IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Generative AI for Cyber Security: Analyzing the Potential of Chatgpt, Dall-E, and Other Models for Enhancing the Security Space

Palve Rutika Popat[1], Prof. A.B. Jadhav[2], Prof. D.K. Gund[3], Dr. H.B. Jadhav[4]
[1,2,3,4]Department of Computer Engineering
Adsul Tecnical campus Chas, India

Abstract: *Generative Artificial Intelligence (GAI) is becoming an important technology in many areas, including cybersecurity. Cyber threats are increasing in number, complexity, and impact, so organizations need security systems that can detect and respond to threats more quickly. GAI can support this requirement by learning patterns from large datasets and generating useful content, predictions, simulations, and security responses. It can assist security teams with activities such as password analysis, phishing detection, malware detection, threat simulation, honeypot creation, threat hunting, data protection, IoT security, deepfake detection, and supply-chain security. It can also help security professionals work with large volumes of security information and automate repetitive tasks. Well-known generative AI technologies such as ChatGPT and DALL-E have increased interest in this area, while security products such as Microsoft Security Copilot, Google Cloud Security AI Workbench, SentinelOne Purple AI, Talon Enterprise Browser, SlashNext Generative Human AI, and Recorded Future AI show how these ideas can be applied in practice. At the same time, GAI has important limitations. It may produce incorrect or unsafe outputs, require costly training and setup, have difficulty understanding context, expose sensitive information, reproduce bias, and be misused by attackers. Therefore, GAI should be used with human supervision, suitable security controls, ethical practices, and continuous evaluation. This paper reviews the applications, practical implementations, and limitations of GAI in cybersecurity, compares the existing and proposed security systems, presents sample outputs of the proposed system, and explains the potential of GAI to strengthen modern security systems.*

Keywords: Generative AI, Cyber Security, Artificial Intelligence, Machine Learning, ChatGPT, DALL-E, Large Language Models, Threat Detection.

I. INTRODUCTION

Generative AI has expanded rapidly and is now being used in many fields. Examples include art and creative work, automatic content generation, recommendation systems, healthcare research, virtual and augmented reality, natural-language applications, and accessibility tools. The common feature of these systems is their ability to learn patterns from large amounts of information and then generate new content that follows those learned patterns.

Cybersecurity is one of the areas where this technology can provide significant support. Modern attacks can be automated, customized, and changed quickly. Security teams therefore need methods that can identify suspicious behavior early and react efficiently. Generative AI can support existing security tools by analyzing large datasets, generating possible threat scenarios, identifying unusual patterns, and helping analysts understand security events.

Traditional security systems generally become active after a suspicious event or attack has been detected. GAI can add a more proactive capability. For example, a model can learn from previous attacks and generate realistic attack

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

scenarios that can be used to test a security system before a similar real attack occurs. It can also generate variations of malicious content so that detection systems can be tested against new forms of threats.

Another advantage is automation. Security organizations receive large amounts of logs, alerts, URLs, malware samples, messages, and other security information. GAI can help summarize this information and assist analysts in finding relevant patterns. This can reduce repetitive work and allow security professionals to spend more time on tasks that require human judgment.

The cybersecurity field also has a large amount of structured and unstructured data. Examples include malicious URLs, indicators of compromise, malware samples, network traffic, security reports, and attack descriptions. Such data can be used to train or support GAI systems. However, the quality of the output depends on the data, model, configuration, and context.

The popularity of ChatGPT and DALL-E has increased public attention toward generative AI. ChatGPT demonstrates how a language model can generate human-like text, explanations, and code, while DALL-E demonstrates generation of images from natural-language descriptions. These technologies illustrate the broader capabilities of GAI. Their security applications must still be considered carefully because the same capabilities can also be misused by malicious actors.

The purpose of this paper is to examine how GAI can be used for cybersecurity, describe the existing and proposed system architectures, present practical applications and real-world security products, show sample outputs of the proposed system, and discuss limitations that must be considered before relying on these systems.

