Open Access License Notice

AI-Cybersecurity Education Through Designing AI-based Cyberharassment Detection Lab

Ebuka Okpala
School of Computing
Clemson University
Clemson, USA
0000-0002-5816-8194

Nishant Vishwamitra
Information Systems and Cyber Security
University of Texas, San Antonio
San Antonio, USA
0000-0002-3728-1921

Keyan Guo
Computer Science and Engineering
University at Buffalo
Buffalo, USA
0000-0001-9961-2442

Song Liao
Department of Computer Science
Texas Tech University
Lubbock, USA
0000-0002-5264-7573

Long Cheng
School of Computing
Clemson University
Clemson, USA
0000-0003-1736-0873

Hongxin Hu
Computer Science and Engineering
University at Buffalo
Buffalo, USA
0000-0001-8710-247X

Xiaohong Yuan
Department of Computer Science
North Carolina A&T State University
Greensboro, USA
0000-0002-1295-9812

Jeannette Wade
Human Health Sciences
University of North Carolina at Greensboro
Greensboro, USA
0000-0001-9126-5554

Sajad Khorsandroo
Department of Computer Science
North Carolina A&T State University
Greensboro, USA
0000-0003-0649-9247

Abstract—Cyberharassment is a critical, socially relevant cybersecurity problem because of the adverse effects it can have on targeted groups or individuals. While progress has been made in understanding cyberharassment, its detection, attacks on artificial intelligence (AI) based cyberharassment systems, and the social problems in cyberharassment detectors, little has been done in designing experiential learning educational materials that engage students in this emerging social cybersecurity in the era of AI. Experiential learning opportunities are usually provided through capstone projects and engineering design courses in STEM programs such as computer science. While capstone projects are an excellent example of experiential learning, given the interdisciplinary nature of this emerging social cybersecurity problem, it can be challenging to use them to engage non-computing students without prior knowledge of AI. Because of this, we were motivated to develop a hands-on lab platform that provided experiential learning experiences to non-computing students with little or no background knowledge in AI. We also discussed the lessons learned in developing this lab for non-computing students. In this lab used by social science students at North Carolina A&T University across two semesters (spring and fall) in 2022, students are given a detailed lab manual and are to complete a set of well-detailed tasks. Through this process, students learn AI concepts and the application of AI for cyberharassment detection. Using pre-and post-surveys, we asked students to rate their knowledge or skills in AI and their understanding of the concepts learned. The results revealed that the students moderately understood the concepts of AI and cyberharassment.

Keywords—*Experiential Learning, Cybersecurity, Cyberharassment, Machine Learning, Hands-on Labs*

I. INTRODUCTION

With the rise of social media, cyberharassment (e.g., cyberbullying and cyberhate) has become more prevalent in daily interactions [1]. It often involves *inappropriate online behavior and deliberate cyber threats* against individuals (such as teenagers [2]), or specific social groups on the grounds of characteristics such as race, sexual orientation, gender, or religious affiliation [3]. Cyberharassment is identified as a critical socially-relevant cybersecurity problem [4], [5], since it can have significant negative impacts on the safety and emotional well-being of targeted groups, especially teens and minority communities. The Cyberbullying Research Center reported that 37% of middle and high school students have been cyberbullied during their lifetime [6], and this number is expected to further increase as teens continue to have an increased online presence. Cyberharassment can even result in catastrophic consequences of increased suicide among the affected teens who are unable to appropriately get away from the harassment [7]. The shift from traditional text-based cyberharassment to *multimodal* (i.e., both texts and images) [8] cyberharassment poses a challenge to effective cyberharassment detection.

Artificial Intelligence (AI)/Machine Learning (ML) has immense potential to solve this critical problem. Automatic

detection methods for text-based and image-based cyberharassment using AI techniques have emerged [9]. Internet companies such as Facebook and Google have also deployed AI algorithms to detect toxic content on social media [10], [11]. Meanwhile, adversaries may exploit vulnerabilities of AI-based classifiers to evade existing cyberharassment detectors [12]–[14]. There exist *social problems*, such as fairness and ethics, in AI models for cyberharassment detection. For example, some particular demographic groups are unfairly treated by AI-based detectors [15], [16]. Concerns have been raised that the vulnerabilities of AI models and their robustness against attacks are biased towards underrepresented groups [17]. As such, an unfair AI-based cyberharassment detection system can perpetuate and aggravate existing societal prejudices and inequalities.

