Information Systems Frontiers (2023) 25:425–429
https://doi.org/10.1007/s10796-023-10372-y

Secure Knowledge Management and Cybersecurity in the Era of Artificial Intelligence

Sagar Samtani[1] · Ziming Zhao[2] · Ram Krishnan[3]

Published online: 18 February 2023

1 Introduction

Over the past half-decade, numerous federal funding agencies such as the National Science Foundation (NSF), National Academies of Science (NAS), and National Security and Technology Council (NSTC) have noted the significant role that Artificial Intelligence (AI)-enabled analytics techniques such as deep learning, machine learning, network science, generative models, reinforcement learning, text analytics, and other techniques could play for Secure Knowledge Management (SKM), and more broadly, cybersecurity. Despite significant interest in the subject, how to design, develop, and evaluate AI-enabled analytics techniques to execute fundamental SKM and cybersecurity tasks, including the systematic process of gathering, synthesizing, organizing data in a secure fashion, malware analysis, incident report generation, password management, risk management, and many other application areas.

In October 2021, the Secure Knowledge Management (SKM) Conference convened many of scholars and practitioners to discuss and share ideas about the use and development of AI for SKM and cybersecurity. Based on the successes of the conference, we launched a special issue to further solidify and attract additional contributions to the rapidly growing and exciting area of research with significant practical impact. In this editorial, we provide a brief background on the role that AI can play in SKM and cybersecurity. We also summarize the papers that were submitted to this special issue and the specific approaches and contributions each accepted paper made to our broader understanding of SKM and cybersecurity.

2 Background on Artificial Intelligence for Cybersecurity and Secure Knowledge Management (SKM)

Recent academic literature has clearly summarized that there are several key areas of cybersecurity and SKM that could be significantly enhanced through the development of novel AI-enabled analytics techniques (Bertino et al., 2021; S. Samtani et al. 2020a; S. Samtani et al. 2020b, 2021, 2022). In summary, the major themes of work include (1) Cyber Threat Intelligence that focuses on identifying emerging threats and key threat actors to help enable effective cybersecurity decision making processes, (2) Disinformation and Computational Propaganda that seeks to identify how fake and/or misleading content proliferates through cyberspace, (3) Security Operations Centers that aims to produce operational cybersecurity capabilities for many organizations and encompasses tasks such as vulnerability management, password management, and others, and (4) Adversarial Machine Learning to Robustify Cyber-Defenses that aims to leverage techniques such as reinforcement learning or generative modeling to synthesize new attack vectors that could evade existing cyber-defenses. Since past research papers and editorials have already described each of these themes in depth, we do not repeat the specific tasks or techniques relevant to each theme. Instead, we point to some promising areas of recent AI development that could have significant benefits for SKM and cybersecurity.

✉ Sagar Samtani
ssamtani@iu.edu

Ziming Zhao
zimingzh@buffalo.edu

Ram Krishnan
Ram.Krishnan@utsa.edu

[1] Department of Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, IN, USA

[2] Department of Computer Science and Engineering, School of Computer Science and Applied Sciences, University at Buffalo, Buffalo, NY, USA

[3] Department of Electrical and Computer Engineering, Klesse College of Engineering and Integrated Design, University of Texas, San Antonio, TX, USA

