IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

Leveraging Generative AI for Cybersecurity: Assessing the Capabilities of ChatGPT, DALL-E, and Other Models to Enhance Security Protocols

Sushma Basavaraju[1], Rajeshwari N[2]

MCA Student, Department of Computer Application, Bangalore Institute of Technology, Bangalore, India[1]

Assistant Professor, Department of Computer Application, Bangalore Institute of Technology, Bangalore, India[2]

ABSTRACT: Generative AI (GAI) is being applied in cybersecurity to tackle increasing cyber threats effectively. Unlike humans, GAI systems can automatically detect and respond to various types of threats with high accuracy. This frees up security professionals to focus on more complex security issues.Big tech companies like Google and Microsoft are integrating GAI into their cybersecurity tools such as Google Cloud Security AI Workbench and Microsoft Security Copilot. These tools use GAI to enhance the detection of new malware and other threats, making security systems more robust.However, GAI systems also have limitations. They can occasionally produce incorrect results, require expensive training, and could potentially be misused by hackers for harmful activities. These drawbacks need to be carefully considered as GAI continues to be adopted in cybersecurity.

I. INTRODUCTION

Generative AI (GAI) is making a big impact across many fields worldwide. It's used in creative industries like art and music to create inspiring works. In content generation, it automates and streamlines the creation process. In healthcare, it simulates biological processes and improves medical imaging. In cybersecurity, GAI helps detect advanced phishing attacks more accurately and can respond automatically to ongoing threats with better success rates. It allows security teams to predict and prepare for attacks proactively, enhancing overall system security. GAI learns from large datasets in cybersecurity to improve its ability to identify and respond to threats effectively. Although GAI is still improving in accuracy and reliability, it's becoming a valuable tool across industries, especially in cybersecurity.

APPLICATIONS

A. **Password Protection:** GAI can analyze large datasets of passwords to understand common patterns, helping to create stronger and harder-to-crack passwords. It can also detect unusual behavior related to password usage, such as suspicious login patterns, to prevent unauthorized access.
B. **Detecting GAI Text in Attacks:**Advanced language models like Google LaMDA and ChatGPT can identify AI-generated text used in phishing emails or malicious codes. This helps in detecting and preventing cyber threats.
C. **Generating Adversarial Attacks:** GAI can create simulated attacks to expose weaknesses in AI text models. By crafting specific texts, it reveals vulnerabilities, aiding in improving model defenses against future attacks.
D. **Simulated Environments:** GAI, security teams can simulate real-world threats to train and test their systems effectively. This enhances readiness and helps in evaluating the effectiveness of security measures.

1.Simulated Attacks:GAI can simulate realistic cyber attacks like phishing emails or social engineering schemes. This helps organizations test their defenses against real-world threats and improve their overall security readiness.
2.Malware and Intrusion Detection: Using GAI, security teams can create synthetic malware samples to test their detection systems. This improves their ability to identify and handle different types of malware, including new variants.
3.Creating Honeypots:GAI can generate convincing decoy systems and content to lure attackers. It helps gather valuable insights into attackers' methods and behaviors, enhancing overall security strategies.
4.Phishing Resilience Training: Advanced language models like ChatGPT can generate simulated phishing messages for training employees. This practical approach helps organizations improve their staff's awareness and response to phishing threats.
5.Synthetic Threat Generation:GAI models learn from real data to simulate diverse cyber threats. This includes creating artificial malware and generating scenarios for testing security systems, aiding in proactive defense strategies.

E.Threat Intelligence: GAI analyzes vast datasets to identify patterns and indicators of compromise, helping detect and handle threats in real-time. It can also predict future security needs by understanding broader threat landscapes, enhancing proactive defense strategies used by tools like Google Cloud AI Workbench and SentinelOne Purple AI.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

F.Security Code Generation and Transfer: Advanced language models (LLMs) like ChatGPT can generate and translate security-related code. For instance, they can assist in querying systems to identify login attempts related to phishing incidents, helping organizations quickly respond and secure compromised accounts.

