LINGUISTIC CUES IN CYBERCRIME: PROFILING CRIMINAL BEHAVIOR THROUGH LANGUAGE ANALYSIS

BY

ASMA MALIK

NATIONAL UNIVERSITY OF MODERN LANGUAGES, RAWALPINDI

August, 2025

Linguistic Cues in Cybercrime: Profiling Criminal Behavior through Language Analysis

By

ASMA MALIK

B.S. English, University of Central Punjab, Lahore, 2022

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF PHILOSOPHY

In **English**

To

FACULTY OF ARTS AND HUMANITIES

NATIONAL UNIVERSITY OF MODERN LANGUAGES, RAWALPINDI

NATIONAL UNIVERSITY OF MODERN LANGUAGES FACULTY OF ARTS & HUMANITIES

THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with the overall exam performance, and recommend the thesis to the Faculty of Arts and Humanities for acceptance.

Thesis Title: Linguistic Cues in Cybercrime: Profiling Criminal Behavior through Language Analysis

Submitted by: Asma Malik **Registration #:** 74/MPhil/EngLing/Rwp/F23

Master of Philosophy
Degree name in full

English Linguistics
Name of Discipline

Dr. Arshad Ali _____
Name of Research Supervisor Signature of Research Supervisor

Dr. Arshad Mahmood _____
Name of Dean (FAH) Signature of Dean (FAH)

_____
Date

AUTHOR'S DECLARATION

I Asma Malik

Daughter of Sajid Ali Khan

Registration # 73/MPhil/EngLing/Rwp/F23

Discipline English Linguistics

Candidate of **Master of Philosophy** at the National University of Modern Languages do hereby declare that the thesis **Linguistic Cues in Cybercrime: Profiling Criminal Behavior through Language Analysis** submitted by me in partial fulfilment of MPhil degree, is my original work, and has not been published or submitted earlier. I also solemnly declare that it shall not, in future, be submitted by me for obtaining any other degree from this or any other university or institution.

I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be cancelled and the degree revoked.

_____

Signature of Candidate

_____

Name of Candidate

_____

Date

ABSTRACT

Title: Linguistic Cues in Cybercrime: Profiling Criminal Behavior through Language Analysis

In recent years, the digital landscape in Pakistan has evolved dramatically, driven by rapid technological advancements and increased internet accessibility. This transformation has brought about numerous opportunities for communication, commerce and information sharing. As individuals increasingly rely on digital platforms for personal and professional interactions, the vulnerabilities associated with these platforms have become more pronounced. It has also led to a surge in cybercrime. This rise of cybercrime has necessitated innovative approaches for understanding and preventing criminal behavior. The present research attempts to formulate an integrated model for criminal profiling for holistic understanding of cybercrime. "Routine Activity theory (RAT)" of criminology has been merged with the "Politeness and impoliteness theories (PIT)" of linguistics. By integrating RAT with PIT, the framework allows for a comprehensive analysis of both the environmental and social dynamics of cybercrime. It considers not only the motivations and behaviors of the offenders but also the linguistic strategies cybercriminal employed in conducting a successful cybercrime. Only cybercrime cases related to "Phishing" and "Cyber harassment" has been selected for this research study. Separate sections of data had been made for the analysis of cybercriminal's sentiments and language. Data collected mainly from the platforms like Twitter, Email, and WhatsApp has been used for analyzing the sentiments of cybercriminals using a pre-trained Roberta model and training it on both Phishing and Cyber harassment datasets separately. On the other hand, Case studies mainly collected from "High court cybercrime judgements" has been used for linguistic analysis. This combined analysis allowed for more nuanced profiling of cybercriminals, as it captures not only what is being said but also how it is said. This can provide a clearer picture of the motivations and psychological states of individuals involved in cybercrime. The results of the study had been validated through expert feedback via an open-ended questionnaire.

Keywords: *Cybercrime, Phishing, Criminal Profiling, Linguistic cues, Routine Activity theory, Politeness Theory, Impoliteness Theory. Sentiment Analysis, Machine Learning.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DOJ	*Department of Justice*
ATM	*Automated Teller Machine*
RAT	*Routine Activity Theory*
PT	*Politeness Theory*
IT	*Impoliteness Theory*
PIT	*Politeness and Impoliteness Theory*
PECA	*Prevention of Electronic Crimes Act*
FIA	*Federal Investigation Agency*
NLP	*Natural language processing*
AI	*Artificial Intelligence*
ML	*Machine Learning*
RoBERTa	*Robustly Optimized BERT Approach*
BERT	*Bidirectional Encoder Representations from Transformers*
NLTK	*Natural language toolkit*

ACKNOWLEDGEMENTS

Words cannot express my gratitude to Allah Almighty Who has showered His countless blessings on me during my journey of compiling this thesis. I am extremely grateful to Him for showing His mercy to me and giving me the strength to keep going, which was the most difficult task in this whole journey.

I owe my thanks to Prof. Dr. Arshad Mahmood, Dean Faculty of Languages for their co-operation in the entire process.

I am also deeply indebted to my supervisor Dr. Arshad Ali and my co-supervisor Ms. Ghazal Anwar for sparing their precious time for feedback and providing useful advices and tips for my thesis.

I would also like to express my gratitude to GAC and FBS committees for refining my thesis by suggesting useful changes that has further polished my research.

This endeavor would also not have been possible without the continuous moral and emotional support of my parents especially my father who generously provided his knowledge and expertise about cybercrime. Moreover, their belief in me has kept my spirits and motivation high during this process.

I have also not been able to undertake this journey without the help of my spouse who has guided me in developing codes which I couldn't have been able to do otherwise, as this was not my area of expertise.

Thank you all.

DEDICATION

To my parents and spouse for their love, endless emotional, moral and financial support, and encouragement.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The phenomenon of cybercrime has become more and more linguistic in nature, with words being utilized as effective instruments to trick, control, harass, and take advantage of people communicating online. The primary motivation for conducting this research stems from the realization that cybercrime relies heavily on language as its primary weapon. Among the most common types, **"phishing assaults"** utilize convincing and misleading language techniques to fool users into disclosing private information, and **"cyber harassment"** uses hostile, intimidating, or coercive language to cause psychological and emotional damage. Even though these crimes are serious, the majority of the research that has been done so far has been on technological detection techniques, paying little attention to the linguistic clues that indicate criminal intent and behavior. This study fills that vacuum by investigating the ways in which language serves as a behavioral marker and a weapon in cybercrime. In particular, it employs sophisticated Natural Language Processing (NLP) model "RoBERTa" to detect language indicators, emotional manipulation, and sentiment patterns in texts that contain Phishing and Harassment. In addition to enhancing cybercrime detection, the overall goal is to show how combining language theory with machine learning models can help us better comprehend the communication tactics used by cybercriminals. The overall aim is to demonstrate how linguistic profiling, when integrated with machine learning, can both enhance the detection of phishing and harassment attempts and contribute to a richer understanding of the communicative mechanisms driving cybercrime.

