Düzce Üniversitesi
Siber Güvenlik ve Dijital Ekonomi

Düzce University
Journal of Cybersecurity and Digital Economics
https://dergipark.org.tr/tr/pub/cdej

Research Article

Examining the Ethical Risks of Generative AI in Cybersecurity: An Experimental Study on Ethical, Gray Area and Unethical Usage Scenarios

Büşra TAKGİL[a]*

[a]Duzce University, Faculty of Engineering, Computer Engineering, Duzce, Türkiye.
*__Corresponding author:__ busratakgil@duzce.edu.tr

Article Information:
Received: 08/05/2025, Revision: 29/05/2025, Accepted: 05/06/2025

ABSTRACT

Generative AI (GenAI) systems, which are among the emerging artificial intelligence technologies, have started to play an important role in the field of cyber security for both defense and attack purposes. This study aims to systematically analyze the ethical risks of GenAI in cybersecurity. In the study, firstly, a literature review on the usage areas of GenAI in cybersecurity was conducted, then ethical, gray area and unethical usage scenarios were developed and applied experiments were carried out. Each scenario was tested separately using large language models (LLM) such as OpenAI ChatGPT and DeepSeek, and metrics such as accuracy rates, false positive rates and ethical risks were analyzed. The applied results have shown that GenAI-based systems can achieve high accuracy rates in early detection of security threats, but at the same time, they can cause serious ethical issues such as individual privacy violations, misleading content production, and malicious use. The findings of the study emphasize the need for stronger policy regulations, technical limitations, and ethical frameworks to ensure the safe and ethical use of Generative AI technologies in cybersecurity. The results also provide significant contributions to the academic literature and practitioners on how GenAI systems should be managed from both defensive and offensive perspectives.

Keywords: *Generative AI, Cybersecurity, Ethical Risks, Gray Area Uses, Artificial Intelligence Security.*

I. INTRODUCTION

In recent years, rapid developments in artificial intelligence technologies have paved the way for the emergence of new application areas in different sectors. Generative AI (GenAI) systems, one of the pioneers of these developments, offer groundbreaking innovations especially in areas such as natural language processing, content generation and automation (Sai et al., 2024). GenAI-based solutions, such as large language models (LLMs), can produce human-like text, code and visuals with high accuracy rates and increase productivity in different sectors (Agrawal et al., 2024). However, the production capacity of these systems also brings various security threats and ethical issues (Hasanov et al. 2024, Kasri et al., 2025).

Cyber security is one of the most critical areas that directly experience the risks of artificial intelligence applications. While GenAI systems offer defensive solutions such as improving intrusion detection systems, automating incident response processes, and detecting malware, they also open the door to unethical uses such as generating malware, targeted phishing attacks, and creating misleading content (Humphreys et al., 2024). This dual nature has caused GenAI to become a “double-edged sword” in cybersecurity and has created a new research need in this field (Pasupuleti et al., 2023).

Studies in the literature address the potential threats and defense capabilities of GenAI systems. For example, Usman et al. (2024) comprehensively demonstrated how GenAI systems change cyber-attack vectors. However, most of the existing research remains at the theoretical level, with limited hands-on experimentation on how GenAI systems perform in different use cases. Therefore, evaluating the behavior of GenAI systems by conducting systematic experiments on both ethical and unethical use cases would fill an important gap in the literature (Novelli et al., 2024).

The contribution of this study to the literature is the detailed experimental analysis of GenAI systems in the categories of ethical use, gray area uses and unethical use in the context of cyber security. Performance measurements were performed using GenAI platforms such as OpenAI ChatGPT and DeepSeek over the scenarios determined for each category, and the results obtained were evaluated in terms of both technical and ethical risks. This structure is not only limited to the description of risks but also provides a framework to guide practitioners and policy makers by offering possible solutions and regulatory recommendations.