II. AN OVERVIEW OF GENERATIVE AI

Generative AI refers to AI systems that learn patterns and structures from large datasets and use the learned information to create new content. Depending on the model, the generated content may be text, images, code, audio, or other forms of information. Generative models commonly use advanced deep-learning approaches such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer-based architectures.

The basic idea is different from simply storing training examples. During training, the model learns relationships and patterns present in the data. After training, it can use those learned patterns to produce new outputs. Because the model has learned from a large amount of information and uses probabilistic generation, its outputs can appear creative or human-like. In cybersecurity, this capability can be used in both defensive and testing activities. A model trained using security data can help identify patterns associated with threats. It can also generate synthetic examples of attacks, malware characteristics, phishing messages, or network activity. These generated examples can be used to test and improve detection systems.

ChatGPT is a well-known example of a large language model that can generate natural-language responses. It can also work with code and explain technical information. DALL-E is an example of a generative system that creates images from natural-language prompts. Although these tools are not themselves complete cybersecurity solutions, their underlying generative capabilities demonstrate how AI can assist security workflows.

GAI can work together with existing security technologies such as firewalls, network segmentation, authentication, endpoint protection, and zero-trust approaches. For example, generated attack scenarios can be used to test network controls, while AI-assisted analysis can help identify unusual activity. Continuous authentication and authorization can also be supported by systems that analyze user and device behavior.

The technology is promising but should not be treated as automatically correct. Generative models can produce incorrect information, sometimes called hallucinations, and can fail when the available context is incomplete. Cybersecurity decisions can have serious consequences, so generated results should be checked by appropriate security controls and human experts.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

III. EXISTING SYSTEM

In the existing approach, cybersecurity is mainly handled by conventional tools such as firewalls, intrusion detection systems (IDS), and antivirus software. These tools collect data from network traffic, system logs, e-mails, and files and compare it with predefined rules and known attack signatures. When a match is found, or when an attack has already been observed, an alert is generated.

The generated alerts are then examined by security analysts. The analyst has to study the alert, search the logs manually, write queries in a specific query language, decide whether the event is a real threat, and then perform the response and patching steps. Figure 1 shows the working of the existing system.

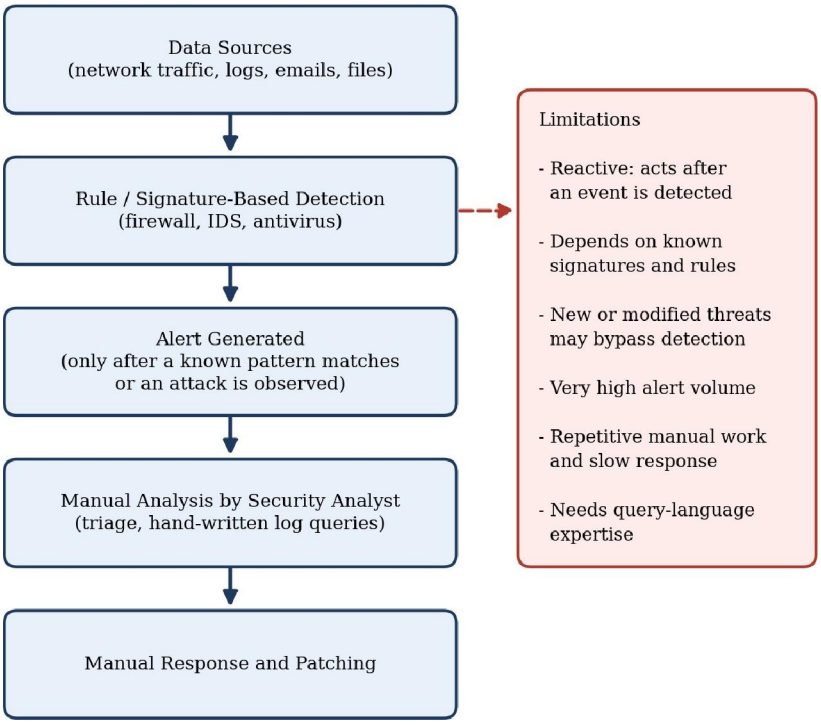


Fig. 1. Block diagram of the existing security system

The main limitations of the existing system are given below.