Cybercrime encompasses any illicit activity that involves a computer, network, or device that is connected to one. (Kate Brush, 2024). Cybercrime or Computer crime is defined as "any criminal law violations that involve a knowledge of computer technology for their perpetration, investigation, or prosecution" in the Department of Justice's (DOJ) guidebook on the subject. ("Computer Crimes", 2021).

Despite its broad scope, this definition is still applicable. In its explanations of the issue, the DOJ divides cybercrime into three categories: 1) Crimes where the target is computer hardware, software, or peripherals, and the criminal has acquired these items illegally; 2) Crimes where the computing device is the direct "subject" or "target" of the crime, meaning that the intended harm is on a computer or system, causing disruption or destruction; and 3) Crimes where identity theft, data theft, money laundering, and the distribution of child pornography are commonly committed, and the tactic or "instrument" is a computer equipment and its connected systems. ("Computer Crimes", 2021).

Cybercrime encompasses a wide range of illegal acts including different digital platforms and technology. The widespread usage of the internet has allowed con artists to easily reach millions of prospective victims worldwide at a low cost and entice them to join fictitious investment schemes. Even with the high risks, most victims want to increase their money fast in hopes of making big returns (Ameiruel Azwan Ab Aziz, 2023). Amid this rapid proliferation of technology, cybercrime has also adapted various different forms, ranging from ransomware attacks and phishing scams to bogus emails and social media activities ("Proofpoint", 2024). They most frequently take advantage of user trust, ignorance, and a propensity to overshare personal information online, however some of these overlap in the ways they are carried out ("Proofpoint", 2024). These cybercrime operations result in lost or corrupted data, financial losses, or damaged reputations.

For instance, identity theft or large financial losses are frequently the outcome of credit card fraud, “Automated Teller Machine” (ATM) skimming, and internet banking schemes. Hacking can have serious political and economic repercussions when it comes to gaining access to confidential information or intellectual property. Numerous risk factors, including software system flaws, insider threats, and poor passwords and cybersecurity procedures, might be linked to data breaches (Hafeez, 2014).

While cybercrimes come in numerous forms, phishing and online harassment crimes are the focus of this study. Small businesses should be particularly concerned about phishing attempts because they are by far the most common and serious cyber security issue. About 90% of data breaches occur as a result of phishing assaults, which have increased in frequency recently ("The biggest cyber threats in Pakistan", 2022).

Cybercriminals employ phishing as a tactic to trick people into giving sensitive information, like bank account details and usernames. Phishing occurs when an employee opens a malicious attachment, clicks on a malicious link, or gives confidential information to someone pretending to be a trustworthy source ("The biggest cyber threats in Pakistan", 2022). These attacks have been more common in Pakistan over the past few years. Cybercriminals are always coming up with new and plausible ways to access sensitive data. Research conducted in October 2022 found that, in comparison to the same period in 2021, the rate of phishing attacks increased by 61% in the six months that ended in October 2022 (voilino, 2023).

When it comes to phishing attempts, attackers use a range of techniques in addition to email-based tactics (asee, 2023). One such technique is vishing, in which con artists use voice recordings to trick victims into giving up personal information over the phone. A tactic used by scammers to send phishing text messages on mobile devices is called "Smishing," also referred to as "SMS phishing." Malicious attachments or URLs are often included in these emails. Page hijacking and calendar phishing are two more ways that attackers might fool users into accessing fraudulent websites or forwarding them fake calendar invites. By educating employees on how to recognize and report phishing schemes, organizations may safeguard confidential information from cybercriminals who attempt to exploit whatever vulnerability they find. People are more vulnerable to these schemas as a result of the growth of online banking and e-commerce, which causes anxiety (Zainab Alkhalil, 2021). Public education on spotting and avoiding phishing scams is essential, as reports suggest that many victims frequently do not notice the signals of the scam until it is too late.

Harassment, on the other side, is defined as intentional aggression towards another person that may or may not have caused harm to the harasser. In the past, harassment was used to describe acts of bullying or harassment towards people on the basis of their caste, gender, religion, or sect. Although the phrase is still used exactly, harassers have discovered new avenues to pursue victims who have the advantage of concealing their true identity thanks to technological advancements (Hafeez, 2014). These new ways of harassment are labelled as online harassment or cyber harassment in today’s world.

Online harassment can take two forms: direct or indirect. In one American case, the victim received thousands of offensive phone calls after her stalker created a fictitious ad on a user-forum, offering her a prostitute service and providing her home address and cell phone number. In America, anti-harassment legislation was developed as a result of this case. Maybe harassment comes from the anonymity of digital media. People are able to communicate without revealing their genuine identify because to the acceptance of internet access under several identities (Hafeez, 2014). The speed at which content spreads and the ease with which one can remain anonymous both contribute to the toxic climate that allows violence and discrimination to flourish (Saleh Mohamed, 2024).

Online harassment cases, which primarily target women and marginalized groups, are on the rise in Pakistan in addition to phishing attempts. According to data gathered by the Digital Rights Foundation's (DFR) helpline, the vast majority of victims of cyber harassment are women (Shmyla Khan, 2019). The Federal Investigation Agency (FIA) received 3027 reports of cybercrime between August 2014 and August 2015, of which 45% concerned electronic abuse against women. Although social media platforms offer a forum for communication and self-expression, they have also turned into havens for harassment and threats.

Previous researchers have explored the role of social media in amplifying discrimination and violence, with a focus on its impact on women, LGBTQ+ individuals, and marginalized communities (Saleh Mohamed, Social Media and Misinformation: Amplifying Discrimination and Violence, 2024). They are of the view that a constant stream of disparaging remarks, threats, and even doxing (the criminal online sharing of personal information) that are directed towards victims can have a severe psychological impact, making people anxious, depressed, and reluctant to interact in online environments (Hafeez, 2014). The anonymity that comes with using the internet often gives offenders more confidence, which makes it difficult for victims to get support and punishment.

