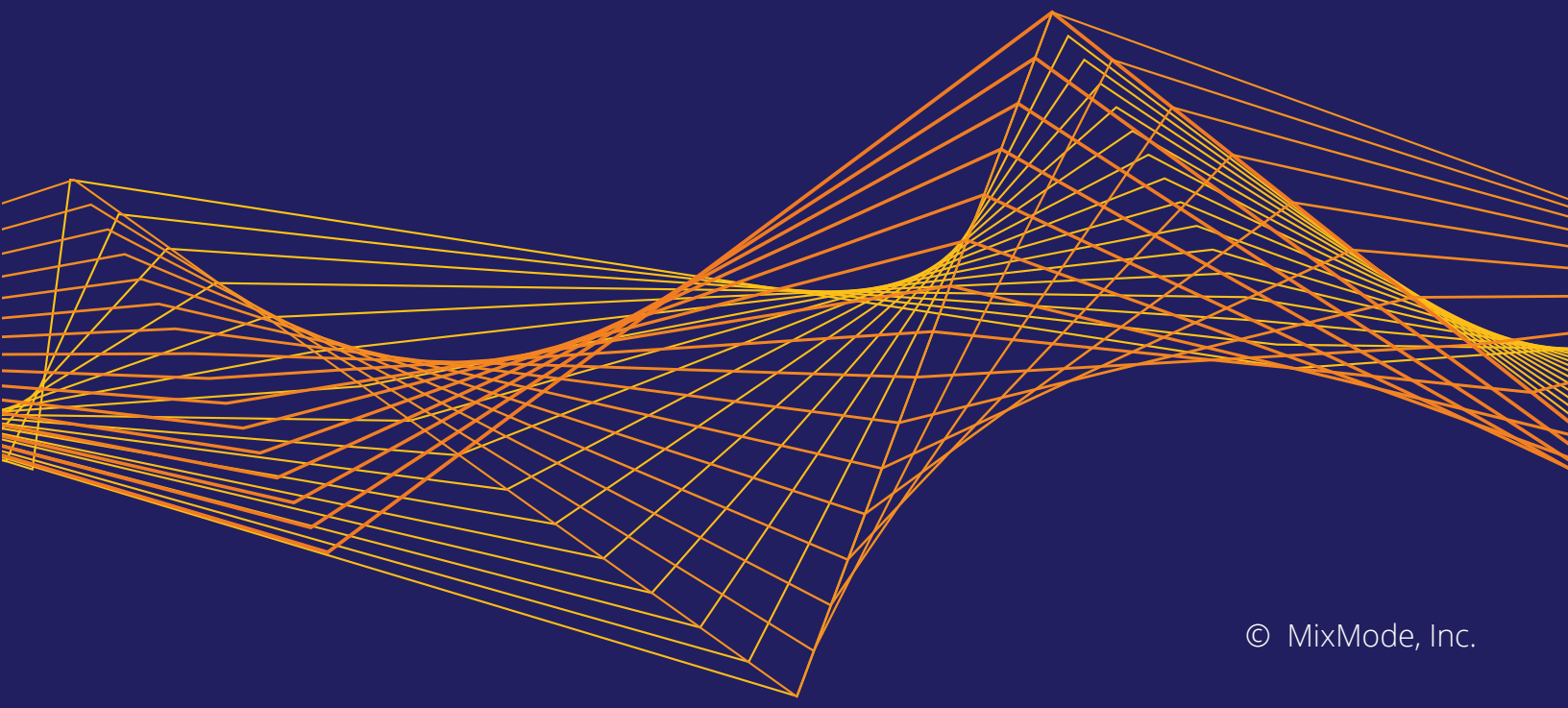
MixMode®

THREAT REPORT

How AI Assistants, Co-Pilots, and Chatbots Create New Cyber Threats

Michael Yelland
MixMode's Architect & Principal Researcher



AI applications are increasingly integrated into everyday workflows, assisting with automation, decision-making, and data processing. However, these same AI tools—ranging from copilots and chatbots to large-scale language models—present significant security risks.