• Reactive nature: the system generally acts only after a suspicious event or attack has been detected.

• Dependence on known patterns: signature and rule-based tools may fail to detect new or slightly modified threats, such as new malware variants or carefully written phishing messages.

• Heavy alert load: analysts receive a very large number of alerts, logs, and messages, which makes manual analysis slow and tiring.

• Manual and repetitive work: writing log queries, preparing summaries, and investigating similar alerts repeatedly consumes time that could be used for tasks requiring human judgment.

• Limited testing and training: realistic attack scenarios for testing controls and training employees are difficult to prepare manually in large numbers.

IV. PROPOSED SYSTEM

The proposed system adds a generative AI layer on top of the existing security tools. It does not remove the firewall, authentication, endpoint protection, or zero-trust controls. Instead, it supports them by learning from large amounts of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

security data, generating useful content, and assisting the analyst. Figure 2 shows the architecture of the proposed system, and its working is explained in the following stages.

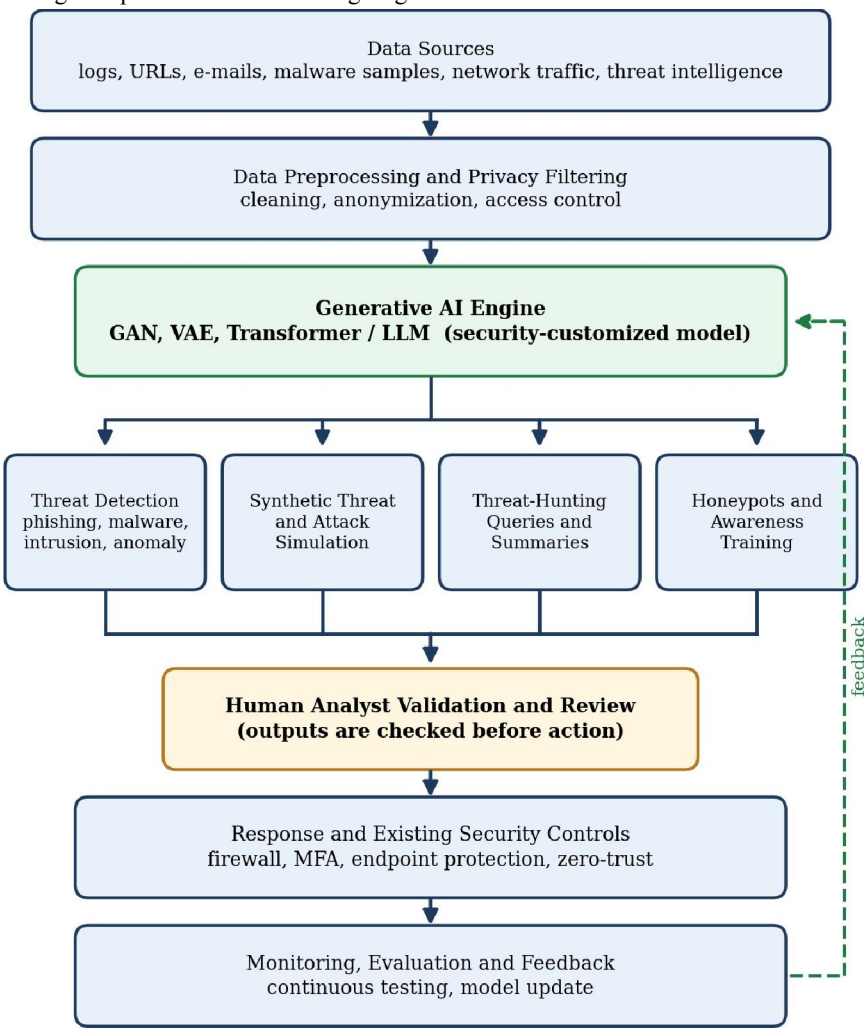


Fig. 2. Block diagram of the proposed GAI-based security system

• Data collection: Security data such as logs, URLs, e-mails, malware samples, network traffic, and threat intelligence are collected from different sources.

• Preprocessing and privacy filtering: The data is cleaned, and sensitive or personal information is anonymized or restricted through access control before it is given to the model, because sending private data to a generative system can itself create privacy risks.

