Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867

Securing Humans from Human-Driven Cyber Threats: AI's Role in Modern Cyber Defence

Aryan Sharma[1], Ms. Ruchi Mishra[2]

Student (BCA), School of Computer Application and Technology , Career Point University, Kota(Raj.)[1]

[1]aryaleena2004@gmail.com

[2]Assistant Professor , School of Computer Applications, Career Point University, Kota

[2]ruchi.mishra@cpur.edu.in

Abstract: Cyber-attacks have evolved into highly sophisticated threats, with human- driven attacks such as phishing, social engineering, and cyberbullying posing significant challenges to both individuals and organizations. This research investigates how artificial intelligence (AI) can effectively mitigate these threats while ensuring accessibility for non-technical users through a minimal- intervention approach. By leveraging AI technologies, this study explores the prevention of diverse human-driven cyberattacks, including insider threats, impersonation attacks, business email compromise (BEC), ransomware facilitated by social engineering, and data breaches caused by human error.

A critical focus of this research is on privacy-preserving methodologies, ensuring that AI solutions safeguard user data and adhere to ethical standards. The study combines a comprehensive literature review with data from a university-wide survey that captures user awareness, experiences with cyber threats, and perceptions of AI-driven security tools. Survey findings are analyzed to evaluate public trust, privacy concerns, and the potential acceptance of AI in cybersecurity.

By integrating theoretical insights with real-world data, this research proposes user-friendly AI strategies that balance technical effectiveness with ethical considerations. The goal is to develop solutions that enhance cybersecurity against human-driven attacks while respecting user autonomy and ensuring broad accessibility.

Keywords: Artificial Intelligence (AI), Cybersecurity, Social engineering, Human-driven cyberattacks, Cybercrime prevention, Privacy-preserving AI, Minimal AI intervention

Introduction:

In the rapidly evolving digital age, cyber threats have become increasingly sophisticated, targeting individuals and organizations alike. Among these, human-driven attacks such

as phishing, social engineering, and cyberbullying stand out due to their reliance on psychological manipulation rather than technological vulnerabilities. These attacks exploit human behavior and trust to gain unauthorized access to sensitive information or systems, often bypassing traditional cybersecurity measures. Their prevalence poses significant risks, emphasizing the need for innovative and adaptable solutions.

Artificial intelligence (AI) has emerged as a powerful tool in the fight against cyber threats. Its ability to analyze vast amounts of data, identify patterns, and detect anomalies in real-time makes it a valuable ally in addressing human- driven cyberattacks. By leveraging machine learning, natural language processing, and other AI-driven technologies, organizations can proactively mitigate risks, automate responses, and reduce the likelihood of successful attacks. However, while AI offers significant potential, its implementation must address ethical concerns, particularly those surrounding privacy and user autonomy.

This research paper investigates the role of AI in combating human-driven cyber threats, focusing on its application to attacks such as phishing, insider threats, vishing, impersonation, business email compromise (BEC), ransomware, and cyberbullying. The study emphasizes the importance of privacy-preserving techniques and minimal AI intervention to ensure accessibility for non-technical users. These principles are crucial in promoting trust and widespread adoption of AI-based cybersecurity solutions.

To provide real-world context, the paper incorporates insights from a university- wide survey. This survey captured participants' awareness of cyber threats, their trust in AI systems, and their privacy concerns regarding AI-driven tools. The findings reveal both opportunities and challenges in implementing AI solutions, highlighting the public's readiness to adopt such tools under certain conditions.

By combining a comprehensive literature review with survey data, this study aims to propose AI-driven strategies that are both effective and user-friendly. The objective is to strike a balance between technological innovation and ethical considerations, ensuring that AI solutions protect users while respecting their privacy and autonomy. As cyber threats continue to evolve, the role of AI in cybersecurity will become increasingly critical, offering a proactive approach to safeguarding digital environments.

Conceptual Framework:

This research explores how AI can combat human-driven cyberattacks like phishing,

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867

social engineering, and cyberbullying while maintaining minimal intervention and prioritizing user privacy. By integrating a literature review and survey analysis, it proposes privacy-conscious, user-friendly AI strategies to protect non-technical users and foster trust in AI-driven cybersecurity solutions.

