When AI Outpaces Security:
A Guide to Regaining Control of Your Agentic Supply Chain

Table of Contents

SECTION I:

The High Cost of Unmanaged Adoption

1. Introducing...the Future

Technology continues to move forward with ever-accelerating speed. Today, we are witnessing a disruptive moment in history with artificial intelligence (AI) and machine learning (ML) becoming increasingly integrated into all aspects of our lives. Nowhere is this more evident than in the world of software application development.

Consumers want their apps to accomplish more, serving up intelligent recommendations and making life easier. Failure to securely incorporate these agentic technologies into your development pipelines will probably result in developers finding their own way to incorporate unvetted AI components into their applications, increasing your governance and security risks in equal measure. The only secure way forward requires controlling AI adoption on a well-governed and trusted path.

The use of AI agents is fast becoming a “must-have” for software developers. Organizations that want to remain competitive must leverage AI-assisted development to build software that integrates AI assets as core components.

Doing so is easier said than done. To keep up with the security demands imposed by AI - specifically as applied to **risk, governance, and traceability** - organizations must swiftly revamp their production pipelines, including the tools, processes, and protocols needed to ensure that they’re building and deploying AI-driven applications as responsibly and ethically as possible, without compromising time-to-value.

AI Adoption by the Numbers

AI is now table stakes — but most organizations can't scale it safely. Enterprise AI adoption has officially crossed the tipping point, and the conversation has shifted from "are we using AI?" to "can we trust it?" **McKinsey's 2026 AI Trust Maturity Survey** captures the moment precisely: AI adoption is accelerating rapidly, with organizations moving beyond experimentation toward scaled deployment of gen AI and, increasingly, agentic AI across core business functions. But as AI systems take on greater autonomy — making recommendations, triggering actions, and interacting with other systems — the consequences of failure grow materially. **Stanford's 2026 AI Index Report** puts the scale in historical perspective: generative AI reached 53% population adoption within three years, faster than either the personal computer or the internet.

The next wave — agentic AI — is arriving even faster. **Gartner** forecasts that 40% of enterprise applications will be integrated with task-specific AI agents by the end of 2026, up from less than 5% in 2025. Demand is matching the supply: in **Gartner's 2026 CIO survey**, only 17% of organizations have deployed AI agents to date, yet more than 60% expect to do so within the next two years — the most aggressive adoption curve of any emerging technology Gartner measured.

But here's where the story turns. Despite near-universal adoption, the bottleneck has shifted decisively from "can we build it?" to "can we trust it in production?" McKinsey found that nearly two-thirds of respondents cite security and risk concerns as the top barrier to fully scaling agentic AI, well ahead of regulatory uncertainty or technical limitations. Organizations know what's coming. They just aren't ready for it.

> *The cost of failing to close that gap is already measurable. According to **IBM**, only 37% have governance policies in place, and Shadow AI adds an average of $670,000 to breach costs. Looking forward, Gartner predicts that 25% of all enterprise GenAI applications will experience at least five minor security incidents per year by 2028, up from 9% in 2025.*

AI adoption is no longer the differentiator. Your ability to govern, secure, and scale your agentic supply chain is. To put it simply: **You need a way to future-proof your agentic supply chain.**

2. The Challenges of Implementing AI-Driven Development Workflows

Effective serious change is always challenging, especially when it comes to processes that have been in place for decades. For better or worse, in today's competitive environment, adaptation has become a necessity rather than an option.

The Tug-of-War: Adopting Fast vs. Staying Secure

Today's organizations are caught in a delicate balancing act. On one hand, there is immense pressure to adopt AI rapidly, allowing innovation and productivity to flourish organically from the bottom up. On the other hand, there is an absolute necessity to stay secure and maintain full control over the software supply chain.

The good news is that AI isn't here to replace your software supply chain wholesale. It is, however, significantly capable of accelerating it. This applies both in terms of the input, as more third-party assets are incorporated, and the output, resulting in a tsunami of binaries from agentic-driven development, while simultaneously exposing applications to more threats than ever before.

The somewhat less good news is that as AI expands your supply chain, it also adds entirely new artifacts, such as AI models, MCP servers, agent skills, plugins, and external AI services to govern. It also introduces new security blind spots, especially surrounding issues such as artifact lineage, model provenance, over-privileged agent access, and visibility across all stages of production.

As organizations move to integrate enterprise-grade AI into their software supply chains, they find themselves encountering challenges that are both organizational and technical in nature.