In addition, law enforcement authorities' responses have been erratic, frequently hampered by a lack of funding, expertise, and understanding of cyber concerns. To address these issues, the Federal Bureau of Investigations (FIA) developed the Cyber Crime Wing.

Nevertheless, because of poor public trust and bureaucratic obstacles, the efficacy of this project is frequently questioned (Jibran Jamshed, 2022). Furthermore, the paper authored by (Hobashia Saleem, 2022) highlights the significance of a robust legal framework in effectively countering cybercrime. Internet crimes continue to pose a serious concern even after the Prevention of Electronic Crimes Act (PECA) was introduced in 2016. The report also emphasized obstacles to enforcement, including lack of public knowledge, data breaches, technological limitations, and terrorist threats. It also emphasizes the value of dedicated courts, public education initiatives, and enhanced technological capacities for law enforcement. Given these difficulties, a cybercriminal profiling model equipped with modern tools of analysis is the need of the hour.

Cybercriminal profiling helps identify offenders and stop future attacks by examining digital evidence and behavioral patterns to comprehend and identify cybercriminals, their motives, and their tactics. Since it builds profiles of cybercriminals by analyzing their digital footprints, it is different from standard criminal profiling. It goes without saying that cybercrimes can be perpetrated from any location in the world. The digital records of their actions are the only proof they leave behind. During the profiling process, it is important to consider that cybercriminals frequently need specialized technical abilities to conceal their digital footprints and camouflage their identities.

Bridging the gap between the technological and human aspects of cybercrime investigation is the main goal of current study. This research study aims to learn more about the identities and motivations of cybercriminals by deciphering linguistic patterns, which emphasizes the value of interdisciplinary cooperation in the battle against cybercrime. Forensic linguistics will continue to be a vital tool as cybercrime develops, assisting investigators in meeting the increasing demand for complex analysis. Natural language processing (NLP) and artificial intelligence (AI) developments enable more rapid and accurate linguistic profiling, which in turn facilitates more efficient cybercrime response.

The relevance of linguistics in protecting the digital world is highlighted by the fact that law enforcement organizations and cybersecurity specialists may both prevent assaults and prosecute offenders by knowing the language used by hackers.

As cyber threats become more sophisticated, so too must our methods, and linguistic analysis will become crucial in determining how digital forensics develops in the future. Hence, this research endeavors to develop an integrated framework by combining insights from the disciplines of linguistics and criminology. In an attempt to do so, this study investigates cybercrimes particularly phishing and online harassment scams that are currently in vogue in Pakistan. By doing so, this research hopes to create a framework for criminal profiling that will help lawyers in scrutinizing a huge list of victims and will ease the investigation process.

1.2 Statement of Problem

In the digital age, the rise of cybercrime, particularly phishing and online harassment, poses significant challenges to individuals, organizations, and society at large. Despite the increasing prevalence of these malicious behaviors, there remains a critical gap in understanding the underlying linguistic cues and emotional expressions that characterize cybercriminal communications. Current approaches to combating cybercrime often focus on technical solutions or legal frameworks, neglecting the nuanced role of language and social dynamics in facilitating or mitigating such behaviors. This research aims to address this gap by investigating the specific linguistic characteristics that indicate intent in cybercriminal behavior, by employing sentiment analysis to predict and identify these activities, and by exploring the influence of social dynamics and online community norms on the prevalence of Phishing and Online harassment.

By unveiling the linguistic and emotional dimensions of digital deviance, this study seeks to contribute to a more comprehensive understanding of cybercriminal behavior, ultimately informing more effective prevention and intervention strategies. Moreover, the lack of a holistic approach that integrates linguistic analysis, emotional context, and social dynamics limits our ability to effectively profile and understand cybercriminals. Therefore, this research is essential for developing targeted strategies to combat cybercrime and enhance online safety in increasingly complex digital environments.

1.3 Research Objectives

- To identify particular linguistic characteristics (e.g. lexical choices, sentence structure etc.) that reflect intent in Phishing and Online harassment communications.
- To analyze how emotions and sentiments expressed in cybercriminal's social media-based communications contribute to the overall understanding of psychological state and behavioral patterns of cybercriminals which ultimately lead to the formation of an estimated criminal profile.
- To investigate the ways in which social variables, such as interpersonal power relations and digital community norms, impact the occurrence and encouragement of cybercriminal practices particularly Online harassment and Phishing.

1.4 Research Questions

1. What particular linguistic characteristics are most indicative of intent in cybercriminal behaviors especially in "Phishing" and "Online harassment"?
2. How can sentiment analysis of Cyber communications mediate through social media enhance the ability to predict and identify cybercriminal activities particularly Phishing and Cyber harassment?
3. What role do social dynamics and online community norms play in facilitating cybercriminal behavior in Phishing and Online harassment?

1.5 Significance of the Study

There are various reasons why this research on criminal profiling and cybercrime is important. First of all, as cybercrime is becoming more and more widespread worldwide, it is essential to comprehend its dynamics in order to create successful preventative and intervention plans. This research looks into how cybercriminals behave and why they do it in an effort to improve law enforcement's investigative methods and increase the success rate of prosecutions. This study aids in the creation of more focused and effective profiling methods by seeing trends in the language, emotional tone, and online activity of cybercriminals.

Furthermore, by using profiling data, public awareness programs can be developed to help people and organizations understand and reduce the risks connected with cybercrimes. In the end, the goal of this research is to make the internet a safer place, increase public confidence in digital technologies, and make society more resilient to cyberattacks. This research helps ensure that user safety remains a major priority as technology advances by fusing insights from behavioral analysis and profiling with practical implementations.

1.6 Delimitations

The main focus of this study will be on two particular types of cybercrime; Online harassment and Phishing. Additionally, the research will focus on diverse range of datasets including forums such as news outlet (Dawn news), chat logs (Emails, WhatsApp) and Twitter trends to gather information for online conversations related to Phishing and Cyber harassment and the High Court official case judgements to identify incidents specifically linked to online harassment and phishing. To further guarantee data relevancy and manageability, this study examines current cybercrime trends and occurrences that have happened within the previous five years.