As cyberharassment grows online, particularly on social media, there is a need to equip computing students with the AI skills and knowledge required to design and develop AI-based systems to detect and remove cyberharassment. As the field of cyberharassment is interdisciplinary, to develop better detection systems, non-computing students, especially social science students, need to have a general understanding of AI and how it is being used in detecting cyberharassment. To teach and engage students in learning cybersecurity and AI-related topics such as data science, instructors have adopted a wide range of pedagogical methods such as flipped classroom [18], [19], project-based learning [20], [21], gamification [22], among others. Experiential learning has been regarded as one of the best ways to train future engineers by engineering educators [23]. Towards this end, experiential learning could be used in teaching AI socially relevant cybersecurity to non-computing students.

Experiential learning, simply put, is learning from experience or learning by doing. More formally, experiential learning is a type of active learning where students learn through experience [24], [25]. Experiential learning is active rather than passive. Instructors have recognized how instrumental experiential learning is in providing students with valuable hands-on experience in an AI/ML-related field such as data science [26]–[29]. In our study, we teach AI-based socially relevant cybersecurity for cyberharassment detection for two semesters using a hands-on experiential lab. Before the introduction of the lab, a questionnaire was used to rate the AI skills and knowledge of the students. After the end of the lab, another questionnaire was used to ask the students to rate their skills and knowledge of AI and AI-based cyberharassment detection covered in the lab.

Our analysis and statistical results (using sample t-test) showed that experiential learning engages students in learning AI-based cyberharassment, and it is viable for teaching AI skills to non-computing students. Also, our findings show that having a theoretical lecture before the experiential lab improves understanding of the lab contents. These findings confirm that the developed lab is viable for teaching AI-driven socially relevant cybersecurity to non-computing students and can be used in other institutions.

II. RELATED WORK

Instructors have mainly employed active learning paradigms such as experiential learning to teach AI/ML-related courses, mostly in engineering and Computer Science. The shortcomings of standard pedagogical methods in data science in online courses and data science specializations are detailed in [30]. Experiential learning mitigates these shortcomings as it focuses on problems to be solved instead of on specific methods being used [30]. Using the experiential learning style theory introduced in [25], they developed a framework to create experiences in a deep learning course.

Understanding the importance of capstone projects in data science courses, [27] developed an interdisciplinary, client-sponsored capstone program in data science and machine learning. In the program, students from different undergraduate and graduate degree programs engage in experiential learning by completing a large-scale data science or machine learning capstone project toward the end of the program.

In [26], the challenges in using capstone projects as experiential learning opportunities in data science courses due to resource constraints and data legalities involved in students working with clients on clients' real-world data sets are emphasized. To tackle this issue, they developed a novel client-facing consulting data science course that provides experiential learning to undergraduate and graduate students.

Using Challenge Based Learning (CBL), an experiential learning method, [31] explored how professional responsibility is understood by engineering students working on a solution to a real-world problem proposed by a client. The authors acknowledge that as technology such as AI/ML advances, it presents complex challenges that require an interdisciplinary approach.

The most recent studies closely related to our work focus on designing and implementing an AI education program for learners with diverse backgrounds at scale [32], and on examining first-year engineering design courses, including their design, projects, challenges, and outcomes [33]. AI education is a challenging task because it is not well studied, making it one of the challenges in engineering education [32]. With this knowledge, the Department of Defense (DoD) and the United States Air Force (USAF) partnered with MIT to design and develop educational research activities that will provide AI training for DoD and USAF personnel with various professional and educational backgrounds from high school to graduate degrees and to the general public [32].

To design a course suitable for teaching data science to students with different backgrounds, a data science course for non-computing students was developed in [21]. Experiential learning has been used in teaching students cybersecurity [34], in [35], K-12 students' self-efficacy in cybersecurity is

improved through hands-on activities performed in a virtual lab. In [36], cybersecurity is integrated into a social science course to enhance the student's cybersecurity awareness.

The popularity of AI/ML has led to the proliferation of research studies on designing better data science education materials, as shown in some of the work reviewed. These works mainly focused on developing data science courses for engineering and computing students, with a few focusing on non-computing students. Most importantly, these works do not focus on hands-on experiential labs that provide students with the experience to bridge the gap between theoretical knowledge and practice. We fill this gap by developing experiential AI/ML hands-on labs for non-computing students, and we discuss the lessons we learned from developing these hands-on labs.

III. DESIGN & DEVELOPMENT OF AI SOCIALLY RELEVANT CYBERHARASSMENT LAB

A. Lab Structure

The experiential learning laboratory consists of one lab where students become familiar with the basics of AI and the AI/ML pipeline for applying ML to a problem.