Security Becomes a Key Factor

At first, AI developers, Data Scientists, and DevOps teams were able to operate just fine while isolated from one another. However, when incorporated into an enterprise-grade high-speed software delivery pipeline, this model becomes untenable. Teams must be de-siloed, a single source of truth must be established, and security becomes a shared priority.

The Rise of Shadow AI

While many organizations are able to get started with AI projects, finding a secure foothold as usage scales proves to be difficult. When internal teams adopt new tools to drive fast, bottom-up innovation, it often results in unmonitored usage.

The lack of centralized control over the AI supply chain exposes the enterprise to a significant potential for operational and regulatory failures, including:

- The AI supply chain becomes less fortified against security threats and compliance risks.
- Internal visibility is gradually lost as AI usage proliferates outside of R&D teams.
- Organizational control over the AI ecosystem begins to slip.
- Governance becomes harder and harder to enforce.
- Scalability becomes a challenge, as internal teams lack the tools, processes, and protocols to implement next-generation technology.
- Access permissions grow increasingly difficult to monitor.

To resolve this tension between innovation and control, organizations must deploy Shadow AI detection tools. These tools provide them with the capability to automatically discover and govern all unmanaged models entering the ecosystem.

By shedding light on these blind spots, security teams can maintain full visibility and enforce compliance without stifling the rapid, bottom-up innovation that makes AI so valuable in the first place.

3. What Integration Entails

As the software supply chain expands, incorporating new teams, tools, and technologies, make it easy for fragmentation and disconnection to undermine existing processes. Not only does it add operational complexity, but more importantly exposes questions of governance and security that must be addressed.

The “Black Box” Problem: Unpacking AI Assets

Even as organizations consider AI agents as a component in their software supply chain, the rush to adopt these technologies introduces a unique technical challenge: The AI assets themselves that comprise these agents are inherently opaque, making them incredibly difficult to scan.

Unlike traditional software, where source code can be reviewed and vetted line-by-line, AI assets, and especially models, are essentially “black boxes” made up of complex weights and binaries. Since standard security tools cannot parse these artifacts and organizations are adopting them so rapidly with minimal technical understanding of their inner workings, attackers are actively seeking ways to leverage this blind spot.

Malicious actors use this lack of visibility to pass poisoned assets, backdoors, and compromised dependencies directly into the **SDLC**. When teams pull AI assets, such as models, MCPs, skills, etc., like any other software artifact, without being able to unpack their context, they open the door to devastating supply chain attacks.

To safely deploy AI agents, organizations need strong, specialized security tools designed to "unpack" these black boxes. Security and development teams must be able to scan these complex binaries, understand their full context, and definitively say it is safe for production. Only then can teams confidently answer the most critical questions about their AI pipeline:

The Key Questions

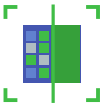
- What is the exact origin and provenance of the AI assets in use?
- What is actually hidden inside the model's architecture and dependencies?
- Which 3rd party assets contain known vulnerabilities or malicious payloads?
- What audit evidence can we offer to prove a model's safety?
- Which MCP servers can be securely integrated without exposing the SDLC?

Answering these critical questions requires moving beyond traditional software security. To safely integrate AI into the SDLC and defend against malicious tampering, organizations must establish a robust framework built on four core pillars:

Traceability: Teams must be able to quickly prove the origin of all AI assets—including first- and third-party models, MCP servers, agent skills and related packages—understand their underlying architecture, and clearly illustrate the source of any change approvals.

Chain of Custody: Models, skills, and MCP servers must always come with a clear chain of custody, illustrating who created, trained, or modified the asset. This ensures no malicious tampering or poisoning occurred mid-pipeline.

Audit-Ready Evidence: Teams must be able to make it clear, upon request, that the inner workings of all assets have been inspected and that they adhere to the requisite security protocols and compliance guidelines.



Secured, Scanned, and Non-Malicious Assets: This is where the black box is finally opened. Automatic and continuous model scanning gives an organization the power to unpack complex binaries, understand their full context, and confidently determine which models, packages, and components present a lower risk and can safely be used at scale.