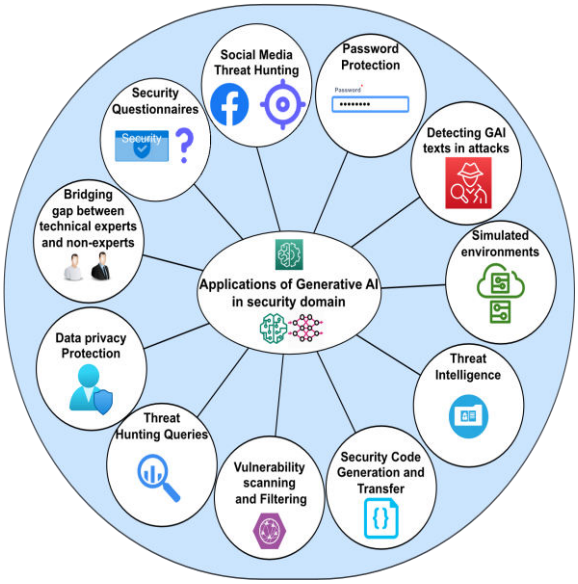
G. Vulnerability Scanning and Filtering: GAI learns from datasets to distinguish real vulnerabilities from false positives, improving the accuracy of vulnerability scans. It can prioritize fixes based on potential impact and exploitability, helping security teams focus on critical issues. Additionally, GAI can scan and analyze code across different programming languages, identifying insecure code and suggesting fixes to strengthen security.

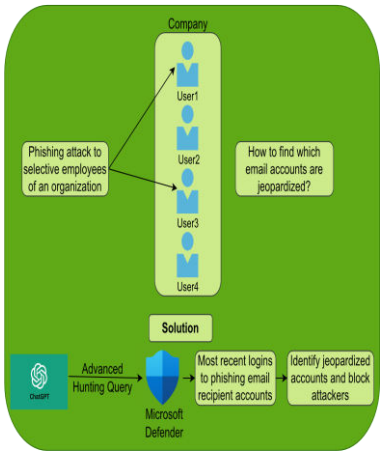
H.Threat-Hunting Queries: Advanced language models (LLMs) like ChatGPT and Google LaMDA can generate complex queries for threat-hunting tools such as YARA, improving response times to potential threats. By learning from historical data, GAI can detect anomalies in system behavior or network traffic, aiding in early detection of security breaches or unauthorized access.

I.Data Privacy Protection: GAI can generate synthetic data that preserves privacy, reducing the need to share sensitive customer information for tasks like fraud detection or personalized recommendations. It also enhances privacy in machine learning by enabling federated learning techniques that keep data local while still training models effectively.

J.Bridging Gap Between Experts and Non-Experts:LLMs can explain technical concepts in plain language, helping non-experts understand technical files and decisions better. This capability fosters collaboration between cyber security experts and other team members, improving overall system security.

K.Security Questionnaires:GAI can automate the creation of security questionnaires by generating questions based on a dataset of security-related queries and responses. This saves time for security professionals and ensures questionnaires are up-to-date with current security threats and standards.