The rest of the paper is organized as follows: Section 2 provides an overview of the related literature and discusses the potential uses of GenAI in cyber security. Section three describes the software; data sets and experimental methods used in the study. In the fourth section, the experimental results for ethical, gray area and unethical use scenarios are presented in detail. The last section summarizes the results of the study and makes recommendations for future research.

II. Literature Review

In recent years, the development of Generative AI (GenAI) technologies has brought both defensive and offensive uses in the field of cyber security. This situation has caused ethical risks and legal issues to gain more importance. Tufan (2024) states that AI technologies enable crimes such as phishing, deep forgery and financial fraud to be committed in a more sophisticated and convincing manner. Such crimes show that AI technologies can be misused, and this can pose serious legal, ethical and security risks. It is emphasized that the information provided by AI may not always be accurate or valid, with serious consequences for legal liability.

Smart and Şimşek (2024) argue that the possibility of using large language models (LLMs) to automate cyberattacks increases the risks and makes it essential to establish comprehensive security measures. In this context, the challenges posed using artificial intelligence technologies in terms of legal and political regulations reveal the necessity of international cooperation and setting standards. Turgut Bilgiç (2024) draws attention to the risks posed by artificial intelligence systems in terms of data privacy and security. It is stated that artificial intelligence systems may have negative effects on the integrity and confidentiality of personal data and this may lead to ethical and legal problems. Karadeniz (2025) states that the European Union's Artificial Intelligence Law aims to protect fundamental rights such as health, safety, democracy and the rule of law against the harmful effects of artificial intelligence systems and in this context, it sets specific requirements and obligations for high-risk artificial intelligence systems. These regulations are considered as an important step to ensure the ethical and safe use of artificial intelligence systems. Ümütlü (2025) states that while artificial intelligence judges offer advantages such as impartiality, speed and legal consistency, they carry serious risks in terms of fair trial rights, accountability and ethical issues. This study examines the advantages and disadvantages of AI judges in terms of international law in detail in the context of human rights.

Usman et al. (2024) extensively examined how GenAI systems change cyber-attack vectors and how these systems can be used by malicious actors. The study shows that GenAI can be used in areas such as social engineering, malware production and system exploitation. Gupta et al. (2023) evaluated the effects of GenAI on cybersecurity and privacy. The study discusses how GenAI can be used on the defensive and offensive sides and what the implications of this use are in terms of social, ethical and privacy implications. Shibli et al. (2024)

examined the misuse of GenAI-based chatbots and how these systems can be used to create smishing campaigns. The study shows how the ethical standards of GenAI systems can be circumvented and what risks this poses to user security. IBM (2024) addressed the ethical aspects of using artificial intelligence in cybersecurity. The study emphasizes that AI systems should be developed in accordance with ethical principles such as transparency, impartiality and human oversight. NTT Data (2024) stated that GenAI systems can lead to ethical issues such as misinformation generation, copyright violations and bias. The study states that developers should take various measures to reduce such inappropriate outputs.

Tabassum et al. (2025) comprehensively examines the ethical and legal issues arising from the combination of GenAI and metaverse technologies. Using a scoping review method, the study analyzes the role of large language models (LLMs) in communication, content generation, translation and game interactions within the metaverse. It details how these technologies give rise to multidimensional problems such as personal privacy, data security, algorithmic bias, misinformation propagation, and intellectual property rights. The authors propose solutions to these problems, such as ethical AI design, transparency and auditability, inclusive data use, and the development of international regulations. The study serves as a strategic guide to shape the impact of the metaverse on future digital societies (Tabassum et al., 2025).