The Evolving Attack Surface

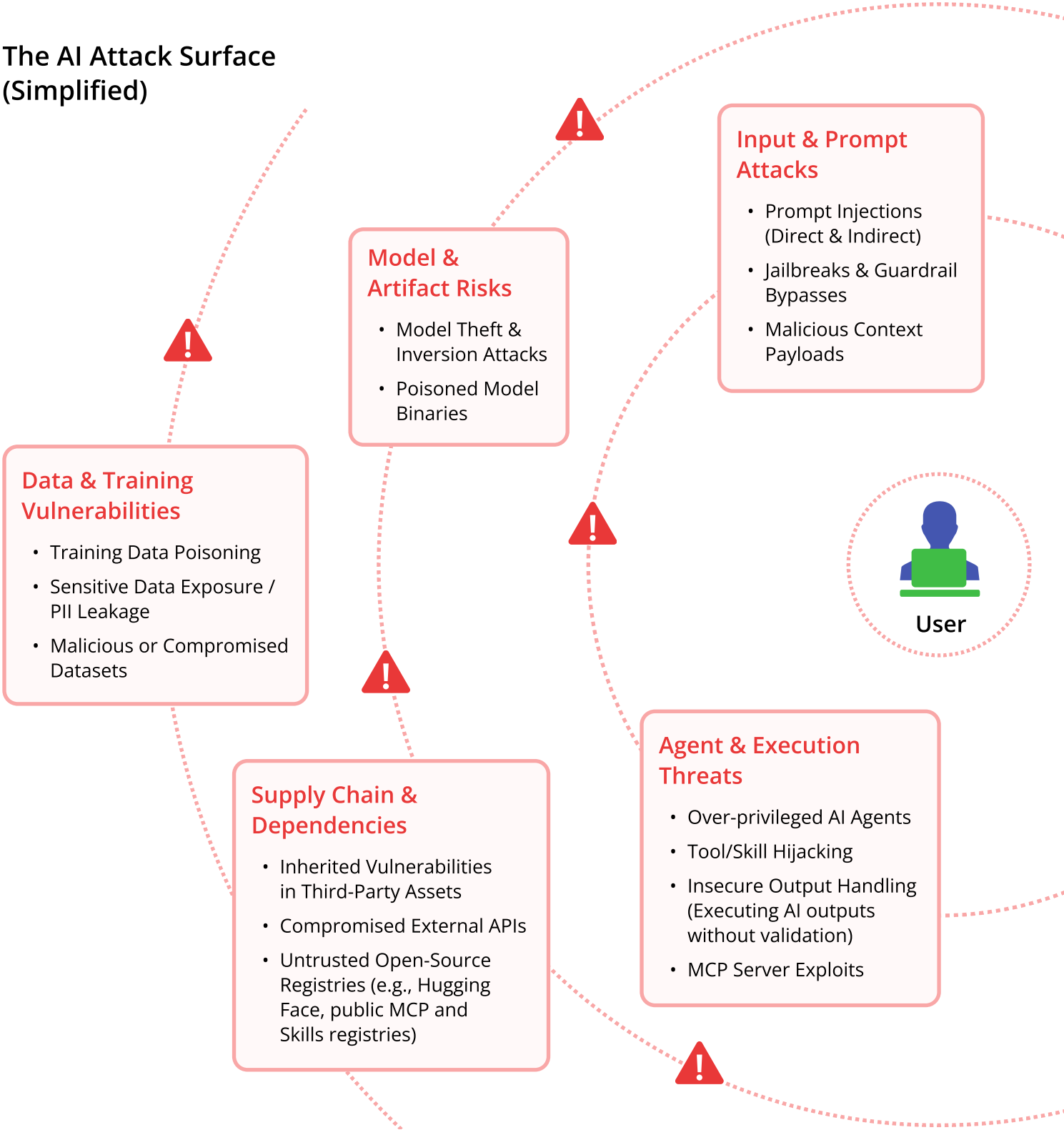
When considering AI as both an assistive tool and a consumer-facing core component in app development, leaders must now contend with a rapidly expanding attack surface within their software supply chain.

For a long time, AI risk discussions centered almost exclusively on the models themselves. But today, the attack surface is no longer just models. It now comes in entirely new forms and objects that organizations have never had to manage before. Security teams must now govern an evolving ecosystem of complex AI artifacts, most notably **MCP servers** and **Agent Skills** or tool-calling capabilities.

When combined with data, algorithms, deployment environments, source code and third-party libraries, these artifacts significantly expand your risk level. They introduce new execution pathways, inherited dependencies, and untrusted elements that traditional security tools simply aren't equipped to parse or monitor.

Risks like unverified inputs, inherited vulnerabilities, and untrusted AI artifacts can severely impact the time-to-value of software development operations. To mitigate these kinds of supply chain risks, organizations now need a more advanced and comprehensive security framework. One that's capable of carrying them into a future of continuous monitoring, rigorous validation, proactive policy gates, and robust control over all artifact types.