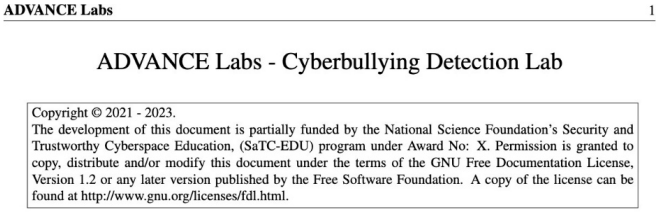
1) Objectives

We designed the experiential learning laboratory with specific learning objectives. Essential for guiding student learning, the learning objectives ensured that our study covered the fundamental elements of AI and different dimensions of AI-driven social cybersecurity. It also demonstrated the interplay between AI and cybersecurity and how AI is used for cyberharassment detection. Specifically, our objective is to develop hands-on experiential labs that will increase general awareness of socially relevant cybersecurity and AI, which is suitable for teaching AI socially relevant cybersecurity to non-computing students.

2) Lab

Our lab adopts a phased design approach. Initially, AI and cybersecurity experts in our team designed the preliminary lab and implemented and integrated the cyberharassment detection code in the lab. The lab was designed considering student profiles, AI and cyberharassment learning objectives, and desired skills. Subsequently, social scientists in our team engaged in dialogues with the AI and cybersecurity scholars to enhance the lab's reach and inclusivity. After the concerns of the social scientists such as the visibility of code and focus on questions that test students' interpretation of the lab assignment results have been addressed in the next iteration of the lab, the researchers collaboratively designed the lecture sessions and lab assignments. The collaborative and interdisciplinary approach ensured that the lab was accessible to a broad range of learners.

Our lab, *Cyberbullying Detection Using AI*, is designed to guide students through a series of learning objectives and the AI development process. The learning objectives include understanding AI (i.e., what is AI), understanding the concept and severity of cyberharassment (e.g., what is cyberharassment and the effects of cyberharassment on people), the importance of using AI in addressing this widespread online social issue (since manual detection is labor intensive and intervention is slow), and introduction to the development and the use of AI systems for cyberharassment detection (i.e., students learn use AI models are developed and used for cyberharassment detection). For the AI development process, the AI experts ensured that the design followed the AI development pipeline (data collection, verification and preprocessing, feature extraction, AI system training, and testing, and use of trained AI systems on a task) to provide students with the fundamentals of the processes followed to develop an AI system. In parallel, the social scientists offered profound insights into the peculiarities of cyberharassment, highlighting its significance and the pressing need for AI intervention. All researchers from diverse academic backgrounds involved in this work cross-verified the lab content to ensure the lab is easily accessible to non-computing students. This process ensures that those new to AI and cyberharassment can quickly grasp and engage with our lab materials. Figure 1 shows a comprehensive instruction manual prepared for the lab to facilitate independent learning and ensure students accomplish the learning objectives outside of the classroom.



ADVANCE Labs 1

ADVANCE Labs - Cyberbullying Detection Lab

1 Lab Overview

In this lab, you will learn about how AI/ML can be used to detect societal issues such as cyberbullying. Cyberbullying is bullying performed via electronic means such as mobile/cell phones or the Internet. The objective of this lab is for students to gain practical insights into online harassment such as cyberbullying, and to learn how to develop AI/ML solutions to defend against this problem.

In this lab, students will be given a starter-code. Their task is to follow the instructions provided in the Jupyter notebook, train an AI/ML model on the given dataset, evaluate their model, and deploy the model by testing it on their own samples. In addition to the attacks, students will also be guided to perform hyperparameter tuning to further improve the performance of their detection models. Students will be asked to evaluate whether their tuning effort improves their detection models or not. This lab covers the following topics:

Fig. 1. A screenshot illustrating Lab 1 instruction manual

3) Lab Delivery

To facilitate a comprehensive understanding of our lab, we implement a three-fold approach that includes a lecture phase, an experiential hands-on experience phase, and a phase dedicated to independent work.

Before the students can work on the hands-on lab, a background lecture designed and developed by the team of researchers is given to the students. In the lecture, the students get acquainted with the nature and nuances of cyberharassment, AI, the AI development process, and the need for utilizing AI for cyberharassment detection.

After introducing the students to the fundamentals needed to complete the lab through the lecture, students are introduced to the hands-on lab and guided through the hands-on experience platform depicted in Figure 2. The lab is developed on the Google Colab platform. The hands-on experience allows students to apply theory in practice. It facilitates a deeper understanding of how AI solutions are developed and, most importantly, how AI can be utilized to mitigate cyberharassment. We have chosen to utilize the Google Colab platform for several compelling reasons. First and foremost, Google Colab provides an interactive environment that integrates text and code cells. This not only enables us to write and execute Python code for deploying AI models and detecting cyberharassment content, but it also allows us to provide clear, step-by-step explanations alongside the code. Moreover, these text cells can be utilized to embed visual aids and explanations such as Figure 3, facilitating a more comprehensive understanding of the concepts and processes involved. Secondly, Google Colab offers access to free GPU resources. This is a significant advantage for our participants as the computational power of GPUs can greatly expedite the execution of AI models, ensuring that experiments are completed within a reasonable time frame. Furthermore, Google Colab's cloud-based nature eliminates the need for complex setups on personal machines, lowering the entry barrier for participants. This easy access, combined with the platform's robust functionality, makes Google Colab an ideal tool for our hands-on experiments in AI education.