1.7 Outline of the thesis chapters

The current research paper is distributed into five chapters. First chapter deals with Introduction. It introduces the topic of the research that is “Linguistic cues in cybercrime: Profiling criminal behavior through language analysis”. It also defines all the key concepts and terminologies related to cybercrimes used in the present research.

It clearly states the primary purpose of the research i.e. locating linguistic cues that indicate criminal intent in the online cybercriminal communications mediated through different social media sites. Moreover, it also demonstrates the research questions that has been answered in the finding section of this study. In addition to this, it also tells us that this research study is delimited to the two particular types of cybercrimes: Phishing and Cyber harassment and involves case studies only from the year 2018 to 2024.

The second chapter, Literature Review, elaborates on the studies that has been previously conducted in this area of research. It gives a detailed account of researches done on profiling cybercriminal behavior through various traditional techniques of analysis and then introduces a research gap that has been filled by this research.

Third chapter, known as Research Methodology, gives the overview of the methodology that has been used in this research study, thoroughly explain the tools and methods that has been used for collecting the data for this research as well as clearly defines the population and sample of the research. Purposive Sampling has been used by this research as it allows to choose only those samples that aligned with the research purpose.

This chapter also deals with stating the overview of the three theoretical frameworks chosen from the disciplines of criminology (Routine Activity Theory) and linguistics (Politeness and Impoliteness Theory) and then provided with an integrated framework that contain insights from these three theories to develop a cybercriminal profiling model to create an estimated criminal profile of cybercriminals by incorporating the environmental and linguistic factors involved in conducting cybercrimes. This chapter also discusses in detail the analytical frameworks used for sentiment analysis (fine-tuning a pre-trained Roberta model on Phishing and Cyber harassment datasets) and linguistic analysis (analyzing the lexical, semantic and syntactic aspects of language manually). Moreover, in the end, this chapter also discusses the third tool i.e. “an open-ended questionnaire” used for getting the validation of the results for this research study.

Data presentation and Analysis has been done in the fourth chapter of this research study. Results obtained from sentiment and linguistic analysis has been discussed separately. Insights from both sentiment and linguistic analysis has been employed in making an estimated profile of cybercriminals. Results obtained from this combination of analysis revealed important emotional and linguistic patterns used by the cybercriminals in online communications that shows how cybercriminals take advantage of the socio-discursive context of digital platforms. Brown and Levinson's theory of impoliteness is consistent with the frequent infractions of politeness norms found in Cyber harassment messages, including the use of imperatives, threats, and face-threatening acts.

Moreover, this chapter also gives an overview of the questionnaire designed for getting feedback about this research study from lawyers who has dealt with the cybercrimes previously or are currently handling such cases as well as the summary of the responses received. In addition to this, discussion section of this research further elaborated on the results of the study.

Last but not the least, fifth chapter of the paper has been dedicated to conclusion. It involves summarizing all the preceding chapters that has been discussed above. The study's conclusion tells us about its significant contributions to the disciplines of criminology, cybersecurity, forensic and computational linguistics, its practical implications in the sectors of cyber security and criminal profiling by integrating language cues into their investigation processes, and an urgent and dire need for ongoing research and development in the realm of cybercrime to keep track of an increasingly changing and evolving nature of cybercrimes. Moreover, it tells us about the limitations of the study, and enhances the reader's insight on the areas where no researches have been done previously and future researchers might be directed on them.

CHAPTER 2

LITERATURE REVIEW

2.1 What is Language?

Language is a set of rules and symbols used for thought and communication. In 1965, Noam Chomsky defined language as "a set of (finite or infinite) sentences, each finite in length and constructed out of a finite set of elements." Its social component was highlighted by Ferdinand de Saussure (1916), who described language as "a system of signs that express ideas," emphasizing its structured and arbitrary nature. By examining language as a component of communicative skill and connecting it to social and cultural circumstances, Dell Hymes broadened the perspective even more. These concepts collectively demonstrate that language is a social and cultural system for meaning-making in addition to being a cognitive instrument. In addition to this, this development in linguistic theory reflects an increasing recognition that language is a result of human interaction influenced by cultural norms, power dynamics, and society conventions in addition to being a mental construct. It is essential for group membership, identity development, and information transfer across generations. Therefore, language is a fundamental component of human cognition, culture, and social existence and much more than just a means of communication.

2.2 Significance of Language

Language is more than just a means of understanding and expressing our ideas to others; it is the selection of words from our mental vocabulary that may either make a conversation completely unpleasant or save our blushes ("Language is more than a means of communication", 2019). Words are what sustain a language. They are crucial to the development of a language and are more than simply sounds. They are a reflection of our culture, customs, and traditions. A language continues to develop by incorporating new words and phrases while monitoring changes in the environment. Each individual, therefore, carries a unique linguistic fingerprint that reflects a lot about the environment in which he or she brought up.

Specific use of vocabulary and communication styles can reveal useful information about a person's age, gender, their degree of education, cultural background and even character (Ilyasova, 2018). Recent studies in linguistics have demonstrated that language influences our emotional experiences and expressions (Stacey, 2015). Emotion regulation is impacted by cultural norms ingrained in language, whereas emotional awareness and expression are influenced by a language's lexicon and structure. The way that bilingual people experience emotions varies based on the language they employ, demonstrating how language shapes emotional reality.

2.3 Impact of Culture and Context on the Use of Politeness and Impoliteness Linguistic Strategies

Two important linguistic techniques for navigating interpersonal interactions and managing social relationships are politeness and impoliteness. The goal of politeness tactics, as described by Brown and Levinson (1987), is to preserve the "face" of the speaker and the listener by demonstrating respect, reducing danger, and fostering social peace. These include the use of honorifics, hedging, indirect language, and other mitigating strategies. On the other hand, impolite tactics intentionally transgress societal standards and expectations, frequently in an effort to challenge authority, show rage, or exert control. According to researchers such as Culpeper (1996), impoliteness can be deliberate and context-dependent, supporting particular communicative objectives in confrontational or competitive situations. Culturally bound and culturally specific, politeness and impoliteness reflect underlying values and social hierarchies ingrained in language use.

Because they can influence the dynamics of interaction, both politeness and impoliteness tactics are important. They have an impact on how people interpret messages, build or preserve relationships, and negotiate power dynamics. Additionally, both are context-dependent and culturally bound; what is deemed polite in one culture could be construed as evasive or fake in another, and what is deemed impolite in one setting might be construed as honesty or assertiveness in another. Communicative competency is improved by knowing and employing these linguistic skills strategically, which enables speakers to modify their language to fit various social contexts, control impressions, and successfully accomplish their communication objectives.