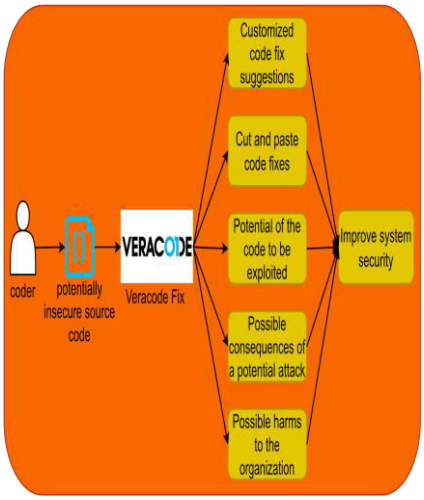


International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |



II. CUSTOMIZED LLMs FOR SECURITY

A. BigID BigAI LLM:

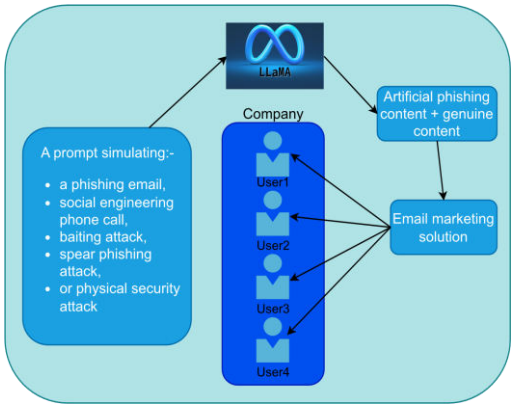
BugID's BigAI is designed to analyze and categorize organizational data to enhance security and improve risk management efforts. It works with both structured and unstructured data stored in various environments (cloud or on-premises). BigAI ensures privacy by processing data within the enterprise's private servers, avoiding exposure to public models. It includes BigChat, a virtual assistant that helps manage compliance challenges by providing guidance based on BigID's extensive resources. Organizations can use BigAI to detect and categorize sensitive data, ensuring compliance with regulations like GDPR and CCPA. By training LLMs on categorized data, organizations can improve accuracy without compromising data security.

B. SlashNext Generative Human AI:

SlashNext Generative Human AI focuses on defending against advanced email threats such as business email compromise (BEC) and phishing attacks. It uses Generative AI (GAI) to create thousands of variations of known threats, allowing it to anticipate and block new attack patterns. The AI leverages natural language processing (NLP) to analyze email content, contextual information, and relationship graphs to detect abnormal communication patterns indicative of threats. It also employs computer vision to inspect URLs in real-time for phishing attempts, enhancing its capability to detect credential phishing webpages and malicious attachments. SlashNext's HumanAI is proactive in identifying and mitigating threats by simulating human behavior and emotional triggers used in social engineering attacks.

C. Sec-PaLM:

Sec-PaLM is an advanced version of Google's PaLM 2 model, specifically tailored for cybersecurity analysis. Built on Google Cloud, Sec-PaLM enhances multilingual understanding and reasoning capabilities, particularly focused on analyzing and interpreting potentially malicious scripts. It aids in detecting and explaining the behavior of scripts that pose risks to individuals and organizations. Sec-PaLM's application in cybersecurity aims to provide proactive detection of threats, leveraging AI to enhance security measures in real-time.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

III. REAL WORLD SCENARIOS

A. SentinelOne Purple AI:

SentinelOne's Purple AI is a threat-hunting tool integrated into their Singularity Skylight platform, designed to enhance cybersecurity efforts using advanced AI techniques, including GAI like OpenAI's GPT-4. Here are its key features:

1. Enhanced Threat Detection:

Purple AI simplifies threat hunting by allowing security analysts to query potential threats using natural language commands. This reduces the complexity and time required for threat detection compared to traditional tools.

2. Integration and Accessibility:

Purple AI is seamlessly integrated into the Singularity Skylight platform, providing users with a user-friendly interface. This integration enhances user flexibility and comfort in using the AI tool.

3.Use of LLMs: SentinelOne employs both proprietary and open-source LLMs, such as GPT-4, which are fine-tuned to effectively operate in cybersecurity contexts. These models improve the efficiency and accuracy of threat analysis.

4. Platform Evolution:

The tool significantly boosts the efficiency of security operations teams, enabling them to scale up threat-hunting activities without needing extensive technical expertise. This simplification empowers organizations to strengthen their security posture without additional human resource investments.

B. Google Cloud Security AI Workbench:

Google's Cloud Security AI Workbench is powered by Sec-PaLM, a specialized LLM designed for cybersecurity. It leverages Google Cloud's infrastructure to enhance security intelligence and streamline threat detection and response. Here are its key features:

1. **Advanced Threat Containment:**

Google Cloud Security AI Workbench combines threat intelligence with AI/ML capabilities to detect and contain potential cyber threats efficiently. Tools like VirusTotal Code Insight and Mandiant Breach Analytics for Chronicle use Sec-PaLM to analyze and respond to malicious scripts and active breaches in real-time.

2. Reducing Operational Burden:

The Workbench aims to simplify security operations by integrating intelligent tools like Assured OSS and Mandiant Threat Intelligence AI. These tools help organizations manage vulnerabilities, analyze threats, and automate security responses, thus reducing manual effort and operational complexity.