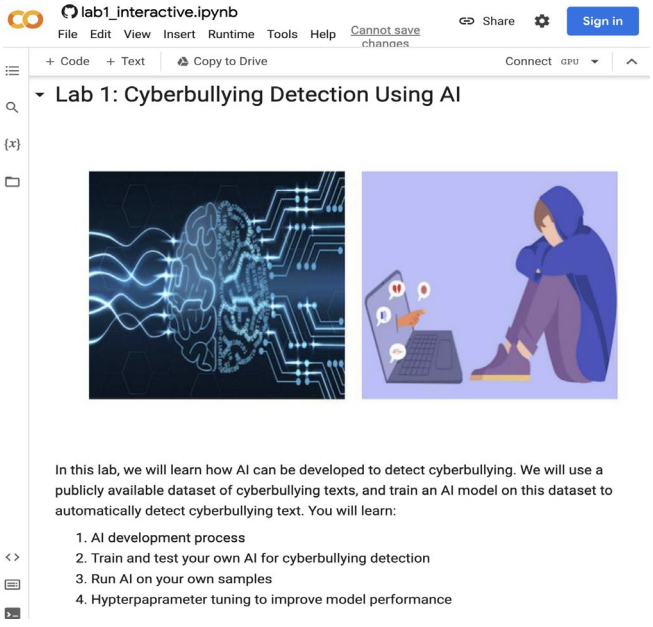


Fig. 2. A screenshot illustrating the lab interface

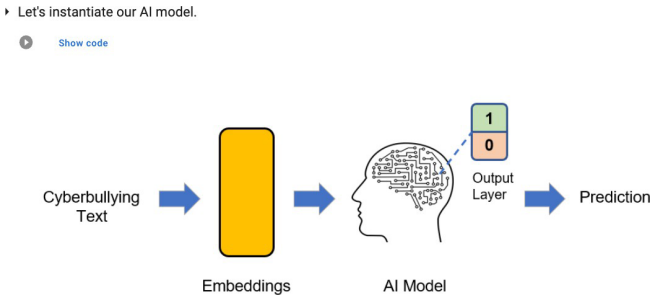


Fig. 3. The visual aid for model explanation in Lab 1

In our labs, we have meticulously curated post-lab assignments that not only enhance the engagement factor but also deepen students' comprehension of the material. For example, as depicted in Figure 4, students are tasked with using arbitrary input statements to test their AI's capability to recognize cyberharassment content. In the Google Colab platform, the correct execution of a cell could depend on the execution of the previous cell, and students are made aware of this, which is also in the lab manual. In the example activity shown in Figure 4, students must complete all prior steps to access the developed AI model for cyberharassment detection. Additionally we have formulated post-lesson discussion questions. Figure 5 depicts an example question. We aim to stimulate students' critical thinking by encouraging them to consider other real-world problems that could be alleviated through AI.

Fig. 4. One example of a lab activity

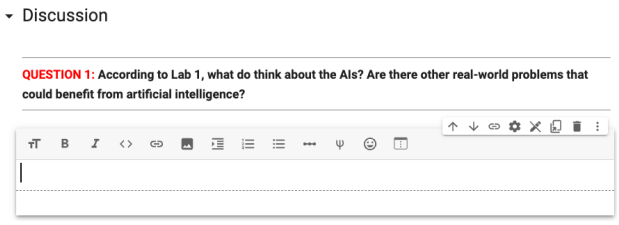


Fig. 5. One example of Lab 1's discussion questions

B. Course Structure

The Sociology program at North Carolina A&T State University includes courses on Social Statistics, parts 1 and 2. Students from Social Statistics 1 were included in this intervention. Students in Social Statistics 1 learn how to interpret and describe data. Students are exposed to topics such as distributions; descriptive statistics (e.g., measures of central tendency and dispersion); statistical null-hypothesis testing, and independent and dependent samples t-tests. They also learn the basic operation of the statistical computing software program SPSS. Throughout the course, students learn the value statistical analysis offers to our attempts to address real-world social problems. This course runs 15 weeks and includes lecture and lab time for analysis. Students attend two 75-minute classes each week. Lectures involve definitions and the demonstration of practice problems. Lab time involves hands-on computing using SPSS. In Table I, we report some basic demographics of the students in the Social Statistics 1 class in the Spring and Fall semesters.