Review of Literature:

The integration of artificial intelligence (AI) into cybersecurity has significantly transformed the ability to detect and mitigate human-driven cyberattacks. Research such as *"Impact of AI on Cybersecurity and Security Compliance"*[1], highlights AI's potential to analyze large datasets, identify anomalies, and respond to threats in real-time. These capabilities are particularly effective in addressing phishing, social engineering, and impersonation attacks. However, the study emphasizes the importance of maintaining compliance with privacy regulations, such as GDPR, to mitigate concerns about data misuse.

Further studies, including *"The Role of AI in Strengthening Cybersecurity: Promise and Risks"*[2], examine AI's dual role as a defensive tool and a potential vulnerability. These works highlight challenges such as adversarial AI, where attackers manipulate AI systems to bypass security, underscoring the need for ethical and transparent AI solutions. Similarly, the research *"Using Artificial Intelligence to Evaluate Detection of Cybersecurity Threats in Ad Hoc Networks"*[3], focuses on technical applications of machine learning algorithms like Random Forest and Convolutional Neural Networks for real- time anomaly detection. While specific to mobile ad hoc networks, these findings are broadly applicable to dynamic cybersecurity contexts.

Privacy and accessibility remain recurring themes in the literature. Studies emphasize the importance of designing AI systems with minimal user intervention, enabling non-technical users to adopt them seamlessly. Additionally, incorporating privacy-preserving technologies, such as localized processing and differential privacy, is identified as critical to building trust and fostering adoption.

This review underscores the evolving role of AI in combating human-driven cyber threats, highlighting both opportunities and limitations. By addressing ethical concerns and prioritizing user-friendly designs, the reviewed literature

provides a strong foundation for exploring AI's transformative potential in cybersecurity.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867

Research Gap Identified:

While existing studies demonstrate AI's potential to mitigate cyber threats like phishing and social engineering, they often overlook the challenges of implementing privacy-preserving solutions that are accessible to non-technical users. Furthermore, limited research explores public perceptions of AI-driven cybersecurity tools, particularly regarding trust and privacy concerns. This research addresses these gaps by proposing minimal-intervention AI strategies tailored for non-technical users while incorporating privacy-conscious frameworks. Additionally, through survey analysis, it provides insights into user awareness, preferences, and acceptance of AI in cybersecurity, bridging the gap between theoretical advancements and practical, user-centric implementations.

Research Methodology:

This study adopts a mixed-methods approach, integrating a literature review and survey analysis to explore the role of artificial intelligence (AI) in addressing human-driven cyberattacks. The methodology is designed to assess the theoretical potential of AI solutions and their practical acceptance among users, focusing on privacy-preserving and user-friendly applications.

A comprehensive review of existing research was conducted to analyse the capabilities, challenges, and limitations of AI-driven cybersecurity tools. Key studies provided insights into AI's ability to detect and mitigate human-driven threats such as phishing, social engineering, and cyberbullying. Additionally, the literature explored ethical considerations, including privacy concerns and adversarial AI risks, providing a theoretical foundation for identifying research gaps and designing user-centric AI strategies.

Survey Design and Data Collection

A university-wide survey was developed to capture real-world insights on public perceptions of AI in cybersecurity. The survey consisted of multiple- choice, multi-select, and text-based questions covering topics such as

familiarity with cyberattacks, trust in AI systems, privacy concerns, and willingness to adopt AI tools with strong privacy guarantees. Participants were informed about the study's purpose and assured of anonymity, encouraging honest responses.

The survey was distributed via digital platforms, including university networks and social media groups, ensuring diverse participation from students, faculty, and staff. Data collection was completed over a two-week period, resulting in a robust dataset representing various technical and non-technical backgrounds.

Data Analysis

Quantitative responses were analysed to identify trends in user trust, comfort levels, and preferences for AI-driven cybersecurity tools. Bar charts, pie charts, and stacked bar graphs were used to visualize key findings, highlighting user concerns and priorities. Qualitative responses were examined to gain deeper insights into privacy-related apprehensions and suggestions for improving AI solutions.

Integration and Interpretation

Findings from the literature review and survey analysis were integrated to develop actionable recommendations. By combining theoretical and empirical data, the research aims to propose AI strategies that balance technical efficacy with ethical considerations, ensuring accessibility and privacy for non- technical users.

