Human Factors in Cybersecurity, Vol. 91, 2023, 83–92
https://doi.org/10.54941/ahfe1003722

Bringing Humans at the Core of Cybersecurity: Challenges and Future Research Directions

Kitty Kioskli[1,2], Haralambos Mouratidis[1], and Nineta Polemi[3]

[1]University of Essex, School of Computer Science and Electronic Engineering, Institute of Analytics and Data Science (IADS), Essex, United Kingdom
[2]Trustilio B.V., Amsterdam, The Netherlands
[3]Department of Informatics, University of Piraeus, Piraeus, Greece

ABSTRACT

The prompt response to successfully adopt good cybersecurity practices from protecting passwords to security incidents' responding to activating a disaster recovery or a business continuity plan depends upon the level of operators' ability in problem solving, resilience, readiness, maturity, observation, and perception. New technologies, such as Artificial Intelligence (AI) can also be helpful to more effectively forecast or respond to serious incidents, especially to massive attacks. However, the cybersecurity operators need to alter their mindsets, adopt new behavioural patterns, and work attitudes to embrace and interact with AI-assistance during cyber defence activities. in addition, when the operators need to assess or mitigate AI socio-technical risks related to bias, transparency and equality, they will base their decisions for estimating or mitigating these risks on their behavioural, social, cultural, and ethical characteristics. In this paper, we are presenting challenges related to human and psychosocial factors of the cybersecurity operators. We also discuss the motives and drivers that impact the cognitive aspects (e.g., focus on operational tasks, attention, objectivity) of the cyber operations. We further identify how the cybersecurity operators' personality traits impact the success of the cybersecurity practices and estimations and analyse research challenges, regarding the impact of operators' profiles on their perceptions and interactions, with AI cyber defending tools and management of AI risks. Finally, we consider the impact these human factors may have on successful cybersecurity operations and practices and provide proposals for interdisciplinary research directions requiring the collaboration of cybersecurity experts, psychologists, and behavioural scientists.

Keywords: Cybersecurity, Human factors, Cognitive factors, Behavioural analytics

INTRODUCTION

The psychological impact of cyberattacks is of paramount importance as it may have crucial operational consequences on the quality of incident handling. Ransomware attacks for example are designed to stress out the nervous system of the operations and guide them to focus their resources on the ransom demand. As a result, major psychological pressure is being applied on the operators (e.g., incident handlers, risk assessors, security team, crisis management team, CSIRTs) that burdens the successful treatment of an

attack, despite the technological advancements, capabilities, and skills of the operators (Brilingaitė et al., 2020).

The psychological dimension of a cyberattack has proven to also have direct consequences (Gross et al., 2016): Firstly, the infrastructure which has been attacked will have an impact on its reputation leading to fear of damage and shame and preventing good quality of communication regarding the incident. The disrupted communication between the teams and their loss of control leave space for the attackers to new attacks because of the stolen data. Furthermore, extreme measures may be taken by lead executives as firing the security teams bringing feelings of frustration, uncertainty, and loss of balance (Snider et al., 2021). Notably, the occurrence of the cyber incidents may result in long-term impact on the physical and psychological health of the employees. It has been reported that during cyberattacks managers and executives experience elevated levels of stress, fatigue, and sleep problems. If the impact of the attack grows in importance, then long-lasting psychological difficulties and depression appear.

A cyber crisis directly affects a working environment by decreasing its quality and resulting in a hit in trust. When this occurs, a technical remedy does not suffice for full re-establishment, as clients, suppliers, providers, and employees will not trust the IT staff or digital equipment while they may constantly worry about another cyberattack. This indicates that cyber crisis management has to create the appropriate conditions for trust restoration (Dwarakanath et al., 2022).

As the number and frequency of cyber-attacks is increasing, the involvement and guidance of the operators are increasingly requested. Meanwhile, this rise of attacks is draining more human and financial resources, so professionals continue to work at a very fast pace. This becomes evident by the exhausting working conditions in incident and crisis management teams since crisis recovery may take several weeks (Demertzi et al., 2023). The operators' phycological resilience (ability to cope with the stress caused by the attack) plays an important role in the effective incident handling and fight against massive cyber-attacks.

Cybersecurity professionals have usually gained their incident handling and crisis management skills through hands-on experience. Their awareness has been acquired in the field, meaning that they may not have received initial training regarding the major psychological impact of cyberattacks and the required psychological practices and social measures that need to be followed. The literature shows that there are no training or specific courses which would help professionals forge their support and knowledge from a psychological and behavioural side.

