TAG

HOW NUDGE SECURITY PROVIDES A SECURITY GUARDRAIL FOR ARTIFICIAL INTELLIGENCE (AI) USE IN SOFTWARE AS A SERVICE (SAAS) APPLICATIONS

DR. EDWARD AMOROSO,
CEO, TAG

nudge

HOW NUDGE SECURITY PROVIDES A SECURITY GUARDRAIL FOR ARTIFICIAL INTELLIGENCE (AI) USE IN SOFTWARE AS A SERVICE (SAAS) APPLICATIONS

DR. EDWARD AMOROSO, CEO, TAG

Nudge Security effectively addresses the growing need for cybersecurity guardrails to protect an organization from unknown and potentially unacceptable use of artificial intelligence (AI) in applicable public Software-as-a-Service (SaaS) applications.

INTRODUCTION:

For the past two years, TAG researchers have held roughly sixty workshops with its Research as a Service (RaaS) customers on the topic of securing their use of artificial intelligence. The workshops involved enterprise team members from security, information technology (IT), legal, public relations, and many other areas, usually commissioned by senior leadership into so-called AI Committees, tasked with advising on AI strategy for the organization.

The TAG team leveraged these workshops to gather the most commonly cited cybersecurity requirements for AI, usually referred to as guardrails. We identified six primary aspects of AI use that demanded deployment and use of security guardrails. The identification of these six areas, listed below, was surprisingly uniform across the workshops, regardless of industry sector or even size of company:

1. **Staff Use of Public and Private AI** – This use case involves staff such as employees and contractors leaking sensitive data or obtaining incorrect results to a Generative AI (Gen AI) system through a Large Language Model (LLM) interface. Obviously, the risk is greater for public platforms such as ChatGPT, but risk also exists for private LLMs that rely on porous enterprise perimeters for protection.

2. **Third-Party Use of AI** – This involves partners, suppliers, and other third-party organizations leaking sensitive data or obtaining incorrect results through use of their own AI-based system. This use of AI is often governed by broad, generic contract language that does not specifically preclude such usage. Organizations thus have the risk of third parties using AI in a manner that is inappropriate to the sponsor's mission.
3. **Platform Use of AI** – This involves a deployed platform using AI, often through live connections to the public cloud, to support certain functional objectives. Admittedly, many buyers select a platform based on their reported use of AI (e.g., CrowdStrike), but the manner in which this is done might not be consistent with the sponsor's policy. The result is that tighter controls are required here.
4. **Merged Entity Use of AI** – When larger companies acquire a smaller company, or when more modest firms are scooped up by a larger entity, the result is often an uneven application of AI policy. This problem is exacerbated by the typically slow integration process that is present when information technology (IT) and security teams try to connect the newly merged groups.
5. **Developer Use of AI** – This case involves developers leaking data or obtaining inaccurate results from AI-based systems or capabilities that are accessible through an application programming interface (API). Guidelines from groups such as OWASP can help, but stronger controls are generally required to ensure that no security incidents occur based on this type of runtime connection.
6. **SaaS Application Use of AI** – This case involves the use of AI by Software-as-a-Service (SaaS) application providers. As with platforms, this might be considered a desirable feature, but it does demand the introduction of proper controls to ensure that no data is being leaked through a SaaS application or that the organization might be getting uneven or incorrect results from an AI accessible through an API.

It is this sixth use case – namely, the SaaS provider leveraging AI for a sponsor through an API connection, that we address in this report. In particular, we were encouraged to discover the strong means by which cybersecurity vendor Nudge Security provides organizations with excellent insights into the use of AI by the major SaaS applications that might be in use. This serves as a strong control for our sixth use-case – and it is the topic of the sections below.

UNDERSTANDING THE SAAS AI THREAT

We should first start by explaining in more depth the type of security threat that we are concerned with regarding SaaS application use of AI. The presumption is that a given enterprise will have the usual assortment of SaaS applications in use across their hybrid network and that it will be typical for sensitive data to be shared with these applications. Some common and popular SaaS examples include Box, Salesforce, SAP, Slack, and Workday.

The AI security threat for SaaS will usually involve improper use of sensitive data. That is, the most common threat will involve a SaaS application using the data of a customer to train their own AI, presumably to reduce costs, improve service, or perform research. The scenarios that emerge include a company sharing proprietary intellectual property with a SaaS application that then learns and passes this learning to a competitor (e.g., Coca-Cola IP supporting Pepsi).

