Spin.ai

REPORT

AI-Driven Browser Extension Risks in 2025

Over **50%** of Extensions are **High Risk**

Authors: Anton Tkachenko, Davit Asatryan, Rainier Gracial

TABLE OF CONTENTS

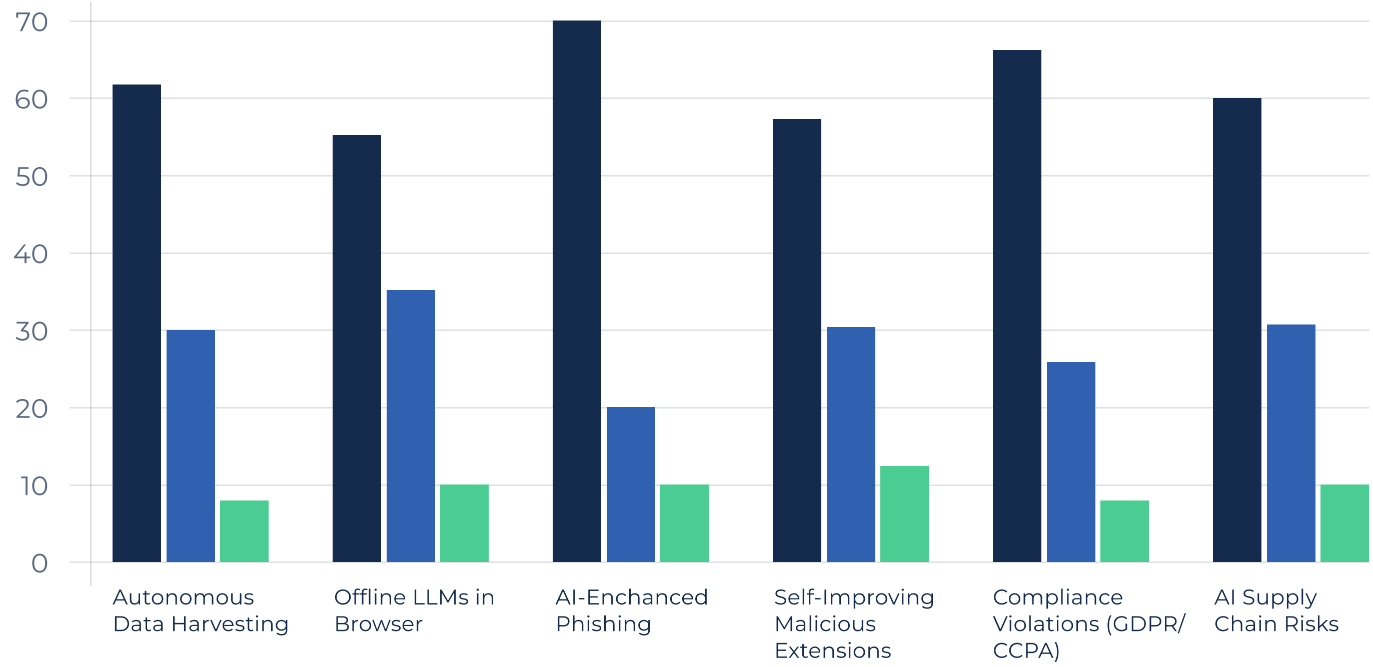
01

AI-DRIVEN BROWSER EXTENSION RISKS IN 2025

The rise of generative AI and machine learning capabilities has led to a sharp increase in AI-powered browser extensions, introducing new and complex threat vectors. While AI-based tools can boost productivity (e.g., AI writing assistants, code generation plugins, AI chatbots, etc.), they also bring heightened security and compliance risks.

Below are key risk areas organizations must consider:

AI-Driven Browser Extensions: Security Risk Distribution by Category (2025)

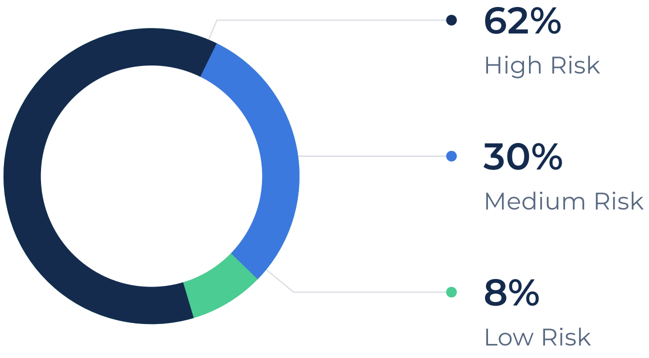


AI Extension Risk Factor	● High Risk	● Medium Risk	● Low Risk
Autonomous Data Harvesting	62%	30%	8%
Offline LLMs in Browser	55%	35%	10%
AI-Enchanced Phishing	70%	20%	10%
Self-Improving Malicious Extensions	58%	30%	12%
Compliance Violations (GDPR/CCPA)	67%	25%	8%
AI Supply Chain Risks	60%	30%	10%

