Human-AI Collaboration for Sustainable Security: Opportunities and Challenges

Wanling Cai
Lero@Trinity College Dublin

Liliana Pasquale
Lero@University College Dublin

Kushal Ramkumar
Lero@University College Dublin

John McCarthy
Lero@University College Cork

Bashar Nuseibeh
Lero@University of Limerick, The Open University

Gavin Doherty
Lero@Trinity College Dublin

Abstract

In the face of increasingly complex and diverse security threats, the advancement of artificial intelligence (AI) brings significant benefits for ensuring cybersecurity, such as AI-powered solutions to identify anomalies and potential threats. However, human input remains critical at different stages of safeguarding systems against threats. Achieving sustainable security requires combining automation and human expertise, the design of which requires collaboration between multiple stakeholders, e.g., software developers, security professionals, and end users. In this poster, we discuss the potential of human-AI collaboration for establishing a resilient and sustainable security ecosystem, as well as the opportunities and challenges for future research on using AI to empower stakeholders to implement sustainable security practices.

1 Introduction

As digital technology continues to advance, the proliferation of cyber-threats and attacks targeting individuals or organizations also increases in both quantity and complexity [45]. To tackle this challenge, AI-powered autonomous and adaptive security approaches, such as self-protecting software systems [36, 44], and autonomous threat detection [21, 28], have been proposed to enhance the security of systems with minimum human interaction. However, to establish a resilient and sustainable security ecosystem, we rely on different stakeholders to contribute their efforts at different stages of securing systems [40]. For example, software developers are expected to secure the development of software [23], and end users should perform security-critical functions (e.g., encryption) [8, 10]. Instead of viewing "human as problem" [1, 15, 35], it is suggested to consider "human as solution" [45] and use AI to augment human capability in implementing security measures and improve resilience [18, 40, 45].

In recent years, HCI researchers have studied human-AI collaboration (i.e., the use of AI to empower people) [16, 32, 38] in different domains, e.g., data science [41], user experience evaluation [13], and qualitative analysis [20], suggesting the potential to increase people's productivity in their tasks. Yet, little research has investigated the collaboration between humans and AI to maintain sustainable security. Inspired by previous human-AI research, in this poster, we review what we know about sustainable security, and discuss the potential of human-AI collaboration for ensuring sustainable security, as well as the opportunities and challenges of empowering different stakeholders in implementing security practices.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2023.
August 6–8, 2023, Anaheim, CA, USA

2 Sustainable Security

In the digital age, sustainable security refers to thinking long-term about digital space, which not only focuses on the development and maintenance of secure systems or software that can adapt to and mitigate evolving threats over a sustained period [2, 33] but also necessitates individual awareness to safeguard online security. Achieving sustainable security requires collaborative efforts from diverse stakeholders to implement security measures. For example, during the system development phase, software developers are expected to integrate security within the development lifecycle and implement correct security controls that address the root cause of existing vulnerabilities (*secure development*) [4, 23]. At run time, end users are expected to behave securely such as using a secure Wi-Fi connection and making a secure mobile payment (*secure interaction*) [11, 12, 39, 45]. Moreover, maintaining system security requires security professionals to detect potential threats, respond to incidents, and protect critical assets and data from potential breaches (*incident response*) [37].

3 Human-AI Collaboration for Sustainable Security: Opportunities and Challenges

Recent advancements in AI show improved capabilities in language generation [14] and cyber-threat detection [3], demonstrating the potential to collaborate with and support people in their decision-making, and performance of tasks, in order to

maintain sustainable security. Drawing on previous research on human-AI collaboration [13, 38, 41], we believe *AI can act as a collaborator* to augment multiple stakeholders' efforts in safeguarding systems against threats at different stages, including secure development, secure interaction, and incident response. *AI can also act as an educator* to provide educational training and guide stakeholders to implement sustainable security practices, so as to create a resilient and sustainable security ecosystem.

