Impact of Generative AI in Cybersecurity and Privacy

Oku Krishnamurthy

Tech Lead Software Engineer- ITRAC, AT&T Services Inc, Automation Platform Department, NJ, USA,0009-0009-4987-5610

[1]*Received: 21 January 2024; Accepted: 12 February 2024; Published: 25 February 2024*

ABSTRACT

This research paper aims to explore the practical applications of Generative AI (GAI) in cybersecurity. As cyber threats' frequency, complexity, and impact continue to escalate in today's world, organizations and security professionals face ongoing challenges in finding practical solutions to combat these threats. GAI technology offers a promising avenue for addressing these challenges by automating processes and enhancing efficiency.

With the implementation of GAI, security professionals can redirect their focus to critical security tasks that require human expertise. Meanwhile, automated systems, with their superior capabilities in identifying novel malware and emerging threats, handle routine threat detection and response. This significantly enhances the overall robustness of cybersecurity measures.

Major tech companies such as Google and Microsoft actively integrate GAI elements into their cybersecurity frameworks to bolster their effectiveness against evolving threats. Notable examples include Google Cloud Security AI Workbench, Microsoft Security Copilot, and SentinelOne Purple AI, which leverage GAI to develop streamlined and resilient approaches to addressing emerging cybersecurity risks.

While GAI in cybersecurity offers significant benefits, it's important to acknowledge the potential drawbacks. These include occasional inaccuracies in results, high training costs, and the risk of malicious actors exploiting GAI for nefarious purposes. This paper thoroughly examines these limitations to provide a comprehensive understanding of the implications of integrating GAI into cybersecurity practices.

INTRODUCTION

The emergence of General Artificial Intelligence (GAI) has sparked a revolution across a multitude of sectors worldwide. Envisioned as a technology with vast potential by influential tech leaders, GAI has already begun to reshape various industries. From igniting innovation in art, music, and literature [1] to completely redefining content creation [2], personalization and recommendation systems [3], healthcare [4], virtual and augmented reality [5], natural language processing [6], and accessibility initiatives [7], the influence of GAI is truly transformative of significant importance is its emerging role in the cybersecurity landscape [8], [9]. GAI holds the promise of bolstering existing security measures by enabling more precise detection of advanced phishing attacks [10], proactive scanning for malicious activity, and automated responses to ongoing threats [11]. Unlike traditional AI-driven security solutions that react after an attack [12], the integration of GAI empowers a proactive approach, equipping security teams to anticipate and counter potential threats more effectively [13]. Furthermore, the wealth of data in cybersecurity provides a robust training ground for GAI, enhancing its effectiveness in protecting systems [14].

[1] *How to cite the article:* Krishnamurthy O. (February, 2024); Impact of Generative AI in Cybersecurity and Privacy; *International Journal of Advances in Engineering Research*, Feb 2024, Vol 27, Issue 2, 26-38

Fig 1: Application of cybersecurity

Combining GAI with network defences like firewalls and segmentation yields manifold benefits, ranging from enhanced visibility into shadow-IT usage to bolstered resilience against ransomware lateral movement [15]. By simulating real-world scenarios, GAI facilitates rigorous security testing, fault identification, and system fortification [16]. Trained on extensive datasets encompassing malicious URLs, indicators of compromise, and malware samples [17], GAI evolves into a potent ally capable of proactive threat detection and response [18].

While GAI's journey towards robustness and predictive accuracy is ongoing, its potential to significantly augment various industries, notably cybersecurity, is undeniable [19].

GAI OVERVIEW

When we notice the pervasive influence of General Artificial Intelligence (GAI) across numerous critical domains, a natural question emerges: What fuels GAI's remarkable power? The answer lies in the meticulous crafting of GAI's outputs, which stem from intricate combinations of the data used to train its algorithms [20]. Thanks to the vast quantities of data employed in algorithm training and the stochastic nature of output selection, GAI's creations exhibit elements of creativity and lifelikeness [21].

GAI models glean patterns and structures from extensive datasets by employing sophisticated deep-learning techniques such as Generative Artificial Networks, Variational Autoencoders, and Transformers [22]. Once trained, these models generate fresh content by sampling from the learned distribution.

Much of the buzz surrounding GAI is attributable to the notable achievements of entities like ChatGPT [23] and DALLE [24]. ChatGPT, a freely available chatbot developed by OpenAI, has garnered significant attention for its prowess in swiftly and adeptly generating unique content across diverse domains, including computer code, academic essays, and poetry. Meanwhile, DALLE, a GAI tool adept at crafting realistic images and art from natural language prompts, is also earning accolades from users.

Undoubtedly, GAI tools like ChatGPT and DALLE have the potential to revolutionize various professions. However, the full scope of their impact and associated risks remains largely uncharted.

APPLICATION VISTAS

GAI finds diverse applications in security (refer to Figure 1). This section delves into a comprehensive exploration of these applications.

A. Fortifying Password Security

Training GAI on extensive password datasets makes it adept at discerning structures and patterns within commonly used passwords. Consequently, it can generate new passwords or prioritize specific combinations during the cracking process, thereby enhancing password guessing efficiency and aiding in password security assessments. Notably, tools like PassGAN [25] leverage Generative Adversarial Networks to autonomously learn the distribution of actual passwords from leaked datasets, obviating manual analysis. GAI can also analyze user behaviour patterns related to password usage, including login patterns, password changes, authentication failures, and devices used for system access. This enables GAI to flag anomalous behaviour indicative of compromised passwords or unauthorized access, facilitating timely mitigation of security breaches.

B. Detecting GAI Text In Attacks

Large Language Models (LLMs) like Google LaMDA and ChatGPT play a crucial role in our cybersecurity arsenal. They enable the detection and watermarking of AI-generated text, a practical application of GAI in identifying malicious activities. By scrutinizing elements like email addresses, domains, and links within the text, these LLMs can flag phishing emails and polymorphic codes, potentially safeguarding users from accessing harmful sites.