Lately, Artificial Intelligence (AI) tools are used by the cybersecurity operators for forecasting, and for more effective management of cybersecurity incidents (in terms of accuracy, reduce time and resources) (Zhang et al., 2022). However, cybersecurity operators need to adopt new behavioural patterns to embrace AI in their cybersecurity operations and retain control of the final decisions.

In this paper, we are presenting challenges related to human factors that impact successful cybersecurity operations and practices, and provide proposals for interdisciplinary research directions that require the collaboration of cybersecurity experts, psychologists, and behavioural scientists. Section 2, analyses the importance of studying the profiling aspects of attackers and operators and the impact and interaction on the cognitive effects of the cybersecurity operations. In Section 3 we discuss the challenge of developing socio-technical cybersecurity scales to measure the severity of the vulnerabilities, the levels of impacts and risks considering human characteristics. In Section 4 the focus turns on behaviour change in AI-based cybersecurity operations. In Section 5 cybersecurity operators' perception of AI-social threats is being considered and discussed. All sections give a set of future directions which may be followed to understand and advance the cognition of cyber operators. Lastly, Section 6 includes the conclusions which are drawn from this piece of work.

PROFILING ASPECTS IMPACT COGNITIVE EFFECTS

Recent research efforts in combating cybercrime study technical and human factors, to determine cybercriminal behaviours, by using a multidisciplinary approach from various scientific domains (e.g., criminology, anthropology, cyberpsychology, behavioural science).

Cyberpsychology, investigative psychology research and behavioural science, have supplied accurate profiling models (Aiken et al., 1996, Common Vulnerability Scoring System, ETSI-TVRA methodology, 2017) for attackers using Fogg's behavioural model (Fogg, 2009; Fogg & Hreha, 2009). Fogg's model (Figure 1) describes that the likelihood of a Behaviour (B) occurring is a product of Motivation (M), Ability (A), and the appropriate Trigger (T).

Based on Fogg's model, an extended attackers' profile was proposed in (Kioskli, & Polemi, 2020; 2022) using five (5) different categories of traits with specific attributes: personality, social- behavioural, technical, motivation and trigger. Profile score was developed, indicating the likelihood of

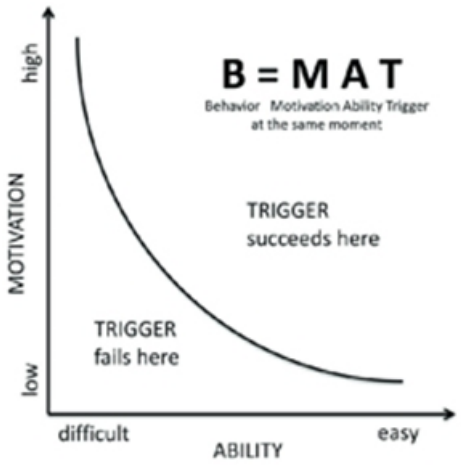

Figure 1: Fogg's model (Fogg, 2009; Fogg & Hreha, 2009).

a person to adopt the behaviour which leads to an attackers' potential (AP) score that reveals the likelihood of carrying out an attack (ETSI-TVRA methodology; MITRE, 2017; NIST, 2012). Fogg's Behavioural Model can be applied to cybersecurity to help individuals and organizations make better decisions about their online security practices. Here's an example of how the model could be applied:

Motivation: People may be motivated to practice good cybersecurity habits because they want to protect their personal information or sensitive data from being compromised. Motivation can also come from a desire to avoid the negative consequences of a security breach, such as financial loss or reputational damage.

Ability: The ability to practice good cybersecurity habits can be influenced by factors such as knowledge, skills, and resources. For example, if someone doesn't know how to create a strong password or isn't aware of the risks associated with using public Wi-Fi networks, they may not have the ability to protect themselves against cyber threats.

Trigger: Triggers can prompt people to take action to improve their cybersecurity practices. Triggers can be internal or external. An internal trigger might be a person's recognition that they are using the same password for multiple online accounts, which could prompt them to take action to create unique passwords for each account. An external trigger might be a reminder from a cybersecurity tool or service to update software or change a password.