3. Closing the Talent Gap: Google Cloud Security AI Workbench makes security more accessible and understandable, even for non-specialists. Solutions like Chronicle AI enable users to search through vast security event data and generate real-time detections without needing extensive cybersecurity expertise. Security Command Center AI provides clear, actionable insights into security, compliance, and privacy issues.

In essence, Google Cloud Security AI Workbench empowers organizations to strengthen their cybersecurity posture by leveraging advanced AI technologies integrated into Google Cloud's secure and compliant infrastructure.

C. Recorded Future AI:

Recorded Future integrates OpenAI's transformer model, GPT, into its Intelligence Cloud to revolutionize the intelligence industry. Trained on extensive threat analysis data and combined with insights from the Recorded Future Intelligence Graph, it uses natural language processing (NLP) and machine learning (ML) to analyze and map insights across vast amounts of data in real-time. This AI provides real-time threat analysis at scale, enhances analyst efficiency, and offers actionable intelligence to businesses to make proactive decisions against cyber threats and vulnerabilities. By automating tedious tasks, it allows analysts to focus more on strategic security initiatives.

D. SecurityScorecard GPT-4 Integration:

SecurityScorecard incorporates OpenAI's GPT-4 into its platform for security assessments. This integration enables cyber security professionals to quickly obtain responses and mitigation suggestions for high-priority cyber risks using natural language processing. Through ScorecardX, Security Scorecard's innovation centre, this solution allows customers to ask complex questions about their business ecosystem and receive prompt, data-driven answers. It enhances risk management decisions by providing insights derived from Security Scorecard's security ratings across various organizations, reducing manual data analysis efforts and improving efficiency over time through continuous learning and development.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

E.Slashnext generative human AI:
SlashNext Generative Human AI is an advanced service using AI to protect against complex cyber threats like supply chain attacks, email fraud, and impersonation schemes. It combines machine learning, computer vision, and natural language processing to mimic human threat researchers. This helps in detecting and preventing various types of attacks across different communication channels.

Key features of SlashNext Generative Human AI include:
1.BEC GAI Augmentation: Generates thousands of new versions of known email threats to anticipate future breaches.
2. Relationship Graphs & Contextual Analysis: Identifies unusual communication patterns by comparing with known good communication styles.
3. Natural Language Processing: Analyzes email content for tone, emotion, and intent to detect social engineering tactics.
4. Computer Vision Recognition: Uses visual analysis to detect subtle differences in phishing websites and block access.
5.File Attachment Inspection: Analyzes attachments for social engineering attributes and malicious code to prevent ransomware.
6. Sender Impersonation Analysis: Evaluates email authenticity and content to prevent impersonation attacks.

IV. LIMITATIONS

A. Wrong/Ethical Concerns:
Generative AI models like GPT can sometimes provide incorrect or misleading information. They may also exhibit biases that could potentially be exploited for unethical purposes, such as manipulating users or facilitating criminal activities.

B. Cost Inefficiency:
Implementing and maintaining GAI systems for security purposes can be very expensive. Only organizations with significant resources and expertise can afford these systems, leaving others vulnerable to cyber threats due to less secure alternatives.

C. High Setup Time:
Training and configuring GAI models takes a considerable amount of time, which can delay organizations wanting to enhance their security quickly.

D. Exploitation by Malicious Actors:
If malicious actors gain access to GAI systems, they can exploit them to discover vulnerabilities, develop sophisticated attacks like malware or phishing campaigns, and create convincing social engineering tactics.

E.Interpretability and Explainability: GAI models often operate as black boxes, making it challenging to interpret their outputs. This lack of transparency can hinder trust and understanding, crucial in security contexts where clear explanations are needed for decisions.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

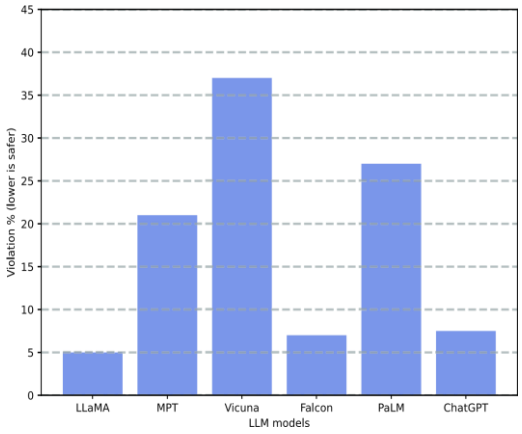
|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

F. Contextual Limitations:
GAI models struggle with maintaining context over long conversations or understanding common sense reasoning. This can lead to nonsensical or incorrect responses, potentially exposing security vulnerabilities.