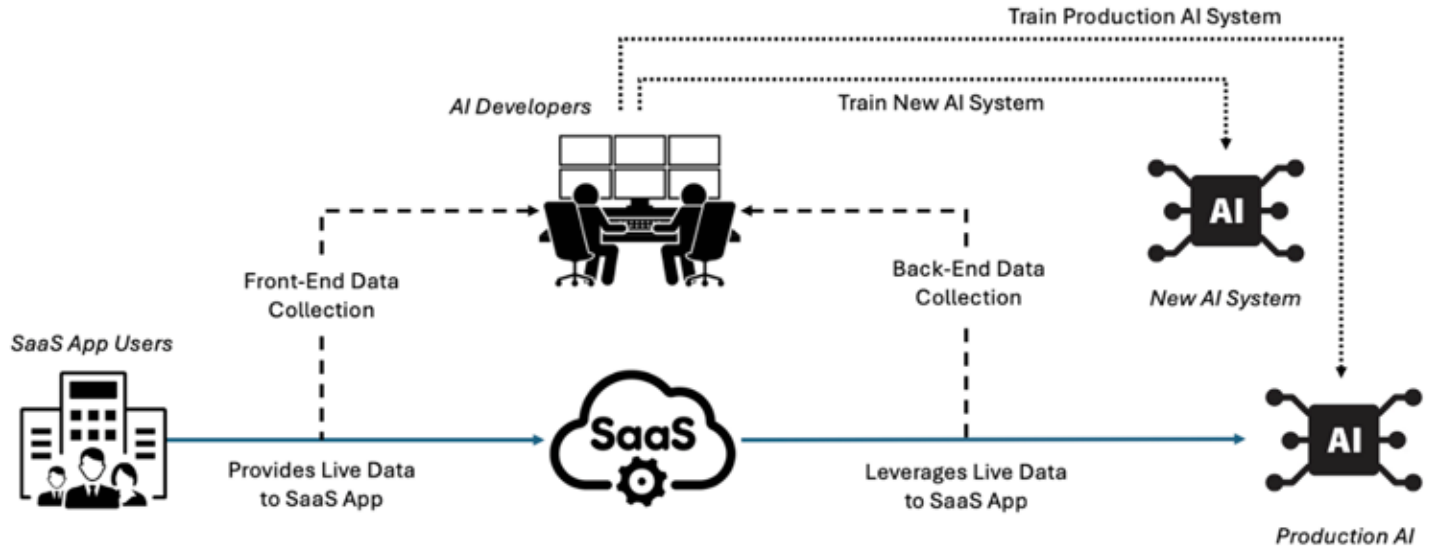

Figure 1. SaaS Threat from AI Usagel

The diagram in Figure 1 suggests several different possibilities in terms of the threats to an enterprise using a SaaS application that is actively engaged in the use of AI. These threats include the following possibilities:

1. **Front-End Data Collection to Train New AI System** – This involves active collection of live data from the enterprise user to train a new AI system, presumably to handle raw input such as prompt entry. Such collection will include sensitive data if the enterprise is not running a data leakage prevention (DLP) filter on the live input channel.
2. **Front-End Data Collection to Train a Production AI System** – This case involves collection of the live data, but with the goal to improve the production AI system handling the enterprise. It's a subtle difference from the previous case, but the motivation is less about research and more about customer experience improvement.
3. **Back-End Data Collection to Train a New AI System** – This case involves using data processed by the SaaS app and being sent to the production AI system. This can be collected and used to train a new AI system that might be attempting a new type of approach (e.g., extending production machine learning to new deep learning).
4. **Back-End Data Collection to Train a Production AI System** – This involves just trying to improve the production AI – and most would agree that this is, in fact, the primary purpose of a production AI. That is, we expect AI systems to learn on input data, so this case is usually considered acceptable.

The overall assumption is that an organization must find some way to address this range of scenarios, and our view is that the initial guard rail must involve visibility into what is actually occurring. This demands that review and monitoring be done for evidence that a SaaS app is using AI, but it also requires ongoing diligence since the use of AI continues to evolve and what a SaaS application does today might not reflect what it is doing tomorrow.

NUDGE SECURITY SaaS AI GUARDRAIL

The commercial Nudge Security platform helps enterprise organizations mitigate the cyber risks associated with SaaS applications. By providing visibility into SaaS, IaaS, PaaS, and generative AI tools being used, Nudge enables IT and security teams to identify and manage potential vulnerabilities effectively. This approach ensures that enterprises maintain a more robust security posture at a time when SaaS adoption is clearly expanding.

A key feature of Nudge involves detection and assessment of the security profiles of AI-powered apps. With the number of unique AI tools growing quickly, organizations face increased risks from unvetted AI systems. Nudge's AI dashboard (see Figure 2 below) provides real-time insights into AI usage by a SaaS app, allowing security teams to evaluate each AI vendor's security and compliance status, including breach histories and data handling practices.