In the context of cybersecurity, Fogg's model can help individuals and organizations understand why some people may not be following good cybersecurity practices, and identify ways to motivate, enable, and prompt them to improve their cybersecurity behavior. For example, by providing education and training on cybersecurity best practices, offering tools and resources to make it easier to implement these practices, and using reminders and alerts to prompt individuals to take action, organizations can help improve overall cybersecurity hygiene.

Cyber operator decision-making becomes more realistic when AP scores have been forecasted. Simulation experiments can reveal the different decisions that need to be undertaken considering a variety of possible attackers' profiles.

Cognitive effects to cybersecurity operations depend on the profiles of the operators as well. For example, cognitive factors like focus on cybersecurity operational tasks, increased attention, accurate perceptions, objectivity, setting priorities, making accurate/correct decisions, problem-solving, high concentration, clear thinking, and increased attention span depend on the cybersecurity operators' profiles. However, based on the authors' knowledge, no research has been conducted in this area.

FURTHER RESEARCH

A research challenge proposed here is to identify personality traits that impact the cognitive effects of cybersecurity operators. Analyze and estimate the cybersecurity operators' profiles. Social bond theory can be explored to increase negative incentives towards misbehaviour and create social

bonds that guide cybersecurity operators towards behaviour change. Further interdisciplinary research needs to be conducted on developing a scoring system (e.g., scales, measurements, KPIs) for the operators' profiles that will be validated in specific operational environments (e.g., defense, health, government, financial).

SOCIO-TECHNICAL CYBERSECURITY SCALES AND MEASUREMENTS

Psychological and cognitive assessments provide useful data which contribute toward the understanding of a person's capabilities and characteristics (Aiken et al., 1996, Groth-Marnat et al., 2009, Selzer et al., 1987). This data is collected and interpreted through various methods, such as rating scales and interviews. NIST (National Institute of Standards and Technology, 2012) and MITRE (MITRE ATT&CK) adopt the rating scale approach and suggest a set of attack factors (characteristics) according to their capability, intention, and target to describe an attacker. However, they do not consider psychological and behavioural characteristics of the attackers as potential attack factors. The existing security vulnerability measurement system, CVSS3.1 (Common Vulnerability Scoring System), is a technologically and industrial driven system and does not consider human factors. CVSS3.1 solely presents the assumption that the attacker is highly skilled; attackers' or risk assessors' profiles are not considered in the estimates. The scoring system CVSS3.1 consists of three metric groups:

- Base: Which represents the intrinsic qualities of a vulnerability that are constant over time and across ICT environments. This is the only public metric group.
- Temporal: Which reflects the characteristics of a vulnerability that is being modified over time due to various changes (e.g., new exploits are published).
- Environmental: Which represents the characteristics of a vulnerability that are unique to the ICT environment. The environment group consists of the affected assets and the implemented control on the assets (attackers are not part of the environment). It considers the effectiveness of the controls and the impacts of the vulnerability on the assets.

Each metric group has metrics that the analyst is assigning to values using the CVSS3.1 calculator. The Base metrics produce a score ranging from 0-10; it reflects the objectivity of the technical severity of the vulnerability. By providing values to the Temporal and Environmental metrics, the analyst can then modify the Base score.

The Environmental Metrics group applies to the vulnerability of an asset hosted in a specific environment and used for specific business purposes. This metric group relates to either the business criticality of the asset that is vulnerable, or to compensating controls or mitigations that might make the enterprise susceptible to the vulnerability. Neither attackers nor operators are included in the environmental metrics (only assets and their controls) and profiles are not taken into account.

In (Kioskli & Polemi, 2022), we enlarge the ICT environment to include the attacker as part of the user (ICT) environment and as a consequence the CVSS3.1 environmental score varies according to the attackers' profile score. As a consequence the overall CVSS3.1 score to become more realistic. It was shown that the score of the severity of the vulnerability decreases as the attacker's profile score decreases. How the CVSS3.1 score will vary according to the operators' profiles has not been studied; The question will be how the strength of the operators defenders' profile and its cybersecurity psychological resilience will influence the vulnerability score?

FURTHER RESEARCH

Simulation scenarios need to be developed to collect data on the impact of attackers' profiles and operators' profiles on the CVSS scores in different environments and economic sectors. Not only attackers but also cybersecurity defenders belong to the ICT environments. Additional investigation ought to be conducted, regarding the impact of the profiles of the cybersecurity operators on the CVSS3.1 scoring system. Interdisciplinary research, utilizing cyber psychologists, behavioural analysts, and cyber professionals, is required to develop measurements for the individual environmental values and advance CVSS3.1 to a key scoring system for evaluating accuracy and objectivity (main cognitive factors) in cyber operations.