TABLE I. Demographics of students who participated in the Social Statistics 1 course in the Spring and Fall 2022 semesters.

	Spring 2022	**Fall 2022**
Demographic	**Percent**	**Percent**
Male	23.81%	14.29%
Female	71.43%	85.71%
Freshman	23.81%	4.76%
Sophomore	33.33%	66.67%
Junior	38.10%	9.52%
Senior	4.76%	14.29%
Graduate	0.00%	4.76%

IV. METHODS

A. Survey Design

To evaluate the impact of our lab on students, we designed two surveys (pre-survey and post-survey) to collect quantitative and qualitative data from the students. Students rate their knowledge about AI and machine learning methods for cyberharassment detection. For both surveys, students rated their knowledge using a 5-point scale, with 1 representing ”Proficient or strongly agree” and 5 representing ”None or strongly disagree”. In the post-survey, students also rate the lab from different perspectives. The post-survey also includes three open questions to understand what helped the students the most in understanding the concepts taught, the difficulties faced in using the lab, and any suggestions for improving the learning experience of the lab.

B. Data Collection

Before our data collection, we obtained our institution’s Institution Review Board (IRB) approval. All the students were notified that the survey would remain confidential and only the research team could view the data. Students were also informed that their participation was optional. Before the class, students filled out the pre-survey to evaluate their knowledge level, and after the lab, students filled the post-survey. Our surveys are conducted using the Qualtrics platform [37].

C. Data Analysis

To analyze the collected data, we first compared the average knowledge score between pre-survey and post-survey. Then we used the sample t-test to determine the existence of statistical significance between the observed differences. We use a significance level of 0.05. If the p-value of the t-test is less than 0.05, we conclude a significant difference exists. Scipy, an open-source Python library for scientific computing, was used to analyze the collected data.

D. AI Model for Cyberharassment Detection

Our AI experts developed an AI model using the pre-trained transformer model BERT [38] via the HuggingFace (huggingface.co) for cyberharassment detection. Students interact with this model by running the cell on the Google Colab platform where the model is defined. After running the cell, the model is available throughout the platform.

V. RESULTS AND DISCUSSION

We analyze the data for the Spring 2022 and Fall 2022 semesters and compare the learning experiences of both semesters because the student feedback from the Spring 2022 semester, as discussed in the lesson learned Section V-E, provided data and insights that were used to inform and refine the lab in the Fall 2022 semester. Before introducing the labs to the students, they were invited to complete a pre-questionnaire, and after completing the labs, they were invited to complete a post-questionnaire. The pre and post-questionnaire contained eleven questions in total. The first three questions were asked in both the pre and post-questionnaire which we group as AI knowledge questions. The remaining eight questions were only asked in the post-questionnaire. We group the first five of these questions as lab experience questions and the last three as qualitative feedback questions. In the initial data analysis, we focus on the first three questions in the pre-and-post questionnaires administered to the students. In the final data analysis, we focus on only the last eight questions in the post-questionnaire that the students completed after completing the labs to gauge their understanding and perception of the lab. The data analysis focused on survey participation, knowledge in automated cyberharassment detection, evaluation of ML classifiers, current state-of-the-art

cyberharassment systems, how ML works, general learning experience and engagement of the labs, and student qualitative feedback. Tables II and III show the Spring 2022 and Fall 2022 survey results.

TABLE II. Spring 2022 semester survey results. M signifies mean.

1 = Proficient, 5 = None	Pre (M)	Post (M)
Automated cyberharassment detection.	3.95	3.56
State-of-The-Art cyberharassment detectors.	4.33	3.67
How machine learning works.	4.10	3.89
The lab engaged me in learning the topic of AI Driven Socially-Relevant Cybersecurity.		2.78
I enjoyed the learning experience of this lab(s).		2.89
I think the learning experience with the lab(s) is effective.	-	2.78
I am satisfied with the level of effort the lab requires for learning this topic.	-	2.78
After using the lab(s), I have better understanding about the concepts learned.	-	3.44

TABLE III. Fall 2022 semester survey results. M signifies mean.