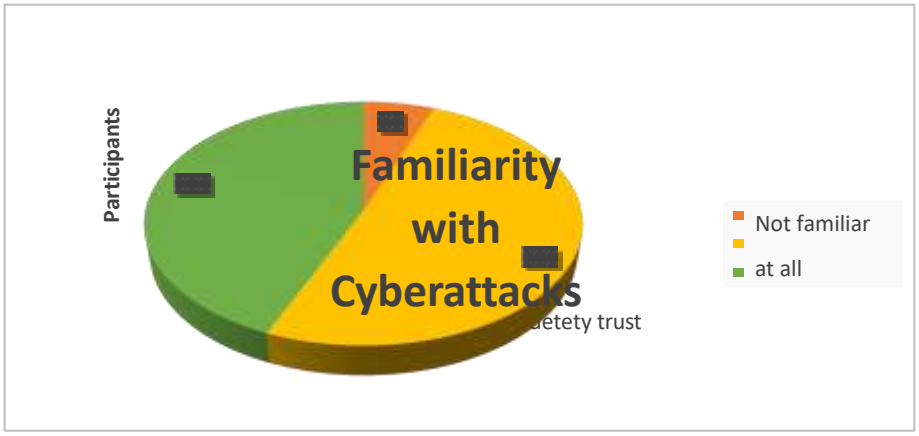
This methodology ensures a holistic understanding of AI's role in cybersecurity, bridging the gap between research and real-world applicability.

Data Analysis & Interpretation:

The survey responses provided valuable insights into user awareness, trust, and preferences regarding AI-driven cybersecurity tools. Quantitative data was analyzed using bar charts, pie charts, and stacked bar graphs, while qualitative responses offered nuanced perspectives on privacy and ethical concerns.

- **Familiarity with Cyber-attacks:**

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867

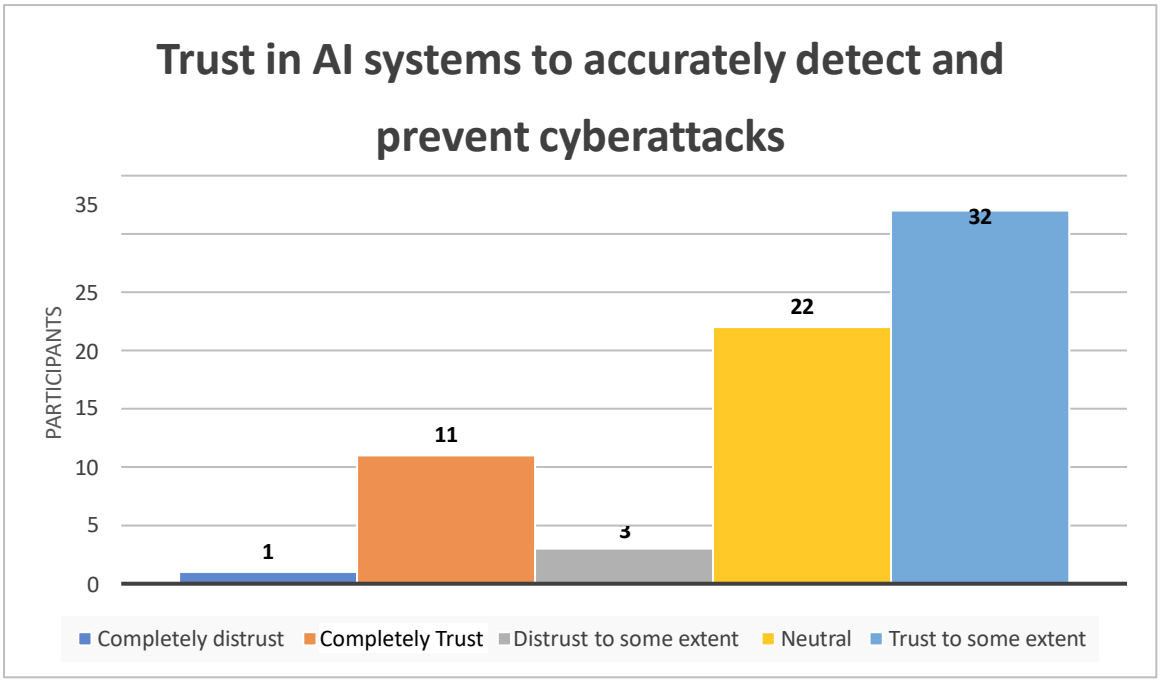


Options -	Number of participants -
Not familiar at all	4
Somewhat familiar	35
Very familiar	30

The majority of respondents (94%) reported being at least somewhat familiar with cyberattacks such as phishing and social engineering. This highlights the relevance of AI solutions for addressing human-driven threats and indicates a baseline awareness among users.

