ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #38019

Integrating Artificial Intelligence into Cybersecurity Curriculum: New Perspectives

AHMET ARIS

Ahmet Aris is a Research Assistant Professor in the Department of Electrical and Computer Engineering at Florida International University. He is conducting research in Cyber-Physical Systems Security Lab (CSL) at Florida International University under the supervision of Dr. A. Selcuk Uluagac. He earned both PhD and MSc. in Computer Engineering from the Graduate School of Science, Engineering and Technology at Istanbul Technical University, Turkey. He also worked at Medianova CDN R\&D Center as an R\&D Analyst. In addition, he conducted research in the Networked Embedded Systems (NES) Group at Swedish Institute of Computer Science (SICS) as a visiting researcher. His research interests include IoT Security, Network Security, Web Security, and Malware.

Luis Puche Rondon

Dr. Luis C. Puche Rondon is a graduate of Florida International University and an alumni of the Cyber-Physical Systems Security Lab (CSL). He received his Bachelors in Computer Science in 2016, and a Masters in Cybersecurity in 2017. Luis has ten years of work experience in Smart home integration and solutions. His research interests include the security of smart environments such as smart homes and offices.

Daniel Ortiz

Computer science major focused on artificial intelligence/machine learning and interested in Natural Language Processing applications. Passionate about diversity in the tech industry!

Monique Ross

Assistant Professor, Knight Foundation School of Computing and Information Sciences and STEM Transformation Institute at Florida International University, research interests include broadening participation in computing through the exploration of: 1) race, gender, and identity in the academy and industry; 2) discipline-based education research that informs pedagogical practices that garner interest and retain women and minorities in computer-related fields. She uses her scholarship to challenge the perceptions of who belong in computing.

Mark Finlayson

Dr. Mark A. Finlayson is Eminent Scholar Chaired Associate Professor of Computer Science and Interim Associate Director in the Knight Foundation School of Computing and Information Sciences (KFSCIS) at Florida International University (FIU). His research intersects artificial intelligence, natural language processing, and cognitive science. He directs the FIU KFSCIS Cognition, Narrative, and Culture (Cognac) Laboratory whose members focus on advancing the science of narrative, including: understanding the relationship between cognition, narrative, and culture; developing new methods and techniques for investigating questions related to language and narrative; and endowing machines with the ability to understand and use narratives for a variety of applications. He received his Ph.D. from MIT in computer science in 2012 under the supervision of Professor Patrick H. Winston. He also holds the M.S. in Electrical Engineering from MIT (2001) and B.S. in Electrical Engineering from the University of Michigan, Ann Arbor (1998). Dr. Finlayson served as a research scientist at the MIT Computer Science and Artificial Intelligence Laboratory for 2½ years before coming to FIU. Dr. Finlayson received an NSF CAREER Award in 2018, an IBM Faculty Award in 2019, and was named the Edison Fellow for AI at the U.S. Patent and Trademark Office for 2019–2021. Dr. Finlayson received FIU's University-wide

Faculty Award for Excellence in Research and Creative Activities, and has received university and departmental awards for Service, Teaching, Mentoring, and Research. His work has been funded by NSF, NIH, ONR, DARPA, DHS, and IBM.

Integrating Artificial Intelligence into Cybersecurity Curriculum: A New Perspective

Abstract

As societies rely increasingly on computers for critical functions, the importance of cybersecurity becomes ever more paramount. Even in recent months there have been attacks that halted oil production, disrupted online learning at the height of COVID, and put medical records at risk at prominent hospitals. This constant threat of privacy leaks and infrastructure disruption has led to an increase in the adoption of artificial intelligence (AI) techniques, mainly machine learning (ML), in state-of-the-art cybersecurity approaches. Oftentimes, these techniques are borrowed from other disciplines without context and devoid of the depth of understanding as to why such techniques are best suited to solve the problem at hand. This is largely due to the fact that in many ways cybersecurity curricula have failed to keep up with advances in cybersecurity research and integrating AI and ML into cybersecurity curricula is extremely difficult. To address this gap, we propose a new methodology to integrate AI and ML techniques into cybersecurity education curricula. Our methodology consists of four components: i) Analysis of Literature which aims to understand the prevalence of AI and ML in cybersecurity research, ii) Analysis of Cybersecurity Curriculum that intends to determine the materials already present in the curriculum and the possible intersection points in the curricula for the new AI material, iii) Design of Adaptable Modules that aims to design highly adaptable modules that can be directly used by cybersecurity educators where new AI material can naturally supplement/substitute for concepts or material already present in the cybersecurity curriculum, and iv) Curriculum Level Evaluation that aims to evaluate the effectiveness of the proposed methodology from both student and instructor perspectives. In this paper, we focus on the first component of our methodology - Analysis of Literature and systematically analyze over 5000 papers that were published in the top cybersecurity conferences during the last five years. Our results clearly indicate that more than 78% of the cybersecurity papers mention AI terminology. To determine the prevalence of the use of AI, we randomly selected 300 papers and performed a thorough analysis. Our results show that more than 19% of the papers implement ML techniques. These findings suggest that AI and ML techniques should be considered for future integration into cybersecurity curriculum to better align with advancements in the field.