The AI Attack Surface (Simplified)



SECTION II:

The Solutions You Need Now

4. Governance + Security + Management = TRUST

In this era of AI app development, trust becomes so much more than a simple best practice. What was once an aspiration is now something that must be thought of, across organizations, as an outright engineering discipline. Organizations that want to stay competitive must set enforceable control levers that span the entire software supply chain, ensuring security and compliance at every stage.

Ultimately, trust is no longer an abstract concept, but preferably defined by tangible evidence from coding to production. Achieving this level of confidence requires organizations to maintain a full end-to-end visibility of the software supply chain, spanning from the developer's initial workspace all the way into runtime execution. If you cannot trace an AI artifact's journey and verify its integrity throughout the SDLC, you cannot inherently trust it.

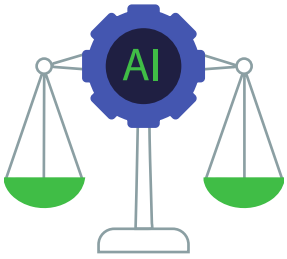
Trust in a Nutshell

While industry frameworks like **Gartner's AI TRiSM** (Trust, Risk, and Security Management) provide a helpful starting point, often these generic definitions fall short of the technical reality. For organizations building enterprise-grade applications, trust less an abstract policy and more a layer that serves as a foundation for the agentic software supply chain.

To truly secure the AI ecosystem, we need to map the core pillars of AI trust directly to the software supply chain, so we can arrive at an actionable framework:

1. **AI Management - The System of Record**
   True management goes far beyond maintaining a simple list or catalog of AI tools. It requires managing models, agent skills, and MCP servers as software binaries within a single, unified system of record. This means establishing immutable provenance and guaranteeing full traceability of every AI component. If you cannot definitively prove who built a model, where it came from, and what dependencies it relies on, you cannot effectively manage your AI supply chain.

2. **AI Security - Continuous Inspection**
   Trust requires deep, continuous inspection of the actual binaries. Instead of simply waiting to catch anomalies at runtime, this layer demands aggressive, automated scanning of AI packages before and during deployment. It ensures that complex binaries are unpacked and analyzed for hidden backdoors, poisoned payloads, malicious code, and licensing compliance issues—mitigating risks long before they can infect the production environment.



3. **AI Governance - The Control Plane**
   If management is your system of record, governance is your control plane. This is where organizations dictate exactly which AI assets are allowed to enter the ecosystem, who is authorized to use them, and how they can be deployed. Governance is about setting strict, enforceable guardrails and access permissions, not just for developers, but for the AI agents themselves and the assets that build them. By enforcing these controls natively, organizations can prevent Shadow AI, stop unauthorized tampering, and keep the entire ecosystem compliant.

5. Cultivating Trust with the Tools of Tomorrow

Integrating AI agents and their building blocks into the software supply chain requires a new approach...one that unifies model governance, lifecycle control systems, and proactively automated security, without slowing your ability to deliver as expected.

With AI's incorporation into the software development supply chain, a new set of challenges and requirements has arisen, each one increasing the pressure on already-stressed delivery pipelines.

As a result, the next-generational approach revolves around four core pillars:

1. Shadow AI Detection

You cannot secure what you cannot see. As developers rapidly adopt new AI tools to accelerate innovation, visibility often drops to zero, creating massive blind spots. A modern security framework must begin with continuous, automated Shadow AI detection. By actively discovering unmanaged models, untrusted external APIs, and unvetted AI assets entering the environment, security teams can bring bottom-up innovation into the light—ensuring every AI asset is accounted for without stifling development speed.

2. Centralized AI Registry

Once AI assets are discovered, they need a secure, governable home. Organizations must rely on a dedicated AI registry that functions as the single, centralized system of record for all AI artifacts. Rather than treating AI as a mystical separate category, this registry manages all first-party and third-party models, MCP servers, agent skills, and related packages exactly like standard software binaries. This establishes an unbroken chain of custody and makes discovery, versioning, and secure approval seamless across the entire enterprise.

3. Automated Policy Enforcement

Trust requires aggressive, proactive governance. Relying on manual security approvals for AI adoption creates a bottleneck that agile teams will inevitably try to bypass. Instead, organizations need automated policy enforcement embedded directly as a supply chain gate. This ensures that every AI model, package, and dependency is continuously scanned for malicious payloads, licensing violations, and known vulnerabilities—automatically blocking non-compliant or poisoned assets before they ever reach production.

