tlv partners

Cybersecurity Blueprint

GenAI Revolution: Navigating Cybersecurity Challenges

Introduction

Welcome to the TLV Partners Cybersecurity Blueprint, a thought leadership series dedicated to exploring the evolving landscape of cybersecurity. In each edition, we dive into critical challenges, emerging trends, and innovative solutions shaping the future of the industry. Drawing on insights from leading security experts, entrepreneurs, and CISOs, this series serves as a strategic guide for organizations and startups navigating the complex world of cybersecurity. Whether you're a seasoned professional or a startup founder, the Cybersecurity Blueprint is here to inform, inspire, and empower your next steps in securing the digital future.

About TLV Partners

TLV Partners is an early-stage venture capital firm investing in Israeli entrepreneurs who are pushing the boundaries of what is possible. With over $1 billion in assets under management, TLV Partners invests in untapped markets and wherever breakthrough technologies may be found, including Cybersecurity, DefenseTech, Data, AI, DevTools, Fintech, Biotech, eCommerce, and more. Since 2015, the firm has backed Israel's most promising companies, including Run:ai (acquired by Nvidia), Granulate (acquired by Intel), Oribi (acquired by LinkedIn), Neosec (acquired by Akamai), Stoke (acquired by Fiverr), Laminar (acquired by Rubrik), SeaLights (acquired by Tricentis), Aqua Security, Next Insurance, Unit, Silverfort, Immunai and more. For more information, visit www.tlv.partners.

Cybersecurity portfolio

SILVERFORT

aqua

oligo

token

aporia

{ } Solvo

Stream.Security

Odo
Acquired by
CHECK POINT

PURESEC
Acquired by
paloalto NETWORKS

neosec
Acquired by
Akamai

Laminar
Acquired by
rubrik

GenAI Revolution: Navigating Cybersecurity Challenges

Generative AI (GenAI) has made incredible progress, yet cybersecurity lags behind. While 92% of Fortune 500 companies already have employees using ChatGPT, 84% of enterprises see cybersecurity risks as the topmost roadblock to GenAI adoption.

This report will analyze the most critical GenAI risks, explore emerging solutions, and address the question: How can enterprises safely leverage GenAI?

Generative AI Security Risks

Generative AI use cases can be classified into three primary categories, with risks varying across each:

1. Employees' direct use of GenAI platforms such as ChatGPT.
2. Deploying homegrown or custom applications that utilize GenAI capabilities, like customer service chatbots.
3. Use of SaaS applications that embed GenAI capabilities - both by internal and/or external users. Common examples: M365 Copilot, Salesforce Einstein, and more.

The challenges and risks an organization faces are largely determined by the nature of its GenAI adoption. Most, if not all, organizations will use all forms of GenAI and therefore are expected to experience the sum of all risks.

After speaking to security experts and entrepreneurs, we've identified the most critical risks and their mitigating (although still emerging) solutions.

Direct GenAI Usage

Employees are using both internal and external large language models (LLMs) to augment and complement their work. ChatGPT and its alternatives (such as Perplexity, Google Gemini, Jasper AI, etc.), are useful across many different tasks: drafting emails and documents, writing software, retrieving enterprise data, marketing content creation, and much more. This introduces numerous challenges, the first of which is governance and enforcement.

Governance, Enforcement, and Shadow AI

In May 2023, just as ChatGPT gained popularity, companies like Apple, Amazon, and Samsung banned ChatGPT. Security concerns include data leakage of confidential information, hallucinogenic content, and compliance issues. Given these concerns, this was an understandable initial reaction. Currently, however, many companies have created policies to allow safe usage of GenAI applications, enabling employees to increase their productivity while following the organization's guidelines.

The challenge lies in enforcing corporate policies. Security teams often lack visibility into which GenAI tools are being employed and how they are being utilized. This phenomenon, known as **Shadow AI,** occurs when AI systems are used without formal approval or oversight from security teams. Shadow AI can manifest in two ways: directly, through employees using GenAI services, or indirectly, when employees unknowingly utilize services that incorporate GenAI.

Solutions focus on both **visibility** over GenAI services and **policy enforcement**, presenting CISOs with usage statistics and allowing them to block certain services. Two Israeli startups tackle this problem with different approaches.

Apex Security offers a unified LLM portal with embedded security features. Its agentless approach allows centralized control but forces employees to use LLMs through the Apex portal. On the other hand, Aim Security's solution allows employees to use GenAI services directly. Aim uses a browser extension to discover, monitor, and enforce policies. This is transparent to the user until a violation is detected.

It's important to note that there are many more risks associated with direct usage of GenAI: information leakage into external LLM, usage of unreliable LLM output, non-privileged information extraction via internal LLMs, copyright violations due to use of protected content, and more.

