Corporate Counsel Guide to AI Security 2025

Contents

Chapter 1

Introduction

Why AI Security Is Now a Legal Imperative

Legal leaders today face a paradox: embrace GenAI to drive efficiency, or risk falling behind.

But according to Thomson Reuters, while 67% of in-house attorneys report confidence in adopting new technologies, only 26% of legal departments with more than 11 attorneys are currently implementing AI.[1]

The reality is that, while sensitive data, regulatory obligations, and reputational risk are all things that must be prioritized, AI doesn't have to be considered as a risk, but rather leveraged as a force to uphold legal standards and drive efficiency.

To achieve this, the legal department must no longer be a passive stakeholder in technology adoption. General Counsels, Chief Legal Officers, and Chief Compliance Officers are now on the frontlines of AI risk management. The main thing to remember is that security is not a barrier or requirement to overcome—it is foundational to the safe, compliant, and successful adoption of AI.

This handbook is designed to equip legal, compliance, and security leaders with the context, considerations, and tools needed to navigate GenAI adoption with confidence.

Chapter 2

The GenAI Opportunity

—and the Double-Edged Sword

AI has the power to completely reshape legal work. From rapid document review and contract comparison to automated summarization and client self-service, the possibilities are real—and so are the risks.

+ Likely Wins

- Legal research
- Contract drafting & redlining
- Due diligence
- Litigation support (summarization, intake tools)
- Document review ..& privilege scanning
- Clause extraction and comparison

— Calculated Risks

- Legal spend analysis
- AI-powered chatbots for legal advice

But these tools operate on highly sensitive data—contracts, regulatory filings, IP, and privileged communications. Without robust security at the point of use, even the most promising application can result in catastrophic exposure.

That's why legal AI enablement requires a balanced, security-first mindset. Legal and compliance leaders don't just need to know *what* AI can do—they must understand *how* it might fail, and *where* their organizations are exposed.

> “ You can’t defend systems effectively when you don’t understand the risks.”

The Risk Landscape: A Snapshot

- **Inference-layer data leaks**
  Sensitive contract terms or regulatory documents exposed through model responses
- **Prompt injection & jailbreaks**
  Malicious prompts override safety features and generate unauthorized or dangerous content
- **Model hallucinations or bias**
  AI outputs inaccurate or discriminatory results that undermine trust or risk liability
- **Unauthorized AI usage**
  Shadow AI tools used without oversight, creating IP and privacy risks
- **Third-party misalignment**
  Vendors deploying GenAI tools without proper security testing or compliance review
- **Emerging Risk – Agentic AI**
  AI agents that act autonomously across workflows—modifying documents, triggering emails, or exposing client data unintentionally

Understanding these risks is not a reason to delay AI—but a prerequisite. to deploying it securely and sustainably.

Chapter 3

The Legal Security Checklist

What to Ask Before You Deploy

Before implementing GenAI tools, legal departments must align with IT and security teams to evaluate risk exposure, regulatory obligations, and downstream business impact. This checklist equips legal leaders with the essential questions to ask vendors, internal stakeholders, and themselves.

Data Governance & Privacy

- [ ] What data is the model trained on? Does it include proprietary or sensitive legal data?
- [ ] Can we prevent our queries or outputs from being retained or used to retrain the model?
- [ ] Is there a way to enforce role-based access to legal data used in AI workflows?

Third-Party Risk & Contracts

- [ ] What are the vendor's obligations for security, audit, and regulatory compliance?
- [ ] Does the contract provide indemnity for model failures or privacy violations?
- [ ] Are there SLAs related to accuracy, availability, and breach notification?

Security Controls & Monitoring

- [ ] Are inference-layer defenses in place to prevent prompt injection and data leakage?
- [ ] How are inputs and outputs being monitored for compliance and risk?
- [ ] What mechanisms exist to audit and log GenAI activity by users and systems?

Compliance & Regulatory Mapping

- [ ] Have we mapped GenAI use to the EU AI Act, GDPR, and relevant U.S. regulations?
- [ ] Can we demonstrate compliance if challenged by a regulator or court?

Model Safety & Explainability

- [ ] Has the model been red-teamed for harmful or biased outputs in legal contexts?
- [ ] Can we explain and justify model outputs used in high-stakes decisions?
- [ ] How are hallucinations or misbehavior flagged and corrected?

Deployment Strategy

- [ ] Are we piloting AI in low-risk use cases first?
- [ ] Do we have a clear plan for human-in-the-loop oversight?
- [ ] How often will we reassess tools and vendors as the tech evolves?

Chapter 4

Building the Right

GenAI Governance Framework

Securing legal AI deployments requires more than just checklists—it demands a governance model rooted in accountability, clarity, and agility. Legal departments must partner with IT, compliance, and business leadership to embed AI governance across people, policies, and platforms.

> “ AI security isn’t a checkbox—it’s a living framework. Your safeguards should adapt as fast as your AI evolves.”