- *Emerging data sources:* To date, substantial emphasis has been placed on using AI-enabled analytics for malware or netflow data. However, there remains significant opportunities to leverage emerging cybersecurity data sources, especially those that are outside of the confines of any one particular organization. Examples include Dark Web data (e.g., hacker forums, DarkNet markets, carding shops, paste sites), open-source coding repositories (e.g., GitHub, GitLab), social media platforms (e.g., Twitter), Internet of Things (IoT) search engines (e.g., Shodan, Censys, Fofsa, ZoomEye), and others. Each data source can help complete a 360-degree view of an organization's external threat landscape.
- *Generative modeling:* The recent years of AI development have been dominated by the successes of reinforcement learning (RL), generative adversarial networks (GANs), and Large Language Models (LLMs; e.g., ChatGPT). Though the successes of these approaches have been largely outside of the context of cybersecurity, these techniques have some applicability to key SKM and cybersecurity tasks. For example, RL and GANs can play an essential role in automatically synthesizing new phishing, honeywords, and other attack vectors to help robustify existing cyber-defenses. Such techniques can also be valuable for developing multi-agent systems to enhance network defenses. LLMs can help to generate incident response reports automatically, or CTI reports for Information Sharing and Analysis Organizations (ISAOs) to help reduce the efforts of CTI and SOC analysts.
- *Multi-modal analytics:* Often due to the lack of data, many AI algorithms are often trained on a single modality of data. However, organizations often have diverse data sources that human analysts need to consider when developing SKM practices carefully. Multi-modal analytics techniques such as multi-view learning and multi-task learning can help to combine multiple data sources together to produce unified representations to support decision-making processes (Cao et al., 2021; Li et al., 2019; Zhang & Yang, 2022). Integrating attention mechanisms into these architectures can help models favor or weight specific categories of data more strongly than others during the learning process or maintain model robustness over time (Vaswani et al., 2017).
- *Self-Supervised Learning (SSL):* Establishing and maintaining gold-standard datasets to train supervised models is often one of the most time-consuming and labor-intensive tasks in SKM. SSL is an emerging paradigm of machine learning research that seeks to automatically process unlabeled data through (1) pretext classification task to attain pseudo-labels for data points to help initialize a model's parameters and (2) leveraging the model for downstream analytics tasks. SSL hold significant promise for SKM and cybersecurity tasks such as anomaly detection, vulnerability detection, bitcoin fraud detection, and others, as many datasets are difficult to manually label and are rapidly evolving.

3 Papers in this Special Issue

Papers for this special issue were primarily drawn from the 2021 Secure Knowledge Management Conference, which was held virtually (originally scheduled for San Antonio, TX) due to the Covid-19 pandemic. In summary, 11 papers were accepted to the SKM conference. 13 total submissions were submitted to the special issue, with seven of the submissions coming from the SKM conference (papers significantly extended). Although each paper was urged to have a component of AI in their work, this was not an explicit requirement. Thus, papers more broadly studying SKM or cybersecurity were also considered. Each submitted paper received at least one round of peer review. Ten papers were ultimately accepted for this Special Issue. The accepted papers for the special issue could be broadly categorized into four major themes: (1) Cyber-Training and Cybersecurity Education, (2) Social Media Analytics for Disinformation Detection and Topic Detection, (3) Permission and Password Management, and (4) Enhancing Model and Cybersecurity Control Robustness. Papers in each theme had a component of AI that was based in one or more of the major themes of extant AI-enabled cybersecurity and SKM research summarized in the previous section. We provide a summary of the papers accepted to each major theme in the following sub-sections.

3.1 Cyber-Training and Cybersecurity Education

Despite the best efforts of academia, industry, and government, the cybersecurity workforce faces a significant shortfall. Moreover, significant efforts are required to define specific cybersecurity roles more carefully for the next generation of cyber threats. The papers in this theme seek to help address these issues.

- In their paper entitled "RQ Labs: A Cybersecurity Workforce Skills Development Framework," Daniel et al. (2023) report their efforts to develop a novel training program framework designed to keep up with the dynamic demands of the cybersecurity workforce. Specifically, the authors discussed their relationship with an acclaimed cybersecurity organization, ReliaQuest, to develop a series of labs to help develop a workforce well-trained in various aspects of cybersecurity, particularly in security operations centers.
- In their paper entitled "The Application of Role-Based Framework in Preventing Internal Identity Theft

Related Crimes: A Qualitative Case Study of UK Retail Companies," Okeke and Eiza (2023) sought to use semi-structured interviews underpinned with a Role-Based Framework to identify the vagueness of roles related to data security responsibilities that could ultimately cause Internal Identity Theft Related Crimes (IIDTRC). This work has implications for policymakers and education programs seeking to mitigate IIDTRC.