Homegrown GenAI Application

Enterprises, seeing the potential that GenAI can bring, are integrating it into their products and services. The most common examples are LLM-based customer service chatbots and GenAI-based content generation features. The specific implementation depends largely on the organization's existing products, but every integration of GenAI introduces potential security risks.

Prompt Injection and Jailbreak

Prompt injection attacks occur when a malicious user crafts an input that leads GenAI to perform unwanted behavior, such as sharing sensitive information. This can be exploited whenever the user has control (direct or indirect) over a part of the prompt. A notable example of this happened in December 2023, when someone convinced a ChatGPT-powered Chevy dealer to sell a $81K Tahoe for just $1.

Early solutions include the idea of an "LLM Firewall" - monitoring prompts and outputs to detect anomalies and block injection attempts. ProtectAI's "Layer" collects all LLM interactions, identifies malicious ones, and responds to violations. ProtectAI has also invested in an open-source project for securing LLMs. Prompt Security offers a similar solution, integrating easily with homegrown GenAI applications via SDK, API, or reverse proxy. It also offers GenAI red teams and an open-source fuzzer to test different attacks on GenAI applications.

Deploying homegrown GenAI applications introduces additional potential risks, the most significant of which are information leakage, misinformation due to hallucinations, and copyright violations.

Development Stage Risks

When it comes to homegrown apps, GenAI introduces new development stage challenges. Dev-stage AI security has been around for a while, tackling challenges like data poisoning and model theft. The new challenge stems from the rise of open-source AI models and a notable organizational change.

Today, organizations rely on AI engineers and data scientists while integrating GenAI into their services. However, these engineers lack the security tools, practices, and methodologies that are commonplace for software engineers. This leads to misconfigurations, the absence of automated tests, and most importantly usage of unsafe open-source models such as Hugging Face.

Robust Intelligence offers an AI validation platform that integrates into CI/CD workflows in the development stage. Its solution scans open-source models, data, and files by performing "algorithmic AI red teaming." HiddenLayer's Model Scanner scans open-source models in the client's environment, attempting to protect the digital supply chain.

It's worth mentioning Noma Security, an Israeli startup still in stealth mode, that has reportedly discovered thousands of infected data science notebooks, and "recently found a keylogging dependency that logged all activities on its customers' Jupyter notebooks." As long as these models work as expected, their malicious nature might remain undetected.

Additional risks in the development stage include training set poisoning, model training sets on sensitive data, and supply chain vulnerabilities.

Usage of SaaS with Embedded GenAI

Anyone using Microsoft Office, Google Slides, Salesforce Einstein, or Notion should be well aware that all of them have integrated GenAI features. As discussed above, this not only raises homegrown application security issues for the respective organizations but also introduces the risks of using SaaS with embedded GenAI.

Organizations of every size utilize SaaS products, and these solutions have access to or contain important enterprise data. Risks are introduced when these applications are boosted with GenAI abilities, often without the proper visibility by the oversight team. App users are often not aware of which applications are running GenAI, what guardrails they must adhere to, and how their use of the app is creating additional risk.

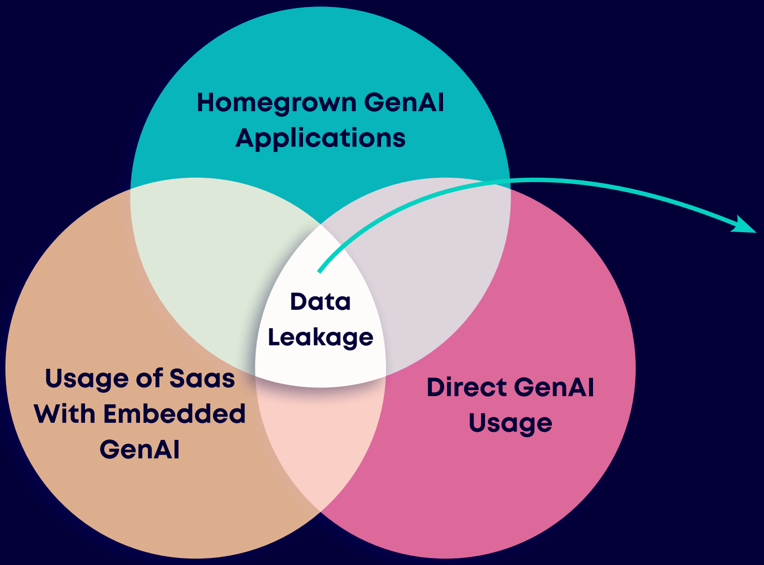
For example, a CRM might take your client's data and use an external LLM to craft an email, without the salespeople or the security team being aware of when this happened and which information was passed on to the LLM. The problem is perpetuated given the number of individuals performing the same exercise and its frequency.