Establish an AI Steering Committee

- Include legal, IT/security, HR, operations, and business leaders
- Define clear responsibilities and escalation paths
- Revisit governance mandates regularly as AI evolves

Formalize Acceptable Use Policies (AUPs)

- Define which AI tools are permitted and prohibited
- Set parameters for usage: what data can be input, by whom, and for what purpose
- Include consequences for violations and procedures for tool vetting

Embed Legal Oversight in AI Development

- Legal teams should be involved in procurement and deployment stages
- Support impact assessments for high-risk AI systems
- Create internal templates and review mechanisms for model evaluations

Monitor Regulatory Change Management

- EU AI Act enforcement is staggered—some provisions already active as of late 2024
- U.S. regulatory patchwork continues to evolve; watch for federal guidance and state-level AI laws
- Treat privacy, ethics, and compliance as core pillars of governance—not legal silos

Before implementing GenAI tools, legal departments must align with IT and security teams to evaluate risk exposure, regulatory obligations, and downstream business impact. This checklist equips legal leaders with the essential questions to ask vendors, internal stakeholders, and themselves.

Establish Vendor Due Diligence & Contract Protocols

- Review vendor security controls and compliance with NIST AI Risk Management Framework (or similar frameworks)
- Demand transparency around model training data and architecture
- Align indemnity, SLA, and audit rights to
- your internal policies and risk appetite

Ensure Cross-Functional Enablement

- Train business stakeholders on what GenAI can and cannot do
- Provide lightweight guidance for safe experimentation
- Balance innovation with guardrails to promote sustainable AI adoption

Recommended Resource

NIST AI Risk Management Framework

Chapter 5

Mapping Legal AI

to Global Regulation

GenAI adoption is not occurring in a vacuum—regulatory bodies across the globe are moving swiftly to define how artificial intelligence can and should be governed. Legal teams must stay ahead of this changing terrain, not only to remain compliant but to build trust and future-proof their AI strategies.

EU AI Act

- **Enforcement timelines are already underway**
  Provisions banning certain use cases (like social scoring) began enforcement in late 2024. GPAI rules go live by June 2025, and high-risk use case rules follow in 2026[2]
- **Broad extraterritorial scope**
  Applies to any organization marketing or operating AI in the EU, even if based elsewhere
- **Tiered risk pyramid**
  Legal, HR, and customer-facing GenAI use cases may be classified as "high risk" depending on context and model behavior
- **Cross-regulatory impact**
  The Act intersects with the GDPR, the EU Data Act, and the EU Copyright Directive, all of which will influence legal AI governance structures [3]
- **AI Literacy Mandates**
  Article 4 requires organizations to train personnel involved in AI oversight and use

China's Evolving AI Regulation

- **Fragmented but accelerating**
  China's regulatory framework spans multiple laws including the Data Security Law and the Personal Information Protection Law, with a draft AI Law in development
- **Prohibited Use Enforcement**
  High scrutiny on alignment with Chinese values; banned GenAI applications tied to politically sensitive or socially disruptive content

U.S. Regulatory Outlook

- **No federal framework—yet**
  But executive orders such as EO 14179 (Removing Barriers to American Leadership in AI) and EO 14192 (Unleashing Prosperity Through Deregulation) support innovation-centric strategies over broad regulation[3]
- **State-level patchwork**
  By 2030, 50% of the U.S. population will be covered by state-level AI regulations, with laws in California, Utah, and Colorado already in force[3]
- **Bias Compliance Considerations**
  EO 14179 calls for investigation into ideological and engineered bias, requiring bias detection and retraining of AI models in sensitive use cases such as legal decisions[3]

50%

By 2030, 50% of the U.S. population will be covered by State-level AI regulations

> “ The pace of AI innovation doesn’t excuse legal inaction. Your best defense is proactive alignment to regulation—before enforcement knocks.”

Recommended Actions for Legal Teams

- Catalog all deployed and procured AI-enabled software and use cases
- Map each application to the relevant tier of risk under applicable regulatory frameworks
- Partner with compliance and security teams to manage third-party risk and update procurement processes
- Prepare for transparency, auditability, and explainability obligations in future litigation or regulatory action
- Develop AI literacy programs for teams managing, procuring, or auditing AI-enabled systems

Core Resources for Legal AI Governance

- The EU Artificial Intelligence Act
- Forrester, How to Approach the EU AI Act
- Gartner, How Global AI Regulations Will Impact Your Enterprise

Chapter 6

Security in Practice

Legal-Focused AI Security

Legal leaders must move beyond theoretical AI risk discussions and operationalize security through clear, proven mechanisms. Fortunately, platforms like CalypsoAI offer solutions tailored for securing legal and compliance use cases across the AI lifecycle.

Real-Time Defense at the Inference Layer

Legal teams interact with AI at the point of inference—where input becomes output. That's why security must focus on the inference layer, where the greatest risk of leakage, manipulation, or misbehavior occurs.