A lot of researches conducted in linguistics and literature revealed that environmental factors such as gender, culture and context can also influence the language use. For Example, an analysis of language used in Jordanian and American TV sitcoms (Bayan B. Rababa'h, 2021) revealed that in both societies, male characters used rudeness tactics more frequently than female characters. Additionally, compared to the American ones, Jordanian characters employed more rudeness techniques. American males and females did not significantly differ in their use of impoliteness techniques, whereas Jordanian males and females did.

In addition to this, a study conducted by Jonathan Culpeper argued that, in intimate contexts, it is less important to be polite than in non-intimate contexts. In other words, being seemingly impolite might encourage intimacy since intimacy is linked to a lack of politeness. This obviously only functions in situations where the rudeness is perceived as being untrue. The use of politeness and impoliteness strategies in communications also vary from culture to culture and context to context. For example, in some cultures "burping" is an acceptable behavior, while in other cultures it is considered impolite. So, it can be generally assumed that an "anti-social" behavior would not be considered impolite if it were accepted as the standard.

Some theorists like Leech (1983), Brown and Levinson (1987) argued that some speech acts are inherently polite or impolite. In other words, they contend that some behaviors such as commands, threats, and criticisms go against one's positive self-image, which is the want for approval, and one's negative self-image, which is the desire for freedom. It is possible to agree in general that certain behaviors are essentially polite while others are essentially unfriendly if one views acts in an abstract manner. But one must remember that context will be taken into consideration in any evaluation of politeness conducted outside of the theorist's sphere of influence. As rightly put by Fraser and Nolan (1981) that there is no such thing as a polite or impolite sentence. We frequently consider some expressions to be rude, yet the context in which they are used—rather than the expressions themselves—determines how polite they are. For instance, a sentence "Go, Eat up", which is meant to be an order (impolite request) will be perceived as a polite request in a context where it involves the benefit of the guest.

2.4 Linguistic Cues in Cybercrime

Information can be conveyed through language in both honest and misleading ways, along with a wide range of feelings and emotions, including fear, rage, joy, and guilt (Adha, 2020). Humans have a history of lying to influence how other people perceive them and to benefit from the deceit (Azianura Hani Shaari, 2019). Scammers may leave language traces behind, but in our technologically reliant age, it is unavoidably difficult to recognize these traces in order to anticipate deceit. Through the internet, knowledge can now be shared without boundaries and communication is accessible to everybody. The widespread usage of the internet has allowed con artists to easily reach millions of prospective victims worldwide at a low cost and entice them to join fictitious investment schemes. In spite of the high risks, the majority of victims want to expand their businesses online in hopes of making significant profits. Unfortunately, especially in asynchronous contacts, deliberate lies have also been spread by computer-mediated communication. As more and more examples of cybercrimes, such as the heinous practice of online investment scams, occur, digital deceit is turning into a common occurrence.

The majority of victims were tricked by the "sweet words" of scammers, which affected their choices and behavior. Scammers typically utilize more positive language (like "guaranteed," "profit," and "success") than negative language (like "dangerous," "suspicious," and "risk") in an attempt to persuade victims that they can trust them and demonstrate the reliability of the good or service (Zafarani, 2020). People are therefore vulnerable to being tricked by fake websites that offer investment frauds. Linguistic cues are frequently used by scammers to present a credible and genuine image of oneself (Adha, 2020). In their deceptive discourse, con artists often refer to groups and favor third-person pronouns (such as us, their, and her) over first-person pronouns (such as I, me, and myself) in an attempt to dissociate themselves from the message they have constructed (Aseel Addawood, 2019). Product trademarks, functional terms, punctuation marks, and tenses are only a few examples of the words used in the advertising materials of scammers to demonstrate the legitimacy of their goods or services. ("Unveiling the cloak of deviance: Linguistic cues for psychological processes in fake online reviews", 2020).

Additionally, it was noted by (Genao, 2021) that con artists refrain from mentioning themselves to their victims.

2.5 Criminal Profiling through Language Analysis

Knowing the identity and intents of cybercriminals is crucial to reducing the threats to information security in the context of cybercrime. Criminal profiling has emerged as a valuable tool that helps identify intruders and their motivations by analyzing their behavior, nature, and mental processes. It also helps to strengthen strategies against potential cyber threats by providing insight into the psychological traits and characteristics of cybercriminals. Criminal profiling, sometimes referred to as offender profiling, is a technique used in criminal investigations to find probable suspects by looking for trends that could indicate potential offenders and victims in the future. It entails a careful analysis of the type of crime committed, the actions of the offender at the scene, and any supporting documentation ("StudySmarter", 2024). Cybersecurity profiling includes a variety of criminological and criminal law components, including as personal characteristics, criminal history, social characteristics, and driving forces. Understanding the characteristics, socioeconomic background, personality attributes, and goals of cybercriminals—including the most elusive ones—is made easier by understanding these factors (Umema Hani, 2024).

Cybercriminals usually display a variety of psychological characteristics that significantly influence their activities and behaviors. These people usually possess an excellent understanding of cyber technology, which they use for malicious ends and a variety of motivations. Financial gain is a prevalent motivation, as seen by actions like data theft and other types of cyber fraud (Li, 2017). Many people want financial gains out of avarice, while others seek power or retaliation against specific individuals, organizations, or groups. Some cybercriminals are thrill-seekers who enjoy the danger that comes with their illegal activity, or they are opportunists who use weaknesses to their advantage for personal gain (Umema Hani, 2024). There exist individuals who choose to completely ignore moral and legal obligations, endangering their standing in the online community. The most common traits are those of fearlessness, indifference to possible outcomes, and lack of empathy.

Additionally, some people show audacity by trying their hacking skills on people and organizations. When taken as a whole, these characteristics provide a complicated picture of the attitudes and actions that lead cybercriminals in different situations (Li, 2017). Psychological profiling is a useful technique that helps identify intruders and their motivations by analyzing their behavior, nature, and mental processes. It also helps build up strategies directed at potential cyber threats by comprehending the psychological traits as well as distinctive features of cybercriminals (all, 2024).