• Generative AI engine: A generative model based on GAN, VAE, or Transformer (LLM) architecture, preferably customized for security tasks, learns the patterns present in the data.

• Security modules: The engine supports four groups of tasks: (i) threat detection, such as phishing, malware, intrusion, and anomaly detection; (ii) synthetic threat and attack simulation for testing; (iii) threat-hunting queries and plain-language summaries of alerts; and (iv) honeypots and phishing-awareness training.

• Human validation: The generated results are reviewed by a security analyst before any important action is taken, because generative models can produce wrong or unsafe output.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, October 2026 Impact Factor: 8.2

• Response and existing controls: After validation, the response is carried out using existing controls such as firewalls, multi-factor authentication, and endpoint protection.
• Monitoring and feedback: The performance of the system is continuously tested and the results are used as feedback to improve the model, as there is no standard measure that fully evaluates all GAI security models.
In this way the proposed system moves security from a purely reactive approach to a more proactive and analyst-assisted approach, while human supervision is kept in the loop.

V. CUSTOMIZED LARGE LANGUAGE MODELS FOR SECURITY

Generic language models may not always have enough cybersecurity-specific knowledge or the required privacy controls. Customized or security-focused language models can be trained or configured using domain-specific information to support security tasks.
Security-focused models can process vulnerability information, incident descriptions, security logs, and other specialized data. The source paper discusses customized security-oriented solutions including BigID BigAI LLM, SlashNext Generative Human AI, and Sec-PaLM. These examples show the movement toward models designed for particular security requirements rather than general-purpose conversation only.
A customized model can be useful when an organization needs controlled access to sensitive data, specialized security terminology, or integration with existing security tools. The design must still consider model accuracy, data protection, access control, and evaluation.

VI. REAL-WORLD SCENARIOS AND PRODUCTS

Several organizations have developed cybersecurity products that incorporate generative AI capabilities. These products illustrate how GAI can move from research concepts into operational security workflows.

A. SENTINELONE PURPLE AI

SentinelOne Purple AI is described in the source paper as a generative-AI-based security assistant intended to support threat hunting and investigation. It can help analysts interact with security information using natural language and assist with finding relevant threats.
The broader idea is to reduce repetitive work and make security analysis more accessible. Analysts can use an AI assistant to investigate events and obtain summaries, while human professionals remain responsible for validating findings and taking important security actions.

B. GOOGLE CLOUD SECURITY AI WORKBENCH

Google Cloud Security AI Workbench is presented as a set of capabilities for applying AI to cloud security. The source paper highlights Chronicle AI and Security Command Center AI.
Chronicle AI can assist users in searching large numbers of security events and generating detections. Security Command Center AI can help explain complex attack information, identify affected assets, and provide mitigation-oriented information. Such capabilities can help users understand large security datasets without manually examining every individual event.

C. MICROSOFT SECURITY COPILOT

Microsoft Security Copilot is described as an AI assistant for cybersecurity professionals. It combines generative AI capabilities with security-specific information to assist with investigations, event summaries, code or URL analysis, and incident-related work.
The system is intended to support analysts rather than replace them. Features such as shared investigation workspaces and reusable prompt sequences can support collaboration and repeatable workflows. The source paper also describes feedback mechanisms that can help users understand and report incorrect AI responses.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, October 2026 Impact Factor: 8.2

D. TALON ENTERPRISE BROWSER

Talon Enterprise Browser is described as integrating enterprise browser capabilities with Microsoft Azure OpenAI services. This approach is intended to provide organizations with more controlled use of generative AI.

The browser can support data-protection controls and can restrict users from entering sensitive information such as credit-card data or source code into an AI interface. Administrative controls and activity reporting can also help organizations manage the use of AI tools in an enterprise environment.

E. SLASHNEXT GENERATIVE HUMAN AI

SlashNext Generative Human AI is presented as a security service focused on threats such as business email compromise, supply-chain attacks, financial fraud, and executive impersonation.