Keywords: AI, Learning, Machine Learning, Cybersecurity Education.

1 Introduction

The ubiquity of computers in everyday life has led to an increased concern with cybersecurity. For instance, recent cyberattacks (e.g., SolarWinds [1], Colonial Pipeline [2], Log4j [3], REvil

ransomware [4]) have brought the necessity for resilient software infrastructure to the forefront of popular works, curriculum expansion, and policy choices. Upon the occurrence of sophisticated and high-profile cybersecurity incidents, the US Government created initiatives with industry leaders to build stronger cybersecurity practices for the nation [5]. Several government organizations and private sector leaders announced their plans to train and build a skilled cyber workforce. The growing body of efforts in this area has resulted in the expansion of techniques and methods that are borrowed from other computer science sub-disciplines, in particular, artificial intelligence (AI) and even more specifically machine learning (ML).

In recent years, AI and ML techniques have become critical technology for cybersecurity researchers and practitioners. For this reason, it is extremely necessary to integrate AI into cybersecurity curricula to build a skilled future cyber workforce. However, integrating AI into cybersecurity is challenging for many reasons. First, AI and cybersecurity are difficult areas of study and appeal to different types of students. Second, AI and cybersecurity both require significant commitments within a fixed number of credit hours. Third, integrating AI into cybersecurity requires time-consuming coursework and curricula design on an already packed cybersecurity curriculum with several prerequisites. In fact, such packed curricula are common in STEM degrees, and pose challenges whenever new material needs to be integrated, such as AI. Moreover, instructors in AI and cybersecurity are not usually cross-trained. In other words, an expert in cybersecurity rarely has expertise in AI, and vice versa. Unfortunately, this results in a few cross-trained researchers and practitioners in the future cyber workforce.

Like most computing disciplines, cybersecurity relies on various computing sub-disciplines such as networking, systems, and infrastructure. AI techniques have been widely used in a growing body of cybersecurity literature to address the complex challenges in cybersecurity [6–13]. Notably, cybersecurity has relied heavily on ML, a subfield of AI which focuses on extracting patterns and statistical associations from data in an automated fashion, often using highly mathematical techniques over large, numerically quantified datasets. As ML heavily relies on mathematical techniques, it appeals strongly to those who already fit the prevailing social norms in computer science, that is, those who are especially interested in highly technical, abstract, and mathematical topics. Therefore, large portions of the population are off-put by ML who do not fit the prevailing social norms in computer science, such as Hispanics [14].

In order to tackle this unique problem, in this paper, we propose a new methodology to integrate AI into the cybersecurity curriculum. Our methodology consists of four components: 1) Analysis of Literature, 2) Analysis of Cybersecurity Curriculum, 3) Design of Adaptable Modules, and 4) Curriculum Level Evaluation. By analyzing the cybersecurity literature, we aim to understand the prevalence of AI and ML in cybersecurity research. By analyzing the cybersecurity curriculum, we intend to determine the materials already present in the curriculum and the possible intersection points in the curricula for the new AI material. Based on the literature and curricula analysis, we aim to design highly adaptable modules that can be directly used by cybersecurity educators where new AI material can naturally supplement or substitute for concepts or material already present in the cybersecurity curriculum. Lastly, our methodology plans to evaluate the effectiveness of the proposed methodology from both student and instructor perspectives.

In this paper, we focus on the first component of our proposed methodology, namely Analysis of Literature. We build a semi-automated analysis pipeline that helps us to systematically analyze

the cybersecurity literature for the prevalence and distribution of AI and ML in cybersecurity research. Our analysis pipeline aims to achive this through the analysis of over 5000 research papers collected from the last five years of the top cybersecurity conferences (i.e., IEEE S&P, ACM CCS, Usenix Security, NDSS, ACSAC, ESORICS). Our analysis of over 5000 cybersecurity research papers published in the last five years in top cybersecurity conferences shows that more than 78% of papers mention AI terminology where the overwhelming majority of papers contain ML keywords. Our results clearly show that AI and ML are being considered/mentioned/used by the majority of the top research papers in the cybersecurity literature. In order to understand how frequently AI and ML are used in cybersecurity papers, we randomly selected 300 papers out of 5000 and performed a manual analysis. Our manual analysis results show that more than 30% of the papers mention ML techniques and ML is the most popular AI technique utilized in the papers. In addition, more than 19% of the papers implement ML techniques. Our analysis of the cybersecurity literature clearly demonstrates that AI and ML techniques are widely prevalent in cybersecurity research, and researchers heavily rely on these techniques to address the current and future complex cybersecurity problems. For this reason, it is vital to integrate AI and ML techniques in cybersecurity education curricula in an efficient way by considering not only the students that have strong mathematical backgrounds but also minorities who do not fit the prevailing social norms in computer science.