3.1 Collaboration for Secure Development

Opportunities: Recent advancements in AI such as ChatGPT show the potential in collaborating with software developers in their secure development process, such as assessing and regenerating more secure code [5, 25] and prioritizing test cases for software testing [24]. To enable effective collaboration between developers and AI, it is crucial for developers to understand AI's capabilities [26] and build trust in AI [27, 32]. To help developers establish appropriate trust and reliance, AI should be transparent and understandable by delivering the suggestion (e.g., generated codes or prioritized test cases) together with explanations [13](e.g., behavior description like how they generate the code), for which explainable AI (XAI) approaches can be used [29, 42]. In addition, the design of learner-centric XAI could also be beneficial in the collaboration process to support inexperienced software developers' learning of secure development practices [22, 31].
Challenge: Inappropriate reliance on AI (e.g., excessively depending on AI without considering its limitations or potential risks) can negatively impact task performance and even diminish the competence of human developers (e.g., coding skills). How can we encourage human developers to think and code when interacting with automated code generation?

3.2 Collaboration for Secure Interaction

Opportunities: AI shows its potential in assisting users in achieving secure interaction with their devices by monitoring user activities and contextual information to identify potential security risks (e.g., phishing attacks) [3] and taking proactive security measures, such as adopting security updates or sending security notifications to alert users to perform actions [43]. However, in security-critical contexts, while a higher level of system automation may avoid certain insecure behaviour [11], it may likely endanger human agency (i.e., the feeling of control in their situations) [19], affecting users' trust and acceptance of the system [6]. Thus, adaptation characteristics could be considered in the design of AI; for instance, the level of automation and human control can adapt to users' context and needs. Moreover, to enable the sustainability of secure interaction, the security should be visible and designed for engaging users in explicit security mechanisms [11]. Here, conversational AI systems [34] may take the role of guiding users in adopting security measures, and also offer adaptive learning experiences, which might help increase end users' security awareness and stimulate their secure behavior.
Challenges: Previous usable security research indicates that security is a secondary task for most end-users [12, 43]. This may lead to questions on how can we engage users, especially those with little security knowledge, in their interaction with AI, and how can we ensure users maintain their awareness of security after the interaction.

3.3 Collaboration for Incident Response

Opportunities: Given the computational power and analysis abilities of AI, it can assist security professionals like security analysts in the process of incident response, e.g., by performing analysis of system logs and identifying anomalies that may indicate potential security threats, and responding to the identified threats [7]. This may help human analysts quickly detect suspicious activities, and also allow them to focus on other challenging issues that require subject matter experts (e.g., identifying new threats and adapting security requirements) [37]. In this collaborative process, effective communication of the analysis process and results between human analysts and AI will be the key to collaborative work performance. Well-established models of human-human communication may be leveraged and adapted to human-AI communication [17]. Providing context-aware explanations [30] and real-time feedback [9] in AI may foster communication and allow more human analysts' understanding of AI capabilities and limitations, which may help achieve appropriate trust and avoid over-reliance on AI. Explanations and feedback mechanisms will also be useful for improving novice analysts' skills and enriching their knowledge.
Challenges: Communication is a complex process. Establishing appropriate trust between human analysts and AI in such high-risk and time-critical conditions will be even more complex. It might become necessary to design new approaches to engineer such a collaboration process and make the communication information and process transparent.

4 Conclusion

We believe human capabilities can be augmented by AI collaboration in order to establish a more resilient and sustainable security ecosystem. In this poster, we outline several potential directions for utilizing AI to empower various stakeholders in implementing sustainable security practices at different stages: secure development, secure interaction, and incident response. Based on past research, we emphasize several crucial considerations in the design of AI, such as fostering appropriate levels of trust and preserving human agency. We hope this may stimulate further discussions surrounding human-AI research for maintaining sustainable security.

Acknowledgments

This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094_P2 and co-funded under the European Regional Development Fund through the Southern Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software.

