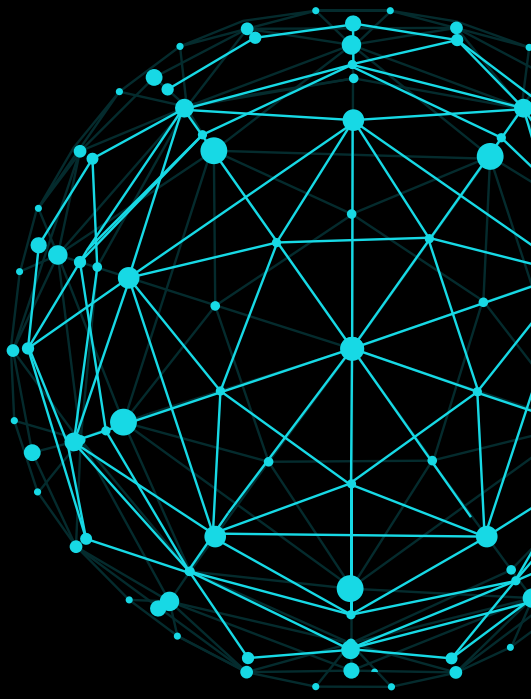
The Hidden Threat of **Shadow AI**



Pat Arvidson, Chief Technology Officer (CTO)
Jennifer Gold, Chief Information Security Officer (CISO)

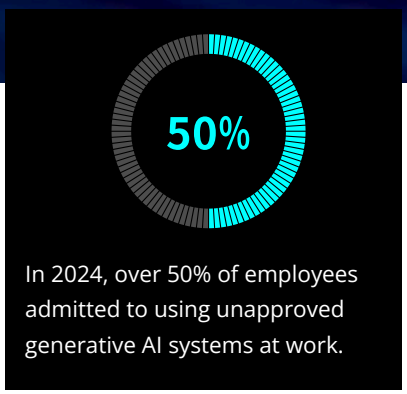
www.risk-aperture.com

About Risk Aperture

Risk Aperture delivers holistic cyber exposure insights, empowering business leaders to effectively manage residual risk and prioritize cybersecurity investments. Leveraging intelligence-driven analysis, research, and data analytics, we provide rapid, actionable insights to support informed decision-making and return on security investment.

Risk Aperture is led by CIA, NSA, and DoD veterans with decades of experience protecting national security and critical infrastructure assets.

To learn more visit us at https://risk-aperture.com/



In 2024, over 50% of employees admitted to using unapproved generative AI systems at work.

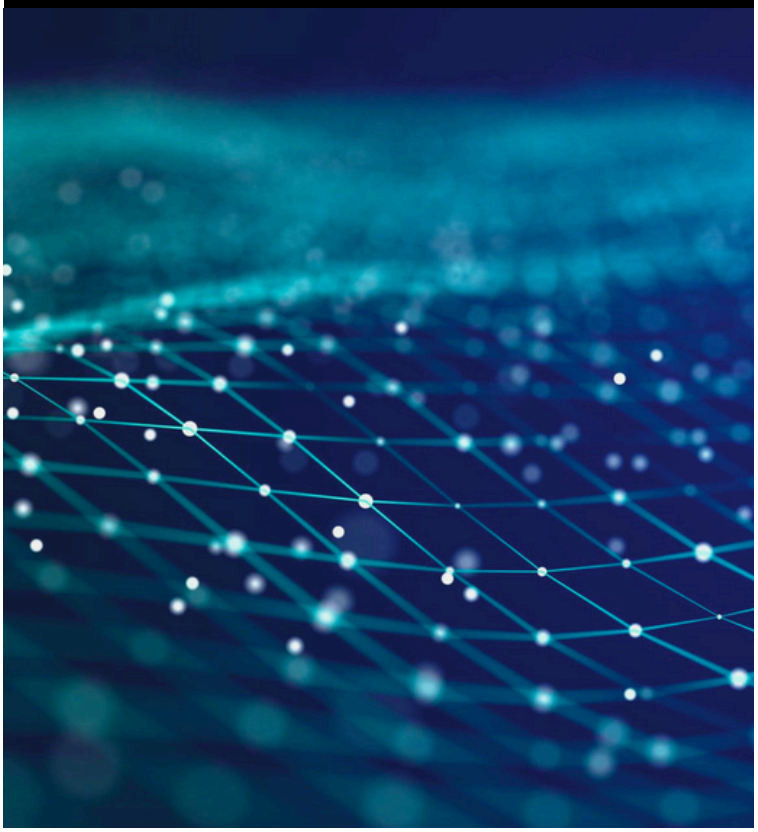
Executive Summary

WHAT IS SHADOW AI?