2.6 Finding and Filling the Research Gap through this Research Study

Despite the fact that a great deal of study has been done on language analysis for criminal profiling and on different kinds of cybercrimes. The multidisciplinary approach used in this research study sets it apart from others and will enhance analysis and offer deeper insights into the language aspects of cybercriminal activity. Recent years have seen a growth in cybercrime due to quick technology developments and easier access to the internet, which has made it necessary to utilize modern theories and models for comprehending cybercriminal behavior. Language is not just a medium of communication but is a reflection of behavior of people. Hence, present study is dedicated to analyze the linguistic cues in social-media based communication of cybercriminals using an integrated framework to grasp an overall understanding of cybercriminal behavior which could ultimately help in developing an estimated criminal profile of a cybercriminal.

Previous researches have been using traditional machine learning approaches that required manual feature engineering techniques to analyze the results. For Example, studies conducted by (R Jayakrishnan, 2018), (Bart Desmet, 2013) have used Support Vector Machine (SVM) classifier to detect emotion at phrase level. Another study has combined SVM with topic model for hierarchical emotion classification (Fan Zhang, 2016). Some researchers (Maryam Hasan, 2018) used combination of Naïve Bayes (NB), Support Vector Machine (SVM) and Decision tree classifiers to categorize emotions. A couple of researches (Bjarke Felbo, 2017) has applied deep learning network that comprises of Bi-LSTM with an attention layer for training language model. Park seo-hui has utilized NLTK and VADER (Lexicon-based sentiment analyzer) for sentiment analysis.

Some (Windy Amelia, 2016) (Sonja Gievska, 2015) even used Hybrid approaches by combining both lexicon-based methods with learning-based methods. On the other hand, present research is focused on training a transformer-based model Roberta which is more powerful, effective and accurate than all the other models that has been previously used for sentiment analysis due to its ability to comprehend complex contextual patterns from large datasets for the classification of sentiments expressed by cybercriminals in their social-media based communications.

The main goal of this research project is to combine established theoretical frameworks from different disciplines with modern emerging concepts that will eventually contribute to the existing knowledge of criminal profiling and computational linguistics. Moreover, it will enable law enforcement organizations and investigators of cybercrime to effectively address the growing activities of cybercriminals.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Research Design

To examine cybercriminal behavior in the contexts of phishing and online harassment, the current study has used a mixed-method approach that combines qualitative linguistic analysis of case studies with quantitative sentiment analysis of cybercriminal's social media-based communications. This methodological research design is grounded in the study's main goal of giving a thorough explanation of how linguistic and emotional characteristics represent the intention, psychological health, and social motivations of those who commit cybercrimes. The results obtained from sentiment analysis has been presented on excel spreadsheets whereas language analysis has been demonstrated through close reading of case studies using concepts from the two theoretical frameworks adopted for this research study. In the end, open ended questionnaires have been circulated among the lawyers handling cybercrime cases for further validation of this research study.

3.1.1 Quantitative Analysis

This study uses a Roberta-based sentiment analysis model to statistically evaluate the psychological cues and emotional tone present in cybercriminal communications. Roberta (A Robustly Optimized BERT Pretraining Approach) is one of the most advanced transformer-based architectures for identifying contextual subtleties in textual input (geeksforgeeks, 2023). Because of its strong pretraining on extensive corpora and flexibility in fine-tuning for specific domains, it is especially well-suited to the intricate and frequently secretive nature of cybercriminal speech. A corpus comprising two specific cybercrime categories—Phishing and Cyber harassment—was first obtained from the **"Hugging Face"** and **"Kaggle"** databases, after which they were annotated properly and was used to refine the model. This allowed for the extraction of sentiment indicators that go beyond surface-level polarity to capture coercion, aggression, manipulation, and implicit emotional purpose.

The study's second research objective, which aims to comprehend how sentiment analysis might improve the predictive identification of cybercriminal behaviors, is addressed by the model by quantifying sentiment trends at scale.

This study's quantitative component is based on assessing machine learning model "RoBERTa" that were trained on the chosen dataset of cyber harassment and phishing texts. In particular, accuracy, precision, recall, and F1-score—standard classification metrics that function as statistical measures of prediction reliability and generalizability—were used to evaluate the RoBERTa model. In order to assess how well the model interpret linguistic cues and differentiate between dangerous and safe digital communication, these metrics offer a solid foundation. The current study does not require the use of inferential statistical tests, such as the t-test, which are used in standard quantitative linguistic studies to examine mean differences among groups. This is due to the fact that models' prediction capabilities—rather than mean value hypothesis testing—are the analytical focus. The categorization metrics themselves serve as quantitative proof, guaranteeing that the analysis of the outcomes is computationally rigorous and reproducible. However, descriptive statistics like frequency counts and percentages of linguistic markers were also used when appropriate to draw attention to patterns of deceit, violence, and emotional manipulation in the data. These descriptive metrics enhance the interpretive connection between computational performance and the linguistic processes being studied, and they supplement the categorization results.

3.1.2 Qualitative Analysis

Sentiment analysis provides computational insights into emotional tone, but it falls short in addressing the pragmatic, rhetorical, and structural aspects of language that are crucial for comprehending purpose and social stance in cybercrime. This is accomplished by incorporating a linguistic case study analysis based on Politeness and Impoliteness Theory (Brown & Levinson, 1987; Culpeper, 2005), which looks at how cybercriminals use language to influence victims, assert authority, execute face-threatening activities, or avoid detection.

This approach focusses on speech acts, grammatical structures, lexical choices, and other linguistic tactics that give qualitative depth to the behavioral profile of offenders. It provides direct support for the first and third research objectives, which examine the language indicators of criminal intent as well as the ways in which social dynamics and online community standards shape deviant behavior. For this purpose, ten case studies (five from Phishing cases and five from cyber harassment cases) have been taken for analysis.

This study's qualitative component enhances the computational assessment by delving further into linguistic cues in texts that contain harassment and phishing. The qualitative study explores how lexical selections, stylistic patterns, and rhetorical techniques are used to influence, mislead, or threaten receivers, drawing on ideas from sociolinguistic approaches to cyber communication. Sampled texts were closely studied and thematically coded in order to find recurrent themes of animosity, coercion, and persuasion. Since they closely resemble linguistic tactics commonly observed in cybercrime speech, categories such emotive appeals, politeness violations, lexical aggressiveness, urgency framing, and identity manipulation were given special attention. By using this two-pronged strategy, the qualitative analysis not only confirms the computational findings but also reveals nuances that automated techniques would miss, like cultural allusions, metaphorical framings, or implied threats. This guarantees that the research findings are computationally sound and linguistically based, offering a thorough grasp of cybercriminals' communication tactics.