BEHAVIOUR CHANGE IN AI-BASED CYBERSECURITY OPERATIONS

Effective testing of cognition and perceptions aspects of how cybersecurity operators can use AI tools (e.g., robots) (James, 2023) to improve cybersecurity practices is a challenge. Research on Human AI Interaction (HAI) concentrate only on general technical challenges aspects and guidelines.

HAI in the cybersecurity operations need to be further studied. The efficiency of the decision-making tasks when various HAI factors take place during an incident is one of the issues to be studied. For example the operators, that use AI assistance during the analysis, and recovery of incidents, need to promptly be able to share, assign, distribute and switch different tasks and information. The effectiveness of the operations will depend upon the trust and confidence that the operators have for the teammates (e.g., operators, members of the team, AI assistance). Also the operators' confidence for their skills, such as not feel insecure that their job will be replaced by AI-assistance, play an important role in the effectiveness of the cybersecurity practices. n additional factor is the confidence for the appropriateness and validity of the actions of the AI-assistance.

Behavioural change processes to improve the effectiveness and acceptance of HAI interaction are often neglected even though research has shown that behavioural interventions are useful in meeting long-term goals.

The cause of the avoidance of behavioural change processes lies on the facts that they are considered value-driven and are not easily implemented in models, compared to technological advancements.

FURTHER RESEARCH

Future research should evaluate the various methodologies that exist to assess the efficiency of the HAI teams while utilizing the Team Effectiveness Model (TEM). Organize interventions with the operators to co-assess the HAI efficiency in cybersecurity operations. Additional, developments of measurements would contribute to the quantifications of the four key factors of the United Theory of Acceptance and Use of Technology (UTAUT): effort expectancy, performance expectancy, social influence and facilitating conditions. Also, the development of behaviour change models, mainly focusing on improving the HAI in cybersecurity operations would be proven helpful. While there are various existing models, such as educational, intrinsic, extrinsic, information-processing, and social, there is not a single model successfully translating theory into practice and providing useful tools to achieve HAI behaviour change for cybersecurity operators. Lastly, the design targeted, innovative interventions should be based on new practices such as appealing to emotion (e.g., promoting awareness via virtual reality), and providing social incentives (e.g., awards) to the operators to improve cognitive factors during cybersecurity practices.

CYBERSECURITY OPERATORS PERCEPTION ON AI-SOCIAL THREATS

Any AI system is a socio-technical system with three (3) types of characteristics: technical (e.g., accuracy, robustness, reliability), socio-technical (e.g., explainability, managing bias, transparency, security, privacy), and guiding principles (e.g., accountability, reliability, environmental well-being, diversity, fairness, traceability) (NIST 2022; ISO/ IEC 24368 2022; ISO/IEC TR 24027, 2021; ISO/IEC AWI TS 12791; ISO/IEC TR 24028, 2020). Thus, attacks on AI systems can exploit: technical threats (e.g., loss of accuracy), socio-technical threats (e.g., loss of explainability), and loss of guiding principles (e.g., loss of accountability). Socio-technical and guiding principles related to incidents cannot be uniformly managed since not all cybersecurity operators have the same level of understanding, learning capacities, behaving or perceptive of notions like bias, fairness, equality, and ethics. The incident handling practices that will be adopted will rely upon the operator's profiles, in particular their behavioural social and ethical values. For example, for cybersecurity operators to collectively reach a consensus on how an attack (e.g., poisoning data) impacts fairness or take a decision on mitigation actions will depend on their common perception of fairness.

FURTHER RESEARCH

Organization of behaviour-change interventions is proposed, using co-design approaches to capture cybersecurity operators' values, requirements, needs and comprehensibility levels, in order to encourage technology acceptance. While a detailed examination of cognitive, social, situational, human and affective factors which may influence the interaction which occurs in HAI-cybersecurity teams would be meaningful. Additionally, the development and

application of methodologies, tools, and measures to easily apply UTAUT are proposed. Lastly, future research ought to ensure the efficiency of the HAI among cybersecurity profiles while utilizing the TEM by analyzing and quantifying the main social factors influencing the team's outcomes, such as, collaborative processes and operators' profiles. Meanwhile, advancement of the TEM processes such as decision-making, problem-solving, communication and coordination is suggested.