Direct GenAI usage and usage of SaaS with embedded GenAI have many risks in common, most notably data leakage, inaccurate output, and copyrighted content.

Spotlight: The Content Risk

Consider a salesperson sharing credit card details with ChatGPT while writing an email, an employee accessing an executive's salary through an internal copilot, or a malicious user extracting PII from a customer service chatbot. One thing remains clear - **GenAI has transformed the way we generate, use, and share data.** We chose to highlight the data risk as it is a major concern for many security professionals, and since it is a shared problem between all GenAI use cases.

GenAI Security Risks and Challenges


Homegrown GenAI Applications
Data Leakage
Usage of Saas With Embedded GenAI
Direct GenAI Usage

Salesperson using ChatGPT to draft emails and sharing credit card information

Malicious **user** extracts PII from enterprise customer support chatbot

Employee discovers boss's salary through internal LLM

Developer sharing private API keys when using copilot to rewrite code

To address this risk, most solutions borrow classic DLP abilities (and limitations) and block the leakage of sensitive information by monitoring the inputs and outputs of LLMs. However, the LLM-DLP solutions will share the same challenges (if not more) of traditional DLPs: inaccuracies (that lead to many false positives and false negatives), limited ability to identify risks (beyond the trivial example of SSN patterns), and high total cost of ownership. Let's reconsider the examples above.

Employees should have access to their own human resource (HR) data, so simply blocking HR data in the internal copilot would not be enough.

LLMs are only a part of an application's pipeline, and LLM outputs are often "enriched" or re-purposed by other systems and human edits. So, even if an LLM is protected by a traditional DLP solution, the output might be later enriched by subsequent edits and manipulation, potentially containing PII.

More advanced solutions are required to solve data risk issues in the age of GenAI. Bonfy.AI tackles data risk directly with an adaptive content analysis platform that self-learns an organization's business content and its user-defined business logic. Rather than focusing on the security of the GenAI tools themselves, Bonfy's solution focuses on the content generated by either AI engines and/or humans, ensuring it adheres to corporate requirements, identifies potential data leakage points, and delivering visibility and mitigation recommendations to the organization prior to the dissemination of the content.

Market Dynamics

The GenAI security market is vibrant, with over 25 startups globally as well as established cybersecurity firms all tackling GenAI security. But before exploring them, let's first look at the AI giant's trajectory.

OpenAI, Google, and Microsoft are aware of the security risks and work to mitigate them. All three offer enterprise packages, declaring they won't train on customers' data. For some security professionals, this might resolve the concerns associated with external GenAI usage, given that enterprise data is already stored on Azure or GCP.

Startups and Large Security Companies

As mentioned, over 25 startups globally are focused on providing security coverage for the risks posed by GenAI and aim to help organizations adopt GenAI safely. Moreover, many leading venture capitalists have made investments in this sector.

Needless to say, startups are showing a wide range of efforts, with some focusing narrowly on one use case (like homegrown apps or employees), and others focus on a specific problem such as data risk.

When it comes to large cybersecurity firms, early adopters have released products, most notably Palo Alto's AI runtime and AI Access products focusing on GenAI, as well as similar releases from Microsoft.

Existing companies are expanding their current products to cover GenAI use cases as well. This includes traditional DLP, browser security, API security, and cloud security solutions. For example, Zscaler announced it provides "visibility and control over your users' interactions with generative AI to help you avoid data loss." Enterprise browsers like Island have released GenAI security features, sparing the need for browser extensions.

It's also worth noting that we haven't yet seen any notable acquisitions in the GenAI security space.

Competitive Landscape

Please see the competitive landscape as we view it today.

Conclusion

It's important to note that we are still in the early stage of GenAI's development. We are only beginning to see real-life GenAI applications. We expect that security solutions will continue to be coupled with GenAI improvements. Recent advancements, such as AI Agents and multi-modal architectures, are still in their early stages but are expected to soon enter the business environment, necessitating new security solutions.

Lastly, it's worth mentioning that malicious actors are advancing their use of GenAI. Hackers are crafting advanced phishing campaigns using ChatGPT, writing (and rewriting) malware with copilots, and creating deep-fake audio recordings of executives. While these are all new ways to perform old tricks, CISOs should be aware of these risks and ensure protection from them.

For any inquiries or feedback regarding this report, please feel free to reach out to Brian Sack at *brian@tlv.partners*

We extend our gratitude to the security experts and industry leaders who contributed their insights to this report. If you are a professional interested in collaborating or being featured in future editions, please contact us.

A special thank you to Yoav Ravid from TLV Partners' Investment Team for his dedicated work on this report.

Brian Sack
Partner

Yoav Ravid
Investment team