C. Generate Examples of Adversarial Attacks

The strategic deployment of GAI is a proactive approach to cybersecurity. It can expose vulnerabilities inherent in GAI text models by creating meticulously crafted input texts. By manipulating these inputs, adversaries can compel the model to reveal its limitations and biases. This process of generating adversarial examples illuminates the weaknesses of GAI models and aids in fortifying defences against future attacks.

D. Simulated Environments

1) SIMULATED ATTACKS

By leveraging GAI's learning capabilities over known attack methodologies, we can empower these models to generate lifelike attack scenarios and strategies. Red team simulations, where GAI models simulate real-world adversaries, help us identify security loopholes within our organizational frameworks. This proactive approach allows our security systems to transition from reactive to proactive, predicting and preempting threats, ultimately enhancing our security measures.

GAI's capacity to fabricate simulated environments or cyber ranges offers practical, hands-on training for cybersecurity professionals. Platforms like Draup leverage GAI to replicate realistic network topologies, traffic patterns, and attack scenarios within these cyber ranges. This controlled environment allows our security personnel to refine their skills, experiment with diverse strategies, and gain practical experience in combating cyber threats, instilling confidence in their abilities.

2) MALWARE AND INTRUSION DETECTION

Generative models can produce realistic representations of malware by analyzing extensive datasets, enhancing the evaluation of malware detection systems' efficacy. Tools like SentinelOne Purple AI and Google Cloud Security AI Workbench utilize GAI techniques to bolster threat-hunting capabilities. Moreover, GAI can discern new or mutated malware variants by discerning standard features among different malware families, strengthening the robustness of security systems.

Furthermore, training GAI algorithms on standard network traffic data facilitates the generation of customary network behaviour representations. These representations enable the detection of anomalies indicative of potential security breaches or intrusive activities, empowering security officials to thwart possible threats.

3) CREATING HONEYPOTS

GAI can fabricate persuasive decoy systems, such as fake websites and applications, to attract potential attackers. Additionally, GAI-generated deceptive content and honeypots can engage attackers in human-like interactions to glean valuable insights into their modus operandi. GAI can dynamically adapt honeypots to current threat

landscapes by training on real-time threat intelligence, ensuring responsiveness to emerging threats. Post-engagement, we can analyze attackers' behavior to fortify our security systems, providing a sense of reassurance about the adaptability and effectiveness of our security measures.

4) PHISHING RESILIENCE TRAINING

LLMs like ChatGPT can generate simulated phishing messages, facilitating frequent and practical resilience training for organizational employees. These simulations, which prompt employees to identify suspicious signs indicative of phishing attempts, serve as adequate replacements for outdated cyber threat awareness programs. Organizations can swiftly deploy phishing resilience training programs by leveraging email marketing solutions, bolstering their defence against phishing attacks.

5) SYNTHETIC THREAT GENERATION

GAI's ability to generate synthetic threat environments enables rigorous testing and evaluation of system security. By learning from real-world datasets, GAI algorithms craft synthetic scenarios closely resembling actual attacks. These scenarios aid in creating synthetic malware samples, which, in turn, train GANs to identify malicious code characteristics. This knowledge facilitates the development of new malware variants for testing security systems, offering invaluable insights into bolstering system security against evolving threats. Additionally, GAI assists in phishing campaign simulation, network traffic simulation, and adversarial attack generation, empowering security teams to assess vulnerability, evaluate security measures, and enhance defence strategies.

E. Threat Intelligence Enhancement

Harnessing the extensive dataset GAI employs for its learning algorithm empowers it to discern patterns and compromise indicators, bolstering its capacity to preemptively detect and manage threats in real-time before breaching frontline defences. Moreover, GAI can forecast supplementary cyber security technologies that could fortify the existing infrastructure. This distinctive aspect of GAI, which we refer to as the proactive approach, involves broadly analyzing threats before delineating relevant features tailored to specific systems. This is in contrast to the organizational predictive threat analysis, which focuses on a more specific and targeted approach. Real-world security solutions like Google Cloud AI Workbench, SentinelOne Purple AI, and SlashNext Generative Human AI are adept at furnishing threat intelligence, facilitating proactive mitigation of constantly evolving threats. GAI's ability to interpret threats through a comprehensive and targeted lens renders it a potent defensive tool against present and future threats [26].

F. Security Code Generation and Transfer

Language Learning Models (LLMs) such as ChatGPT offer seamless code generation and transfer capabilities to bolster system security [26]. For instance, consider a scenario where a phishing attack compromises numerous employee credentials within a company. While the phishing email recipients are identified, discerning whether they inadvertently executed code designed to steal their credentials necessitates clarification. In such instances, a Microsoft 365 Defender Advanced Hunting query can be deployed to ascertain the ten most recent logins performed by phishing email recipients within thirty minutes of receiving malicious emails. These queries facilitate the identification of login activity associated with compromised credentials. ChatGPT, leveraging its Codex model, can provide the Microsoft 365 Defender Advanced Hunting query to scrutinize login attempts on compromised email accounts, expediting the identification and thwarting of attackers while providing users with guidance on password updates. The process of code generation and transfer involves ChatGPT's Codex model analyzing the original code and generating an equivalent version in the desired programming language. This simplification enhances query execution efficiency, reducing response time to cyber-attack incidents.