Raman et al.'s (2024) study presents a comparative analysis of generative AI tools that use large language models (LLMs), such as ChatGPT and Bard, on their proficiency in the Certified Ethical Hacker (CEH) exam. In the study, 218 multiple-choice questions taken from the CEH exam were used to measure the ethical hacking knowledge of these two AI tools, and the answers were evaluated in terms of qualities such as accuracy, comprehensiveness, clarity, and conciseness. According to the findings, Bard achieved a higher accuracy rate of 82.6%, while ChatGPT showed better clarity, comprehensiveness, and conciseness performance with an accuracy rate of 80.8%. Interestingly, it was observed that confirmation questions such as "Are you sure?" increased the accuracy of the answers. It was also stated that Bard avoided answering on some sensitive topics, while ChatGPT included ethical references, which revealed different security policies among the developers. The study provides valuable insights into the usability of AI-based tools in cybersecurity and highlights the need to evaluate their integration into professional certification processes such as CEH (Raman et al., 2024). Al-kfairy et al. (2024) presented a comprehensive study that systematically examines the ethical issues raised by GenAI technologies and integrates the perspectives of different disciplines. The study highlights major ethical concerns such as privacy violations, data security, intellectual property rights issues, misleading content (deepfake and disinformation), algorithmic biases, and the reinforcement of social inequalities. The impacts of these technologies in sectors such as health, education and media have been discussed, emphasizing the need to develop frameworks based on individual rights, transparency and fairness. It also calls for a multidisciplinary dialog between policy makers, software developers and researchers. The paper makes a significant contribution to the literature by providing both theoretical foundations and practical solutions for ethically responsible GenAI development (Al-kfairy et al., 2024).

Nadella et al. (2025) presented a holistic framework that addresses how GenAI technologies can be used to ensure corporate data privacy and detect cyber threats. In the study, models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are used to generate synthetic datasets that mimic the real world and combined with anomaly detection, differential privacy, encryption and data masking techniques. The proposed system has been tested in financial, healthcare, and smart city applications, with impressive results such as 94-96% accuracy, 90-95% F1 score, and low processing time (1.5 s) for real-time applications. In addition, the system works in compliance with legal regulations such as HIPAA, GDPR and CCPA with techniques such as AES-256 encryption, TLS 1.3, differential privacy ($\varepsilon = 0.1$). In this respect, the study makes a significant contribution to the literature by providing both a practical and scalable solution for generative AI-supported cybersecurity applications (Nadella et al., 2025).

Hagendorff (2024) presented a comprehensive scoping review systematically mapping the ethical dimensions of GenAI. Based on 179 academic publications, the analysis examines the ethical implications of large language models (LLMs) and text-to-image models. The study reveals trends in the literature by classifying 378 ethical issues under 19 headings such as justice, security, harmful content generation, hallucinations, privacy violations, interaction risks, cybercrime, labor displacement, property rights, governance and sustainability. It also emphasizes that a significant number of these issues are based on hypothetical foundations and are not supported by sufficient empirical data and argues that ethical debates should be conducted in a more balanced, data-driven and multidimensional manner. This study provides a critical and instructive framework for the methodological and contextual aspects of research on GenAI ethics (Hagendorff et al., 2024).

This literature review reveals that GenAI technologies pose ethical risks in cybersecurity and that comprehensive regulations are needed to manage these risks. In this context, the study aims to contribute to the literature by systematically analyzing the ethical risks of GenAI in cybersecurity.

III. METHOD

In this study, an experimental approach is adopted to evaluate the effects of Generative AI (GenAI) systems in ethical, gray area and unethical use scenarios in cyber security. The study process consisted of literature review (obtaining relevant articles), scenario generation, tool setup and testing, model training and evaluation. Each step was structured in a systematic way and analyzed in line with the experimental findings. Figure 1 shows the systematic approach schematically.

Both generative models and cyber security analysis tools were used in the experiments. OpenAI's ChatGPT API and DeepSeek models were used for GenAI-based content generation and development of attack/defense scenarios. Wireshark (network traffic analysis), Metasploit Framework (penetration tests) and Python programming language (pandas, scikit-learn, matplotlib libraries) were used for simulation of cybersecurity environments and data analysis. Code development and model training were performed using Python version 3.11 and the Jupyter Notebook environment.