References

[1] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[2] Ross Anderson. Making security sustainable. *Communications of the ACM*, 61(3):24–26, 2018.

[3] Giovanni Apruzzese, Pavel Laskov, Edgardo Montes de Oca, Wissam Mallouli, Luis Brdalo Rapa, Athanasios Vasileios Grammatopoulos, and Fabio Di Franco. The role of machine learning in cybersecurity. *Digital Threats*, 4(1), mar 2023.

[4] Hala Assal and Sonia Chiasson. Security in the software development lifecycle. In *SOUPS@ USENIX Security Symposium*, pages 281–296, 2018.

[5] Brett A Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. Programming is hard-or at least it used to be: Educational opportunities and challenges of ai code generation. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 500–506, 2023.

[6] L Jean Camp. Designing for trust. In *Trust, Reputation, and Security: Theories and Practice: AAMAS 2002 International Workshop, Bologna, Italy, July 15, 2002. Selected and Invited Papers 5*, pages 15–29. Springer, 2003.

[7] Marcello Cinque, Raffaele Della Corte, and Antonio Pecchia. Contextual filtering and prioritization of computer application logs for security situational awareness. *Future Generation Computer Systems*, 111:668–680, 2020.

[8] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, UPSEC'08, USA, 2008. USENIX Association.

[9] Maarten de Laat, Srecko Joksimovic, and Dirk Ifenthaler. Artificial intelligence, real-time feedback and workplace learning analytics to support in situ complex problem-solving: A commentary. *The International Journal of Information and Learning Technology*, 37(5):267–277, 2020.

[10] Verena Distler, Gabriele Lenzini, Carine Lallemand, and Vincent Koenig. The framework of security-enhancing friction: How ux can help users behave more securely. In *New security paradigms workshop 2020*, pages 45–58, 2020.

[11] Verena Distler, Marie-Laure Zollinger, Carine Lallemand, Peter B Roenne, Peter YA Ryan, and Vincent Koenig. Security-visible, yet unseen? In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13, 2019.

[12] Shamal Faily, Lizzie Coles-Kemp, Paul Dunphy, Mike Just, Yoko Akama, and Alexander De Luca. Designing interactive secure system: chi 2013 special interest group. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2469–2472. 2013.

[13] Mingming Fan, Xianyou Yang, TszTung Yu, Q Vera Liao, and Jian Zhao. Human-ai collaboration for ux evaluation: Effects of explanation and synchronization. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–32, 2022.

[14] Albert Gatt and Emiel Krahmer. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170, 2018.

[15] Matthew Green and Matthew Smith. Developers are not the enemy!: The need for usable security apis. *IEEE Security & Privacy*, 14(5):40–46, 2016.

[16] Hongyan Gu, Chunxu Yang, Mohammad Haeri, Jing Wang, Shirley Tang, Wenzhong Yan, Shujin He, Christopher Kazu Williams, Shino Magaki, and Xiang'Anthony' Chen. Augmenting pathologists with navipath: Design and evaluation of a human-ai collaborative navigation system. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2023.

[17] Andrea L Guzman and Seth C Lewis. Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, 22(1):70–86, 2020.

[18] Allyson I Hauptman, Beau G Schelble, Nathan J McNeese, and Kapil Chalil Madathil. Adapt and overcome: Perceptions of adaptive autonomous agents for human-ai teaming. *Computers in Human Behavior*, 138:107451, 2023.

[19] Jeffrey Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019.

[20] Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R Brubaker. Supporting serendipity: Opportunities and challenges for human-ai collaboration in qualitative analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23, 2021.

[21] Anupam Joshi, Barbara Carminati, Rakesh M Verma, Elisa Bertino, Murat Kantarcioglu, Cuneyt Gurcan Akcora, Sagar Samtani, Sudip Mittal, and Maanak Gupta. AI for Security and Security for AI. *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, pages 333–334, 2021.