1.1 AI-DRIVEN BROWSER EXTENSION RISKS IN 2025

Modern AI-driven extensions often feature autonomous data scraping capabilities. These AI models can:

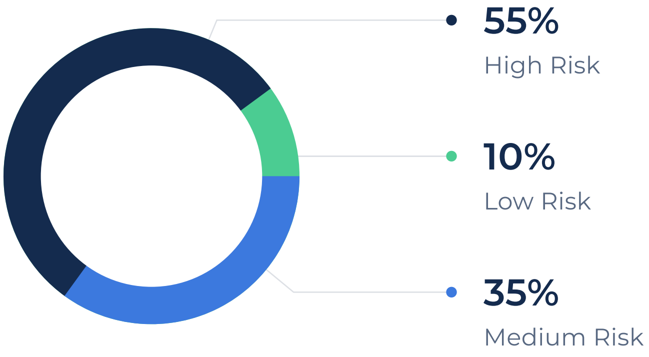
- **Automatically extract PII and sensitive business data** from web pages, SaaS platforms (e.g., CRM or HR software), and browser session cookies.
- **Summarize or reprocess user data** (e.g., through AI-generated summaries) and transmit outputs to third-party servers without user consent.



1.2 OFFLINE LLMS OPERATING IN THE BROWSER

In 2025, some extensions now embed lightweight large language models (LLMs) directly into the extension itself, capable of operating offline. This:

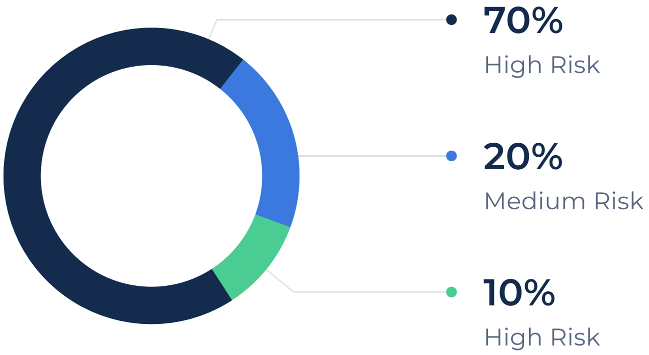
- **Bypasses perimeter defenses and DLP tools** that typically monitor network traffic for data exfiltration.
- Increases the likelihood of **"shadow AI" risks**, where sensitive corporate data is ingested into opaque AI models on unmanaged or non-compliant extensions.



1.3 AI-ENHANCED PHISHING AND SOCIAL ENGINEERING

AI-driven extensions can:

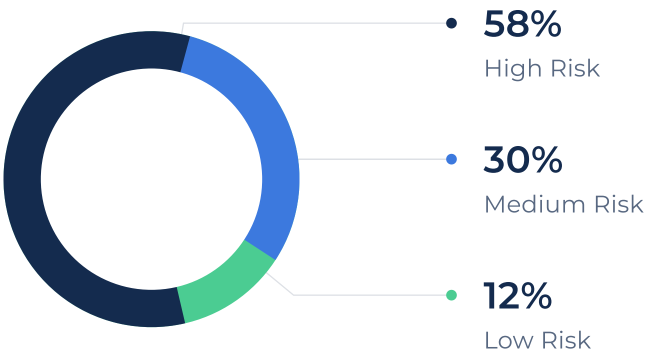
- **Generate real-time phishing payloads or social engineering scripts** customized to the user's behavior and browsing patterns.
- Interact with SaaS platforms on behalf of the user, **injecting fraudulent comments, emails, or messages** automatically within CRM, communication, or project management tools.



1.4 SELF-IMPROVING MALICIOUS EXTENSIONS

Through AI/ML, some malicious extensions:

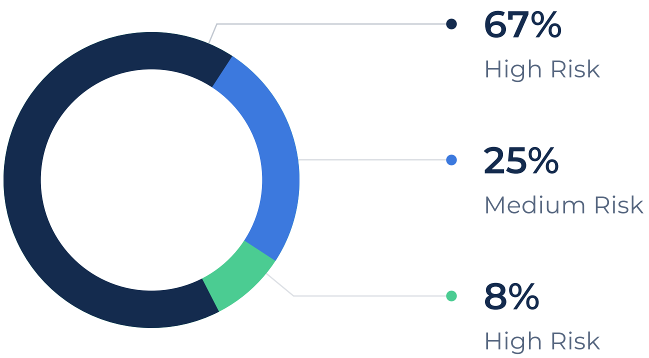
- **Adapt their behavior dynamically** based on detection signals (e.g., adjusting activity when inside a corporate environment).
- Use **reinforcement learning** to optimize attack patterns, reducing the likelihood of detection by traditional signature-based defenses.