Recent concerns around DeepSeek have highlighted the dangers of AI data collection, but the issue extends far beyond a single AI tool. The real risk lies in the broader AI ecosystem, where many platforms operate with minimal transparency, unknown data handling practices, and potential geopolitical influences.

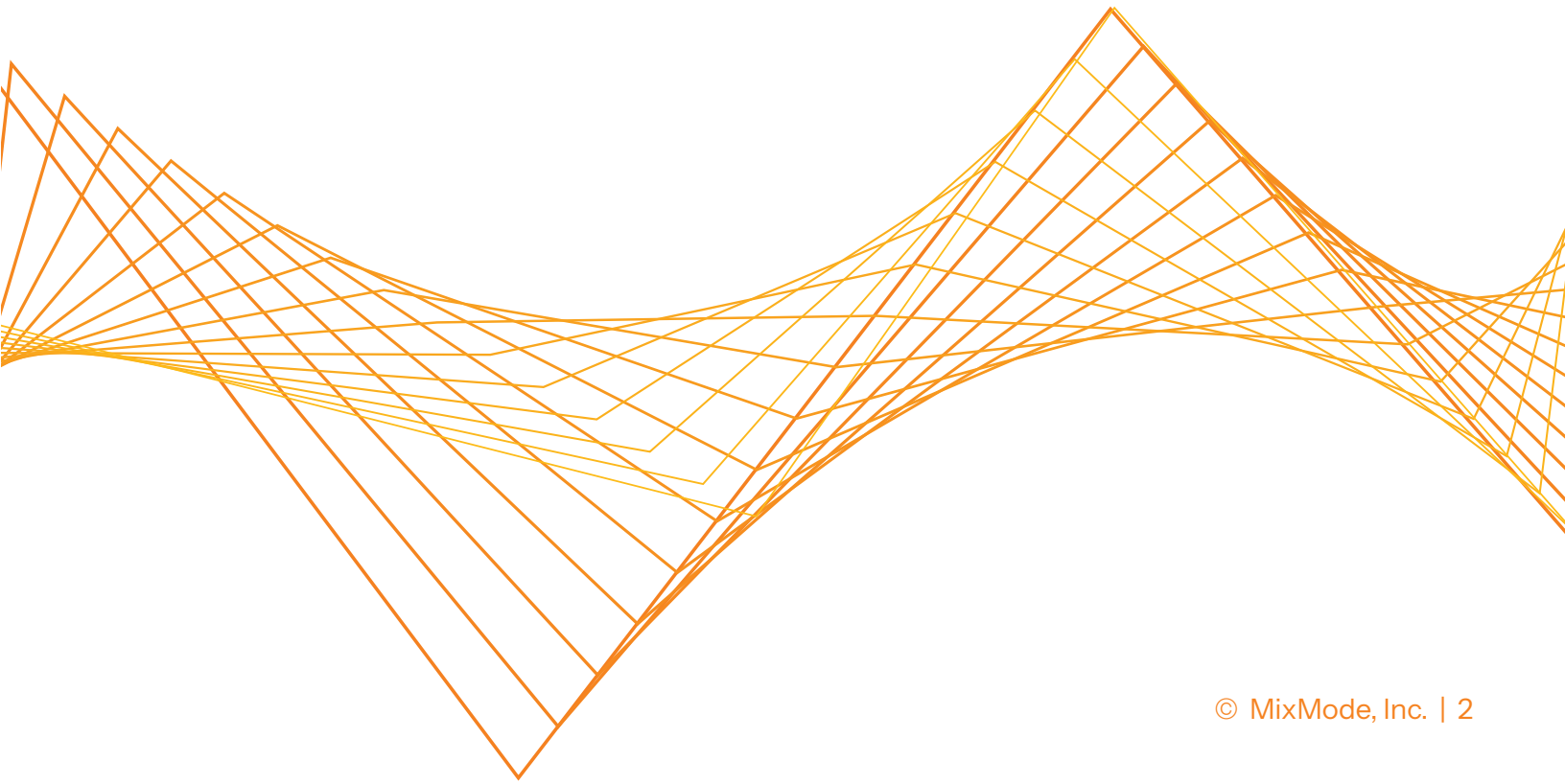
This report examines how AI-powered applications can become insider threats, how domain filtering is an inadequate security solution, and how MixMode's AI-driven detection provides a more effective defense against AI-assisted cyber risks.