Artificial Intelligence (AI) is transforming business operations, enabling automation, generating data-driven insights, and accelerating innovation. Yet, as AI systems become more integrated into everyday processes, a new risk has emerged: Shadow AI. Shadow AI refers to unauthorized, unapproved, and undocumented AI tools and models used without the knowledge or authorization of IT, compliance, or security teams. These tools, ranging from simple automation scripts to advanced generative AI models, pose significant threats to organizations. If left unchecked, Shadow AI can lead to security breaches, data leaks, operational inefficiencies, and regulatory violations creating multiple business risks and ultimately jeopardizing a company's intellectual property (IP) and competitive advantage.Shadow AI encompasses any AI tools, algorithms, or systems used within an organization without proper authorization or oversight. Shadow AI is often adopted by individuals or departments seeking efficiency but bypassing formal governance processes. It includes a wide range of applications, from machine learning models to generative AI tools. Shadow AI is also known as Bring Your Own AI (BYOAI). The availability of generative AI models has contributed significantly to the increase in Shadow AI usage, which can expose organizations to risks, including data privacy violations, intellectual property exposure, and risks to operational integrity.

Futurum Group. "More than 50% of Workers Admit to Using Unapproved Generative AI Tools." 2024.

> *Shadow AI refers to unauthorized, unapproved, and undocumented AI tools and models.*



Introduction

THE RISE OF SHADOW IT AND SHADOW AI

Shadow IT has been a challenge since the 1990s, when employees began using unauthorized software and hardware to boost productivity. The rise of personal computing and external applications like email and file-sharing services fueled its growth. By the early 2000s, with the cloud computing boom and Bring Your Own Device (BYOD) movement, Shadow IT became common in enterprises. Tools like Dropbox, Google Docs, and Slack offered flexibility and speed but also introduced risks such as data breaches, regulatory violations, and governance gaps (Gartner, 2024).

A new challenge has emerged in 2024, Shadow AI. As AI platforms become widely accessible, employees and departments across industries are adopting AI tools—ranging from machine learning models to advanced generative AI systems—without proper oversight. Like Shadow IT, Shadow AI promises operational efficiency but brings security and governance concerns. Shadow AI is especially risky because these models interact with sensitive data and make decisions that affect core business processes. Without proper oversight, they can lead to biased outcomes, data exposure, and violations of regulations like GDPR and CCPA (Futurum Group, 2024).

Key Characteristics of Shadow AI

Shadow AI introduces a variety of risks because it operates without formal approval or oversight. Unlike sanctioned AI systems, these tools bypass established governance frameworks, creating significant vulnerabilities that can impact an organization across security, compliance, and operational domains.

Unapproved
Shadow AI tools are implemented without approval from essential departments like IT, security, or compliance. This lack of formal vetting opens the door to unmonitored vulnerabilities, outdated algorithms, and insecure protocols, which can result in unforeseen security risks and operational disruptions.

Unauthorized
These tools are deployed outside existing security protocols and regulatory frameworks. Without compliance checks, Shadow AI often lacks basic security measures like encryption, access controls, and data privacy protections, making the organization more vulnerable to cybersecurity threats, data breaches, and non-compliance with laws such as GDPR and CCPA.