[22] Anna Kawakami, Luke Guerdan, Yang Cheng, Anita Sun, Alison Hu, Kate Glazko, Nikos Arechiga, Matthew Lee, Scott Carter, Haiyi Zhu, et al. Towards a learner-centered explainable ai: Lessons from the learning sciences. *arXiv preprint arXiv:2212.05588*, 2022.

[23] Raees Khan. Secure software development: a prescriptive framework. *Computer Fraud & Security*, 2011(8):12–20, 2011.

[24] Muhammad Khatibsyarbini, Mohd Adham Isa, Dayang NA Jawawi, and Rooster Tumeng. Test case prioritization approaches in regression testing: A systematic literature review. *Information and Software Technology*, 93:74–93, 2018.

[25] Raphaël Khoury, Anderson R Avila, Jacob Brunelle, and Baba Mamadou Camara. How secure is code generated by chatgpt? *arXiv preprint arXiv:2304.09655*, 2023.

[26] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. " help me help the ai": Understanding how explainability can support human-ai interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023.

[27] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

[28] Jonghoon Lee, Jonghyun Kim, Ikkyun Kim, and Kijun Han. Cyber threat detection based on artificial neural networks using event profiles. *Ieee Access*, 7:165607–165626, 2019.

[29] Q Vera Liao and Kush R Varshney. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*, 2021.

[30] Brian Y Lim, Anind K Dey, and Daniel Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2119–2128, 2009.

[31] Karina Cortiñas Lorenzo and Gavin Doherty). Using learning theories to evolve human-centered xai: Future perspectives and challenges. *ACM CHI Workshop Human-Centered Perspectives in Explainable AI*, 2023.

[32] Kazuo Okamura and Seiji Yamada. Adaptive trust calibration for human-ai collaboration. *Plos one*, 15(2):e0229132, 2020.

[33] Liliana Pasquale, Kushal Ramkumar, Wanling Cai, Gavin Doherty, John McCarthy, and Bashar Nuseibeh. Sustainable adaptive security. *arXiv preprint arXiv:2306.04481*.

[34] Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, et al. Conversational ai: The science behind the alexa prize. *arXiv preprint arXiv:1801.03604*, 2018.

[35] Lena Reinfelder, Robert Landwirth, and Zinaida Benenson. Security managers are not the enemy either. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2019.

[36] Mazeiar Salehie, Liliana Pasquale, Inah Omoronyia, Raian Ali, and Bashar Nuseibeh. Requirements-driven adaptive security: Protecting variable assets at runtime. In *2012 20th IEEE international requirements engineering conference (RE)*, pages 111–120. IEEE, 2012.

[37] Bruce Schneier. The future of incident response. *IEEE Security & Privacy*, 12(5):96–96, 2014.

[38] Isabella Seeber, Eva Bittner, Robert O Briggs, Triparna De Vreede, Gert-Jan De Vreede, Aaron Elkins, Ronald Maier, Alexander B Merz, Sarah Oeste-Reiß, Nils Randrup, et al. Machines as teammates: A research agenda on ai in team collaboration. *Information & management*, 57(2):103174, 2020.

[39] Peter Story, Daniel Smullen, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. From intent to action: Nudging users towards secure mobile payments. In *Proceedings of the 16th Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 379–415, 2020.

[40] Susan M Tisdale. Cybersecurity: Challenges from a systems, complexity, knowledge management and business intelligence perspective. *Issues in Information Systems*, 16(3), 2015.

[41] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–24, 2019.

[42] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. Explainable ai: A brief survey on history, research areas, approaches and challenges. In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*, pages 563–574. Springer, 2019.

[43] Ka-Ping Yee. Aligning security and usability. *IEEE Security & Privacy*, 2(5):48–55, 2004.

[44] Eric Yuan, Naeem Esfahani, and Sam Malek. A systematic survey of self-protecting software systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 8(4):1–41, 2014.

[45] Verena Zimmermann and Karen Renaud. Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies*, 131:169–187, 2019.