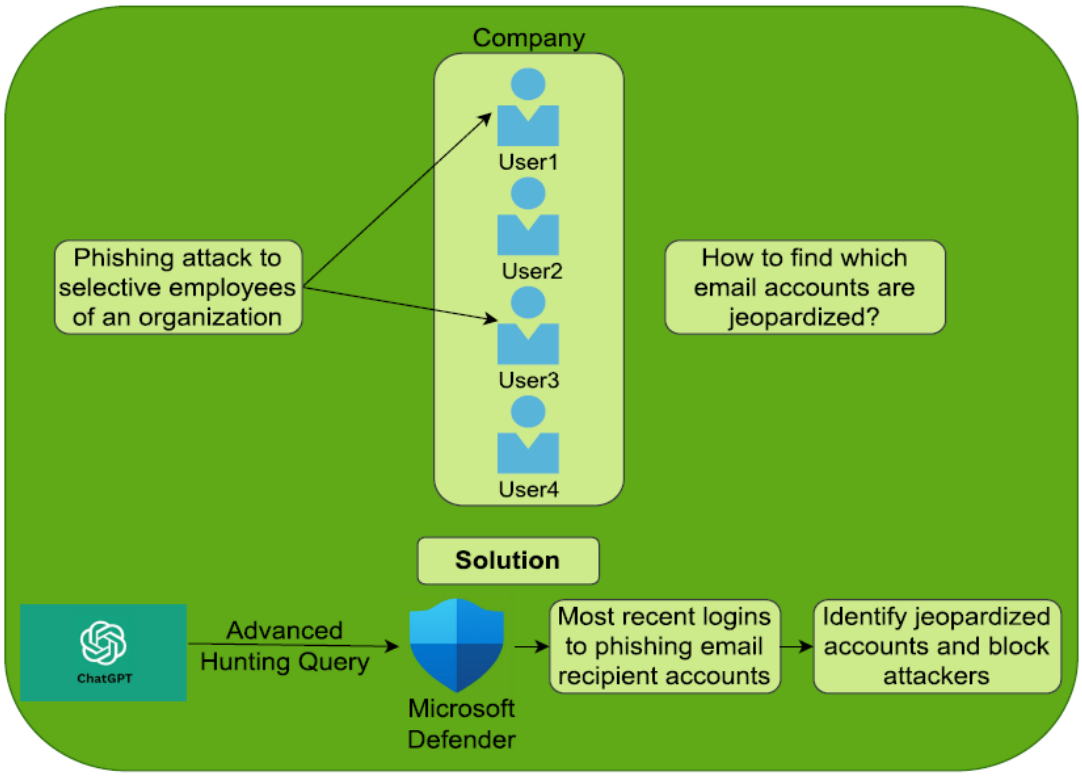

Fig 2. Security code generation using OpenAI's ChatGPT.

G. Vulnerability Scanning and Filtering

GAI models, trained on datasets inclusive of false positives, acquire the ability to formulate filters or rules distinguishing genuine vulnerabilities from benign anomalies, thus aiding in reducing false positives during vulnerability scanning. This feature helps security teams prioritize addressing authentic vulnerabilities while minimizing time spent investigating false positive alerts. Additionally, GAI models can be trained to contextualize vulnerability assessments by incorporating system configurations, network topology, user access privileges, and asset criticality. This means that GAI can understand the unique environment in which the vulnerabilities exist, and prioritize them based on their potential impact and exploitability within that specific environment. This contextualization enables GAI to prioritize vulnerabilities based on their potential impact and exploitability within specific environments, empowering security officials to focus on mitigating risks posing the greatest threat to the organization [27].

Moreover, GAI models proficiently scan various programming languages for vulnerabilities, detecting insecure code samples and providing developers with insights to address potential susceptibilities preemptively. These models offer contextual information about existing code, including customized fix suggestions, potential exploit scenarios, and the associated risks to the organization. Veracode Fix exemplifies a specialized tool harnessing GAI technology for code vulnerability detection and remediation.

H. Threat-Hunting Queries

LLMs like ChatGPT, Meta LLaMA, and Google LaMDA facilitate the creation of threat-hunting queries, augmenting system response time and efficiency. These queries, tailored for malware research and detection tools like YARA, enable cyber security officials to concentrate efforts on critical security aspects. At the same time, GAI swiftly identifies and addresses potential threats in the background. The role of GAI models in the creation of threat-hunting queries involves analyzing historical data and generating queries that can identify deviations and anomalies indicative of security breaches or unauthorized access. By dynamically adapting to evolving threat landscapes, GAI-generated queries reflect the latest indicators and emerging threat trends, empowering cyber security teams to combat nascent threats proactively. Newly introduced threat-hunting tools such as SentinelOne Purple AI, Google Cloud Security Workbench, and Microsoft Security Copilot leverage LLMs to enhance the productivity of security analysts.

I. Data Privacy Protection

GAI models are adept at producing shareable synthetic data, serving various organizational purposes such as ML model training for fraud detection and customized product recommendations. By minimizing the need for sharing customer data, GAI-driven synthetic data generation addresses privacy concerns and compliance with data protection laws, thereby enhancing data privacy protection [20]. The process of producing shareable synthetic data involves GAI models analyzing the original data and generating a synthetic version that retains the key characteristics of the original data without exposing any sensitive information. This approach ensures that the privacy of the original data is maintained, while still allowing for the use of the synthetic data in various organizational processes. Additionally, GAI facilitates the development of privacy-preserving machine learning models, refining federated learning techniques to generate synthetic data reflecting local patterns on edge devices. This decentralized approach ensures collaborative learning model training without compromising sensitive data, preserving user privacy [28].

Furthermore, GAI models trained on user interface designs acquire insights into visual elements and interaction patterns, enabling the creation of privacy-aware user interfaces. These interfaces minimize exposure to sensitive information by masking or obscuring sensitive data fields and providing privacy-centric options during data sharing. Talon enterprise browser exemplifies an emerging solution embedding Microsoft Azure OpenAI service to bolster organizational security.