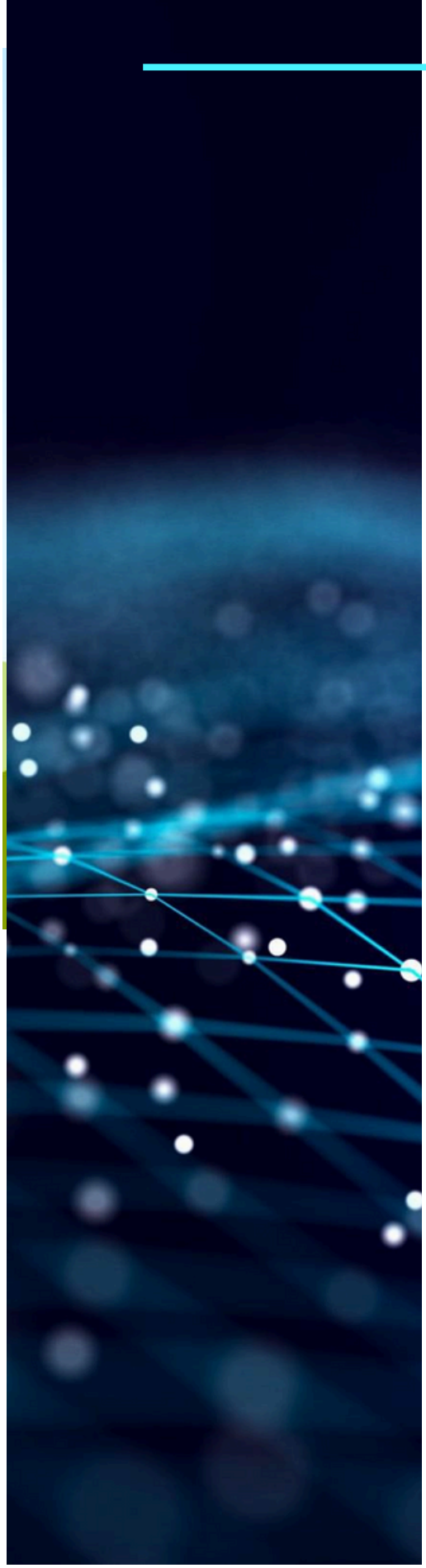
Undocumented
Shadow AI systems frequently lack proper documentation and integration with official IT systems, making them difficult to track, manage, or maintain. This absence of oversight leads to operational inefficiencies, an increased risk of model drift, and security blind spots that make it challenging to decommission or secure the tools when necessary.

With the growing dependence on AI to enable efficiency, employees are circumventing official authorization procedures, often without realizing the associated risks.

Comparison of Shadow IT vs. Shadow AI

Shadow AI and Shadow IT both involve unsanctioned technologies. While Shadow IT refers to unauthorized hardware and software within an organization, Shadow AI highlights the use of AI tools operating outside approved governance, posing greater risks due to opaque decision-making and unmonitored security vulnerabilities.

Aspect	Shadow IT	Shadow AI
Scope and Technology	Unauthorized use of general IT resources, software,and hardware	Specifically involves the unsanctioned use of AI tools and models
User Base	Typically adopted by tech-savvy employees or developers	Can be used by a wide range of employees, including non-technical users
Governance Challenges	Managed through security practices like encryption and access controls	Involves non-deterministic models, making traditional security measures less effective
Data Risks	Risks tied to unauthorized data access and storage	Additional risks such as data leakage, model bias, and sensitive data exposure
Regulatory Compliance	Well-understood compliance risks	Novel compliance challenges related to evolving AI regulations
Detection and Monitoring	Can be detected via network monitoring tools	Harder to detect, especially in cloud-based AI environments



The Risks of Shadow AI

Shadow AI introduces risks that are harder to detect and mitigate than traditional unauthorized software.

Security Vulnerabilities
AI systems process large amounts of sensitive data, making them prime targets for cyberattacks. Shadow AI tools, often lacking encryption, access controls, or monitoring, can lead to data breaches.

Data Privacy and Compliance Violations
AI tools often handle personal data, subject to regulations like GDPR and CCPA. Shadow AI's unregulated use increases the risk of non-compliance.

Model Drift and Accountability
AI models degrade over time as their training data changes, a phenomenon called model drift. Shadow AI, without proper oversight, is prone to drift, leading to biased outputs and legal liabilities, especially in compliance-sensitive tasks.