- **Trust in AI Systems:**

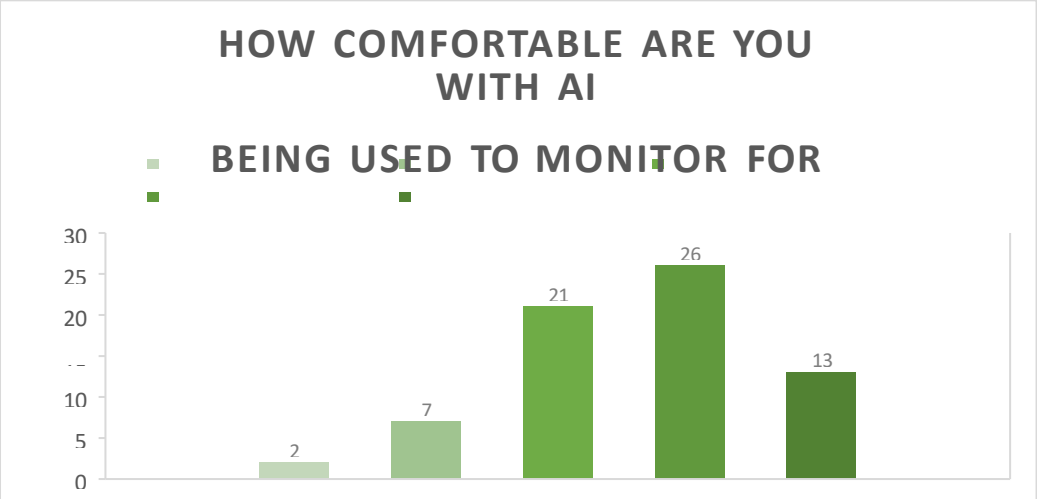
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867



Options -	Number of participants -
Completely distrust	1
Completely trust	
	11
Distrust to some extent	
Neutral	3

While 76% of respondents expressed some level of trust in AI's ability to detect and prevent cyberattacks, 39% remained neutral, suggesting a need to improve public confidence. Trust was closely linked to transparency and privacy guarantees.