J. Bridging Gap Between Technical Experts and Non-Experts

LLMs like OpenAI ChatGPT, DeepMind Chinchilla AI, and Meta LLaMA can articulate their reasoning, facilitating the translation of technical files into plain language. This ability enables users with limited technical proficiency to comprehend the functionality and implications of various technical documents, aiding in informed decision-making and preventing inadvertent cyber security breaches. Moreover, GAI serves as a conduit for bridging the communication gap between cyber security experts and non-experts, offering easily understandable explanations for cyber issues and fostering collaboration towards enhancing system security.

K. Security Questionnaires

The prevalent threat posed by third parties significantly impacts cyber security posture, with compromised third-party vendors contributing to approximately 60% of data exposures. Utilizing security questionnaires represents an efficient method for identifying security risks across vendor networks. Leveraging LLMs like ChatGPT to draft these questionnaires expedites the process, as models trained on comprehensive security-related questions and responses can generate tailored inquiries covering compliance, cyber security, and risk management aspects. This approach saves time for security officials, who can focus on refining and adjusting pre-generated questionnaires based on evolving threat scenarios. Organizations can uphold robust security standards and mitigate potential hazards by adapting to ongoing security trends.

L. Social Media Threat Detection

In social media, threat detection involves analyzing data from various platforms to pinpoint vulnerabilities and potential risks. Scanning social media channels for specific keywords deemed sensitive by an organization can identify potential exposure points for sensitive data or phishing attempts. Advanced language models (LLMs) such as ChatGPT, LLaMA, Chinchilla AI, or LaMDA can be employed. These models analyze collected data from social media channels, allowing intelligent prompts to extract desired information efficiently.

Furthermore, General Artificial Intelligence (GAI) models can be trained to contextualize social media posts comprehensively, considering text, images, and user interactions. By grasping the context, sentiment, and intent behind social media content, GAI models like ChatGPT can detect potential threats like hate speech, extremist content, or illegal activities. This proactive monitoring aids in risk mitigation.

Moreover, GAI models can be trained on datasets featuring trustworthy or malicious users. This training enables them to assess user reputation or trustworthiness on social media platforms, providing credibility metrics for social media accounts. Such metrics help organizations identify and flag suspicious or potentially malicious accounts for further investigation.

These applications of GAI in social media threat detection enhance efficiency and facilitate timely risk mitigation, ensuring a safer online environment.

M. IoT Security

The Internet of Things (IoT) connects many devices, from smart home appliances to industrial sensors, allowing them to exchange information over the internet. However, the rapid expansion of IoT devices has raised significant security concerns because many need more robust security measures.

GAI is crucial in bolstering IoT security by conducting behavioural analysis and anomaly detection. By establishing models of normal behaviour for various IoT environments and devices, GAI can detect deviations from standard patterns, signalling potential security breaches or unauthorized access. This enables organizations to address cyber threats swiftly, thus promoting robust security.

Furthermore, GAI aids in IoT threat detection and prevention by simulating attack scenarios on IoT systems to identify vulnerabilities. Additionally, it assists in analyzing firmware and software for potential security flaws, generating synthetic code to test devices' responses to different attack vectors. GAI enhances overall IoT security by improving firmware and software security measures based on analysis results.

Additionally, GAI facilitates designing and testing secure authentication methods for IoT devices, including biometric or multi-factor authentication. It also supports anonymization and aggregation of IoT data to protect user privacy while allowing meaningful analysis.

Leveraging GAI in IoT security enhances device resilience against cyber threats, enables proactive risk identification, and fosters a safer environment for interconnected devices and users.

N. Deepfake Detection And Mitigation

Deepfakes, highly realistic artificial media created using GAI techniques, present significant concerns about misinformation, fraud, and privacy violations. GAI tools offer effective means to detect and mitigate the adverse impacts of deepfakes.

Algorithms developed using GAI can analyze visual and audio cues to identify inconsistencies in deepfake content, such as unnatural facial expressions or mismatched lip movements. Training these algorithms on diverse datasets helps differentiate between authentic and manipulated media.

Additionally, GAI aids in developing forensic tools to trace digital media's origin, edit history, and verify content authenticity. Visualizations generated by GAI highlight alterations and anomalies in deepfake content, aiding in identifying tampered elements. Furthermore, GAI contributes to developing techniques like watermarking, content authentication, and media fingerprinting to counteract deepfake effects.

GAI also facilitates the creation of simulated deepfake examples for educational purposes, raising awareness of potential risks and helping individuals recognize signs of manipulated media.

As deepfake technology advances, GAI remains pivotal in developing sophisticated detection methods and mitigation strategies, ensuring the accuracy and reliability of digital media.

O. Supply Chain Security

Ensuring the integrity and authenticity of products within the supply chain is vital to prevent counterfeiting and tampering. GAI offers innovative solutions to enhance supply chain security.

GAI assists in designing unique identifiers like QR codes or holograms for each product, enabling consumers to verify product authenticity. It creates holographic labels or seals with intricate patterns, enhancing product packaging security. Additionally, GAI optimizes package designs to deter tampering or unauthorized access.

Furthermore, GAI analyses historical supply chain data to predict security risks, enabling proactive mitigation. By creating digital maps of supply chain routes and intermediaries, GAI helps track goods' journeys and identify vulnerabilities.

By applying GAI techniques, organizations can build a more secure and transparent supply chain ecosystem, maintaining consumer trust and regulatory compliance.

P. Blockchain Security

Blockchain technology, known for its decentralized and secure nature, can be further fortified using GAI techniques to address security challenges.

GAI assists in analysing intelligent contracts to identify vulnerabilities before deployment, simulating scenarios to ensure contract functionality without security loopholes. It also generates highly secure private keys for users and recommends vital storage and backup methods.

Furthermore, GAI monitors blockchain network activities to detect unusual patterns indicating security breaches. It generates predictive models to recognize potential threats, triggering proactive security measures. Additionally, GAI ensures blockchain data integrity by creating digital signatures or hashes and verifying data consistency across distributed nodes.