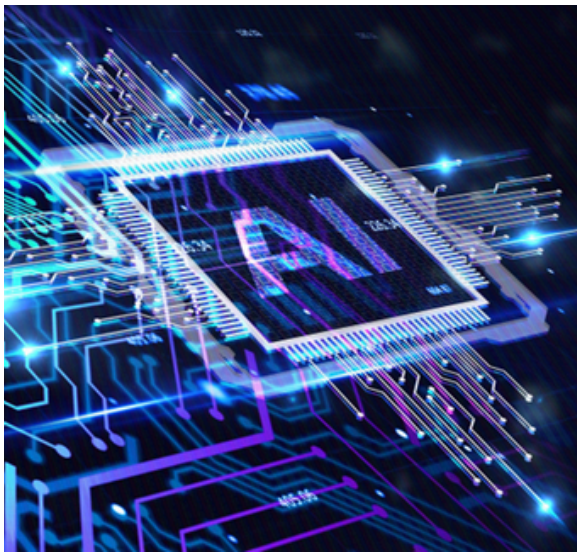
Exposure of Intellectual Property (IP)
Generative AI tools may require access to proprietary data. Without strict privacy controls, this can expose IP. In 2023, employees of a tech company accidentally exposed internal code using generative AI (Forbes, 2023), illustrating the risks of unregulated AI use.

Operational Inefficiencies
Unauthorized AI tools can fragment workflows and create inconsistent outputs. For instance, a marketing team using AI may produce content misaligned with company branding or compliance, leading to inefficiencies and duplicated efforts.

Generative AI Cybersecurity Threats

The rise of generative AI has unlocked immense potential for businesses, but it has also created new vulnerabilities that extend far beyond organizational walls. One of the most alarming developments is the infiltration of darknet forums, where generative AI tools and datasets have become prime targets for cybercriminals. As these AI models become more accessible, a troubling pattern has emerged, underscoring the urgent need for stronger governance and security measures.

In 2023, unauthorized "jailbroken" versions of a generative AI tool began circulating on darknet forums. These models bypassed built-in content moderation systems, allowing users to generate outputs that violate ethical and legal standards (Darktrace AI Research, 2023)..



The ease with which these models can be modified reveals how cybercriminals exploit generative AI when safeguards are inadequate. Without proper oversight, AI tools that are meant to innovate and streamline business operations can be weaponized for illicit purposes

Generative AI has also been turned into a tool for creating advanced malware. Cybercriminals are using AI to generate custom malware code, expanding both the scale and sophistication of their attacks (Security Affairs, 2023). The ability to produce malicious software quickly and efficiently illustrates the dangerous potential of unregulated AI. When AI falls into the wrong hands, it can dramatically amplify the capabilities of cybercriminals, threatening organizations of all sizes. tatur?

In addition to AI-generated malware, stolen credentials for popular generative AI platforms are now a hot commodity on darknet marketplaces. **Each day, roughly 400 generative AI account credentials are put up for sale,** offering unauthorized access to powerful tools. Some of these accounts are sold for as little as $15, well below the legitimate cost of usage (CSO Online, 2024). This underground trade exposes the platforms to abuse and also increases the risk of security breaches for the organizations using them.

Malicious AI Tools Circulating in Darknet Markets

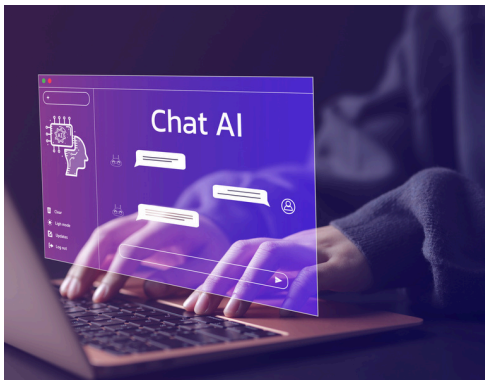
Darknet forums have become active markets for malicious AI tools built to facilitate cybercrime. These tools enable cybercriminals to launch increasingly sophisticated attacks, threatening individuals and entire industries.

Recent breaches involving AI platforms demonstrate an urgent need for stronger security measures. In one instance, a flaw in a generative AI tool's source code exposed users' chat histories and personal information. In another, employees at a major tech company unintentionally leaked sensitive corporate data, including source code and meeting transcripts, while using an AI tool for internal operations.