3.1.3 Rationale for using a mixed-method approach

The methodological justification for the combination of Roberta-based sentiment analysis and qualitative language analysis is triangulation, a technique that improves the validity and reliability of results by bringing together several data sources and analytical viewpoints. Linguistic analysis gives depth and interpretive richness, while sentiment analysis delivers scalability and breadth, guaranteeing a more comprehensive understanding of cybercriminal behavior. Additionally, this mixed approach is consistent with the study's theoretical framework, which incorporates both Politeness/Impoliteness Theory (to understand social and communicative behavior in online contacts) and Routine Activity Theory (to contextualize opportunity and motivation in cybercrime).

3.2 Tools and Methods for Data Collection

The data for training the sentiment analysis model Roberta has been taken from two well-known communities of datasets: Hugging face and Kaggle. Kaggle is an online community of data scientists and machine learning engineers which allows users to find datasets they want to use in building AI models whereas Hugging Face Datasets is a library focused on natural language processing (NLP) technologies which provides a collection of pre-processed and ready-to-use datasets for various NLP tasks. These datasets contained cybercriminal communications mediated through social media and other platforms including Twitter, WhatsApp, spam Emails and Dawn news. These datasets are then carefully curated, processed, and standardized to ensure consistency and ease of use. This dataset is then divided into two sets: training and testing sets. 80% of the data has been used for training phase while the rest 20% has been utilized for testing phase. On the other hand, data for linguistic analysis has been gathered from case studies obtained from various sources mainly from High Court case judgements and partly from social media datasets used in the sentiment analysis. Open-ended questionnaires have been formulated to collect responses from lawyers handling cybercrime cases in courts.

3.3 Population of the Study

There are no direct population or people involved in this research. However, an open-ended questionnaire conducted at the end for the validation of this study involves ten lawyers who are dealing with the cybercrime cases in High Court. The main purpose of circulating the questionnaire among the lawyers is to get feedback about this research study and its findings from the persons who are directly involved in investigation of the cybercrimes.

3.4 Sample of the Study

The case studies that have been chosen for examination in this research has centered on instances of “Online harassment” and “Phishing”. Only ten case studies have been taken for this research. Purposive sampling has been used in the selection of the case studies. In qualitative research, purposeful sampling is a method used to choose a certain set of people or units for analysis (team, 2023).

Purposive sampling has been employed by the researcher to choose study-relevant samples that will ultimately aid in achieving the intended goals of study.

3.4.1 Sample size

For sentiment analysis, sample size has been kept large for accurate and detailed training of Roberta model whereas for linguistic analysis only ten case studies have been selected (five from each cybercrime). Moreover, five open-ended interviews have been conducted in the end for the validation of results.

3.5 Theoretical Framework

By utilizing an integrated theoretical framework, this study draws on valuable insights from three theories, including linguistics' "Politeness and Impoliteness theory (PIT)" and criminology's "Routine Activity theory (RAT). The integration of RAT and PIT in this paradigm enables a thorough examination of the social and environmental aspects of cybercrime. It takes into account the language tactics used by offenders in addition to their intentions and actions. By recognizing how environmental factors impact the behavior of both the perpetrator and the victim, the framework also highlights the context in which cybercrime occurs. After independently defining each model, this study proposed an integrated framework and explained the importance of applying it to the current investigation.

3.5.1 Routine Activity theory (RAT)

Routine activity theory is a criminological paradigm that emphasizes the ecological process of crime and links it closely to the environment (cohen, 1979). According to this, crime happens when a motivated criminal meets a suitable victim without the presence of a responsible guardian. Instead of concentrating on crime, it looks at how exposure, proximity, and availability impact crime rates by altering the regular activities of potential both victims and perpetrators (Perera, 2024). The macro-level analytical approach of routine activities theory draws attention to the broad changes in victim and offender behavior patterns (Cohen, 1981).

The theory makes the noteworthy assumption that everybody who has the chance to commit a crime can do so. Furthermore, according to this theory, victims have the ability to decide whether or not to become victims. For instance, by avoiding risky situations, they can lessen their chances of becoming victims.

A motivated perpetrator, an appropriate target, and the lack of capable guardians are the three requirements for a crime to occur, according to routine activity theory. When these three elements come together, criminality results (Perera, 2024).

Figure 1

Elements of crime

i. **Motivated Offender:** People who are both capable and willing to perpetrate wrongdoing are considered motivated offenders (miro` llinares, 2014).

ii. **Suitable Target:** An individual or piece of property that the motivated criminal can easily locate and interact with could be an ideal target (Nickerson, 2022). However, it appears that the main determinant as to whether the offense would be done in this case is the target's suitability. The acronym VIVA represents the following factors that a motivated criminal would consider when deciding whether a potential target is appropriate (miro` llinares, 2014)

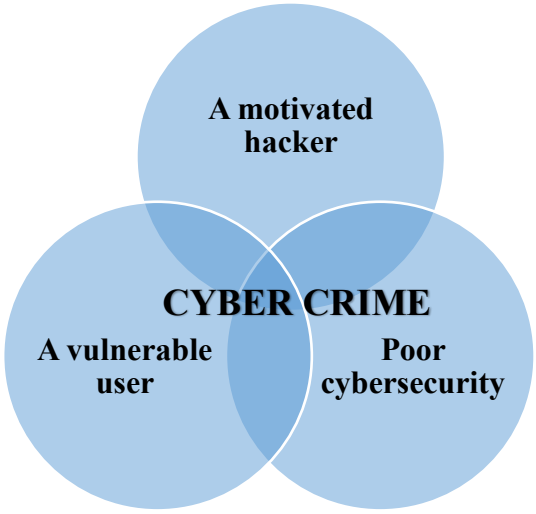
- **Value** = the actual or symbolic value [from the offender's point of view] of engaging the target;

- **Inertia** = the ideal target's weight, size, shape, or other physical characteristics that act as barriers to the motivated offender.
- **Visibility** = Visibility occurs when an offender sees a target, indicating that it is a good target for assault.
- **Access** = the location of the target and the layout of the website make it more attacking-prone.

iii. **Absence of Capable Guardians:** People or things that successfully discourage criminal activity just by existing in time and space can serve as capable guardians. This kind of guardianship, whether formal or informal, can stop misbehavior even when a motivated perpetrator selects a good target (Nickerson, 2022).