Incorporating GAI into blockchain security strengthens its resilience against attacks and vulnerabilities, fostering greater trust in blockchain solutions across industries. These applications empower organizations to overcome security challenges and fully leverage the benefits of blockchain technology.

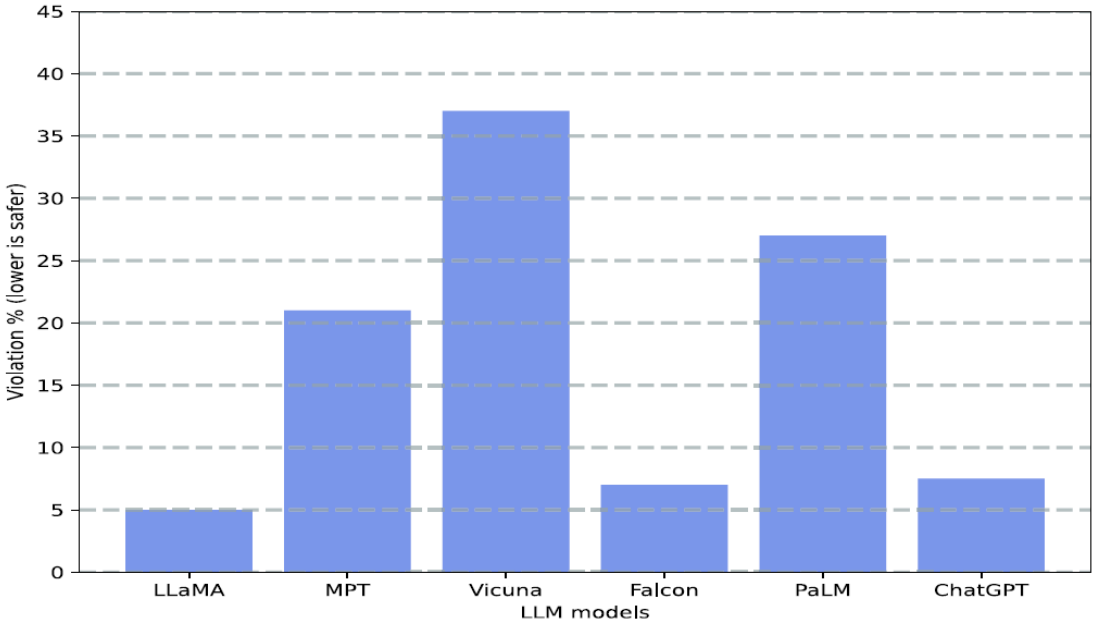

Fig 3. Safety human assessment outcomes for LLaMA 2-Chat in comparison to other models - both open and closed-source.

TAILORED LLMs FOR SECURITY

This section delves into three LLMs customized specifically to bolster security measures.

A. BigID BigAI LLM

BigID's BigAI is an LLM meticulously engineered to scrutinize and classify organizational data with unparalleled precision. This precision enhances security measures and fortifies risk management protocols. Capable of scanning both structured and unstructured data, whether housed in cloud environments or on-premises, BigAI employs a fusion of ML-driven classification and GAI techniques. It streamlines data organization by suggesting titles and descriptions for tables, columns, and clusters, facilitating easier retrieval through search functionalities.

One of the most notable features of BigAI is its commitment to safeguarding consumer data privacy. It achieves this by harnessing proprietary servers and models within the enterprise's domain, thereby protecting sensitive information from public exposure. BigChat, a virtual personal assistant integral to BigID's ecosystem, plays a crucial role in compliance management. Unlike services like Google Cloud Security AI, which primarily focus on threat identification and management, BigChat draws on a wide range of resources to comprehensively address user inquiries, from documentation and guides to forum posts and PDFs.

For instance, users can leverage BigAI to identify and categorize sensitive data, ensuring adherence to privacy regulations such as GDPR and CCPA.

Table 1. A comparison between GPT-4 and PaLM LLM models.

Features Of Model	GPT-4	PaLM 2
Developed By	OpenAI	Google
Model flow	Transformer Style Architecture	Transformer-based stacked layers are used in small, medium, and large versions of the Palm 2 model architecture; the parameters change according to the model's size.
Model Size	Parameter size of 1.5 tr	Parameter sizes of 340 billion
Pre-training	Supervised learning self	Self- attention
Attention mechanism	Yes	Yes
Fine tuning	Supervised learning	Supervised learning
Multimodal training data	Yes	yes

Training an LLM on regulated customer data outside its designated purpose can jeopardize user privacy and heighten data security risks. BigID BigAI mitigates such risks by enabling precise categorization, designation, and tagging of data based on regulations, sensitivity levels, and intended use across structured and unstructured datasets. This allows organizations to selectively curate datasets, excluding sensitive information like human resources data, thereby avoiding privacy infringements.

Moreover, BigAI empowers organizations to identify, filter, and manage structured data for conventional AI applications and unstructured data for emerging conversational AI technologies. By extending data governance and security measures to modern conversational AI and LLMs, BigID ensures responsible data handling practices, simplifying data identification, analysis, and fortification.

B. SlashNext Generative Human AI

SlashNext Generative Human AI represents a pioneering Artificial Intelligence solution that takes a proactive stance in combating sophisticated cyber threats. Leveraging Generative Adversarial Intelligence (GAI), it effectively counters advanced business email compromise (BEC), supply chain attacks, executive impersonation, and financial fraud.

Building upon SlashNext's existing HumanAI capabilities, which emulate human threat researchers, this solution amalgamates natural language processing, computer vision, machine learning, relationship graphs, and deep contextualization to thwart multi-channel messaging attacks. By anticipating AI-generated BEC threats through techniques like AI data augmentation and cloning, HumanAI can generate myriad variants of core threats, enabling self-training on potential variations.