1 = Proficient, 5 = None	Pre (M)	Post (M)
Automated cyberharassment detection.	4.38	3.31
State-of-The-Art cyberharassment detectors.	4.71	3.44
How machine learning works.	4.43	3.13
The lab engaged me in learning the topic of AI Driven Socially-Relevant Cybersecurity.	-	2.56
I enjoyed the learning experience of this lab(s).	-	2.44
I think the learning experience with the lab(s) is effective.	-	2.13
I am satisfied with the level of effort the lab requires for learning this topic.	-	2.56
After using the lab(s), I have better understanding about the concepts learned.	-	2.31

A. Survey Participation

The survey was presented to students in the Spring 2022 and Fall 2022 semesters. Of the 21 students enrolled in the course in Spring 2022, 21 (100%) students completed the pre-survey, and 9 (42.9%) students completed the post-survey. In the Fall 2022 semester, 21 students were enrolled in the class, of which all students completed the pre-survey, and 16 (76.1%) students completed the post-survey—showing that the Fall 2022 session of the course had more student participation than when the lab was first introduced in Spring 2022.

B. AI Knowledge

The first three questions: automated Cyberharassment detection, state-of-the-art Cyberharassment systems, and how ML works presented to the students in the pre and post-surveys. Most students had little or no knowledge in these areas before enrolling in the class and after participating in the hands-on lab, as there is no significant difference ($p > 0.05$) after comparing the means of pre-survey and post-survey results of Spring 2022. In Fall 2022, there is a significant difference ($p < 0.05$) in all these areas, indicating that the improvements made after the Spring 2022 semester helped improve students' knowledge.

C. Lab Experience

The student's responses to how the lab engaged them in learning about AI-driven socially relevant cybersecurity, the overall learning experience, the effectiveness of the lab, lab difficulty, and understanding of concepts were positive. Leaning towards *Somewhat agree* and *Neither agree nor disagree* in the Spring semester and towards *Somewhat agree* in the Fall semester except in the understanding of concepts where the mean response by students in the Spring and Fall semesters was 3.44 (*Neither agree nor disagree and Somewhat disagree*) and 2.31 (*Somewhat agree and Neither agree nor disagree*). Results indicate that the modifications made after the Spring semester improved students' experience with the lab in the Fall semester.

D. Qualitative Feedback

Our last three survey questions were open-ended questions about what has been the most helpful for learning, what has caused the most difficulty when using the lab, and how the lab can be improved. We used these as the qualitative data source, providing insights into students' perceptions and preferences. Overall, the students understood the purpose of the lab and cyberharassment, found the terminology confusing, had technical difficulty, and wanted the lab to be more fun and engaging. In the Spring 2022 semester, in response to *"What has been most helpful for your learning in using the lab so far?"* notable student responses were: *"I understand the purpose of cyberbullying and its purpose and how it is designed to be successful"* and *"The guest speakers coming in to help."* For the Fall 2022 semester, the notable student responses were: *"The most helpful part for my learning has been the hands-on activity, being able to ask questions while going through the work and having a guest speaker gave some*

new insight," and *"I learned a lot about cyber bullying that I did not know about and the different forms it can come in."*

For the question *"In terms of your learning, what has caused you the most difficulty in using the lab so far?"* the notable student responses in the Spring 2022 semester were: *"I had a hard time understanding how to actually complete on my own," "The terminology," "I could not stay focus and lacked engagement,"* and *"Being online."* Notable responses in the Fall semester were: *"It was difficult that when there was a troubleshooting error that I could not walk through it with someone like I could in person.", "The most difficult part is not usage of the lab itself; it is remembering certain aspects of what to do and what to look for when in the lab.", "The most difficulty experienced in the labs is facing errors and technical difficulties."*

Finally, for the question *"What suggestion(s) can you make that would enhance your learning experience with the lab?"* notable student responses in the Spring 2022 semester were: *"Make the lab more engaging / fun," "Break down steps on how to actually complete the activity," "I would say, trying using a different online platform for this lab, to make everything a little bit easier for students to understand.",* and *"Better terminology."* In the Fall 2022 semester, notable responses were: *"The instructor was helpful; it's just hard to learn over the computer — such a difficult thing to do," "I cannot think of anything at the moment. I really enjoyed learning about this lab and how it worked," "An in person option," "provide a tutorial video," "I would say slowing down the directions,"* and *"Teach more about how to face technical difficulties."*

E. Lesson Learned

The authors learned the following lessons in implementing a cloud-based laboratory experience. We outline tips for developing a cloud-based laboratory for teaching AI socially relevant cybersecurity.

1) Code Dilution

The lab is implemented on Google Colab, Google's cloud-based jupyter notebook platform. The initial implementation of the lab on Google Colab presented the students with the raw code implementation of cyberharassment detection using PyTorch, an open-source Python framework for developing ML, especially deep learning systems. From the Spring 2022 feedback from the students, we observed that the students were not positive towards the code implementation since they have very little programming experience. This frustrated students and slowed interest and learning. With this knowledge, in preparation for the Fall 2022 semester, we improved the learning experience by re-implementing the lab with the code hidden to enable the students to think about social issues and focus on understanding how AI can be used to approach social issues such as cyberbullying.