The system combines machine learning, computer vision, natural-language processing, contextual analysis, and relationship information. It can generate many variations of known business-email-compromise threats to improve defensive preparation. It can also analyze communication style, context, and possible manipulation signals.

F. RECORDED FUTURE AI

Recorded Future AI is described as using AI to organize and analyze large amounts of threat intelligence. It combines information from different sources and can help security teams identify relationships between adversaries, victims, technical indicators, and other objects.

The benefit of such systems is the ability to process information at a scale that is difficult for analysts to manage manually. AI-generated summaries and relationships can support investigation, but analysts should still verify important findings against reliable evidence.

VII. OUTPUT

This section presents sample outputs of the proposed system for three common security tasks: alert summarization, conversion of a natural-language question into a threat-hunting query, and phishing e-mail analysis. These are illustrative sample outputs prepared to show the kind of result the proposed workflow produces. They are not measured experimental results, and all names, addresses, and domains used in them are fictitious. In every case the result is marked for analyst review, in line with the human-validation stage of the proposed system.

A. Sample Output 1: Alert Summarization

Input: raw security events received within 10 minutes

```
[10:02] 14 failed logins      user=a.sharma    src_ip=203.0.113.45
[10:07] 1 successful login    user=a.sharma    src_ip=203.0.113.45
[10:08] New mailbox forwarding rule created    user=a.sharma
[10:09] Large file download from shared drive user=a.sharma
```

Output: AI-generated alert summary

```
Summary    : Possible compromise of account 'a.sharma'.
Reason     : Repeated failed logins were followed by a successful login from
             the same unfamiliar IP, then a forwarding rule and a bulk download.
Suggested  : 1) Reset password and revoke active sessions
             2) Review and remove the new forwarding rule
             3) Enforce multi-factor authentication
Status     : PENDING ANALYST REVIEW
```

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

B. Sample Output 2: Natural-Language Threat-Hunting Query

Analyst request and generated query

```
Analyst : Show users with more than 10 failed logins and at least one
          successful login from the same IP in the last 24 hours.

Generated query (pseudo-SQL):
  SELECT user_name, source_ip,
         SUM(CASE WHEN result = 'FAILED'  THEN 1 ELSE 0 END) AS failed_count,
         SUM(CASE WHEN result = 'SUCCESS' THEN 1 ELSE 0 END) AS success_count
  FROM   auth_logs
  WHERE  event_time >= NOW() - INTERVAL '24 hours'
  GROUP  BY user_name, source_ip
  HAVING SUM(CASE WHEN result = 'FAILED'  THEN 1 ELSE 0 END) > 10
     AND SUM(CASE WHEN result = 'SUCCESS' THEN 1 ELSE 0 END) >= 1;

Status  : Query shown for analyst review BEFORE execution.
```

C. Sample Output 3: Phishing E-mail Analysis

Input: received e-mail (fictitious)

```
From    : HR Department <hr-support@company-payroll.example>
Subject : Urgent: Verify your payroll account today
Body    : Your salary will be stopped. Click the link and enter your login
          details within 2 hours.
```

Output: phishing analysis

```
Verdict    : SUSPICIOUS - phishing likely
Indicators : 1) Look-alike sender domain ('payroll' uses the digit 1)
             2) Urgency and threat of salary stoppage
             3) Request for login credentials through a link
             4) Link domain does not match the company domain
Suggested  : Quarantine the e-mail, warn recipients, report to security team
Status     : PENDING ANALYST REVIEW
```

D. Comparison of Existing and Proposed System

Table I summarizes the difference between the existing system and the proposed system.

TABLE I. COMPARISON OF EXISTING AND PROPOSED SYSTEM

Feature	Existing System	Proposed System
Detection approach	Predefined rules and known signatures	Learned patterns plus existing rules
Timing	Mostly reactive, after an event	More proactive, using generated test scenarios
New or modified threats	May be missed	Variants can be generated for testing and detection
Alert analysis	Manual reading of alerts and logs	AI-generated summaries for the analyst
Log queries	Written by hand in a query language	Generated from natural-language questions
Training and testing	Few manual scenarios	Generated attack and phishing scenarios

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Role of human	Performs all analysis	Reviews and validates AI output
Main risk	Missed unknown threats, alert overload	Wrong output, privacy risk, cost, misuse

VIII. LIMITATIONS OF GENERATIVE AI IN CYBER SECURITY

Although GAI offers many security benefits, it also introduces technical, operational, ethical, and privacy challenges. These limitations are important because an incorrect security output can lead to missed threats or unnecessary actions.