Critical features of HumanAI include:

1. BEC GAI Augmentation: Generating diverse BEC variants to preempt future attacks.

2. Relationship Graphs & Contextual Analysis: Establishing baseline communication patterns to detect abnormal behaviours.

3. Natural Language Processing: Analyzing email content for topic, tone, emotion, intent, and social engineering strategies.

4. Computer Vision Recognition: Utilizing LiveScan to identify visual anomalies in URLs, such as phishing webpages.

5. File Attachment Inspection: Analyzing attachments for social engineering traits and malicious code.

6. Sender Impersonation Analysis: Evaluating headline details and email authentication to thwart impersonation attacks.

SlashNext sources threats from a vast database, enabling zero-hour detections and analyzing approximately 700,000 new threats daily. HumanAI's efficacy lies in its ability to discern human emotions exploited by threat actors and simulate them in detention.

C. Sec-PaLM

Sec-PaLM, a breakthrough in cybersecurity analysis, is a specialized version of Google's PaLM 2. It stands out with its advanced multilingual, reasoning, and coding capabilities, which are trained on datasets tailored for security use cases.

Accessible through Google Cloud, Sec-PaLM elucidates the behaviour of potentially malicious scripts and enhances the detection of scripts posing risks to individuals and organizations during critical times.

CHALLENGES AND CONSTRAINTS

While General Artificial Intelligence (GAI) presents promising avenues for enhancing cybersecurity, its implementation is limited. As outlined in Table 2, the regulatory landscape poses several challenges when integrating GAI into cybersecurity practices. The limitations associated with employing GAI in security contexts include:

TABLE 2. Regulatory challenges in implementation of GAI in security

Challenges of Regulatory	Brief Description
Academic Property	An LLM may produce content identical to privately held cyber security literature or study, thus infringing intellectual property rights.
Quality control & Standardization	Regulation is necessary to maintain AI-generated cyber security advice at an appropriate level because the consistency and reliability of the advice can vary depending on the data used to train the GAI model.
Data Ownership	Identifying and controlling who has access to the data the LLMs use to learn can get challenging, especially when that data relates to cyber security.
Continuous Monitoring & Validation	Determining continuous performance, accuracy, and validity across time and various datasets is an essential regulatory task.

A. Risk of Incorrect or Unethical Outputs:

Given the novelty of GAI models, their long-term ramifications still need to be determined. Consequently, their utilization has inherent risks, including known and unknown factors. For instance, language learning models (LLMs) like ChatGPT may provide misleading information that appears convincing yet inaccurate. Additionally, social biases inherent in these models can be exploited for illicit or unethical purposes.

B. Cost Inefficiency:

Implementing GAI systems for security purposes can be a substantial financial commitment. Only organizations with the financial capacity to cover the significant expenses and expertise required to establish and maintain such systems can ensure adequate security for their data and assets. This economic barrier may leave other entities vulnerable to evolving threats, necessitating the adoption of alternative, less secure measures. In response to these ethical dilemmas, some entities may consider seeking subsidies for GAI tools, particularly to assist non-profit organizations in safeguarding personal data.

C. Time Intensive Setup:

GAI models demand significant time for training, spanning from weeks to months. This prolonged setup period can impede organizations seeking swift responses to security needs, potentially hindering their agility.

D. Susceptibility to Malicious Exploitation:

A significant drawback of GAI systems is their susceptibility to exploitation by malicious actors. In the wrong hands, these systems can be leveraged to identify vulnerabilities, develop malware, and orchestrate convincing phishing schemes, compromising system integrity.

E. Interpretability and Explainability:

The complex, black-box nature of GAI models presents challenges in interpreting and explaining their outputs. This lack of transparency limits their applicability in critical security systems where interpretability is paramount.

F. Contextual Limitations:

GAI may struggle to understand and generate coherent responses in certain contexts, leading to nonsensical or irrelevant outputs. Factors contributing to this limitation include difficulty tracking extensive conversations, lack of common-sense reasoning, ambiguity resolution, and maintaining coherence in multi-turn dialogues. Such limitations may result in misunderstandings between security officials and GAI tools, heightening system vulnerability.

G. Difficulty with Long-range Dependencies:

Maintaining coherence and consistency in generating long sequences poses challenges for GAI models. Factors contributing to this limitation include finite short-term memory, operating on fixed-length token sequences, and encountering vanishing gradient problems during training. These challenges may compromise system security efficiency.

H. Data-related Concerns:

GAI tools introduce various risks to data privacy, including potential breaches, inadequate anonymization, perpetuation of biases, lack of consent and transparency, and insufficient data retention and deletion practices.

I. Lack of Control:

Users exert minimal control over GAI model outputs, mainly when models autonomously generate content without specific user instructions. This lack of control complicates identifying, classifying, and mitigating subtle threats, necessitating thorough inspection by cybersecurity officials.

J. Need for Empirical Evaluation:

The lack of standardized metrics for empirically evaluating GAI models or commercial off-the-shelf GAI-based security products introduces uncertainty when selecting appropriate solutions. It's crucial to develop benchmark datasets and evaluation frameworks to facilitate easy comparison of GAI system performance. This empirical evaluation is a key step in ensuring the effectiveness and reliability of GAI in cybersecurity.