Contributions: The contributions of this work are as follows:

- We propose a new methodology to integrate AI and ML into cybersecurity education curricula.
- We build and openly share[1] a semi-automated analysis pipeline that analyzes the cybersecurity literature for the prevalence of AI and ML techniques.
- We show the prevalence of AI and ML techniques in over 5000 cybersecurity research papers that were published by the top cybersecurity conferences in the last five years.

Organization: The remainder of this paper is structured as follows. We provide the background information and motivation in Section 2. Section 3 explains our new methodology and an example institutional context to apply the methodology. In Section 4, we provide the details of the data collection and analysis steps for the analysis of cybersecurity literature. Section 5 gives the analysis results and our findings. Section 6 provides discussions and future work. Finally, Section 7 concludes the paper.

2 Background and Motivation

2.1 Cybersecurity and AI Education

Cybersecurity Education. Cybersecurity education is challenging and the landscape of cybersecurity education is very broad as there exist several educational frameworks and perspectives. The Joint Task Force on Cybersecurity Education published the curriculum guidelines for four-year institutions teaching post-secondary degree programs in cybersecurity in 2018 [15]. The published guidelines included a set of knowledge areas in cybersecurity

[1] https://gitlab.com/lcpdev/ai-analyzer

curriculum that could be used as a framework (e.g., knowledge units, topics, essentials, and learning outcomes) with industry perspectives. Following these guidelines, ACM Committee for Computing Education in Community Colleges (CCECC) published their curriculum guidelines for two-year associate degree programs in cybersecurity in 2020 [16, 17]. Likewise, the NIST National Initiative for Cybersecurity Education (NICE) proposed the Workforce Framework for Cybersecurity Framework [18] as a reference document to share and describe cybersecurity work in cybersecurity education, training, and workforce development. In addition to the educational frameworks developed by these task forces and institutions, several researchers have actively worked on cybersecurity education and provided different perspectives. These studies include but not limited to the comprehensive survey of Švábenský et al. [19], taxonomy of curricula by Mouheb et al. [20], the recent non-traditional approaches of gamification [21], behavioral aspects [22], multidisciplinary techniques [23], and ethical aspects [24].

AI Education. AI is a challenging topic for beginners to learn due to complex fundamental theories (e.g., machine learning, game theory) [25]. In order to motivate learners and help them learn, researchers proposed several methods to teach AI to students including the cumulative way to teach AI components [26], the use of games [25, 27–29], emotional intelligence [30], and consideration of ethical aspects [31, 32]

Cybersecurity and AI Education. In terms of the studies that consider both cybersecurity and AI education, there exists only one study in the literature. Farahmand [33] shared the initial results of a research project that aims to integrate AI and cybersecurity research into the cybersecurity curriculum. In his study, he developed a module that aims to teach students the difference between causal analysis and traditional correlation analysis using real-world examples from cybersecurity applications. Although promising, this study is somewhat limited when compared to the wide scope of the cybersecurity and AI education field. The lack of research in this field presents an opportunity for interdisciplinary work that considers both cybersecurity and AI education in a unified way.

2.2 Motivation

AI techniques have become a critical technology for cybersecurity researchers and practitioners. Integrating AI into cybersecurity curricula is increasingly necessary to better prepare a future cyber workforce. However, AI and cybersecurity are difficult areas of study, appeal to different types of students, and individually require significant commitments within a fixed number of credit hours. Both AI and cybersecurity are already quite challenging disciplines, requiring curricula already packed with prerequisites and difficult and time-consuming coursework. Such packed curricula are common in STEM degrees and pose challenges whenever new material needs to be integrated. Furthermore, instructors in AI and cybersecurity are not usually cross-trained: an expert in cybersecurity rarely has expertise in AI and vice versa. This deficit is then reflected in the study body, resulting in few cross-trained researchers and practitioners.

Cybersecurity, like most computing disciplines, draws on various computing sub-discipline expertise. It leverages networking, systems, and infrastructure, as well as newly emerging paradigms such as the Internet of Things (IoT). A growing body of cybersecurity literature has demonstrated the use of different AI techniques to address the complex challenges in

cybersecurity. Notably, cybersecurity has relied heavily on ML, a subfield of AI. ML focuses on extracting patterns and statistical associations from data in an automated fashion, often using highly mathematical techniques over large, numerically quantified datasets. While an often necessary integration, ML's heavily mathematical focus is off-putting for large portions of the population; in other words, ML appeals strongly to those who already fit the prevailing social norms in computer science, that is, those who are especially interested in highly technical, abstract, and mathematical topics.

Critically, these two challenges—integrating more material into the same amount of curriculum time, and focusing on heavily mathematical integrations—directly erect further barriers for students. (e.g., Hispanics) [14]. The landscape of cybersecurity education and AI education also shows that these two fields of education have mostly been studied in isolation. However, these two disciplines could be explored as an opportunity to draw on other subdisciplines to advance each other.