1.5 DATA PRIVACY AND COMPLIANCE VIOLATIONS

AI-driven extensions can:

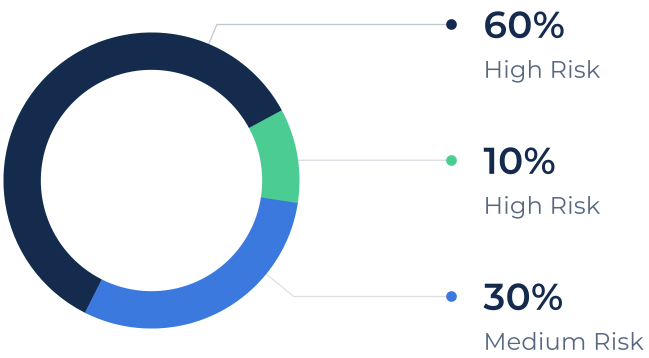
- AI-driven extensions frequently **collect behavioral analytics**, screen captures, and keystroke patterns to improve model performance.
- This creates **GDPR, CCPA, and HIPAA violations,** especially when extensions transmit anonymized or pseudonymized data across borders or sell telemetry to third-party brokers.



1.6 SUPPLY CHAIN RISK FROM AI MODEL PROVIDERS

AI model providers can:

- Many AI extensions **depend on third-party AI APIs or models** (e.g., APIs provided by open-source AI platforms or cloud LLM providers).
- This introduces **supply chain vulnerabilities** if the AI provider itself is compromised or if model updates introduce unexpected data-handling behavior (e.g., via trojanized model weights or poisoned datasets).



02

EMERGING AI BROWSER RISK EXAMPLES (2024-2025)

The cybersecurity landscape has seen significant evolution in how AI-driven browser extensions are used — both as productivity tools and attack vectors. Below are some recent examples and trends observed in 2024 and early 2025:

1.1 AI-AUGMENTED PHISHING AUTOMATION

In early 2024, a popular AI browser extension originally marketed as a productivity enhancer for Salesforce and HubSpot users was found exfiltrating sensitive CRM data. The extension leveraged its AI model to automatically extract lead lists, customer PII, and internal sales data, sending it to an external cloud server without user consent. This breach went undetected for months because the extension's offline LLM allowed it to perform data extraction even without internet connectivity, queuing data for later upload when the user was online.

1.2 AI-POWERED DATA EXTRACTION TOOLS TURNED MALICIOUS

In mid-2024, threat actors deployed AI-powered browser extensions masquerading as marketing automation tools. These extensions used embedded generative AI models to:

- Analyze user activity within corporate Google Workspace and Microsoft 365 accounts.
- Automatically generate targeted phishing emails or Slack messages that mimicked the tone and structure of internal communications.
- Deliver these payloads directly within logged-in sessions, bypassing traditional email security gateways.

This trend demonstrates how AI can weaponize native browser sessions and automate highly convincing social engineering campaigns.

1.3 CHATGPT-INSPIRED MALICIOUS EXTENSIONS

Following the surge of generative AI in 2023, by 2024 dozens of rogue AI-powered browser extensions mimicking ChatGPT and similar AI services began proliferating. A particularly damaging case involved a fake AI assistant extension that hijacked session cookies and OAuth tokens from logged-in users on platforms like Salesforce, ServiceNow, and Microsoft Teams. The stolen tokens enabled lateral movement within enterprise SaaS ecosystems, leading to significant data leakage events across multiple industries.

1.3 AI-SUPPORTED CREDENTIAL STUFFING VIA BROWSERS

In late 2024, security researchers uncovered extensions with embedded AI models trained to:

- Identify login forms across websites automatically.
- Preload common credential stuffing lists and attempt automated logins across SaaS platforms like Atlassian, Trello, and Confluence.
- Obfuscate login attempts to appear as legitimate user activity, making detection by traditional SIEM and UEBA solutions more difficult.

This highlights how AI is enabling "intelligent automation" in credential theft campaigns directly from within the browser.

1.3 CHATGPT-INSPIRED MALICIOUS EXTENSIONS

An incident in early 2025 involved a widely used AI extension for legal document summarization. Attackers infiltrated the supply chain of a third-party AI API provider used by the extension. Malicious payloads were embedded into model updates, allowing silent injection of rogue JavaScript into browser sessions that targeted data-rich environments, such as corporate legal portals and contract management SaaS tools.