3.2 Social Media Analytics for Disinformation Detection and Topic Detection

Papers in this theme sought to leverage computational approaches to mine the rich, yet largely unstructured data in social media platforms to derive insights for various social media analytics applications. Papers in this theme largely focused on detecting and/or categorizing disinformation, fake news, or other content on social media platforms with automated AI-enabled analytics methods.

- In their paper entitled "A Theory-based Deep-Learning Approach to Detecting Disinformation in Financial Social Media," Chung et al. (2023) sought to detect disinformation in social media through a validated deep learning approach entitled TRNN that is based on social and psychological theories and leverages a series of temporal and contextual information and multiple series of Long Short-Term Memory units to identify disinformation in over 745 K financial social media messages for four US high-tech company stocks.
- In their paper entitled "Do Fake News in Different Languages Tell the Same Story? An Analysis of Multi-level Thematic and Emotional Characteristics of News about COVID-19," Zhou et al. (2023) sought to explore the role of capturing thematic and emotional characteristics of fake news at different levels. The proposed topic modeling approach introduces a divergence measure design to ascertain the importance of thematic characteristics for fake news detection in multiple languages.
- In their paper entitled "Heterogeneous Information Fusion based Topic Detection from Social Media Data," Rani and Kumar (2023) developed a topic detection framework with a transformer-based approach for topic modeling and topic-based video retrieval that specifically leverages textual metadata to find web video topics.

3.3 Permission and Password Management

In this theme, papers sought to help tackle the age-old SKM problem of managing permissions and passwords.

- In their paper "DyPolDroid: Protecting Against Permission-Abuse Attacks in Android," Rubio-Medrano et al. (2023) present a semi-automated security framework that allows for users and administrators to design and enforce a user-friendly abstraction (counter-policies) to restrict the set of permissions granted to malicious applications. The source code of their framework is publicly accessible and open-source.
- In their paper "Password and Passphrase Guessing with Recurrent Neural Networks," Nosenko et al. (2023) developed a rule-based approach that delegated rule derivation, classification, and prediction to a Recurrent Neural Network to guess passwords in a dataset containing 28.8 million users and their 61.5 million passwords. The predictions of their approach can succeed in under 5,000 attempts, a 100% improvement over existing algorithms.

3.4 Enhancing Model and Cybersecurity Control Robustness

An increasing body of research is seeking to help robustify cyber-defenses and controls by leveraging deep reinforcement learning, adversarial learning, and automated perturbations. Research in this theme seeks to make contributions to these areas.

- In their paper entitled "Towards Adversarially Superior Malware Detection Models: An Adversary-Aware Proactive Approach using Adversarial Attacks and Defenses," Rathore et al. (2023) proposed two false negative evasion attacks to help expose vulnerabilities in android malware detection models. The proposed work also defined two defense strategies, Adversarial Retraining and Correlation Distillation Retraining, to help protect against adversarial attacks.
- In their paper entitled "Deep Reinforcement Learning in the Advanced Cybersecurity Threat Detection and Protection," Sewak et al. (2023) reviewed the state of the art in leveraging deep reinforcement learning for cybersecurity threat detection and protection. The paper offers valuable insights into the newer generation of reinforcement learning techniques for various cybersecurity applications, including hybrid edge-cloud detection, spoofing attacks in 5G networks, software-defined networks, and more.
- In their paper entitled "Visualizing Convolutional Neural Network Models' Sensitivity to Nonnatural Data Order," Klepetko and Krishnan (2023) investigate the issue of how to best order training data for input into the prevailing convolutional neural network model to help enhance malware infection analysis performance. The paper presents a novel functional algorithm for order importance that is model-dependent and visualization tools to enhance their analysis.