A. CREATING A DATA SET

The data set used in the study was not taken directly from ready-made sources but was created through existing literature and original scenario generation. In this context, current academic articles and sectoral reports were analyzed, and potential usage scenarios of Generative AI in the field of cyber security were extracted. Table 1 shows the list of articles analyzed for three different categories. The dataset is organized under three main categories:

- Ethical Use Cases: The use of GenAI for ethical purposes in cyber security defense (e.g. malware detection, authentication systems).
- Gray Area Use Cases: Areas whose use is legally or socially controversial (e.g. social media surveillance, content analysis).
- Unethical Use Cases: Use of GenAI in abusive, illegal or unethical activities (e.g. social engineering attacks, deepfake production).

In the scenario creation phase, a total of 53 academic articles and industry reports were analyzed. Table 1 presents a representative subset of 20 publications, selected based on their direct relevance to at least one of the scenario categories: ethical, gray area, or unethical use. These articles were chosen to reflect diversity in geographical scope, methodological rigor, and thematic coverage. Each scenario in the dataset was inspired or supported by specific articles. For example, the article "Is Generative AI the Next Tactical Cyber Weapon For Threat Actors?" directly informed the construction of the phishing and social engineering scenarios. Similarly, "Mapping the Ethics of Generative AI" helped shape the conceptual boundaries for gray-area applications involving content moderation and user profiling.

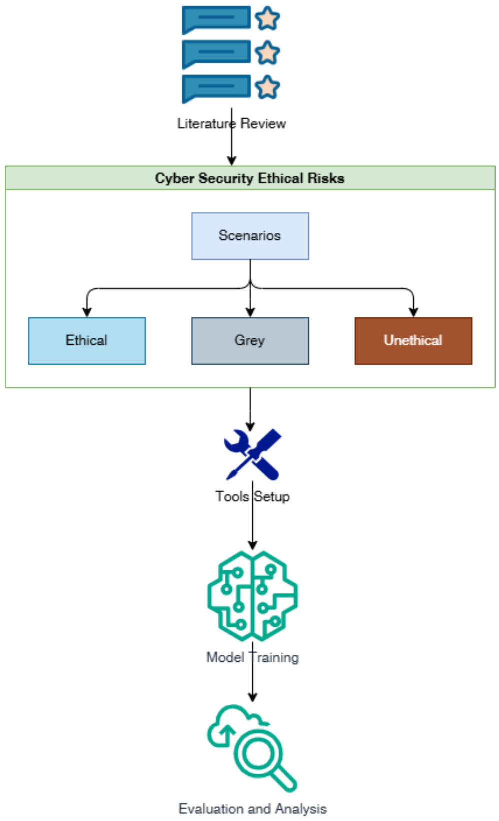

Figure 1. Block diagram of the study.

After analyzing the articles in Table 1, scenarios were developed for each category. For each scenario, the relevant use case, purpose of use, potential risks and impacts were detailed. In total, 17 different scenarios were included in the dataset. The scenarios were categorized into three main categories based on expert opinions and criteria obtained from the literature. The following evaluation criteria were considered in the process of placing the scenarios in ethical, gray area or unethical classes:

- The intended use's potential for social benefit or harm
- Risk that the use violates individual rights and freedoms
- Compliance of the use with legal regulations (GDPR, KVKK, etc.)

Table 1. Representative list of selected articles (20 out of 53) analyzed for scenario development.

Article Name
Is Generative AI the Next Tactical Cyber Weapon for Threat Actors?
The Role of Generative AI in Cyber Security
A Survey on the Application of GANs in Cybersecurity
Üretken Yapay Zekâdaki Etik Sorunlar: Sistematik Bir İnceleme
Artificial Intelligence Focused Cyber Risk and Security Management
Cybersecurity Maturity of Turkey: An Assessment with ENISA's NCAF
GenAI against humanity: nefarious applications of generative AI and LLMs
Artificial intelligence (AI) cybersecurity dimensions
Mapping the Ethics of Generative AI: A Comprehensive Scoping Review
Generative AI for Pentesting: The Good, The Bad, The Ugly
Gizlilik ve Güvenlik Endişeleri Üretken AI: Kapsamlı bir anket
From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy
Enhancing Cyber Security Enhancement Through Generative AI
Generative AI in Cybersecurity