In order to tackle this unique problem, in this paper, we perform a holistic analysis of research literature and the cybersecurity curriculum to identify natural insertion points for AI material. Understanding the prevalence of ML techniques in cybersecurity research will enable us to design highly adaptable modules that can be directly used by cybersecurity educators. In these modules, new AI material can naturally supplement or substitute concepts or material already present in the cybersecurity curriculum. This way cybersecurity educators will explore the use of other types of AI such as natural language processing that have a broader appeal to those outside the mainstream of computer science, including minoritized groups.

3 Methodology and An Example Institutional Context

In this section, we firstly explain our proposed methodology to integrate AI into cybersecurity curriculum and its components. Afterwards, we provide an example institutional context to implement our methodology in, namely the cybersecurity currilum of Florida International University.

3.1 Methodology

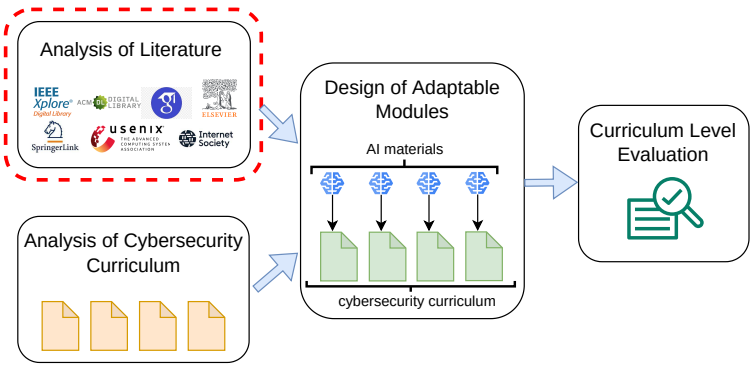


Figure 1: Our proposed methodology to integrate AI into cybersecurity curriculum.

The methodology we use to integrate AI into the cybersecurity curriculum is shown in Figure 1. Our methodology consists of four components: Analysis of Literature, Analysis of Cybersecurity Curriculum, Design of Adaptable Modules, and Curriculum Level Evaluation. By Analysis of Literature, we understand the prevalence and distribution of machine learning techniques in cybersecurity literature. By Analysis of Cybersecurity Curriculum, we determine the materials already present in the curriculum and find out where we can supplement or substitute new AI material to the material already present in the cybersecurity curriculum. By Design of Adaptable Modules, we design modules for the existing cybersecurity curriculum where the determined new AI material will be supplemented or substituted into the existing curriculum. Specifically, we focus on natural language processing (NLP), which is the subfield of AI that focuses on enabling computers to understand and use human language. Importantly, the study of language naturally leads to topics that appeal to social, cultural, and humanistic concerns, which have been shown to have more appeal to minoritized groups. Luring students with this premise provides an opportunity to draw more explicit connections to the impact of cybersecurity globally, but also to their specific communities. We think that NLP can serve as a topical bridge that engages a broader population and garners their interest in the more expansive field of cybersecurity. Finally, by Curriculum Level Evaluation, we perform a rigorous assessment of the effectiveness of the effort both from the student perspective and the instructor perspective. In this paper, we focus on the first component of the methodology, namely Analysis of Literature as highlighted in red in Figure 1. In the following section, we demonstrate the implementation of this component through a cybersecurity education curriculum use-case in Florida International University (FIU).

3.2 An Example Institutional Context

FIU's cybersecurity curriculum begins with foundational subjects of Introduction to Databases (CGS 1540), Programming in Java (COP 2250), Intermediate Java (COP 3804), Computer Operating Systems (CGS 3767), Secure C Programming for Engineers (COP 2270), etc. Then upper-level courses with a strict emphasis on cybersecurity are introduced, Computing and Network Security (CNT 4403), Enterprise Cybersecurity (CIS 4365), Mobile and IoT Security (CNT 4182). Electives include Introduction to Digital Forensics Engineering (EEL 4802), Data Communications (CNT 4513), Introduction Malware Reverse Engineering (EEL 4804). Current enrollment consists of 310 students of which 36% are first-time in college (FTIC) and 63% are transfer students (community college or other four-year colleges). The demographic composition (self-identified) of enrollment is as follows: 1% Asian, 11% Black or African American, 70% Hispanic, 7% Nonresident Alien, 2% not reported, 2% two or more races, and 7% White. The self-reported gender of the students' is 64% male and 16% female.

4 Analysis of Cybersecurity Literature

In order to analyze the literature to understand the prevalence and distribution of machine learning techniques in cybersecurity literature, we collected research papers from various resources and analyzed them in a systematic way. In this subsection, we first explain our data collection process and then provide details of the data analysis process.

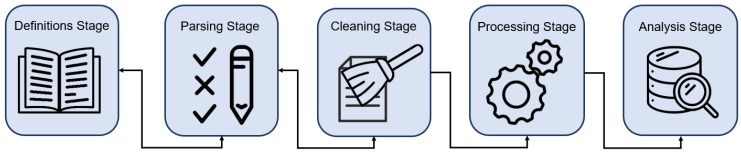


Figure 2: Data analysis pipeline to process and analyze cybersecurity literature for the prevalence and distribution of machine learning techniques.