2) Lab Instructions

Developing the lab instructions in the lab manual is crucial to enhancing the student experience. If the lab instructions are not very detailed, with step-by-step instructions on how to complete the lab activities, the students struggle with understanding and completing the lab independently. From the Spring 2022 student feedback, students complained that the instruction manual needed more detail and felt the instructors assumed they were familiar with AI and AI terminology. In the Fall 2022 semester, we improved the lab manual by toning down the terminology and providing more step-by-step descriptions so the students could complete the labs independently and understand the purpose and rationale behind each step.

3) Pre-lab Lecture

Before allowing the students to complete the lab independently, we prepared lecture slides about the lab and introduced them to the problem and the activities they would be completing. The students found this pre-lab lecture particularly helpful.

Other best practices include recording the pre-lab lecture so that the students can refer back to it when working on the lab activities, anticipating possible technical difficulties, and including steps to solve the issues in the lab manual. Having an in-person option where the students can complete the labs during class could help with engagement and technical issues they might encounter.

VI. CONCLUSION AND FUTURE WORK

We have developed an AI socially relevant cybersecurity lab for cyberharassment detection for non-computing students. We introduced a cyberharassment detection lab development, its implementation, and assessment. The development process has been guided by the learning objective of introducing hands-on experiential labs that will increase general awareness of socially relevant cybersecurity and AI and is suitable for teaching AI socially relevant cybersecurity to non-computing students. Our lab offers meaningful experiential learning opportunities that allow students to work on real-world social issues such as cyberharassment. After incorporating student feedback in the redesign of the lab used in the Fall semester, the knowledge or skills of most students in automated cyberharassment detection and how ML works improved significantly compared to the Spring semester. Also, students found the detection of cyber-harassment helpful and understood the purpose of using AI for social issues. Cybersecurity experts should collaborate with non-cyber experts to pinpoint ways that knowledge in other fields can be used to improve overall cybersecurity awareness and more robust approaches to cybersecurity issues. Also, from the lessons learned in developing our experiential, hands-on labs for non-computing students, for cybersecurity to be more interdisciplinary, the instructional or introductory materials must be very clear and motivated by real-world issues to increase engagement as the real-world problems are relatable. Finally, cybersecurity education needs to be integrated into early education curricula, such as in middle and high school; this will influence

the perception of cybersecurity as students decide on majors later in their careers and will influence their attitude toward cybersecurity research collaborations if they are in non-cybersecurity fields. In the future, we plan on continuing to refine the lab and use the knowledge gained in developing three labs currently under development that cover multi-modal (text and image) cyberharassment detection, adversarial attacks on cyberharassment systems, and bias mitigation in cyberharassment systems. Additionally, we plan on developing these labs for computer science and engineering students in the future.

REFERENCES

[1] K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. Kelley, D. Kumar, D. McCoy, S. Meiklejohn, T. Ristenpart, and G. Stringhini, "Sok: Hate, harassment, and the changing landscape of online abuse," pp. 473–493, 2021.

[2] D. E. S. Swearer, "Research on school bullying and victimization: What have we learned and where do we go from here?" *School Psychology Review*, p. 365–383, 2013.

[3] S. Wachs, M. F. Wright, and A. T. Vazsonyi, "Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition," *Criminal Behaviour and Mental Health*, vol. 29, no. 3, pp. 179–188, 2019.

[4] "Center for Informed Democracy & Social-Cybersecurity," https://www.cmu.edu/ideas-social-cybersecurity/.

[5] "Social-Cybersecurity," http://www.casos.cs.cmu.edu/projects/projects/social_cyber_security.php.

[6] J. W. Patchin, "Cyberbullying Statistics," https://cyberbullying.org/2019-cyberbullying-data, 2019.

[7] D. Ducharme, "Machine learning for the automated identification of cyberbullying and cyberharassment," Ph.D. dissertation, University of Rhode Island, 2017.

[8] "Hateful Memes Challenge and Data Set," https://ai.facebook.com/tools/hatefulmemes/.

[9] H. Zhong, H. Li, A. Squicciarini, S. Rajtmajer, C. Griffin, D. Miller, and Caragea, "Content-driven detection of cyberbullying on the instagram social network," p. 3952–3958, 2016.

[10] "AI advances to better detect hate speech," https://ai.facebook.com/blog/ai-advances-to-better-detect-hate-speech/.

[11] "Google's Hate Speech Detection A.I. Has a Racial Bias Problem," https://fortune.com/2019/08/16/google-jigsaw-perspective-racial-bias/.