Acknowledgements We would like to thank the Co-Editors-in-Chief and the Information Systems Frontiers (ISF) editorial offices for their support throughout the process. We are appreciative to all the authors who submitted their work to this special issue at ISF. Special thanks to Dr. Bhavani Thuraisingham from the University of Texas, Dallas, for her invaluable support and guidance throughout the special issue process. We are deeply indebted to the reviewers who took the time to provide constructive reviews for the submitted papers. This material is based in part upon work supported by the National Science Foundation under grant #2133980 and under Grant Numbers DGE-1946537 (IU SFS), OAC-1917117 (CICI), CNS-1850362 (CRII SaTC), and DGE-2038483 (SaTC-EDU).

Data Availability The data for this paper is available if needed.

References

Bertino, E., Kantarcioglu, M., Akcora, C. G., Samtani, S., Mittal, S., & Gupta, M. (2021). AI for Security and Security for AI. *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy* (pp. 333–334). Association for Computing Machinery. https://doi.org/10.1145/3422337.3450357

Cao, G., Iosifidis, A., Gabbouj, M., Raghavan, V., & Gottumukkala, R. (2021). Deep multi-view learning to rank. *IEEE Transactions on Knowledge and Data Engineering, 33*(4), 1426–1438. https://doi.org/10.1109/TKDE.2019.2942590

Chung, W., Zhang, Y., & Pan, J. (2023). A theory-based deep-learning approach to detecting disinformation in financial social media. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10327-9

Daniel, C., Mullarkey, M., & Agrawal, M. (2023). RQ labs: A cybersecurity workforce skills development framework. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10332-y

Klepetko, R., Krishnan, R. (2023). Visualizing convolutional neural network models' sensitivity to nonnatural data order. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10330-0

Li, Y., Yang, M., & Zhang, Z. (2019). A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering, 31*(10), 1863–1883. https://doi.org/10.1109/TKDE.2018.2872063

Nosenko, A., Cheng, Y., & Chen, H. (2023). Password and passphrase guessing with recurrent neural networks. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10325-x

Okeke, R. I., Eiza, M. H. (2023). The Application of role-based framework in preventing internal identity theft related crimes: A qualitative case study of UK retail companies. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10326-w

Rani, S., & Kumar, M. (2023). Heterogeneous information fusion based topic detection from social media data. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10334-w

Rathore, H., Samavedhi, A., Sahay, S.K., & Sewak, M. (2023). Towards adversarially superior malware detection models: An adversary aware proactive approach using adversarial attacks and defenses. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10331-z

Rubio-Medrano, C. E., Soundrapandian, P. K. D., Hill, M., Claramunt, L., Baek, J., Geetha S., & Ahn, G.-J. (2023). DyPolDroid: Protecting against permission-abuse attacks in android. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10328-8

Samtani, S., Kantarcioglu, M., & Chen, H. (2020b). Trailblazing the artificial intelligence for cybersecurity discipline: A multi-disciplinary research roadmap. *ACM Transactions on Management Information Systems, 11*(4), 1–19. https://doi.org/10.1145/3430360

Samtani, S., Chen, H., Kantarcioglu, M., & Thuraisingham, B. (2022). Explainable artificial intelligence for cyber threat intelligence (XAI-CTI). *IEEE Transactions on Dependable and Secure Computing, 19*(4), 2149–2150. https://doi.org/10.1109/TDSC.2022.3168187

Samtani, S., Abate, M., Benjamin, V., & Li, W. (2020a). Cybersecurity as an industry: a cyber threat intelligence perspective. In T. J. Holt & A. M. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 135–154). Springer International Publishing. https://doi.org/10.1007/978-3-319-78440-3_8

Samtani, S., Yang, S., & Chen, H. (2021). ACM KDD AI4Cyber: The 1st Workshop on Artificial Intelligence-Enabled Cybersecurity Analytics. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 4153–4154). Association for Computing Machinery. https://doi.org/10.1145/3447548.3469450

Sewak, M., Sahay, S. K. & Rathore, H. (2023). Deep reinforcement learning in the advanced cybersecurity threat detection and protection. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10333-x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Accessed 12 Jan

Zhang, Y., & Yang, Q. (2022). A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering, 34*(12), 5586–5609. https://doi.org/10.1109/TKDE.2021.3070203