4.1 Data Collection

We collected over five thousand papers from the largest international peer-reviewed cybersecurity conferences. The papers were collected manually from every single conference, as they are all freely available to universities, with ranges from the years 2016 to 2021 (as the pandemic caused many not to have workshops in 2021). We chose these specific conferences based on relevancy to our study, as well as the quality of accepted submissions. Among these include ACM Computer and Communications Security (CCS), IEEE Symposium on Security and Privacy (S&P), Network and Distributed System Security (NDSS), USENIX Security, Annual Computer Security Applications Conference (ACSAC), Asiacrypt, European Symposium on Research in Computer Security (ESORICS), Eurocrypt, Privacy Enhancing Technologies Symposium (PETS), Public-Key Cryptography (PKC), and Research in Attacks, Intrusions and Defenses.

4.2 Data Analysis

After the collection of papers, we needed to understand the prevalence and distribution of Ml techniques in cybersecurity literature. To accomplish this, we built a semi-automated analysis pipeline that is shown in Figure 2. Our analysis pipeline consists of five stages: Definitions, Parsing, Cleaning, Processing, and Analysis stages respectively. Each stage was refined by input from our subject matter experts - cybersecurity, machine learning, and education. We implemented the pipeline using freely-available, open-source libraries. The following software and libraries were used as part of this project: Python 3.9.0, PDFMiner 4, and JSON Pickle. To support reproducible and open science, we share our implementation openly[2]. In the following paragraphs, we give a detailed explanation of each stage of the data analysis pipeline.

Definitions Stage. This stage aims to build a dataset of AI and machine learning keywords to be searched for in the collected papers. We implemented this stage by gathering a word-bank of keywords associated with AI and Machine learning from various resources such as the keywords of the 2021 AAAI Conference on Artificial Intelligence[3] and the glossary of open ML ebook project[4]. This data bank was organized into three distinct data structures. First, a complete data bank of all the keywords, synonyms, and acronyms. Second, the keywords were organized into lists of synonyms and acronyms. For instance, the list for 'support vector machine' included 'svm' and 'support vector machine' under the same keyword. Third, was a map of topics and the associated keywords with each topic.

[2]We will share our implementation in camera-ready version of the paper.
[3]https://aaai.org/Conferences/AAAI-21/aaai21keywords/
[4]https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html

Parsing Stage. In this stage, PDFs of the collected research papers are converted into plaintext for easier processing and filtering. We implemented this stage using PDFMiner[5] and libraries included in the Python 3.9.0 distribution. Using PDFMiner, individual PDFs were processed and then exported into raw text files. Further, information of each PDF was exported into a separate file including the title, page count, word count, and metadata. The raw text and data were exported using JsonPickle as individual files ending in _rawtext and _raw respectively.

Cleaning Stage. This stage was implemented using standard functions included in the Python distribution. This step was necessary as automated PDF parsing into text is not perfect and some symbols are not processed correctly. First, some strings (e.g., `fi` and `fl`) are not processed correctly in some PDF files. To address this issue, a table was created that converted all the incorrect parsing into the correct strings with replacement functions. Additionally, all strings were converted to their lowercase equivalent. These parsed documents were then exported using Python I/O functions with a filename ending in _cleaned.

Processing Stage. This stage was implemented using standard Python libraries. For the processing stage, all the keywords in the data bank that was built in the Definitions Stage were used. In this processing stage, for each cleaned document, the processing code finds words in the document that match keywords in the data bank. When a match is found, each keyword found is counted and recorded. The map of the number of found keywords was exported along with the page count and word count of each individual document.

Analysis Stage. This stage consists of primary and secondary analysis processes. The primary analysis stage uses standard Python libraries. In the primary analysis, the software first calculates the number of times a specific term was used in all papers. Additionally, this analysis determines the number of papers that have ML terminology and the distribution of each term. Even though the primary analysis can give us clues on the prevalence and distribution of machine learning in cybersecurity literature, it cannot exactly show if machine learning is used in a research paper or only mentioned. In order to tackle this issue, after the primary analysis, we performed a secondary analysis where a random sampling of 300 papers was chosen using a pseudo-random number generator (PRNG). Then a manual analysis of these papers was conducted to determine whether a machine learning technique is used or only mentioned within the paper.

5 Prevalence of ML in Cybersecurity Literature

In this section, we provide the Primary and Secondary Analysis results in understanding the prevalence and distribution of machine learning and AI techniques in cybersecurity literature.

5.1 Primary Analysis Results

In the primary analysis, our semi-automated analysis pipeline analyzed the prevalence of AI and ML terminology in our paper database that consists of 5263 papers. Our results show that 4120 papers (over 78%) out of 5263 contain AI terminology. Our analysis results are depicted in Figure 3. As shown in the figure, AI and ML terminologies are widely prevalent in cybersecurity

[5]https://pypi.org/project/pdfminer/

research and researchers mention or use these techniques to address the complex problems in the cybersecurity domain.

