Industry Report, 2026. AI adoption among legal professionals doubles to 69%; 43% of firms have no formal AI policy and no plans to create one; only 9% have a written, actively enforced policy; 54% of firms provide zero AI training; 11% mandate training for all staff. *lawnext.com/2026/03/ai-adoption-among-legal-professionals-has-more-than-doubled-in-a-year*

Clio Legal Trends Report, 2025. 79% of legal professionals use AI; 53% say their firm has no AI policy or they are unaware of one; 40% of legal professionals use legal-specific AI solutions. *clio.com*

Thomson Reuters Generative AI in Professional Services Report, 2025. Share of legal organizations actively integrating generative AI rose from 14% in 2024 to 26% in 2025; 45% of law firms use AI or plan to make it central to their workflow within one year. *thomsonreuters.com*

ABA Legal Technology Survey Report, 2024. 30.2% of attorneys reported their offices were currently using AI-based technology tools; usage highest at firms with 500 or more lawyers (47.8%). *americanbar.org*

Embroker Law Firm AI Survey, 2024–2025. 41% of lawyers surveyed cited data privacy as a fundamental concern about AI adoption; 39% cited security vulnerabilities and cyberattacks targeting AI systems. *embroker.com*

AffiniPay Legal Industry Report, 2025. Survey of 2,800+ legal professionals on AI adoption, firm-level usage, and governance readiness across practice areas and firm sizes. *mycase.com/blog/ai/ai-adoption-in-law-firms*

8am / LlamaLab 2026 Analysis. 59% of employees use unapproved AI tools; 75% share sensitive data with them. *llamalab.ai/blog/legal-ai-adoption-doubles-2026-8am-report*

About SecurIT360

SecurIT360 is a full-spectrum cybersecurity firm providing AI governance consulting, adversarial security testing, and managed detection and response services to organizations across the legal industry and beyond. Our team combines deep expertise in security frameworks, offensive security, and SOC operations to deliver the complete AI security picture that modern organizations require.

securit360.com