4. Secure AI Gateway

Managing the serving and runtime of AI requires a robust, secure control point. A Secure AI Gateway acts as the critical intermediary between your internal applications, AI agents, and external LLM providers. Crucially, it also serves as the central connection and enforcement tool for all MCP servers and AI agent skills. By sitting at this intersection, the gateway routes every single request, rigorously checks access permissions, and automatically blocks any action if the agent or user is not explicitly authorized. As the final line of defense, it prevents sensitive data leakage, thwarts prompt injection attacks, and ensures that all complex agentic workflows and tool-calls remain secure, compliant, and fully governed at scale.

6. Security Concerns and Value Perspectives

Security concerns take on a unique shape from role to role. The solution you need, however, is a universal one—a single, controlled, auditable software supply chain that treats AI with the same rigor as traditional software.

CISOs

As the attack surface rapidly expands beyond just models to include new, complex artifacts like MCP servers and agent skills, CISOs face the daunting task of securing an entirely new ecosystem. When "Shadow AI" proliferates through bottom-up innovation, organizational control slips, and unmonitored usage creates massive security blind spots. To ensure the organization is protected against everything from data poisoning to prompt injections, CISOs need verifiable, end-to-end trust. This requires a comprehensive framework featuring automated Shadow AI detection to bring hidden assets into the light, coupled with a Secure AI Gateway that acts as the final line of defense to enforce access controls and govern agentic workflows at scale.

AppSec Managers

Proper application security management requires solving the "black box" problem. AppSec teams can no longer rely on traditional security tools that fail to parse complex model binaries, as this leaves the SDLC highly vulnerable to poisoned assets, inherited dependencies, and supply chain attacks. To secure the pipeline without creating bottlenecks for development teams, AppSec requires automated policy enforcement and continuous inspection embedded directly into the supply chain. This approach ensures that every AI package and dependency is automatically unpacked, scanned for malicious payloads, and vetted for compliance before it ever reaches production.

Engineering & DevOps Leaders

If engineering or DevOps is your domain, your primary directive is balancing delivery speed with operational reliability—managing the intense tug-of-war between adopting fast and staying secure. Dealing with a tsunami of new agentic binaries and fragmented toolchains significantly increases operational complexity and pipeline stress. DevOps leaders need a Centralized AI Registry that functions as a single source of truth, treating all AI models, external services, and MCP servers exactly like standard, first-class software binaries. This de-silos teams, unifies the workflow, and provides a seamless, frictionless path to operationalizing AI at scale.

GRC Leaders

As a leader in the GRC space, you carry the heavy burden of compliance, control, and verifiable accountability. Untracked model lineage, untrusted third-party assets, and over-privileged agents present massive regulatory and compliance risks. GRC leaders require a robust governance control plane that guarantees immutable provenance, a clear chain of custody, and deep traceability across all AI components. With automated governance serving as a centralized system of record, GRC teams can effortlessly monitor access permissions and call up audit-ready evidence proving that all inner workings of the organization's AI assets adhere to requisite security protocols.

SECTION III:

The New Way Forward

7. JFrog's Platform Approach to Trusted AI Delivery

As with so many things, the way forward is unity.

By unifying software and AI management, security, and governance under a single platform, every stage of your development pipeline sits harmoniously under the same set of control and operational principles. Outmoded "best practices" are left in the past, where they belong, as you move forward with a scalable, clearly defined set of guidelines and protocols.

Moving AI-driven development protocols from experimentation to viable production requires unifying everything on a single platform—ideally, one that normalizes the management of AI artifacts under the same tried-and-tested principles that you're already using to effectively govern your software delivery pipeline.

With the **JFrog Platform**, AI is integrated fully and treated as part of the overall development lifecycle in a way that effectively de-silos teams and makes cross-departmental collaboration easier and more effective than ever before.

The **JFrog AI Catalog** is the single system of record for your enterprise AI supply chain. It provides centralized governance and proactive security for all AI workloads, from internal and third-party models to agent skills and MCP servers, enabling you to eliminate Shadow AI and deliver trusted applications with speed and control.

8. Future-Proofing Your Supply Chain

When we talk about future-proofing your supply chain, we're talking about building and maintaining an agentic supply chain that establishes and proves trust from one end of the software supply chain to the other, encompassing all types of software development - especially AI. This requires an end-to-end layer of trust that scales at an enterprise level, which is exactly what the JFrog AI Catalog provides.

Ready to secure your AI supply chain? Book a **personalized consultation** with our Solutions Engineers to map out your trusted path to production.