Although our results provide significant findings, they do not exactly show whether cybersecurity researchers are really using AI and ML techniques in their papers or not. For instance, a paper may mention a keyword in reference to other work (e.g., ...in contrast with our approach, Smith et al. used *deep learning*...), may explicitly deny use (e.g., ...we did not use deep learning in our approach...), or may not even refer to the technique at all (e.g., ...we found the problem deep, learning that it required more investigation than originally supposed...). Distinguishing all these different cases from an actual indication of use is in itself a complex natural language processing task beyond the scope of this work. To address this issue, we sampled 300 papers and performed manual analysis.

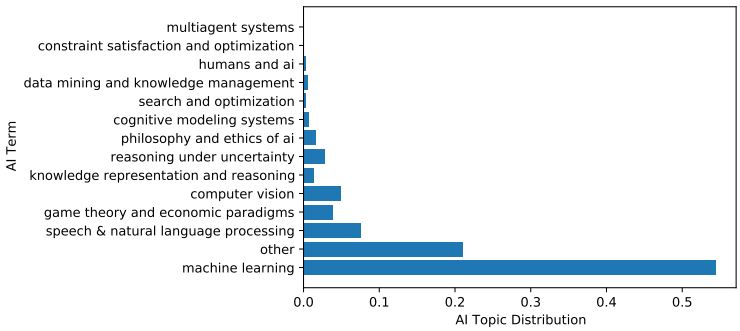


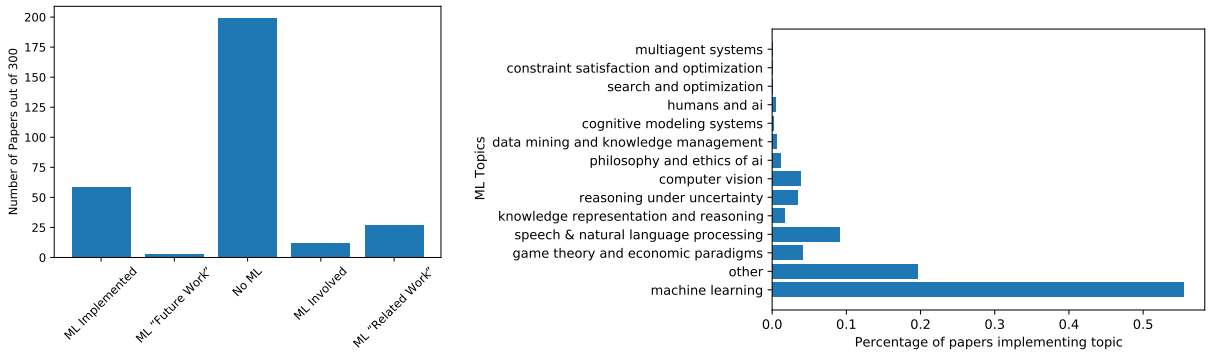
Figure 3: Distribution of ML topics by percentage that contain AI terminology.

5.2 Secondary Analysis Results

Our secondary analysis of 300 papers show that 59 out of 300 random paper samples directly leveraged AI or ML techniques. Our results are shown in Figure 4 (a). In the figure, *ML Implemented* states the number of cybersecurity papers directly implementing AI/ML as part of their research, *ML Future Work* shows the number of papers where ML was part of the future work to be done, *No ML* papers do not use or mention ML/AI techniques, *ML Involved* includes papers where ML was present within the research, but not implemented or was not the focus of research. The criteria for involvement was whether an attack or study vector itself uses ML relevant to the study or not. *ML 'Related Work'* includes papers where the study does not include ML but it mentions ML as its related work uses it. In essence, almost 20% of the analyzed papers implemented ML techniques.

In addition to the prevalence of ML in the manually analyzed papers, we performed an additional analysis for specific categories of topics under the umbrella of AI. Figure 4 (b) displays the distribution by percentage of papers that had significant AI presence. Our findings show that machine learning, game theory, and natural language processing are the primary domains that cybersecurity researchers benefited from while conducting their studies.

As ML was used in a significant portion of papers, further investigation was conducted into specific ML topics. Our results show that Neural Networks dominate the research, representing 19% of all the keywords we assessed. We think that this is not surprising as neural networks are



(a) Distribution of ML content

(b) Distribution of ML topics

Figure 4: Distribution of ML content (a) and ML topics (b) in the randomly chosen 300 papers.

widely used in deep learning and conventional ML techniques due to their accuracies when properly implemented.