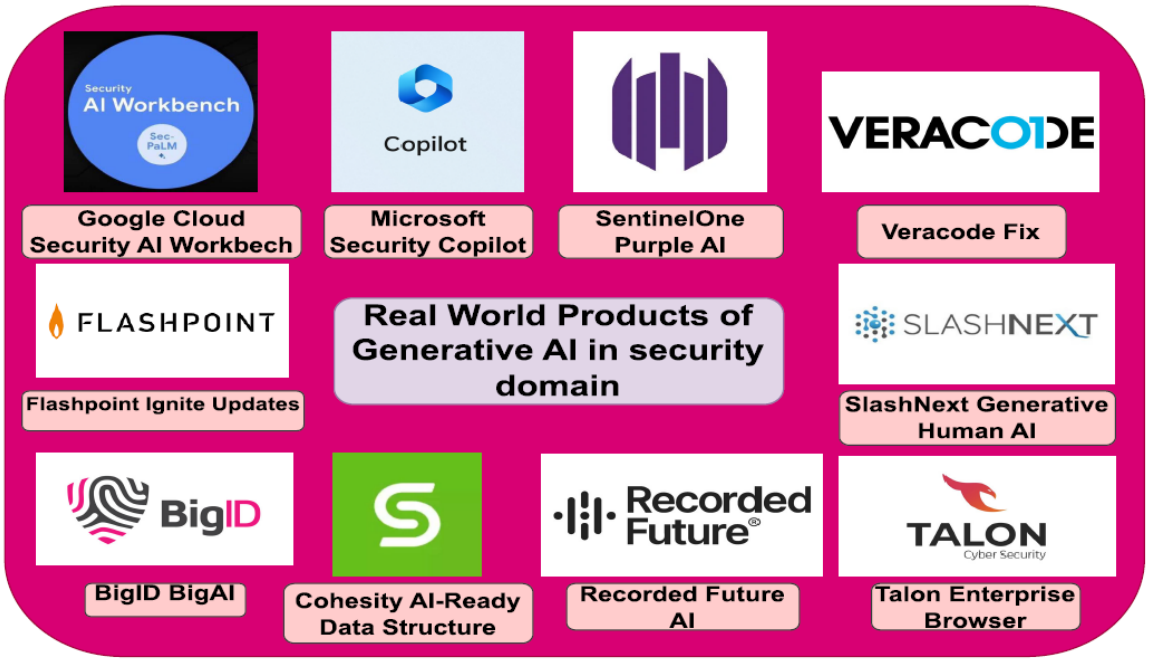

Fig 5. Real world products of GAI in cyber security domain.

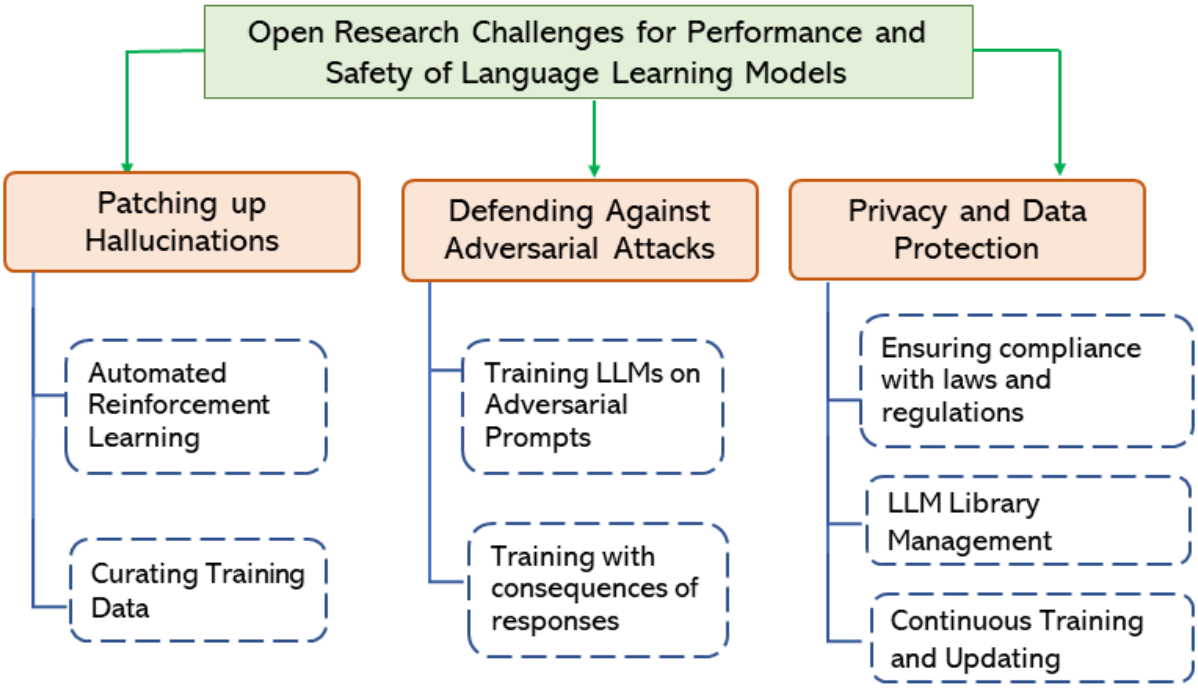

Fig 6. Open research challenges and potential future directions for LLMs performance and security.

CONCLUSION

Artificial Intelligence (AI) holds immense promise in fortifying cybersecurity capabilities. AI-powered models present many opportunities to augment threat detection, bolstering defences against various cyber threats. These models can identify anomalies within systems, craft diverse threat scenarios for meticulous analysis, and even automate responses to security breaches. Moreover, they exhibit proficiency in cracking passwords, detecting phishing attempts, and pinpointing malware infiltrations, enhancing overall security resilience.

Tech behemoths are actively harnessing AI's potential to roll out innovative security products that promise robust protection to end-users. This amalgamation of AI prowess and cybersecurity imperatives has spawned a wave of cutting-edge solutions designed to counter evolving threats in cyberspace. However, amid the excitement surrounding AI's transformative impact on cybersecurity, it remains imperative to exercise caution.

Indeed, the capabilities that make AI a potent ally in safeguarding digital assets also render it susceptible to exploitation by malicious actors. The advent of deep fakes and hyper-realistic phishing campaigns underscores the potential for AI-driven attacks to wreak havoc on unsuspecting targets. Thus, while embracing AI's transformative potential, it is paramount to remain vigilant and proactive in mitigating associated risks.