AI as an Insider Threat: The Risks of Unchecked AI Data Collection

Many organizations unknowingly allow AI tools to access sensitive data, assuming they operate within a controlled environment. However, these AI platforms often log interactions, process keystrokes, and even analyze local files—all without clear user oversight.

Common security risks associated with AI-powered tools include:

- **Session Hijacking & Persistent Authentication Risks:** AI chatbots and assistants store user credentials for seamless interaction, but this can lead to unauthorized data access.
- **Data Harvesting & AI Training Risks:** Many AI models collect user-generated content to refine their responses, creating potential intellectual property (IP) and data leakage risks.
- **Geopolitical Exposure:** AI platforms hosted in foreign cloud infrastructures may be subject to external government access or legal obligations.



To illustrate these risks, below is a table analyzing major AI platforms and their associated concerns:

AI Platform	Organization (Country & Cloud Infrastructure)	Data Collection Risk	Domain
Google Gemini	Google (USA, Google Cloud)	AI model integrates deeply with cloud services, collecting extensive user data.	gemini.google.com
Claude	Anthropic (USA, AWS)	Stores conversation logs, raising privacy concerns.	claude.ai
Meta AI	Meta (USA, Meta Cloud / AWS)	Integrated with Facebook and Instagram, gathering personal user data.	meta.ai
Meta AI	Meta (USA, Meta Cloud / AWS)	Integrated with Facebook and Instagram, gathering personal user data.	meta.ai
Microsoft Copilot	Microsoft (USA, Azure)	Can access and suggest code from private repositories, raising IP security concerns.	copilot.microsoft.com
Grok	X/Twitter (USA, Likely X Internal Cloud / AWS)	Collects real-time social media interactions and user data.	x.ai
Perplexity AI	Perplexity AI (USA, AWS)	Actively crawls and stores web interactions, raising data retention concerns.	perplexity.ai

AI Platform	Organization (Country & Cloud Infrastructure)	Data Collection Risk	Domain
Poe	Quora (USA, AWS)	Aggregates multiple AI models, raising ambiguity over data ownership.	poe.com
OpenAI Play-ground	OpenAI (USA, Azure (Microsoft Partnership)	Processes API requests with stored user input logs.	platform.openai.com
GitHub Copilot	GitHub (USA, Azure)	Generates and stores coding suggestions, potentially exposing proprietary code.	github.com
Hugging Chat	Hugging Face (France, AWS)	Open-source, but logs API interactions and model usage.	huggingface.co
Character. AI	Character.AI (USA, Google Cloud / AWS)	Builds chat personalities using user conversation data.	character.ai
Mistral Chat	Mistral AI (France, AWS / European Cloud Providers)	Open-source, but requires external API access, logging user data.	mistral.ai
GroqChat	Groq (USA, AWS)	Optimized for low-latency inference, raising concerns about real-time data processing.	groq.com

AI Platform	Organization (Country & Cloud Infrastructure)	Data Collection Risk	Domain
Phind	Phind AI (USA, AWS)	Tailored for developers, indexing search queries and interactions.	phind.com
You.com	You.com (USA, Google Cloud / AWS)	Integrates AI with search results, tracking browsing activity.	you.com
Pi (Inflection AI)	Inflection AI (USA, Google Cloud / AWS)	Marketed as a personal assistant but logs user interactions.	heypi.com
DeepSeek	DeepSeek (China, Alibaba Cloud / Tencent Cloud)	Servers in China, with concerns over data access by the Chinese government.	deepseek.com
Perplexity Assistant	Perplexity AI (USA, AWS)	Enables deep web search and voice command processing.	perplexity.ai
ChatGPT Gov	OpenAI & Microsoft (USA, Azure Government Cloud)	Restricted to government agencies but operates within Microsoft's cloud.	openai.com
Colosseum 355B	iGenius (Italy, European Cloud Providers)	Targeted at regulated industries, collecting sensitive data.	igenius.ai