6 Discussion & Future Work

One of the most notable challenges of this work was the initial parsing of publications and PDF documents. Before any analysis could be performed on publications, a computer-readable format needed to be created. Automated PDF parsing has been a known development challenge as the PDF format is designed as a human-readable format and not designed for automated parsing. Manual analysis of several samples was required as well as custom code to clean-up invalid characters in some documents. Finally, processing thousands of manuscripts required hours for the parsing to text, cleaning, and further conversion into an analysis-friendly format. Another challenge we faced was regarding the analysis process. When analyzing, we realized determining the appropriate vocabulary for keyword detection was important for both accuracy and precision. This is a problem encountered by all NLP projects, the approach we decided on was maintaining only the specific words that describe either a technique, model, or process in AI. This narrowed down the amount of papers containing AI implementation, instead of papers that only reference an idea not relevant to AI, but grammatically similar.

During the rest of our study, we will continue implementing the rest of our methodology given in Figure 1. Specifically, in the analysis of cybersecurity Curriculum, we will determine the materials already present in the FIU cybersecurity curriculum and find out where we can supplement or substitute new AI material to the material already present in the cybersecurity curriculum. Following that, in the Design of Adaptable Modules, we will design modules for the existing cybersecurity curriculum where the determined new AI material will be supplemented or substituted into the existing curriculum. In this part, we will also explore the use of other types of AI that have a broader appeal to those outside the mainstream of computer science, including minoritized groups. Specifically, we will focus on natural language processing (NLP). Finally, in curriculum level evaluation, we will perform a rigorous assessment of the effectiveness of the effort both from the student perspective and the instructor perspective.

7 Conclusion

In this paper we proposed a new methodology to integrate AI and ML techniques into cybersecurity education curricula. Our methodology consists of four components: i) Analysis of Literature, ii) Analysis of Cybersecurity Curriculum, iii) Design of Adaptable Modules, and iv) Curriculum Level Evaluation. Specifically, we focused on the first component of our methodology - Analysis of Literature, and systematically analyzed over 5000 papers that were published in the top cybersecurity conferences during the last five years. Our results clearly showed that more than 78% of the cybersecurity papers mention AI terminology. To determine the prevalence of the use of AI, we randomly selected 300 papers and performed a secondary analysis. Our results showed that more than 19% of the papers implement ML techniques. These findings suggest that AI and ML techniques should be considered for future integration into cybersecurity curriculum to better align with advancements in the field and to better consider minoritized groups.

Acknowledgements

This work was partially supported by the U.S. National Science Foundation (Award: NSF-CAREER CNS-1453647 and NSF-2039606). The views expressed are those of the authors only, not of the funding agencies.

References

[1] I. Jibilian and K. Canales, "The us is readying sanctions against russia over the solarwinds cyber attack. here's a simple explanation of how the massive hack happened and why it's such a big deal," https://www.businessinsider.com/solarwinds-hack-explained-government-agencies-cyber-security-2020-12, 2021, [Online; accessed 11-February-2022].

[2] S. Kelly and J. Resnick-ault, "One password allowed hackers to disrupt colonial pipeline, ceo tells senators," https://www.reuters.com/business/colonial-pipeline-ceo-tells-senate-cyber-defenses-were-compromised-ahead-hack-2021-06-08/, 2021, [Online; accessed 11-February-2022].

[3] J. Korn, "The log4j security flaw could impact the entire internet. here's what you should know," https://www.cnn.com/2021/12/15/tech/log4j-vulnerability/index.html, 2021, [Online; accessed 11-February-2022].

[4] CNBC, "In private conversation, hackers behind massive ransomware outbreak lower demand to $50 million," https://www.cnbc.com/2021/07/05/revil-hackers-behind-massive-ransomware-outbreak-drop-demand-to-50m.html, 2021, [Online; accessed 11-February-2022].

[5] T. W. House, "Fact sheet: Biden administration and private sector leaders announce ambitious initiatives to bolster the nation's cybersecurity," https://www.whitehouse.gov/briefing-room/statements-releases/2021/08/25/fact-sheet-biden-administration-and-private-sector-leaders-announce-ambitious-initiatives-to-bolster-the-nations-cybersecurity/, 2021, [Online; accessed 11-February-2022].

[6] M. A. N. Abrishamchi, A. H. Abdullah, A. David Cheok, and K. S. Bielawski, "Side channel attacks on smart home systems: A short overview," in *IECON 2017 = 43rd Annual Conference of the IEEE Industrial Electronics Society*, Oct 2017, pp. 8144–8149.

[7] A. Acar, H. Fereidooni, T. Abera, A. K. Sikder, M. Miettinen, H. Aksu, M. Conti, A.-R. Sadeghi, and S. Uluagac, "Peek-a-boo: I see your smart home activities, even encrypted!" in *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2020.

[8] A. I. Newaz, A. K. Sikder, M. A. Rahman, and A. S. Uluagac, "Healthguard: A machine learning-based security framework for smart healthcare systems," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2019.

[9] L. P. Rondon, L. Babun, A. Aris, K. Akkaya, and A. S. Uluagac, "Survey on enterprise internet-of-things systems (e-iot): A security perspective," *Ad Hoc Networks*, vol. 125, p. 102728, 2022.

