BTAC BULLETIN

BEHAVIORAL SCIENCE | LAW ENFORCEMENT & COUNTERINTELLIGENCE | CYBERSECURITY | EMPLOYEE MANAGEMENT RELATIONS | THREAT ASSESSMENT & MANAGEMENT

EMERGING INSIDER RISK IN THE AGE OF AI

The rapid adoption of Artificial Intelligence (AI) in the workplace presents a paradigm shift that has fundamentally altered the insider threat landscape. Driven by a desire for increased productivity, well-intentioned personnel are adopting unsanctioned AI tools and services. This widespread use of "Shadow AI" is rooted in the unique ways users integrate these systems into their daily workflows. Shadow AI encompasses two distinct threat vectors: the intentional use of external commercial or private LLMs, and the activation of embedded AI features within existing government and sensitive networks that have not yet been fully evaluated for security risks. When bypassing traditional security controls, both vectors create a massive, unmonitored attack surface where new risks outpace current technical safeguards and governance. For Insider Threat programs, this lack of oversight into data leakage pathways means Shadow AI is no longer a future concern; it is a primary vector for Unauthorized Disclosure (UD), data spillage, and potential operational security (OPSEC) failures.

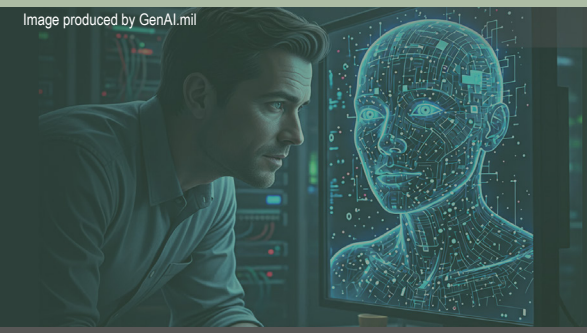

Image produced by GenAI.mil

The Rise of Shadow AI

$10.3M ANNUALLY

17% FROM PREVIOUS YEAR

Making negligence the costliest insider risk category

Source 1

The Governance Visibility Gap[1]

There is a significant disconnect between AI adoption and oversight. This gap creates significant blind spots for security programs.

- **92%** Say generative AI has fundamentally changed how employees access and share information
- **13%** Have formally integrated AI into their business strategies
- **18%** Have fully integrated AI governance into their insider risk management programs
- **44%** Have minimal to no visibility into AI agent activity

Data Leakage Pathways[3]

Leakage Pathway	Threat Description & UD Risk
Public LLMs	Inputting sensitive data into commercial AI tools (e.g., ChatGPT, Claude) to draft memos or write code, which directly transmits proprietary information to external servers.
Intimate Self-Disclosures	Oversharing personal vulnerabilities, financial struggles, or work frustrations with unregulated AI companions (e.g. Character, Nomi, Botify.AI). This creates severe counterintelligence, profiling, and blackmail vulnerabilities.
Oversharing for Productivity	Using public AI to process performance evaluations or disciplinary documents, inadvertently exposing PII and CUI to external AI training models.
AI Meeting Notetakers	Inviting unsanctioned AI assistants (e.g., Otter.ai, Fireflies) to internal conferences. These tools record sensitive verbal discussions, completely bypassing Data Loss Prevention (DLP) controls.
Agentic AI & Extensions	Using unauthorized browser extensions (e.g., grammar checkers, PDF summarizers) that continuously read and transmit on-screen text to external servers, expanding the threat boundary to the desktop.

Securing the Network: A Three-Pillar Strategy to Counter Shadow AI

Policy and Governance[1]

Update Acceptable Use Policies (AUPs):

Explicitly address the risks of both personal AI companion services and the use of public LLMs for official work. Prohibit the input of any sensitive, non-public, or CUI information.

Mandate Approved Tools:

Prohibit unauthorized LLMs and direct personnel to use only approved, secure government platforms like GenAI.mil.

Technical Monitoring & Controls[2]

User Activity Monitoring (UAM):

Tune UAM strategies to detect behavioral indicators associated with high-risk AI use, such as concerning conversations with AI chatbots and companions.

Network Hardening:

Implement countermeasures to detect and block unauthorized AI-enabled tools that could introduce malware or backdoors.

Workforce Education & Training[2]

Reframe the Threat:

Educate the workforce that inputting data into a public AI prompt is the exact equivalent of posting that data on a public web page.

Promote Secure Practices:

Train personnel on the appropriate use of secure AI, proper data handling, and the critical importance of OPSEC in the age of generative models.

1. DTEX, & Ponemon Institute. (2026). 2026 Cost of insider risks global report. DTEX Systems. 2. Narayan SB (2025) What is One Effective Way Organizations Can Reduce the Risk of Insider Threats Without Disrupting Productivity. J Eng Artif Intell Vol.1 No.2: 15. 3. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.