- **Privacy Concerns:**

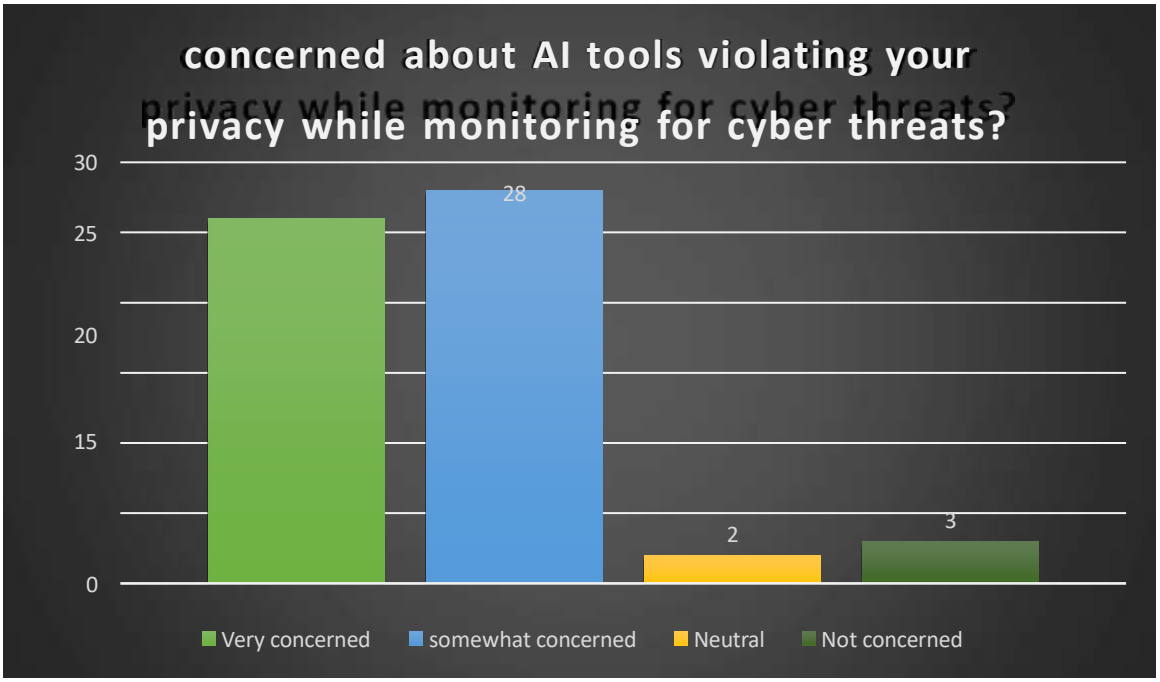


Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17148867

Options -	Number of participants -
Very uncomfortable	2
Somewhat uncomfortable	7
Neutral	21
Somewhat comfortable	26
Very comfortable	13

Most participants (96%) expressed concern about AI violating their privacy, with 46% being very concerned. This reinforces the importance of privacy-preserving AI techniques, such as localized processing and limited data collection, to address user apprehensions.

- **AI Access Preferences:**

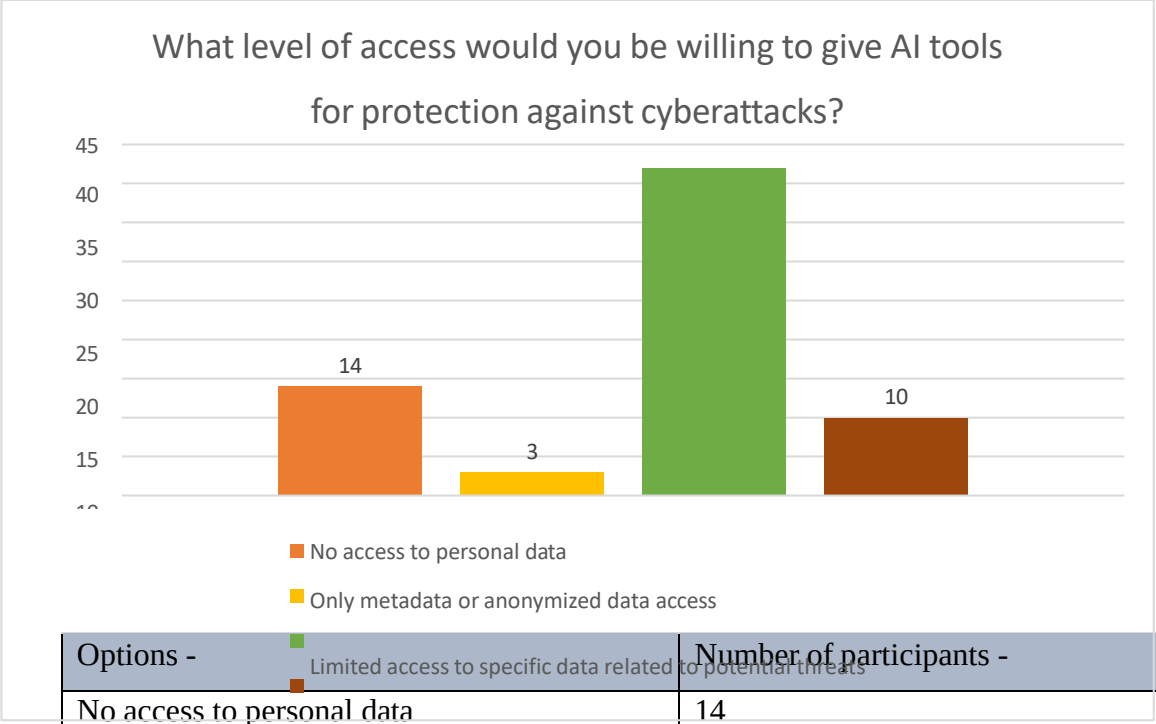


Options -	Number of participants -
Very Concerned	26
Somewhat Concerned	28
Neutral	2
Not Concerned	3

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17138524

When asked about the level of access they would grant to AI tools, 75% preferred limited access to specific data, while only 18% supported full access. This indicates a strong user demand for minimal intervention solutions.