Generative AI: A New Challenge for Cybersecurity
Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective
Generative AI with GANs in IDS and Obfuscation Attacks
Mapping Global Cybercrime Trends: A Kohonen Map Approach (2016-2023)
Threats and Opportunities with AI-based Cybersecurity Intrusion Detection: A Review
Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity

The distribution of scenarios in the whole data set is given in Figure 2. In the study, LLMs used prompt masking, obfuscation, jailbreak instructions, chained commands, and obfuscation techniques with code or command masking to produce unethical content.

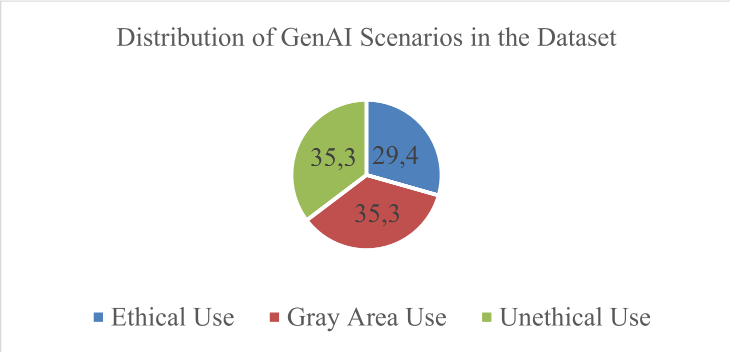

Figure 2. Distribution of scenarios in the data set.

IV. Experiments and Results

The scenarios developed in this study were implemented in a testbed to evaluate the potential use cases of GenAI systems in cybersecurity. Each scenario was categorized as ethical, gray area, or unethical use, focusing on three key outcomes for each category: technical capability, ethical compliance, and legal risk. The scenario analyses were evaluated on both qualitative content generation and GenAI's suggested outcomes.

In the experimental environment, LLM models such as ChatGPT (OpenAI), DeepSeek and similar LLM models were tested for the accuracy, attack potential and vulnerability to unethical manipulation of textual responses to cybersecurity scenarios. In total, 17 scenarios were empirically tested according to their respective categories:

- Ethical Use Scenarios:

LLM-based systems achieved high accuracy rates and meaningful defense outputs in cyber threat analysis, intrusion prevention and behavioral biometrics scenarios. In these scenarios, it was observed that the model did not produce unethical routing.

- Gray Area Usage Scenarios:

The models produced effective content on topics such as social media surveillance, content analysis and persuasion techniques, but the risk of privacy violations was observed. Especially the content produced against "modify" commands contains limited censorship mechanisms, which points to risky areas of use.

- Unethical Use Scenarios:

In direct illegal use requests such as phishing email generation, deepfake text generation and malware sample generation, the models showed protective filters most of the time, but with some special instructions, unethical outputs were observed.

The models produced output at different sensitivity levels as shown in Table 2:

Table 2. Sensitivity levels of the models.

Scenario Type	Response Type of Models	Ethical Filtering Success	Open Risk Monitoring
Ethical Use	Supportive, Constructive	High	None
Grey Area Usage	Content creation without questioning	Middle	Partial
Unethical Use	Open in some jailbreak prompts	Low	High

As a result of the experimental observations, each scenario was evaluated according to individual risk, societal benefit and legal compliance criteria. Table 3 summarizes this evaluation. Table 3 has been adapted according to KVKK and GDPR legislation.

Table 3. Evaluation results of the models.