Ethical considerations loom large in this discourse. As AI assumes an increasingly pivotal role in cybersecurity, stakeholders must establish ethical guidelines governing its deployment and usage. Responsible utilization of AI entails safeguarding against its misuse and ensuring that its benefits are equitably distributed across society.

REFERENCES

1. Z. Epstein, A. Hertzmann, M. Akten, H. Farid, J. Fjeld, M. R. Frank, M. Groh, L. Herman, N. Leach, R. Mahari, A. S. Pentland, O. Russakovsky, H. Schroeder, and A. Smith, "Art and the science of generative AI," Science, vol. 380, no. 6650, pp. 1110–1111, 6650.
2. Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," 2023, arXiv:2303.04226.
3. W. Wang, X. Lin, F. Feng, X. He, and T.-S. Chua, "Generative recommendation: Towards next-generation recommender paradigm," 2023, arXiv:2304.03516.
4. P. Zhang and M. N. Kamel Boulos, "Generative AI in medicine and healthcare: Promises, opportunities and challenges," Future Internet, vol. 15, no. 9, p. 286, Aug. 2023.

5. Y. Hu, M. Yuan, K. Xian, D. Samitha Elvitigala, and A. Quigley, "Exploring the design space of employing AI-generated content for augmented reality display," 2023, arXiv:2303.16593.
6. Ö. Aydin and E. Karaarslan, "Is ChatGPT leading generative AI what is beyond expectations?" Academic Platform J. Eng. Smart Syst., vol. 11, no. 3, pp. 118–134, Sep. 2023.
7. L. A. Bauer and V. Bindschaedler, "Generative models for security: Attacks, defenses, and opportunities," 2021, arXiv:2107.10139.
8. O. S. Striuk and Y. P. Kondratenko, Generative Adversarial Networks in Cybersecurity: Analysis and Response. Cham, Switzerland: Springer, 2023, pp. 373–388.
9. H. Chen, Y. Zhang, Y. Cao, and J. Xie, "Security issues and defensive approaches in deep learning frameworks," Tsinghua Sci. Technol., vol. 26, no. 6, pp. 894–905, Dec. 2021.
10. M. E. Bonfanti, "Artificial intelligence and the offence-defence balance in cyber security," in Cyber Security: Socio-Technological Uncertainty and Political Fragmentation. Evanston, IL, USA: Routledge, 2022, pp. 64–79.
11. D. B. Rawat, R. Doku, and M. Garuba, "Cybersecurity in big data era: From securing big data to data-driven security," IEEE Trans. Services Comput., vol. 14, no. 6, pp. 2055–2072, Nov. 2021.
12. J. Le, A. Viswanathan, and Y. Zhang, "Generating high-fidelity cybersecurity data with generative adversarial networks," in ASCEND Reston, VA, USA: American Institute of Aeronautics and Astronautics, Nov. 2020.
13. N. Tihanyi, T. Bisztray, R. Jain, M. A. Ferrag, L. C. Cordeiro, and V. Mavroeidis, "The FormAI dataset: Generative AI in software security through the lens of formal verification," in Proc. 19th Int. Conf. Predictive Models Data Analytics Softw. Eng. New York, NY, USA: Association for Computing Machinery, Dec. 2023, pp. 33–43, doi:10.1145/3617555.3617874.
14. D. Noever and S. E. Miller Noever, "Virus-MNIST: A benchmark malware dataset," 2021, arXiv:2103.00602.
15. V. Mallikarjunaradhya, A. S. Pothukuchi, and L. V. Kota, "An overview of the strategic advantages of ai-powered threat intelligence in the cloud," J. Sci. Technol., vol. 4, no. 4, p. 1, Aug. 2023.
16. K.-B. Ooi et al., "The potential of generative artificial intelligence across disciplines: Perspectives and future directions," J. Comput. Inf. Syst., pp. 1–32, Oct. 2023.
17. A. Bandi, P. V. S. R. Adapa, and Y. E. V. P. K. Kuchi, "The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges," Future Internet, vol. 15, no. 8, p. 260, Jul. 2023.
18. S. Kumar, D. Musharaf, S. Musharaf, and A. K. Sagar, "A comprehensive review of the latest advancements in large generative AI models," in Communications in Computer and Information Science. Cham, Switzerland: Springer, 2023, pp. 90–103.
19. J. Babcock and R. Bali, Generative AI With Python and Tensorflow 2: Create Images, Text, and Music with Vaes, Gans, LSTMs, Transformer Models. Birmingham, U.K.: Packt, 2021.
20. ChatGPT. Accessed: Jun. 22, 2023. [Online]. Available: https://chat.openai.com/
21. Dall Now Available Without Waitlist. Accessed: Jun. 22, 2023. [Online]. Available: https://openai.com/blog/dall-e-now-available-without-waitlist
22. B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, "Passgan: A deep learning approach for password guessing," in Proc. 17th Int. Conf., 2019, pp. 217–237.
23. M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy," IEEE Access, vol. 11, pp. 80218–80245, 2023.
24. 6 Ways Generative Ai Chatbots and Llms Can Enhance Cybersecurity|CSO Online. Accessed: May. 25, 2023. [Online]. Available: https://www.csoonline.com/article/575377/6-ways-generative-aichatbots-and-llms-can-enhance-cybersecurity.html
25. A. Yan, X. Feng, X. Zhao, H. Zhou, J. Cui, Z. Ying, P. Girard, and X. Wen, "HITTSFL: Design of a cost-effective HIS-insensitive TNUtolerant and SET-filterable latch for safety-critical applications," in Proc. 57th ACM/IEEE Design Autom. Conf. (DAC), Jul. 2020, pp. 1–6.