CalypsoAI Inference Defend Offers:

- Real-time protection against prompt injection and jailbreak attempts
- Customizable controls that prevent sensitive data (e.g., PII, contractual language) from being shared or exposed via model outputs
- Adaptive policies that respond to user role, risk threshold, and AI system behavior

Red-Teaming for AI

Security testing of AI systems before launch is still rare across most industries. According to Forrester, fewer than 20% of organizations have centralized, funded AI governance programs—leaving most AI deployments under-tested and under-secured.[4]

CalypsoAI Inference Red-Team Provides:

- Scenario-based testing of legal GenAI systems (e.g., redlining assistants, clause generators)
- Proactive attacks to evaluate susceptibility to data leakage or biased output
- Audit-ready reports that accelerate deployment timelines and satisfy compliance mandates

Monitoring & Auditing for Legal Defensibility

- Ensure activity logging for every prompt and output involving legal GenAI us
- Maintain clear records for model testing, validation, and red-teaming
- Provide cross-functional access for legal, security, and audit stakeholders

By embedding these proactive AI security solutions directly into legal AI workflows, organizations unlock GenAI's benefits while maintaining control, compliance, and trust.

Use Case

Document Review Firewall for E-Discovery

A multi-national legal team used Inference Defend to safeguard model-assisted e-discovery workflows. Sensitive contract terms, jurisdictional language, and PII were identified and filtered in real-time, preventing unauthorized model behavior during litigation prep.

Use Case

Risk-Aware Redlining Bot

A financial services firm piloting a GenAI clause comparison tool used Inference Red-Team to run injection attacks and discriminatory bias. Findings were used to retrain their model and add policy constraints before launch.

Chapter 7

Your First 90 Days

Legal AI Action Plan

For legal leaders ready to move from planning to practice, these are the priority moves to enable secure, defensible GenAI adoption.

> “ Securing AI in legal means taking fast, informed action—before someone else makes the decision for you.”

Month 1

Establish Baseline Awareness & Oversight

- Inventory all existing and proposed GenAI tools in use across the department
- Identify high-risk workflows (e.g., litigation, contracts, investigations)
- Form or join a cross-functional AI governance group
- Review AI vendor contracts for data, indemnity, and security terms

Month 2

Design & Implement Controls

- Draft and distribute Acceptable Use Policies (AUPs) for legal AI use
- Begin red-teaming critical GenAI tools (or schedule external audits)
- Test inference-layer controls in legal workflows using tools like CalypsoAI Inference Defend
- Evaluate access controls and ensure proper logging of model interactions

Month 3

Operationalize & Educate

- Finalize and document AI policies, risk maps, and deployment procedures
- Launch AI literacy training for legal, compliance, and procurement teams
- Set a review cadence for evaluating new tools and updating governance structures
- Develop reporting mechanisms for tracking AI tool performance and incidents

Glossary

Agentic AI *– AI systems capable of taking autonomous actions in pursuit of goals, often interacting with multiple systems via APIs.*

Bias *– Systematic and repeatable error in AI decision-making, often resulting in unfair or discriminatory outcomes.*

Data Exfiltration *– Unauthorized removal or transfer of data from within an organization, often through AI model queries or leaks.*

DPIA (Data Protection Impact Assessment) *– A mandatory process under GDPR to identify and mitigate privacy risks in data processing activities.*

Inference Layer *– The point at which an AI model receives a user prompt and produces an output. This is where the most immediate risk of misuse or leakage occurs.*

Generative AI (GenAI) *– A class of AI that generates new content (text, images, code, etc.) in response to prompts, based on large-scale training data.*

Hallucination *– When a GenAI model outputs incorrect, misleading, or fabricated content presented as factual.*

Jailbreak *– A prompt-based technique that circumvents safety filters or governance layers embedded within an AI system.*

Model Misalignment *– When an AI system's behavior diverges from intended or authorized outputs, often leading to regulatory or reputational risk.*

NIST AI RMF *– The National Institute of Standards and Technology's AI Risk Management Framework, a U.S.-based guide for managing AI-related risk.*

Prompt Injection *– A type of attack where a user manipulates an AI prompt to change the model's behavior or gain unauthorized access to information.*

Role-Based Access Control (RBAC) *– A security principle that restricts system access based on a user's role within the organization.*

Shadow AI *– Use of GenAI tools outside of approved channels or without security oversight, creating compliance, privacy, and IP risks.*

Third-Party Risk *– Potential exposure that arises from using external AI vendors, especially when tools are poorly governed or noncompliant.*

Sources

1. *Thomson Reuters: Ready or not: artificial intelligence and corporate legal departments*
2. *Forrester: How To Approach The EU AI Act*
3. *Gartner: How Global AI Regulations Will Impact Your Enterprise*
4. *Forrester: Global Commercial AI Software Governance Market Forecast, 2024 to 2030*

Additional Resources

FTI Consulting, Use Cases for AI in the Legal Function

NIST AI Risk Management Framework

The EU Artificial Intelligence Act

The Forrester Wave™: Contract Lifecycle Management Platforms, Q1 2025

While CalypsoAI may reference related legal issues, we do not provide legal advice or services, and our guidance should not be construed or used as a specific guide to action.