Scenario	Individual Risk	Social Benefit	Legal Compliance
Malware Detection	Low	High	Appropriate
Deepfake Production	High	Low	Inappropriate
Social Media Monitoring	Middle	Middle	Controversial
Phishing Email Generation	High	None	Inappropriate
Authentication Systems	Low	High	Appropriate

In this study, 17 scenarios created to evaluate the ethical impacts of Generative AI in the field of cybersecurity are classified into three main groups as "Ethical Use", "Gray Area" and "Unethical Use" according to the purpose of use. Ethical use scenarios are cases where artificial intelligence is used in defensive and socially beneficial areas such as network security, malware detection, vulnerability scanning and cyber security education. Gray area scenarios include technically feasible uses that pose a risk to privacy, freedom of expression and manipulation; applications such as social media surveillance, identity verification with behavioral biometrics and content censorship are included in this group.

In unethical use scenarios, AI is used directly for malicious purposes, such as personalized phishing attacks, deepfakes, fake news propagation and automated malware development. This classification highlights that GenAI technologies carry serious ethical, legal and societal risks as well as potential benefits, and therefore the context of use should be clearly defined and supported by auditable structures. Table 4 presents the scenarios.

V. CONCLUSIONS AND FUTURE WORK

This study aims to systematically analyze the potential uses of Generative AI (GenAI) systems in the field of cybersecurity and the ethical risks they pose. A unique dataset based on the literature was created, scenarios were classified under three main categories (ethical use, gray area use and unethical use) and analyzed through experimental tests. The experiments revealed that GenAI systems have high potential for defensive uses but may pose serious risks in gray area and unethical scenarios.

In ethical use cases, GenAI systems have been shown to produce effective and secure outputs in areas such as malware detection, cyber threat prediction and authentication. This shows that GenAI can be used as a valuable tool in cyber security defenses. However, it was found that there are significant risks in gray area usage scenarios, especially in user privacy and data security. In applications such as social media surveillance and content moderation, potential rights violations may occur due to the limited ethical filtering capacity of the models.

Table 4. Ethical risk scenarios in cyber security.

Category	Scenario	Explanation
Ethical Use	Cyber Threat Analysis with Artificial Intelligence	Detection of malware and attacks by examining network traffic and file activities with AI.
Ethical Use	Vulnerability Detection and Security Testing	Automatic detection of vulnerabilities in network and software systems with AI.
Ethical Use	Cyber Security Simulations for Educational Purposes	Production of cyber-attack and defense simulations with AI.
Ethical Use	Prediction of Possible Cyber Threats	Prediction of new threats from historical attack data.
Ethical Use	Authentication Systems (Behavioral Biometrics)	Authentication by analyzing user behavior with AI.
Grey Area Usage	Detection of Disinformation Content	Detecting disinformation in social media content.
Grey Area Usage	Social Media Perception Management	Directing social media users' perceptions with AI.
Grey Area Usage	Automatic Moderation of Sensitive Topics	Automatic censorship/filtering of content on sensitive topics.
Grey Area Usage	AI-Powered Social Media Surveillance	Detecting harmful content with AI, but there is a privacy risk.

Grey Area Usage	Automatic Content Analysis	Threat analysis in emails and social media content.
Grey Area Usage	Artificial Intelligence-Enhanced Persuasion Techniques	Creating persuasive messages specific to user profiles.
Unethical Use	Malware Generation	Automatic malware development with AI.
Unethical Use	Identity Spoofing by Creating Deepfakes	Fraud by creating fake audio and video.
Unethical Use	Misleading the Public with False Information	Generating fake news and propaganda with AI.
Unethical Use	Malware Development (Scenario 2)	AI-powered advanced malware generation and propagation.
Unethical Use	Social Engineering Attacks	Creating personalized phishing attacks using AI.
Unethical Use	Misleading Content and Propaganda	Misleading the public with deepfake and fake news content.

