

The Journal of The
Colloquium for
Information Systems
Security Education
(CISSE)

Paper Presented
By

Keyan Guo
11/13/2024

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

Ebuka Okpala, Nishant Vishwamitra, **Keyan Guo**, Song Liao, Long cheng, Hongxin Hu, Yongkai Wu, Xiaohong Yuan, Jeannette Wade, and Sajad Khorsandroo

Introduction

- Artificial intelligence (AI) Education across disciplines
 - Call for AI-ready workforce by global organizations
 - For example, World Economic Forum
- Experiential learning hands-on lab
 - For socially relevant cybersecurity education
- Previous works
 - Cybersecurity education for engineering and computing students
- Hate speech (cyberharassment/cyberbullying) is interdisciplinary
 - Lack of education materials



Design & Development of Cyberharassment Lab

- Lab - Cyberharassment detection using AI
- Learning objectives
 - Basics of AI
 - Understanding the severity of cyberharassment
 - Importance of using AI for cyberharassment detection
 - AI/ML pipeline for applying ML to a problem



Design & Development of Cyberharassment Lab

- Lab delivery – three-fold approach
 - Lecture phase
 - Experiential hands-on experience phase
 - Independent work phase



Design & Development of Cyberharassment Lab

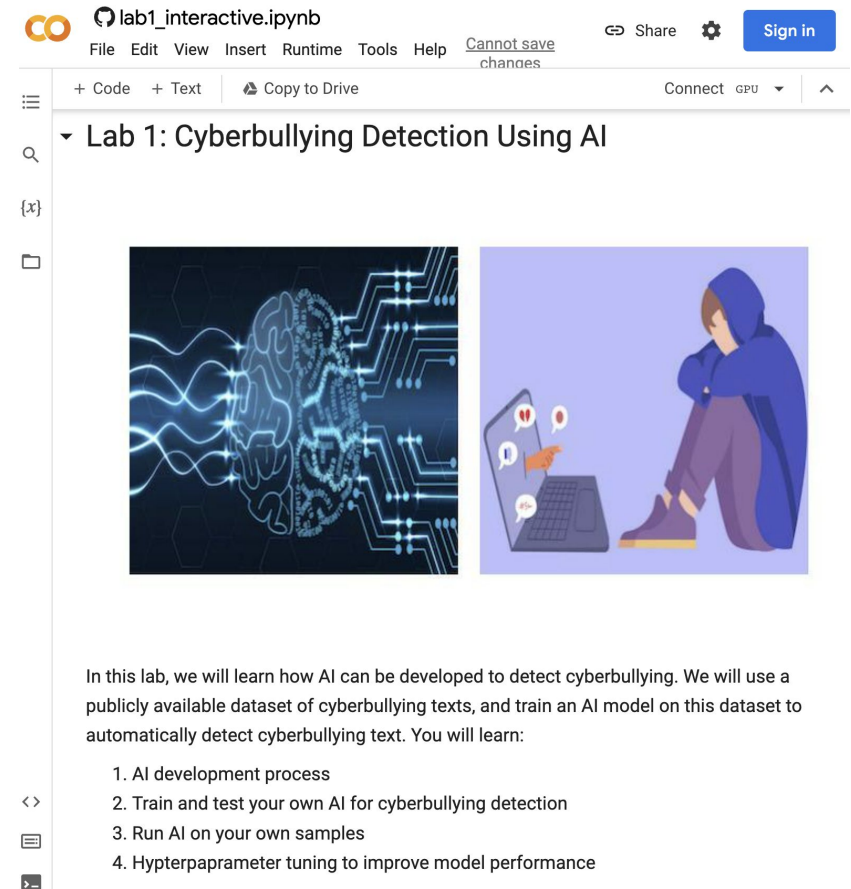
Lecture Phase

- Background lecture
 - Nature and nuances of Cyberharassment
 - AI and AI development process
 - Need for AI for cyberharassment detection

Design & Development of Cyberharassment Lab

Hands-on Phase

- Students are guided through the hands-on experience platform (Google Colab)
 - Interactive environment
 - Visualization (images and text)
 - Free GPU
 - Cloud-based
 - Complex setups are eliminated


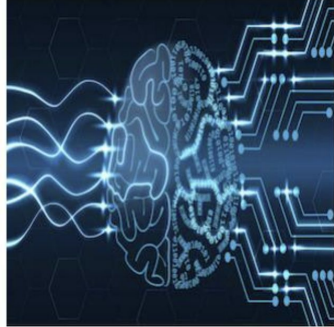


lab1_interactive.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes Share Settings Sign in

+ Code + Text Copy to Drive Connect GPU

Lab 1: Cyberbullying Detection Using AI



In this lab, we will learn how AI can be developed to detect cyberbullying. We will use a publicly available dataset of cyberbullying texts, and train an AI model on this dataset to automatically detect cyberbullying text. You will learn:

1. AI development process
2. Train and test your own AI for cyberbullying detection
3. Run AI on your own samples
4. Hyperparameter tuning to improve model performance

Lab Platform

Design & Development of Cyberharassment Lab

Independent Work Phase


- Post-lab assignment
- Post-lesson discussion

QUESTION 1: According to Lab 1, what do think about the AIs? Are there other real-world problems that could benefit from artificial intelligence?



Task 7: You can try your own sentence to see if your AI is working well!

▶ My inputs

✓
0s  `testing_sentence:` " Let's test our AI with any sentence here! "

`my_label:` Non-cyberbully

[Show code](#)

▶ AI's prediction

✓
0s [50] [Show code](#)

AI prediction: Cyberbullying not detected. Confidence: 90.83%
We can see the prediction is: correct!

Design & Development of Cyberharassment Lab

Course Structure

- Social statistics class at North Carolina A&T University
 - Spring 2022
 - Fall 2022
- Class syllabus include
 - Distributions
 - Descriptive statistics
 - Hypothesis testing, etc.,

Demographic	Spring 2022	Fall 2022
	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%

Student demographics of Social Statistics across semesters

- Survey design
 - Pre-and-post surveys using Qualtrics
 - 5-point scale
 - 1 means proficient or strongly agree
 - 5 means none or strongly disagree
- Data collection
 - IRB approval
 - Surveys were confidential
 - Survey participation was optional
- Data analysis
 - Compared the average knowledge score between pre-survey and post-survey
 - Test for statistical significance using the sample t-test

Survey Participation

Semester	Survey	# Students	Participated
Spring 2022	Pre	21	21 (100%)
	Post		9 (42.9%)
Fall 2022	Pre		21 (100%)
	Post		16 (76.1%)

Results

- Lab iteration
 - Students' knowledge or skill in cyberharassment detection and how ML works improved significantly compared to the Spring semester
- Lessons learned
 - Code dilution
 - Lab instruction must be very detailed
 - Pre-lab lecture with slides
 - Introducing the lab and the problem is helpful

Pre and post surveys

TABLE II: Spring 2022 semester survey results.

1 = Proficient, 5 = None	Pre	Post
Automated Cyber Harassment Detection.	3.95	3.56
State-of-The-Art Toxic Content 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.

1 = Proficient, 5 = None	Pre	Post
Automated Cyber Harassment Detection.	4.38	3.31
State-of-The-Art Toxic Content 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

Other Labs

- Image based cyberbully detection
- Adversarial attacks in cyberbully detection
- Debiasing word embeddings
- Hateful memes detection
- LLMs for cyberbullying detection