A. TENDENCY TO PRODUCE WRONG OR UNSAFE RESULTS

Generative models do not guarantee that every answer is correct. They may produce inaccurate information, incorrect code, misleading explanations, or unsafe recommendations. In cybersecurity, this can be serious because analysts may make decisions based on generated information.

For this reason, AI output should be treated as assistance rather than unquestionable truth. Important results should be checked using security logs, trusted intelligence, automated controls, and human expertise.

B. COST INEFFICIENCY

Training large generative models can require significant computing resources, large datasets, specialized hardware, and skilled personnel. Running these models can also create continuing infrastructure and operational costs.

Organizations therefore need to consider whether a GAI solution provides sufficient value for the specific security task. Smaller or specialized models may be more practical for some applications.

C. HIGH SETUP AND IMPLEMENTATION TIME

Deploying GAI in a security environment is not only a matter of connecting a model to existing tools. Data preparation, integration, access control, testing, monitoring, and user training can require substantial time.

The system must also be evaluated before being trusted in operational workflows. Poor integration can introduce new risks instead of reducing existing ones.

IX. CONCLUSION

Generative AI has significant potential to support cybersecurity by improving threat detection, generating attack scenarios, identifying abnormal behavior, assisting with password analysis, detecting phishing and malware, supporting threat hunting, and automating repetitive security tasks. The reviewed applications show that GAI can be used not only for general language generation but also for specialized security workflows.

Overall, GAI can strengthen cybersecurity when it is implemented responsibly and evaluated continuously. Future work should focus on better security-specific datasets, stronger evaluation benchmarks, improved context handling, privacy-preserving techniques, and safeguards that reduce misuse while retaining the benefits of generative technology.

REFERENCES

[1] Z. Epstein et al., “Art and the science of generative AI,” Science, vol. 380, no. 6650, pp. 1110–1111, 2023.

[2] Y. Cao et al., “A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT,” arXiv:2303.04226, 2023.

[3] W. Wang, X. Lin, F. Feng, X. He, and T.-S. Chua, “Generative recommendation: Towards next-generation recommender paradigm,” arXiv:2304.03516, 2023.

[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, 2023.

[5] Y. Hu et al., “Exploring the design space of employing AI-generated content for augmented reality display,” arXiv:2303.16593, 2023.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

[6] Ö. Aydin and E. Karaarslan, "Is ChatGPT leading generative AI what is beyond expectations?" Academic Platform Journal of Engineering and Smart Systems, vol. 11, no. 3, pp. 118–134, 2023.
[7] I. Prieto and B. Blakely, "Proposed uses of generative AI in a cybersecurity-focused SOAR agent," Proc. AAAI Symposium Series, vol. 2, no. 1, p. 386, 2024.
[8] L. A. Bauer and V. Bindschaedler, "Generative models for security: Attacks, defenses, and opportunities," arXiv:2107.10139, 2021.
[9] O. S. Striuk and Y. P. Kondratenko, Generative Adversarial Networks in Cybersecurity: Analysis and Response. Cham, Switzerland: Springer, 2023, pp. 373–388.
[10] H. Chen, Y. Zhang, Y. Cao, and J. Xie, "Security issues and defensive approaches in deep learning frameworks," Tsinghua Science and Technology, vol. 26, no. 6, pp. 894–905, 2021.
[11] M. E. Bonfanti, "Artificial intelligence and the offence-defence balance in cyber security," in Cyber Security: Socio-Technological Uncertainty and Political Fragmentation. Routledge, 2022, pp. 64–79.
[12] D. B. Rawat, R. Doku, and M. Garuba, "Cybersecurity in big data era: From securing big data to data-driven security," IEEE Transactions on Services Computing, vol. 14, no. 6, pp. 2055–2072, 2021.