CONCLUSION

This paper adopts the view that the effectiveness of cybersecurity operations will be advanced by: building bridges between cyber engineers, cyberpsychology researchers, and behavioural and social scientists, to develop advanced holistic socio-technical security management, incident handling techniques, and measurement systems.

Figure 2 outlines the main proposals for future research directions provided in this paper with the objective to bring humans (attackers and operators) in the loop of cybersecurity operations and practices.

Adversaries and cybersecurity professionals (CISOs, engineers, operators, developers, auditors, trainers) are the people that are the heart of cybersecurity.

The effectiveness of cybersecurity practices and operations will be feasible if human factors are considered in their actions, from measuring profiles

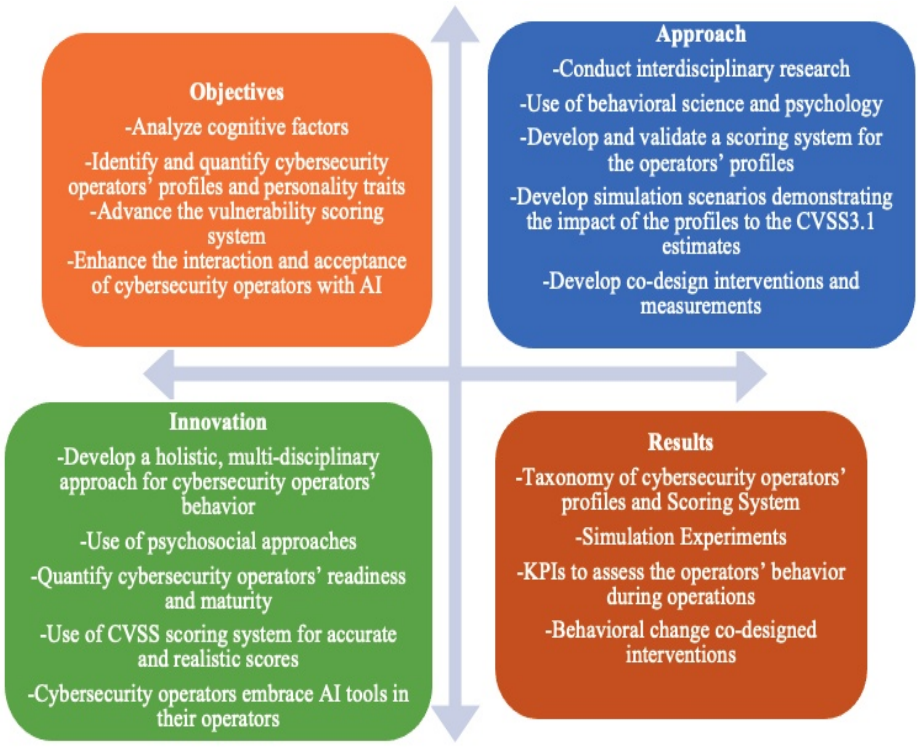

Figure 2: Quad chart.

and risks to managing security incidents to embracing security policies, to implementing controls, to adopting procedures and to auditing security mechanisms.

AKNOWLEDGEMENT

The research conducted in this paper was funded by the project 'A Dynamic and Self-Organized Artificial Swarm Intelligence Solution for Security and Privacy Threats in Healthcare ICT Infrastructures' (AI4HEALTHSEC) under grant agreement No 883273. The project was funded by the European Union's Horizon 2020 research and innovation programme. The authors are also grateful for the financial support provided for the 'Collaborative, Multimodal and Agile Professional Cybersecurity Training Program for a Skilled Workforce In the European Digital Single Market and Industries' (CyberSecPro) project. This project has received funding from the European Union's Digital Europe Programme (DEP) programme under grant agreement No 101083594. Special thanks to the partners of these projects and their contributions. The authors would also like to thank the European Union Agency for Cybersecurity (ENISA), especially Dr Fabio di Franco and the members of the AHWG, for developing the ECSF, where the first author had the honour to serve as rapporteur. The authors would also like to express their gratitude to the University of Piraeus Research Center (UPRC) for its continuous support. Lastly, the views expressed in this paper represent only the views of the authors and not of the European Commission, or the partners in the above-mentioned projects, or University of Piraeus, or UPRC, or University of Essex, or trustilio B. V.

REFERENCES

Aiken LR. (1996). Rating scales and checklists: Evaluating behaviour, personality, and attitudes. Oxford, England: John Wiley and Sons.