AI Platform	Organization (Country & Cloud Infrastructure)	Data Collection Risk	Domain
Operator (OpenAI)	OpenAI (USA, Azure)	Automates web interactions, potentially exposing session data.	openai.com
Sound-Hound Voice AI	SoundHound (USA, AWS)	AI integrates with smart devices, collecting voice data.	soundhound.com
LockedIn AI	LockedIn AI (USA, AWS)	Assists with interviews but stores responses and analytics.	lockedin.ai
Baiduspi-der	Baidu (China, Baidu Cloud)	China's primary web crawler, indexing data for AI training.	baidu.com
Bytespider	ByteDance (China, ByteDance Internal Cloud / Alibaba Cloud)	TikTok's AI-driven web crawler, harvesting user data.	bytedance.com
Yandex Bot	Yandex (Russia, Yandex Cloud)	Russian search engine bot indexing global content, raising espionage concerns.	yandex.com

These AI applications—and many others—operate with varying levels of transparency and security controls, making them a critical area of concern for enterprises and government agencies alike.

Legacy Domain Filtering: An Outdated Security Measure

For years, organizations have relied on domain filtering to block access to malicious or unauthorized services. While this approach provides a basic level of defense, it is increasingly ineffective against AI-powered threats. Attackers leveraging AI can easily circumvent domain filtering through:

- Dynamic Domain Generation (DGA): AI can generate thousands of alternate domains to avoid detection.
- Legitimate AI Service Exploitation: Cybercriminals can use trusted AI chatbots or copilots as command-and-control (C2) servers.
- Rapid Domain Pivoting: Threat actors can quickly switch to new, unlisted domains before blacklists are updated.

Traditional cybersecurity measures struggle to keep up with AI-driven evasive techniques, requiring a more adaptive approach to AI security monitoring.

How MixMode Detects AI-Powered Threats Before They Escalate

MixMode takes a proactive approach to AI-based threats by leveraging real-time, self-learning AI to monitor and detect suspicious activity. Unlike traditional security tools that rely on predefined rules, MixMode continuously adapts to emerging threats by analyzing behavior across networks and cloud environments.

Key Capabilities of MixMode's AI-Driven Detection:

- **Behavioral Detection Instead of Static Signatures**
  - Analyzes AI-generated traffic patterns to detect automated or malicious AI activity.
  - Flags unexpected AI-driven C2 domains before they are widely recognized as threats.
- **Real-Time AI-Powered Traffic Correlation**
  - Correlates DNS logs, proxy data, and network traffic to detect unauthorized AI usage.
  - Identifies AI services acting as unauthorized data exfiltration channels.
- **Subdomain & Alternate Domain Monitoring**
  - Tracks new AI-related subdomains and lookalike domains that may indicate emerging threats.
  - Monitors AI model interactions for suspicious behavioral anomalies.
- **Identifying AI-Powered User Behavior Anomalies**
  - Detects deviations in normal user behavior that suggest AI-assisted automation.
  - Differentiates between legitimate AI-enhanced workflows and unauthorized AI use.
- **Hunting for AI-Generated Network Traffic**
  - Recognizes unexpected interactions with large language model APIs.
  - Flags stealthy AI-assisted reconnaissance traffic.

By focusing on real-time behavioral monitoring and anomaly detection, MixMode provides security teams with a decisive advantage against AI-powered cyber threats.

Why MixMode's AI-Driven Security is Essential for Today's Threat Landscape

The increasing adoption of AI-powered applications presents a dual challenge: balancing productivity benefits with security risks. While AI can be used to enhance efficiency, it also introduces new attack vectors that traditional security measures fail to address.

MixMode's AI-driven security approach ensures that:

- AI threats are detected **before** they can exploit enterprise infrastructure.
- Organizations gain real-time **visibility** into AI-powered data exfiltration attempts.
- Security teams receive **proactive alerts** rather than relying on outdated blacklists.

As AI continues to evolve, so must the security strategies that protect organizations from its misuse. **With MixMode's self-learning AI, enterprises can stay ahead of AI-powered threats—before they escalate into full-blown breaches.**