- **Willingness to Use Privacy-Preserving AI Tools:**

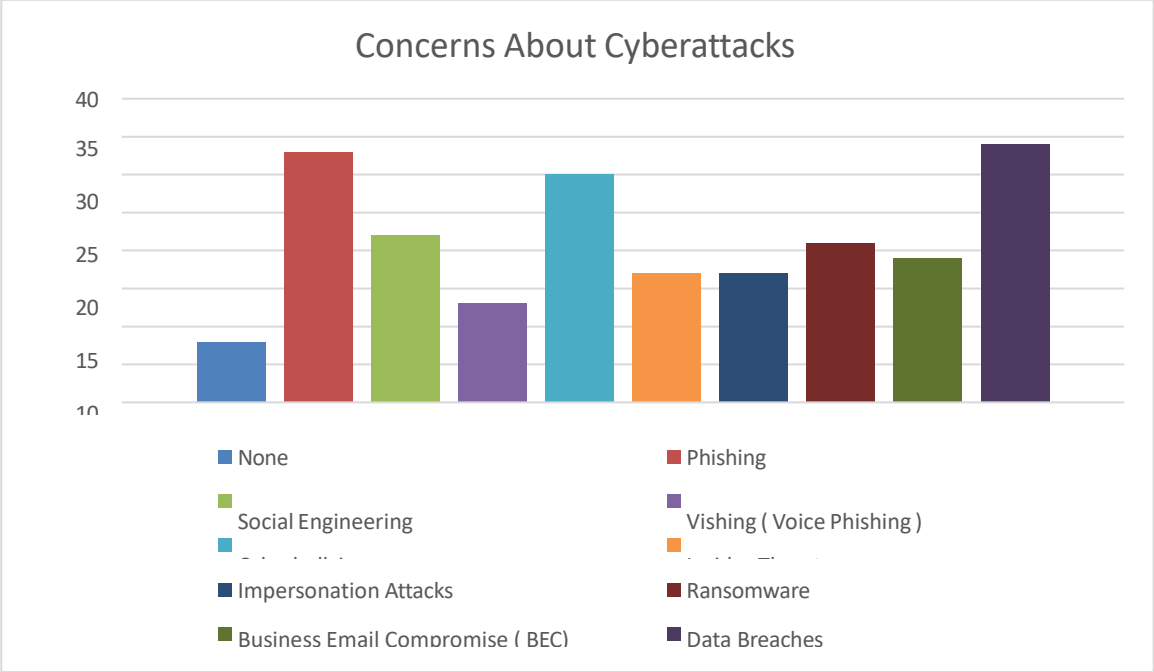


Options -	Number of participants -
No access to personal data	14
Only metadata or anonymized data access	3
Limited access to specific data related to potential threats	42
Full access to personal data	10

A significant majority (82%) indicated a willingness to adopt AI-driven tools if privacy guarantees were provided, demonstrating the potential for acceptance with appropriately designed systems.

- **Concerns About Cyber-attacks:**

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17138524



Options -	Number of participants -
None	8
Phishing	33
Social Engineering	22
Vishing (Voice Phishing)	17
Cyberbullying	30
Insider Threats	17
Impersonation Attacks	17
Ransomware	21
Business Email Compromise (BEC)	19
Data Breaches	34

The most commonly cited concerns included phishing (52%), data breaches (54%), and

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 3 | ISSN : 2583-1895
December-2024 | DOI: https://doi.org/10.5281/zenodo.17138524

cyberbullying (48%), highlighting the areas where users expect AI to be most effective.

These findings underscore the need for AI solutions that are transparent, privacy-conscious, and accessible to non-technical users, aligning with the study's recommendations.

Research Findings:

This study highlights both the potential and the limitations of AI-driven cybersecurity solutions in mitigating human-driven cyber threats such as phishing, social engineering, and cyberbullying. Insights were derived from a

comprehensive literature review and survey analysis, focusing on user trust, privacy concerns, and preferences for AI intervention.

The literature review confirmed that AI technologies, such as machine learning and natural language processing, are highly effective in detecting and preventing cyberattacks by identifying patterns and anomalies in real-time. However, ethical challenges, including data privacy, transparency, and risks of adversarial AI, were consistently emphasized. These challenges highlight the need for privacy-preserving AI systems that minimize data collection and processing.