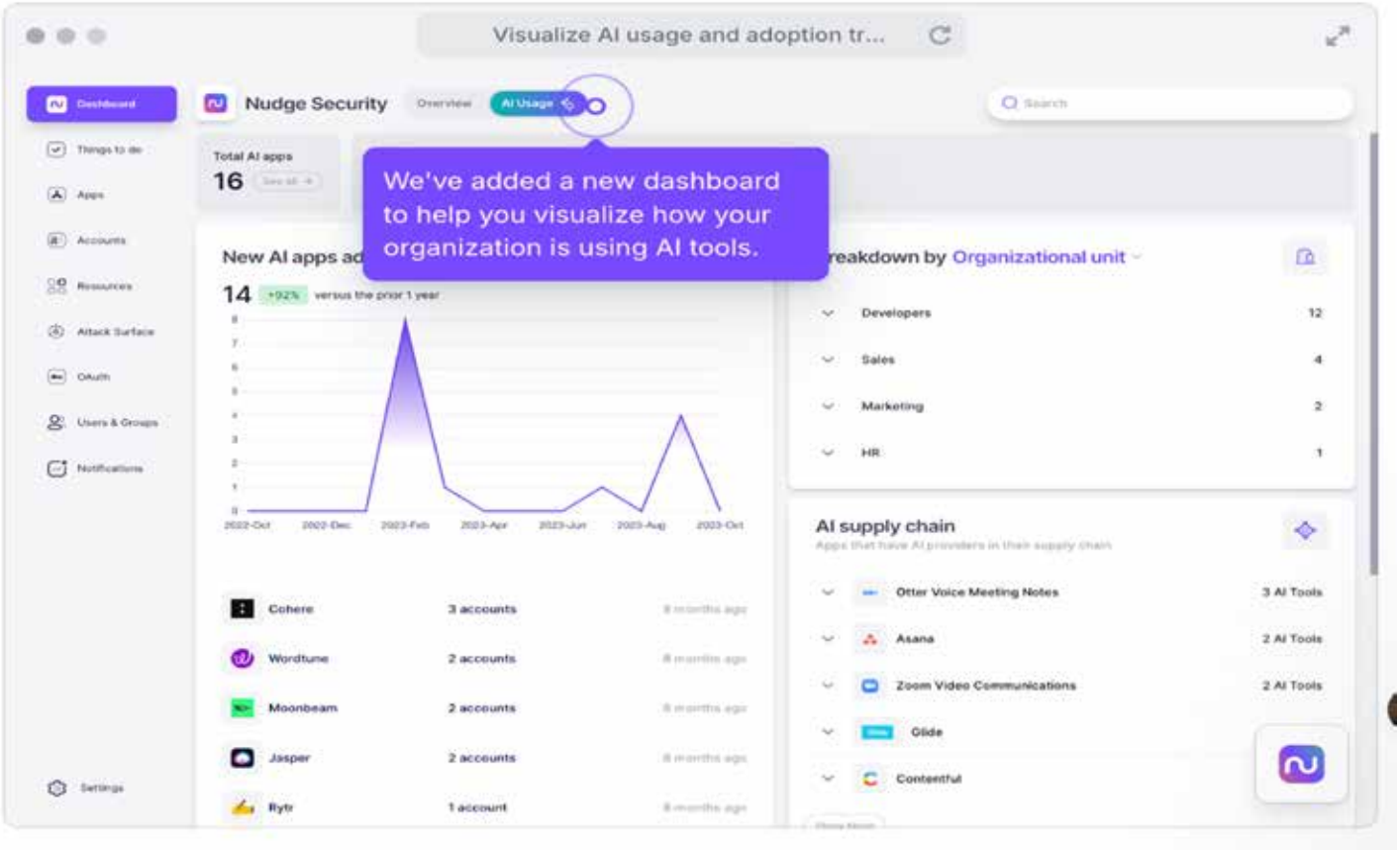

Figure 2. Nudge Dashboard Showing SaaS Evidence of AI Usage

To address the supply chain risks of AI integration, Nudge offers automated playbooks that deliver just-in-time security nudges to employees. These "nudges" educate users about safe and compliant AI usage, ensuring that employees are aware of the potential security implications when interacting with AI tools by reviewing the organization's AI acceptable use policy as they begin to adopt AI tools. By engaging the right users at the right time, Nudge helps to drive a just-in-time AI governance model without hindering productivity.

It is also worth mentioning that Nudge's platform includes OAuth SaaS integration risk management, which is essential for monitoring app-to-app integrations that may involve AI components. By providing a full inventory of integrations: native marketplace apps, API tokens, webhooks, OAuth connections, scopes, and risk scores, organizations can proactively manage third-party data access and revoke risky integrations with ease. This is crucial in preventing unauthorized data sharing between SaaS apps and AI tools.

NEXT STEPS

Our team at TAG was excited to learn about Nudge's ability to provide a meaningful guardrail for SaaS use of AI. Most readers will agree that guardrails are emerging around direct generation AI usage such as ChatGPT. Reverse proxies, for example, serve as acceptable means for implementing filtering and DLP for such usage – and commercial vendors such as Zscaler and others are actively marketing such capability.

Guardrails for SaaS use of AI have been lacking, however – at least, until now. To that end, we strongly recommend that enterprise security teams and AI committees engage with Nudge for more information on how the system works. TAG Research as a Service (RaaS) customers can also reach out to a TAG analyst through their TAG RaaS portal account. We can help to sort out how guardrails from Nudge can integrate with broader AI security initiatives.

ABOUT TAG

Recognized by Fast Company, TAG is a trusted next generation research and advisory company that utilizes an AI-powered SaaS platform to deliver on-demand insights, guidance, and recommendations to enterprise teams, government agencies, and commercial vendors in cybersecurity and artificial intelligence.