Experiments on unethical usage scenarios have shown that existing LLM (Large Language Model) systems can bypass security filters with some special prompt engineering. In scenarios such as phishing email generation, deepfake content generation, and malware development, it was observed that although GenAI models sometimes apply protective filters, they can produce unethical content through manipulation techniques. This finding suggests that existing AI systems need to be supported not only by technical limitations, but also by social approaches such as usage policies and user education. This study strongly emphasizes that GenAI systems should be evaluated not only in terms of technical performance but also in terms of ethical and legal dimensions. It is inevitable for model developers, policy makers and academia to work in cooperation to ensure that artificial intelligence technologies can be used safely and in a manner that respects human rights.

The findings of this study lay the groundwork for more comprehensive future research. First, larger data sets that include different cultural contexts should be created to analyze the responses of GenAI systems to different user profiles. In this way, the effects of cultural, linguistic and sociodemographic variables on the ethical performance of GenAI can be more deeply understood. Secondly, the scenarios used in this study were evaluated with qualitative content analysis. In future studies, quantitative analyses (e.g., calculating ethical risk scores, scoring attack potential) can be performed on the outputs given to the scenarios. Thus, the ethical compliance levels of GenAI models according to different scenarios will become measurable.

Thirdly, special attention should be paid to the situations where GenAI systems can be used as attack vectors. It is important to systematically classify methods for bypassing security filters such as jailbreak prompts and develop countermeasures. In addition, the development of real-time ethical filtering and anomaly detection mechanisms that can be integrated into GenAI systems constitutes a critical area for future research. There is a need for in-depth studies on the legal aspects of GenAI use, especially the responsibility and ownership of AI-generated content. In this regard, interdisciplinary studies on the integration of technology law, cybersecurity policies and ethical regulations are recommended.

DECLARATIONS

Acknowledgements: Any person and/or institution can be acknowledged in this section.

Author Contributions: All work was done by B.T.

Conflict of Interest Statement: Author declares no conflict of interest.

Supporting/Supporting Organizations: This research has not received any external funding.

Ethical Approval and Participant Approval: This article does not contain any studies on human or animal subjects. Scientific and ethical principles were followed during the preparation of this study and all studies used are given in the references.

Plagiarism Statement: This article was scanned with a plagiarism program. No plagiarism was detected.

Availability of Data and Materials: Data sharing is not valid.

Use of AI Tools: The Author declare that they did not use Artificial Intelligence (AI) tools in the creation of this article.

REFERENCES

Agrawal, G., Kaur, A., & Myneni, S, 2024. A Review of Generative Models in Generating Synthetic Attack Data for Cybersecurity. Electronics, 13(2), 322. https://doi.org/10.3390/electronics13020322 .

Akıllı, M., & Şimşek, M., 2024. Dijital Diplomaside Büyük Dil Modelleri: Fırsatlar ve Riskler. *İnsan ve Toplum*, 12(1), 1-20. https://dergipark.org.tr/tr/download/article-file/4422215Home+4Home+4Home+4.

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O., 2024. Ethical challenges and solutions of generative AI: An interdisciplinary perspective. Informatics, 11(3), 58. https://doi.org/10.3390/informatics11030058.

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L., 2023. *From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy*. arXiv preprint arXiv:2307.00691.arXiv+1Wikipedia+1.

Hagendorff, T., 2024. Mapping the ethics of generative AI: A comprehensive scoping review. Minds and Machines, 34(39). https://doi.org/10.1007/s11023-024-09694-w .

Humphreys, D., Koay, A., Desmond, D. et al., 2024. AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business. AI Ethics 4, 791–804. https://doi.org/10.1007/s43681-024-00443-4.

I. Hasanov, S. Virtanen, A. Hakkala and J. Isoaho, 2024. Application of Large Language Models in Cybersecurity: A Systematic Literature Review, in IEEE Access, vol. 12, pp. 176751-176778, doi: 10.1109/ACCESS.2024.3505983.

IBM. 2024. *Navigating the ethics of AI in cybersecurity*. IBM Think. https://www.ibm.com/think/insights/navigating-ethics-ai-cybersecurityIBM - United StatesNTT Data. (2024). *Security Risks of Generative AI and Countermeasures*. https://www.nttdata.com/global/en/insights/focus/2024/security-risks-of-generative-ai-and-countermeasuresnttdata.com.