Brilingaitė, A., Bukauskas, L. and Juozapavičius, A. (2020) "A framework for competence development and assessment in hybrid cybersecurity exercises," Computers & Security, 88, p. 101607. Available at: https://doi.org/10.1016/j.cose.2019.101607.

Common Vulnerability Scoring System (CVSS) Version 3.1 Calculator (no date). Available at: https://www.first.org/cvss/calculator/3.1.

Demertzi, V., Demertzis, S. and Demertzis, K. (2023) "An overview of cyber threats, attacks and countermeasures on the primary domains of Smart Cities," Applied Sciences, 13(2), p. 790. Available at: https://doi.org/10.3390/app13020790.

Dwarakanath, S., Ravi, K. and Vijayakumar, R. (2022) "A study on the emotions of an employee after a cyber security attack in their organization." Available at: https://doi.org/10.31234/osf.io/gjqe9.

ETSI-TVRA (2017). CYBER: Methods and protocols. Part 1: Method and pro forma for Threat, Vulnerability, Risk Analysis (TVRA) Available at: https://www.etsi.org/deliver/etsi_ts/102100_102199/10216501/05.02.03_60/ts_10216501v050203p.pdf.

Fogg BJ, Hreha J. (2009). Behaviour Wizard: A Method for Matching Target Behaviours with Solutions. In: Ploug T., Hasle P., Oinas-Kukkonen H. (eds) Persuasive Technology. Lecture Notes in Computer Science;vol 6137. Springer, Berlin, Heidelberg.

Fogg BJ. (2009). A behaviour model for persuasive design. In Proceedings of the 4th international Conference on Persuasive Technology p. 40.

Gross, M. L., Canetti, S. and Vashdi, D. R. (2016) "The psychological effects of cyber terrorism," Bulletin of the Atomic Scientists, 72, pp. 1–20. Available at: https://doi.org/10.1080/00963402.2016.1216502.

Groth-Marnat G., (2009). Handbook of psychological assessment. Hoboken, NJ: John Wiley and Sons.

ISO/ IEC 24368, (2022). Artificial Intelligence-overview of ethical and societal concerns. Available at: https://www.iso.org/standard/78507.html.

ISO/IEC AWI TS 12791, (no date) - Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks - Under development.

ISO/IEC TR 24027, (2021). Information technology - Artificial intelligence (AI) - Bias in AI systems and AI aided decision making. Available at: https://www.iso.org/standard/77607.html.

ISO/IEC TR 24028 (2020). Information technology - Artificial intelligence - Overview of trustworthiness in artificial intelligence. Available at: https://www.iso.org/standard/77608.html.

James, K. (2023) AI tools and robots for cybersecurity operations. Available at: https://cybersecurityforme.com/artificial-intelligence-for-cybersecurity/.

Kioskli K., Polemi N. (2020). A socio-technical approach to cyber risk assessment. International Journal of Electrical and Computer Engineering. 14(10): pp. 305–309 (received the best paper award).

Kioskli K., Polemi N. (2020). Measuring psychosocial and behavioural factors improves attack potential estimates. In Proceedings of the 15th International Conference for Internet Technology and Secured Transactions, pp. 216–219.

Kioskli K., Polemi N. (2020). Psychosocial approach to cyber threat intelligence. International Journal of Chaotic Computing, 7 (1): pp. 159-165.

Kioskli K., Polemi N. (2022). Estimating attackers' profiles results in more realistic vulnerability severity scores. Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE2022)/Track: 'Human Factors in Cybersecurity', New York, New York, USA, 53 (1): pp. 138–150. Springer, Elsevier, CRC.

MITRE Adversarial Tactics, Techniques, and Common Knowledge (no date). Available at: https://attack.mitre.org.

NIST, (2012). Attackers' Profiles. Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf.

NIST, 2022 AI RMF Available at: https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.

Selzer MA., Kernberg P, Fibel B, et al. (1987). The personality assessment interview: Preliminary Report. Psychiatry, 50: pp. 142–152.

Snider, K. L. et al. (2021) "Cyberattacks, cyber threats, and attitudes toward cybersecurity policies," Journal of Cybersecurity, 7(1). Available at: https://doi.org/10.1093/cybsec/tyab019.

Zhang, Z. et al. (2022) "Explainable artificial intelligence applications in cyber security: State-of-the-art in research," IEEE Access, 10, pp. 93104–93139. Available at: https://doi.org/10.1109/access.2022.3204051.