As AI becomes more integrated into business processes, its exploitation on darknet forums continues to grow. Some of the most dangerous tools in these underground markets include:

- **WormGPT**, based on the GPT-J model, which facilitates business email compromise (BEC) attacks.
- **FraudGPT**, a toolkit used to create malware, phishing sites, and hacking utilities.
- **XXXGPT,** deployed to control botnets and install remote access trojans (RATs), giving attackers full access to compromised systems.
- **PoisonGPT**, a proof-of-concept tool designed to manipulate open-source AI for spreading disinformation

The ease with which these models can be modified reveals how cybercriminals exploit generative AI when safeguards are inadequate. Cybercriminals are increasingly exploiting vulnerabilities in AI platforms, creating a thriving underground market for stolen credentials, malicious AI tools, and jailbroken models. This trend emphasizes the need for businesses to establish strong AI governance, secure deployments, and remain vigilant against emerging threats.



Managing the Risks of Shadow AI

Establish Clear Governance Policies

AI is both a driver of innovation and a source of hidden risks, with Shadow AI—unauthorized and unmonitored technology usage—among the most pressing concerns. Without oversight from IT or security teams, Shadow AI can compromise data security, breach regulatory standards, and erode a company's competitive advantage. Addressing this issue requires a proactive and structured approach centered around comprehensive governance. The following seven strategies can help organizations mitigate the risks of Shadow AI.

Establish Clear Governance Policies
Effective governance begins with well-defined policies emphasizing security, privacy, and ethics across all tools. Every system within the organization must comply with these guidelines to ensure alignment with regulatory and internal standards. A centralized registry is essential for maintaining visibility into all technology assets. This registry tracks tools and provides insights into their purposes and deployment, allowing organizations to identify and address risks before they escalate.

- Conduct Continuous Audits and Monitoring
- Develop a Response Plan for Incidents

Conduct Continuous Audits and Monitoring

Technologies evolve dynamically, unlike traditional software, which can lead to unforeseen vulnerabilities or misalignments with business goals. Regular audits are essential, but ongoing monitoring is equally critical to minimizing risk. Automated tools can continuously detect unauthorized usage and flag potential threats, such as data leaks or regulatory violations. By integrating these tools into daily operations, organizations can spot risks early and prevent minor issues from escalating into major threats. Organizations can implement monitoring frameworks to respond faster to vulnerabilities.

Develop a Response Plan for Incidents

Even with vigilant monitoring, unauthorized technology-related incidents can still occur. Developing a specific incident response plan that includes clear protocols for identifying, containing, and remediating breaches is essential. Quick action is necessary to prevent regulatory violations and operational disruptions.

An effective response plan includes:

- Defined roles for IT, security, compliance, and legal teams.
- Escalation procedures for critical breaches.
- Post-incident reviews to refine governance and prevention strategies.

By incorporating response actions into the overall incident management process, organizations can reduce the impact of breaches and strengthen future governance.

- Establish Cross-Departmental Collaboration
- Build a Culture of Transparency

Establish Cross-Departmental Collaboration

Shadow systems often proliferate in organizations where departments work in silos. The needs for various technologies differ widely across business units. For example, marketing, HR, finance, and development may require tailored solutions. When IT cannot meet these needs promptly, employees may seek unauthorized alternatives. To prevent this, IT must collaborate closely with all departments to understand their requirements and deliver compliant, efficient tools.

Proactive collaboration ensures that solutions align with business goals and meet compliance standards, reducing the need for employees to turn to unauthorized tools. Open communication between departments fosters trust and simplifies the implementation of governance practices across the organization.

Build a Culture of Transparency

Alongside collaboration, building a culture of transparency around technology usage is crucial for mitigating risks. Employees should feel comfortable reporting their needs and challenges, knowing that IT and security teams will value their input and take prompt action. Assigning department-specific liaisons can bridge the gap between IT and other teams. These liaisons serve as contact points for queries and ensure that tools are properly vetted and compliant with governance policies.

Organizations can create structured frameworks that reduce unauthorized usage and promote responsible technology deployment by encouraging open communication and setting clear accountability.

Align with Industry Standards and Strengthen Data Strategy

Align with Industry Standards and Strengthen Data Strategy

Adopting industry standards like the NIST Risk Management Framework and ISO/IEC 27001 supports secure governance practices. These frameworks offer clear guidelines on privacy, security, and ethics. However, governance is only as strong as the data strategy supporting it. Models rely on the quality of the data they process. A strong data strategy ensures systems use accurate, relevant, and secure data, minimizing risks such as data drift.