03

RECOMMENDATIONS FOR ORGANIZATIONS (2025)

As browser extensions continue to evolve in complexity, particularly with the integration of AI/ML technologies, organizations must adopt a layered and proactive security strategy. Below are comprehensive recommendations for mitigating the risks associated with both AI-driven and conventional extensions:

1.1 BUILD A REAL-TIME AI-AWARE EXTENSION INVENTORY

- Maintain a **continuous and automated inventory** of all extensions deployed across corporate-managed and BYOD endpoints.
- Ensure this inventory also identifies **AI-enabled extensions** or those utilizing external AI APIs (e.g., generative AI tools, summarization plugins, AI chatbots).
- Classify extensions based on business justification, operational criticality, and vendor trustworthiness.

1.2 INTEGRATE BROWSER SECURITY INTO YOUR ZERO TRUST STRATEGY

- Implement **browser-level Zero Trust policies** that enforce least-privilege access, including restrictions on data access, clipboard, screen capture, and cookie permissions for extensions.
- Consider deploying **secure enterprise browsers** (e.g., Talon, Island, or Chrome Enterprise with enhanced controls) that allow granular management of browser plugins/extensions.
- Use **browser isolation technology** (remote browser isolation) for high-risk workflows or SaaS access.

1.3 AI-SPECIFIC EXTENSION POLICIES

- Develop a dedicated **AI Governance Framework** that includes security and compliance guidelines for AI-powered extensions.
- Require **AI transparency** from extension developers, including disclosures on AI model usage, data processing activities, and third-party AI APIs leveraged.
- Prohibit or restrict extensions using **local/offline AI models** unless approved by the security team.

1.4 CONTINUOUS RISK SCORING & BEHAVIORAL MONITORING

- Adopt AI-powered SSPM (SaaS Security Posture Management) and **Browser Security Platforms (BSP)** to provide **real-time risk scoring** for all extensions.
- Enable **behavioral monitoring** to detect suspicious actions such as automated data scraping, unauthorized API calls, or AI-powered automation within the browser.
- Ensure that risk scoring models account for **AI-specific risks** like model drift, shadow AI behavior, and unauthorized AI-API usage.

1.5 AUTOMATE POLICY ENFORCEMENT

- Enforce **automated allow/deny policies** based on continuous risk assessments:
  - Automatically block high-risk or unknown AI extensions from accessing critical SaaS environments.
  - Dynamically restrict permissions (e.g., limit **webRequest, identity**, or **clipboardRead**) based on contextual factors such as device type, user role, or geo-location.
- Integrate browser extension controls with your **CASB, SWG, or ZTNA solutions** for a unified security policy across endpoints and browsers.

1.6 STRENGTHEN THIRD-PARTY RISK MANAGEMENT (TPRM)

- Include browser extension developers and AI API providers in your **third-party risk management program**.
- Perform **due diligence and regular security audits** on vendors supplying AI-based browser extensions, looking for red flags such as:
  - Lack of security certifications (e.g., SOC 2, ISO 27001).
  - Insufficient disclosure of AI model behavior and data flows.
  - History of poor security practices or past breaches.

1.7 EDUCATE EMPLOYEES ON MODERN EXTENSION THREATS

- Launch **targeted security awareness training** programs focused on modern extension risks, especially around AI-enabled plugins and shadow extensions.
- Provide **visual guides** and playbooks on how to recognize potentially malicious extensions, including risks like AI phishing bots and data-mining assistants.
- Teach users to report extensions showing unusual behavior, such as unexpected automation or unauthorized pop-ups.

1.4 PREPARE FOR COMPLIANCE AUDITS

- Map your extension ecosystem to **regulatory frameworks** such as GDPR, HIPAA, CCPA, and the EU AI Act.
- Ensure that sensitive data processed by AI extensions does not violate **data sovereignty rules** or cross-border data transfer restrictions.
- Keep **audit-ready logs** of extension usage, risk scores, and remediation actions taken by security teams.

04 PRO TIPS

Consider deploying **sandbox environments** to safely evaluate new AI-driven extensions before production rollout.

Prioritize AI anomaly detection: Invest in tooling that detects abnormal AI model behavior inside extensions (e.g., LLMs making unsolicited network requests or dynamically rewriting DOM elements).

About Spin.AI

Spin.AI is an innovative provider of SaaS security solutions for mission-critical SaaS apps (Microsoft 365, Google Workspace, Salesforce, and Slack). Our all-in-one SpinOne platform helps organizations mitigate risk, save time, reduce downtime, and improve compliance.

BOOK A DEMO

info@spin.ai

www.spin.ai