G. Difficulty with long-range dependencies: Generative AI may have difficulty maintaining coherence when generating long sequences of text, which could lead to fragmented or inconsistent outputs, compromising security effectiveness.

H. Data-related Concerns:
Using GAI tools may pose risks to data privacy, such as breaches, inadequate anonymization, perpetuation of biases from training data, lack of consent transparency, and improper data retention practices. These issues can lead to privacy violations and undermine user trust in the system.



I. Lack of Control:
Users have limited control over the outputs generated by AI models like GPT-3.5. These models create content based on prompts provided by the user, but users can't fine-tune the specifics of the generated content. This lack of detailed control can be a problem for cybersecurity professionals who need to identify and manage subtle threats that require careful examination.

J.Need for Empirical Evaluation:
There's no standard way to measure and compare the performance of AI models or security products. This makes it hard to choose the best one for specific needs. To solve this, we need to develop standard datasets and evaluation methods to consistently measure how well these AI systems perform.

V. CONCLUSION

Generative AI has the potential to significantly enhance cybersecurity efforts. It can improve threat detection, create various scenarios for analysis, spot system irregularities, crack passwords, detect phishing attempts and malware, and automate security responses. Major tech companies are actively developing real-world applications that harness GAI to provide effective security solutions.

However, there are risks associated with GAI. Malicious individuals could exploit these technologies for sophisticated attacks such as deep fakes or convincing phishing scams. To mitigate these risks, it's crucial to establish ethical guidelines for the use of GAI, ensure responsible implementation of these technologies, and continuously advance cybersecurity measures alongside GAI development While Generative AI holds promise for strengthening cybersecurity defenses, it must be approached with caution regarding ethical implications. Robust safeguards need to be continually developed to prevent potential misuse of this technology.

REFERENCES

1. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.URL: https://arxiv.org/abs/2005.14165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207060 |

2. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.URL: https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf
4. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092.URL: https://arxiv.org/abs/2102.12092
5. Chaudhary, S., Kolhe, R., & Rana, J. (2022). An empirical analysis of AI-driven cybersecurity for detecting attacks. Journal of Cyber Security and Mobility, 11(3), 423-447.URL: https://journals.sagepub.com/doi/abs/10.1177/23266053211006734
6. Pieters, W., & Gutwirth, S. (2017). Privacy and security in the cyber age: The challenges of cybersecurity. Information Polity, 22(4), 301-308.URL: https://content.iospress.com/articles/information-polity/ip160044
7. Wang, W., Wang, W., Chen, X., Zhang, X., & Lu, W. (2019). HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE Access, 6, 56854-56864.URL: https://ieeexplore.ieee.org/document/8581437
8. Uchida, K., Tsuji, M., & Manabe, Y. (2020). Automatic network intrusion detection: A survey of the state of the art. Journal of Network and Computer Applications, 156, 102581.URL: https://www.sciencedirect.com/science/article/pii/S1084804520302151
9. Mahdavifar, S., & Ghorbani, A. A. (2019). Application of deep learning to cybersecurity: A survey. Neurocomputing, 347, 149-176.URL: https://www.sciencedirect.com/science/article/pii/S0925231219309757
10. Liu, Y., Li, Y., & Zhang, H. (2020). Secure and efficient communication over the Internet of Things: A survey. IEEE Internet of Things Journal, 7(6), 5043-5064.URL: https://ieeexplore.ieee.org/document/8945573
11. Zhang, C., Li, Y., Wang, J., & Zhang, T. (2019). Enhancing cybersecurity with machine learning: A comprehensive review. IEEE Access, 7, 34314-34337.URL: https://ieeexplore.ieee.org/document/8646257

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details