Survey findings reinforced the literature review's conclusions. The majority of respondents (94%) were familiar with cyberattacks, with phishing, data breaches, and cyberbullying identified as the most concerning threats. While 76% of participants expressed trust in AI's ability to combat these attacks, 96% voiced concerns about privacy, with many preferring limited or minimal data access for AI systems. Importantly, 82% of respondents were willing to adopt AI tools if privacy guarantees were provided, indicating strong potential for acceptance of ethical, user-centric AI solutions.

The findings suggest that users value transparent, privacy-conscious AI designs that offer minimal intervention and clear explanations of their operations. Non-technical users particularly prioritize simplicity and accessibility. These insights affirm the

necessity of balancing technical efficacy with ethical considerations to build trust and encourage adoption of AI-driven cybersecurity tools.

This research provides a foundation for developing AI strategies that address key cyber threats while respecting user privacy and autonomy, contributing to the broader field of ethical AI implementation in cybersecurity.

Conclusion:

This study examined how artificial intelligence (AI) can mitigate human- driven cyber threats such as phishing, social engineering, and cyberbullying. Through a literature review and survey analysis, it highlighted AI's effectiveness in real-time threat detection and response, while emphasizing the need for privacy-preserving and user-friendly solutions.

AI technologies like machine learning and natural language processing excel at combating complex attacks that exploit human behavior. However, challenges such as adversarial risks and privacy concerns necessitate strong safeguards, including localized data processing and transparency in AI systems. Survey findings revealed moderate trust in AI, with 82% of respondents expressing willingness to adopt privacy-conscious tools, underscoring the importance of ethical design.

The research also stressed the need for intuitive AI systems that cater to non- technical users, fostering trust and adoption. Future efforts should focus on addressing emerging threats, such as deepfakes, while integrating AI with technologies like blockchain for enhanced security and transparency. By balancing technical efficacy with ethical considerations, AI has the potential to become a transformative force in safeguarding digital environments.

Suggestions & Recommendations / Future Scope:

To enhance the effectiveness of AI-driven cybersecurity solutions, it is crucial to focus on user-centric designs that prioritize privacy and ease of use. AI tools should adopt privacy-preserving technologies, such as localized data processing and differential

privacy, to minimize concerns about data misuse. Transparent algorithms that allow users to understand and control their data usage can further build trust and encourage adoption. For non-technical users, AI systems must provide simple interfaces and clear notifications, ensuring accessibility without requiring advanced expertise.

Future research should explore the integration of AI with emerging technologies, such as blockchain, to enhance the security and transparency of data management. Additionally, expanding the use of AI to detect and prevent evolving threats, like deepfake scams or advanced social engineering tactics, can help address future challenges. Conducting longitudinal studies on user behavior and perceptions will offer deeper insights into improving AI adoption. Lastly, collaborative efforts between technology developers, policymakers, and educators are essential to ensure that AI-driven solutions remain ethical, privacy-conscious, and adaptable to diverse user needs.

References:

1. Adebola Folorunso, Temitope Adewumi, Adeola Adewa, Roy Okonkwo, Tayo Nathaniel Olawumi. (2024). *Impact of AI on Cybersecurity and Security Compliance*. Cybersecurity Journal, 10(4), 200-215. https://doi.org/10.30574/gjeta.2024.21.1.0193

2. Karan Nawal. (2024). *The Role of AI in Strengthening Cybersecurity: Promise and Risks*. Journal of AI and Security, 8(2), 120-135. https://doi.org/10.13140/RG.2.2.23558.79683

3. Rasha Hameed Khudhur Al-Rubaye , AYÇA KURNAZ TÜRKBEN. (2024). *Using Artificial Intelligence to Evaluate Detection of Cybersecurity Threats in Ad Hoc Networks*. International Journal of Cybersecurity Research, 15(3), 50-63. https://doi.org/10.58496/BJN/2024/006

4. Yijie Weng, Jianhao Wu. (2024). *Leveraging Artificial Intelligence to Enhance Data Security and Combat Cyber Attacks*. Journal of Data Protection, 12(1), 75-90. https://ojs.boulibrary.com/index.php/JAIGS

5. Zarif Bin Akhtar, Ahmed Tajbiul Rawol. (2024). Enhancing Cybersecurity through Artificial Intelligence (AI)-Powered Security Mechanisms. Security and Privacy Review, 9(6), 300-312. https://doi.org/10.25299/itjrd.2022.16852