Organizations that implement data governance frameworks are better positioned to protect their data, enhance the quality of outputs, and safeguard sensitive information from exposure.

Managing Shadow AI requires a comprehensive, proactive approach that integrates governance, continuous monitoring, cross-department collaboration, and adherence to industry standards. By adopting these seven strategies, organizations can effectively mitigate the risks associated with unauthorized technology usage, ensuring secure, compliant, and ethical practices across all departments.

Proactive Governance and Security Drive Secure Innovation

Shadow AI, unauthorized or unmanaged AI systems, poses significant risks to data security, operational integrity, and regulatory compliance. These unmanaged systems expose organizations to potential breaches, intellectual property theft, and compliance violations under frameworks like GDPR and CCPA. To mitigate these risks and ensure secure AI innovation, proactive governance and security measures are essential.
Real-time monitoring, regular audits, and strict adherence to industry standards—such as encryption protocols like AES-256—are critical for protecting valuable data assets and fostering responsible AI development. By adopting strategic governance, organizations can not only address immediate risks but also create a foundation for long-term innovation, securing their position as leaders in an AI-driven world.

Key Takeaways:

- Secure all AI-related tools with multi-layered encryption (e.g., AES-256) and perform regular compliance audits to ensure alignment with regulatory standards like GDPR and CCPA.
- Enforce strict alignment between AI tools, company policies, and regulatory frameworks to prevent non-compliance and ensure ethical outputs.

By prioritizing proactive governance today, organizations can unlock the full potential of AI while safeguarding their operational integrity, compliance, and long-term competitive advantage.

Sources

- Futurum Group. "More than 50% of Workers Admit to Using Unapproved Generative AI Tools." 2024.
- Gartner. "AI-Driven Security Incidents in 2024." 2024.
- GDPR Enforcement Report. "GDPR Enforcement Report." 2024.
- Forbes. "Samsung's ChatGPT Mishap: When IP Meets AI." 2023.
- Darktrace AI Research. "Jailbroken AI Models on the Dark Web." 2023.
- Security Affairs. "AI-Generated Malware: How Generative AI Is Being Exploited." 2023.
- Infosecurity Europe. "Threat Vectors: Generative AI and Dark Web Bots." Accessed 2024. https://www.infosecurityeurope.com/en-gb/blog/threat-vectors/generative-ai-dark-web-bots.html.
- CSO Online. "Hacked Generative AI Accounts: The Hottest Selling Product on the Darknet." Accessed 2024. https://www.csoonline.com/article/3479476/hottest-selling-product-on-the-darknet-hacked-genai-accounts.html.
- Markets and Markets. "Generative AI Breach: OpenAI Takes Action, Bug Patched." Accessed 2024. https://www.marketsandmarkets.com/industry-news/Generative-AI-Breach-Openai-Takes-Action-Bug-Patched.
- WithSecure Labs. "Generative AI: An Attacker's View." Accessed 2024. https://labs.withsecure.com/publications/generative-ai-an-attackers-view.
- Resecurity. "Massive Dump of Hacked Salvadorean Headshots and PII Highlights Growing Threat Actor Interest in Biometric Data." Accessed 2024. https://www.resecurity.com/blog/article/massive-dump-of-hacked-salvadorean-headshots-and-pii-highlights-growing-threat-actor-interest-in-biometric-data.
- Flare. "Dark Web and ChatGPT: The Generative AI Connection." Accessed 2024. https://flare.io/learn/resources/blog/dark-web-chatgpt-generative-ai.
- Dark Reading. "100K Infected Devices Leak ChatGPT Accounts on Dark Web." Accessed 2024. https://www.darkreading.com/application-security/100k-infected-devices-leak-chatgpt-accounts-dark-web.
- CSO Online. "Data of 300K DigiDirect Customers Leaked in Alleged Attack." Accessed 2024. https://www.csoonline.com/article/3479476/hottest-selling-product-on-the-darknet-hacked-genai-accounts.html.

Risk Aperture
risk-aperture.com