Zhou, L., Tao, J. & Zhang, D. (2023). Does fake news in different languages tell the same story? An analysis of multi-level thematic and emotional characteristics of news about COVID-19. *Information Systems Frontiers*, *25*(2). https://doi.org/10.1007/s10796-022-10329-7

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Sagar Samtani is an Assistant Professor and Grant Thornton Scholar in the Department of Operations and Decision Technologies at the Kelley School of Business at Indiana University. Samtani graduated with his Ph.D. from the Artificial Intelligence (AI) Lab in the University of Arizona's Department of Management Information Systems, where he served as an NSF CyberCorps Scholarship-for-Service Fellow. Samtani's research aims to develop AI-enabled analytics based on deep learning, network science, and text mining approaches for Dark Web analytics, vulnerability assessment for advanced cyberinfrastructure, open-source software security, cyber threat intelligence, AI risk management, and mental health applications. Samtani has published over sixty journal, conference, and workshop papers on these topics in leading Information Systems venues such as *MIS QuarterlyInformation Systems ResearchJournal of MIS*, and *ACM Transactions on MIS,* cybersecurity venues such as *ACM TOPS, IEEE TDSC,* IEEE S&P, USENIX, and *Computers and Security*, and machine learning venues such as ACM KDD, IEEE ICDM, *IEEE Intelligent Systems*, and others. He has received over $4 M (in PI and Co-PI roles) in funding from the NSF SaTC, CICI, SFS, and CRII programs, as well as private sources. He currently serves as an Associate Editor for *ACM TMISACM DTRAP*, and *Information and*

Management. He holds leadership positions in leading industry entities, including spots on the CompTIA ISAO Executive Advisory Council and the DEFCON AI Village board of directors. Samtani has won several awards for his scholarly activities, including induction into the NSF/CISA CyberCorps SFS Hall of Fame in 2022, the AIS Early Career Award in 2022, being named as a Top 50 Undergraduate Professor by Poets and Quants in 2022, the ACM SIGMIS Doctoral Dissertation award in 2019, Runner-Up for the INFORMS Nunamaker-Chen Dissertation Award in 2018, and multiple teaching awards for his courses on AI for cybersecurity, CTI, and business analytics. Samtani has been cited in the *Associated PressMiami HeraldFoxScience MagazineAAASThe Penny Hoarder,* and other venues. He is a member of INFORMS, AIS, ACM, IEEE, and INNS.

Ziming Zhao is an Assistant Professor at the Department of Computer Science and Engineering (CSE) and the director of the CyberspACe securiTy and forensIcs lab (CactiLab), University at Buffalo. His current research interests include system and software security, trusted execution environment, formal methods for security, and usable security. His research has been supported by the U.S. National Science Foundation (NSF), the U.S. Department of Defense, the U.S. Air Force Office of Scientific Research, and the U.S. National Centers of Academic Excellence in Cybersecurity. He is a recipient of an NSF CRII Award. His research outcomes have appeared in IEEE S&P, USENIX Security, ACM CCS, NDSS, ACM MobiSys, ACM TISSEC/TOPS, IEEE TDSC, IEEE TIFS, etc. He is also a recipient of best paper awards from USENIX Security 2019, ACM AsiaCCS 2022, and ACM CODASPY 2014. He received the Ph.D. degree in Computer Science from Arizona State University, Tempe, AZ, in 2014.

Ram Krishnan is a Professor of Electrical and Computer Engineering at the University of Texas at San Antonio, where he holds Microsoft President's Endowed Professorship. His research focuses on (a) applying machine learning to strengthen cybersecurity of complex systems and (b) developing novel techniques to address security/privacy concerns in machine learning. He actively works on topics such as using deep learning techniques for runtime malware detection in cloud systems and automating identity and access control administration, security and privacy enhanced machine learning and defending against adversarial attacks in deep neural networks. He is a recipient of NSF CAREER award (2016), the University of Texas System Regents' Outstanding Teaching Award (2015) and the UTSA President's Distinguished Award for Research Achievement (2016). He received his PhD from George Mason University in 2010.