[10] A. K. Sikder, H. Aksu, and A. S. Uluagac, "6thsense: A context-aware sensor-based attack detector for smart devices," in *26th USENIX Security Symposium, USENIX Security 2017*, 2017.

[11] H. Oz, A. Aris, A. Levi, and A. S. Uluagac, "A survey on ransomware: Evolution, taxonomy, and defense solutions," *ACM Comput. Surv.*, jan 2022, just Accepted.

[12] F. Naseem, A. Aris, L. Babun, E. Tekiner, and A. S. Uluagac, "Minos: A lightweight real-time cryptojacking detection system," in *NDSS*, 2021.

[13] J. Franco, A. Aris, B. Canberk, and A. S. Uluagac, "A survey of honeypots and honeynets for internet of things, industrial internet of things, and cyber-physical systems," *IEEE Communications Surveys Tutorials*, vol. 23, no. 4, pp. 2351–2383, 2021.

[14] A. Q. Gates, H. Thiry, and S. Hug, "Reflections: The computing alliance of hispanic-serving institutions," *ACM Inroads*, vol. 7, no. 4, p. 69–73, nov 2016.

[15] J. T. F. on Cybersecurity Education, *Cybersecurity Curricula 2017: Curriculum Guidelines for Post-Secondary Degree Programs in Cybersecurity*. New York, NY, USA: Association for Computing Machinery, 2018.

[16] C. T. Group, *Cybersecurity Curricular Guidance for Associate-Degree Programs*. New York, NY, USA: Association for Computing Machinery, 2020.

[17] C. Tang, C. Tucker, C. Servin, M. Geissler, and M. Stange, "Curricular guidance for associate-degree cybersecurity programs," in *Proc. of the 51st ACM Tech. Symposium on Computer Science Education*, 2020.

[18] R. Petersen, D. Santos, M. C. Smith, K. Wetzel, and G. A. Witte, "Workforce framework for cybersecurity (nice framework)," 2020.

[19] V. Švábenský, J. Vykopal, and P. Čeleda, "What are cybersecurity education papers about? a systematic literature review of sigcse and iticse conferences," in *Proceedings of The 51st ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '20. New York, NY, USA: ACM, 2020.

[20] D. Mouheb, S. Abbas, and M. Merabti, *Cybersecurity Curriculum Design: A Survey*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, pp. 93–107.

[21] C. Irvine, M. Thompson, and K. Allen, "Cyberciege: gaming for information assurance," *IEEE Security Privacy*, vol. 3, no. 3, pp. 61–64, 2005.

[22] W. Patterson, C. E. Winston, and L. Fleming, "Behavioral cybersecurity: A needed aspect of the security curriculum," in *SoutheastCon 2016*, 2016, pp. 1–7.

[23] M. Lukowiak, S. Radziszowski, J. Vallino, and C. Wood, "Cybersecurity education: Bridging the gap between hardware and software domains," *ACM Trans. Comput. Educ.*, vol. 14, no. 1, mar 2014.

[24] Z. Trabelsi and W. Ibrahim, "A hands-on approach for teaching denial of service attacks: A case study," *Journal of Information Technology Education : Innovations in Practice*, vol. 12, pp. 299–319, 2013.

[25] H. Zhou, H. Zhang, Y. Zhou, X. Wang, and W. Li, "Botzone: An online multi-agent competitive platform for ai education," in *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 2018.

[26] P. Langley, "An integrative framework for artificial intelligence education," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 9670–9677, Jul. 2019.

[27] M. Giannakos, I. Voulgari, S. Papavlasopoulou, Z. Papamitsiou, and G. Yannakakis, *Games for Artificial Intelligence and Machine Learning Education: Review and Perspectives*. Springer, 2020.

[28] D.-M. Yoon and K.-J. Kim, “Challenges and opportunities in game artificial intelligence education using angry birds,” *IEEE Access*, vol. 3, pp. 793–804, 2015.

[29] W. Li, H. Zhou, C. Wang, H. Zhang, X. Hong, Y. Zhou, and Q. Zhang, “Teaching ai algorithms with games including mahjong and fightthelandlord on the botzone online platform,” in *Proceedings of the ACM Conference on Global Computing Education*, 2019.

[30] S. O. Sood, “Emotional computation in artificial intelligence education,” in *AAAI 2008*, 2008.

[31] L. Sijing and W. Lan, “Artificial intelligence education ethical problems and solutions,” in *2018 13th International Conference on Computer Science Education (ICCSE)*, 2018, pp. 1–5.

[32] N. Garrett, N. Beard, and C. Fiesler, *More Than "If Time Allows": The Role of Ethics in AI Education*. New York, NY, USA: Association for Computing Machinery, 2020, p. 272–278.

[33] F. Farahmand, “Integrating cybersecurity and artificial intelligence research in engineering and computer science education,” *IEEE Security Privacy*, vol. 19, no. 6, pp. 104–110, 2021.