Karadeniz, S., 2025. Avrupa Birliği Yapay Zekâ Kanunu'nun Risk Grupları ve İlgililerin Yükümlülükleri. *Hukuk Fakültesi Dergisi*, 29(1), 273-338. https://dergipark.org.tr/tr/download/article-file/4181598Home+2Home+2Home+2.

Kasri, W., Himeur, Y., Alkhazaleh, H. A., Tarapiah, S., Atalla, S., Mansoor, W., & Al-Ahmad, H., 2025. From Vulnerability to Defense: The Role of Large Language Models in Enhancing Cybersecurity. Computation, 13(2), 30. https://doi.org/10.3390/computation13020030 .

Nadella, G. S., Addula, S. R., Yadulla, A. R., Sajja, G. S., Meesala, M., Maturi, M. H., Meduri, K., & Gonaygunta, H., 2025. Generative AI-enhanced cybersecurity framework for enterprise data privacy management. *Computers, 14*(2), 55. https://doi.org/10.3390/computers14020055.

Novelli, C. et al., 2024. Generative AI in EU law: liability, privacy, intellectual property, and cybersecurity. Comput. Law Secur. Rev. 55, 106066.

R. Pasupuleti, R. Vadapalli and C. Mader, 2023. Cyber Security Issues and Challenges Related to Generative AI and ChatGPT. International Conference on Social Networks Analysis, Management and Security (SNAMS), Abu Dhabi, United Arab Emirates, 2023, pp. 1-5, doi: 10.1109/SNAMS60348.2023.10375472.

Raman, R., Calyam, P., & Achuthan, K., 2024. ChatGPT or Bard: Who is a better Certified Ethical Hacker? *Computers & Security, 140*, 103804. https://doi.org/10.1016/j.cose.2024.103804.

S. Sai, U. Yashvardhan, V. Chamola and B. Sikdar, 2024. Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E, and Other Models for Enhancing the Security Space, in IEEE Access, vol. 12, pp. 53497-53516, 2024, doi: 10.1109/ACCESS.2024.338510.

Shibli, A. M., Pritom, M. M. A., & Gupta, M., 2024. *AbuseGPT: Abuse of Generative AI ChatBots to Create Smishing Campaigns*. arXiv preprint arXiv:2402.09728.arXiv.

Tabassum, A., Elmahjub, E., Padela, A. I., Zwitter, A., & Qadir, J., 2025. *Generative AI and the metaverse: A scoping review of ethical and legal challenges*. IEEE Open Journal of the Computer Society, 6, 348–359. https://doi.org/10.1109/OJCS.2025.3536082 .

Tufan, B. N., 2024. Yapay Zekâ ve Suç: Gelecek Açısından Hukuksal ve Etik Tehditler. *Medeniyet Belleten*, 12(1), 1-20. https://dergipark.org.tr/tr/download/article-file/4367943.

Turgut Bilgiç, E., 2024. Genel Veri Koruma İlkelerinin Yapay Zekâ Karşısında Uygulanabilirliği. *Hukuk ve Adalet Eleştirel Hukuk Dergisi*, 15(57), 273-290. https://dergipark.org.tr/en/download/article-file/3654085.

Usman, Y., Upadhyay, A., Gyawali, P., & Chataut, R., 2024. *Is Generative AI the Next Tactical Cyber Weapon For Threat Actors? Unforeseen Implications of AI Generated Cyber Attacks*. arXiv preprint arXiv:2408.12806.

Ümütlü, A. Y.,2025. Algoritmik Adalet: Uluslararası Hukukta Yapay Zeka Hakimliği. *Selçuk Üniversitesi Hukuk Fakültesi Dergisi*, 33(1), 777-815. https://dergipark.org.tr/tr/pub/suhfd/issue/91009/1637446.