[12] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2015. [Online]. Available: https://arxiv.org/abs/1412.6572

[13] J. Li, S. Ji, T. Du, B. Li, and T. Wang, "Textbugger: Generating adversarial text against real-world applications," 2019. [Online]. Available: https://www.ndss-symposium.org/ndss-paper/textbugger-generating-adversarial-text-against-real-world-applications/

[14] W. E. Zhang, Q. Z. Sheng, A. A. F. Alhazmi, and C. Li, "Adversarial attacks on deep-learning models in natural language processing: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 3, pp. 1–41, 2020.

[15] M. Sap, D. Card, S. Gabriel, Y. Choi, and N. A. Smith, "The risk of racial bias in hate speech detection," pp. 1668–1678, Jul. 2019. [Online]. Available: https://www.aclweb.org/anthology/P19-1163

[16] E. Okpala, L. Cheng, N. Mbwambo, and F. Luo, "Aaebert: Debiasing bert-based hate speech detection models via adversarial learning," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 1606–1612.

[17] V. Nanda, S. Dooley, S. Singla, S. Feizi, and J. P. Dickerson, "Fairness through robustness: Investigating robustness disparity in deep learning," *CoRR*, vol. abs/2006.12621, 2020. [Online]. Available: https://arxiv.org/abs/2006.12621

[18] J. C. Nwokeji, R. Stachel, and T. Holmes, "Effect of instructional methods on student performance in flipped classroom," pp. 1–9, 2019.

[19] S. Eybers and M. Hattingh, "Teaching data science to postgraduate students: A preliminary study using a" flip" classroom approach." *International Association for Development of the Information Society*, 2016.

[20] J. C. Nwokeji and P. S. T. Frezza, "Cross-course project-based learning in requirements engineering: An eight-year retrospective," pp. 1–9, 2017.

[21] Y. Velaj, D. Dolezal, R. Ambros, C. Plant, and R. Motschnig, "Designing a data science course for non-computer science students: Practical considerations and findings," pp. 1–9, 2022.

[22] R. Matovu, J. C. Nwokeji, T. Holmes, and T. Rahman, "Teaching and learning cybersecurity awareness with gamification in smaller universities and colleges," pp. 1–9, 2022.

[23] L. Samavedham and K. Ragupathi, "Facilitating 21st century skills in engineering students," *The Journal of Engineering Education*, vol. 26, no. 1, pp. 38–49, 2012.

[24] A. Y. Kolb and D. A. Kolb, "Learning styles and learning spaces: Enhancing experiential learning in higher education," *Academy of management learning & education*, vol. 4, no. 2, pp. 193–212, 2005.

[25] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[26] A. Barman, S. Chen, A. Chang, and G. Allen, "Experiential learning in data science through a novel client-facing consulting course," pp. 1–9, 2022.

[27] G. I. Allen, "Experiential learning in data science: Developing an interdisciplinary, client-sponsored capstone program," pp. 516–522, 2021.

[28] S. Rosenthal and T. Chung, "A data science major: Building skills and confidence," pp. 178–184, 2020.

[29] P. Anderson, J. Bowring, R. McCauley, G. Pothering, and C. Starr, "An undergraduate degree in data science: curriculum and a decade of implementation experience," pp. 145–150, 2014.

[30] E. Serrano, M. Molina, D. Manrique, and L. Baumela, "Experiential learning in data science: From the dataset repository to the platform of experiences," pp. 122–130, 2017.

[31] D. A. Martin and G. Bombaerts, "Enacting socio-technical responsibility through challenge based learning in an ethics and data analytics course," pp. 1–7, 2022.

[32] A. F. Salazar-Gomez, A. Bagiati, N. Minicucci, K. D. Kennedy, X. Du, and C. Breazeal, "Designing and implementing an ai education program for learners with diverse background at scale," pp. 1–8, 2022.

[33] H. A. Hashim, C. Tatarniuk, and B. Harasymchuk, "First year engineering design: Course design, projects, challenges, and outcomes," pp. 1–9, 2022.

[34] T. Lowe and C. Rackley, "Cybersecurity education employing experiential learning," 2018.

[35] A. Konak, "Experiential learning builds cybersecurity self-efficacy in k-12 students," *Journal of Cybersecurity Education, Research and Practice*, vol. 2018, no. 1, p. 6, 2018.

[36] C. M. B. Turner and C. F. Turner, "Analyzing the impact of experiential pedagogy in teaching socio-cybersecurity: Cybersecurity across the curriculum," *Journal of Computing Sciences in Colleges*, vol. 34, no. 5, pp. 12–22, 2019.

[37] "Qualtrics," https://www.qualtrics.com/

[38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.