Figure 2

Elements of Cybercrime



3.5.1.2 Rationale for using RAT

Since Cohen and Felson (1979) initially proposed the Routine Activity Theory, criminologists have used it extensively to describe the circumstances surrounding criminal activity. Because of its focus on the convergence of three factors—a motivated offender, an appropriate victim, and the lack of a skilled guardian—it is particularly pertinent to digital contexts, as criminals take advantage of online spaces to target vulnerable people.

By emphasizing offender desire and situational opportunity, RAT offers an explanatory model for why specific communications—such as phishing attempts or harassing messages—are generated in the context of cybercrime.

RAT is modified for this study in order to demonstrate how hackers utilize language as a tool to take advantage of opportunity structures. The victim is made a "suitable target" for phishing by means of language techniques like urgency, authority impersonation, and appeals to trust. Moderation and other social "guardianship" mechanisms are less successful in harassment when there is overt hostility or covert rudeness. The dynamics of everyday activities are thus reframed in linguistic terms by mapping linguistic cues onto the offender-target-guardian triad of RAT.

3.5.2 Politeness Theory (PT)

The theory makes extensive use of Erving Goffman's **"face theory,"** which defines **"face"** as a person's social identity or self-image (Goffman, 1967). The underlying premise of politeness theory is that individuals have two distinct "faces": positive and negative (Levinson, 1987). A **positive face** symbolizes a person's wish to be liked and valued by others. Consider this to be a person's sense of self-worth. A **negative face** is a person's wish to defend their individual liberties, including their right to free speech and action.

Positive and negative politeness are the two forms of politeness that we appeal to when we treat others with courtesy (studysmarter, 2023). By appealing to someone's Positive Face, you can boost their self-esteem and make them feel more confident about themselves. This is known as **positive politeness**. For example, we might agree with someone's statement, congratulate them on their accomplishments, or comment on their attire. We steer clear of arguments, insults, and criticisms when we want to preserve someone's good reputation. Similar to this, in phishing situations, perpetrators may speak to victims in a polite or flattering manner in an attempt to establish rapport and increase the likelihood that they will comply with strange requests, like clicking on a harmful link. On the other hand, appealing to someone's negative side, or giving them the impression that they haven't been coerced or exploited, is known as **negative politeness**.

To make phishing attempts appear less dangerous and more authentic, criminals may, for instance, use hedging or indirect inquiries in cybercrime instances to avoid direct demands (studysmarter, 2023).

According to politeness theory, decisions about which politeness tactic to use are influenced by the social context of the speech act. Three sociological factors determine whether a politeness strategy is effective. These factors are rank, distance, and power (Universalclass, n.d.). **"Power"** describes the perceived power relationship between the hearer and the speaker. The social distance between the speaker and the hearer is referred to as **"distance." "Rank"** describes how sensitive the topic is within a given culture, or how the subject is ranked culturally. In the United States, for instance, a woman's age, weight, and income are all considered highly sensitive topics. In several other cultures, however, these topics are simply considered facts that should be spoken. In cases of cybercrime, especially those involving phishing and cyber harassment, context analysis can be useful in determining why particular tactics work better in different circumstances (Universalclass, n.d.). Four essential politeness techniques that people employ for controlling social encounters and preserve their dignity were put forth by Brown and Levinson (Levinson, 1987). These consists of:

i. **Bald on Record:** This tactic involves saying things clearly and bluntly without holding back on being courteous. It is direct and could come out as rude or direct. It is employed when a speaker wishes to make a point quickly and clearly, frequently in circumstances where being courteous is not a top concern. For instance, issuing directives or submitting requests.

ii. **Positive Politeness:** This technique aims to improve the listener's positive face, which is connected to their need for social acceptance and sense of self-worth. It is demonstrating camaraderie and solidarity. Using inclusive language, displaying care in the listener's welfare, or offering compliments are a few examples. As an illustration, "We all want to succeed, right?" The goal of this tactic is to give the listener a sense of worth and appreciation.

iii. **Negative Politeness:** By protecting the listener's negative face, this tactic seeks to address their need for independence and freedom from coercion. Indirectness, hedging, and deference are all part of it. "I'm sorry for bothering you, but..." and "If you don't mind, could you...?" are two examples. By minimizing imposition, this approach respects the listener's right to decline.

iv. **Off-Record:** The speaker can avoid direct confrontation or accountability for the message by using this tactic, which involves indirectness and ambiguity. Rather than making direct claims, it depends on clues, recommendations, or implications. Saying "It would be fantastic if someone would assist with this," for instance, suggests a request without really asking for assistance. By using this tactic, the speaker can continue to avoid taking responsibility for their statements.

3.5.3 Impoliteness Theory (IT)

Impoliteness, according to Culpeper, is "...communicative strategy aimed for attacking face, and consequently trigger social conflict and discord" (Aydınoğlu, 2013). Culpeper ((Jonathan Culpeper, 1996)) suggested the following impoliteness tactics, which somewhat resemble Brown and Levinson's (1987) politeness methods:

i. **Bald on Record:** This tactic entails speaking clearly and bluntly without compromising your politeness (Rudianto, 2023). Criminals may use bald on-record to directly give orders or make threats. For instance, "Now give me your password!" A sense of pressure and urgency may result from this strategy.

ii. **Positive Impoliteness:** The goal of this tactic is to damage the listener's positive façade, which is connected to their sense of self-worth and need for social acceptance. Behaviors that show contempt or a lack of consideration for the hearer's feelings are frequently included. Making someone feel alienated, dismissing their requests, or excluding them from a discourse are a few examples.

iii. **Negative Impoliteness:** This tactic overtly puts the hearer's face in danger, frequently by acting impolitely. Threats of account suspension if demands are not fulfilled are examples of negative impoliteness in phishing. As an instance, "You will lose access to your account if you don’t comply". Direct insults or threats may be used in online harassment, for instance, "I'll ruin your life."

iv. **Off-record Impoliteness:** Off-record impolite behavior is when someone uses vague or indirect language to avoid taking responsibility for their words or facing direct criticism. Because it uses inferences, clues, or recommendations instead of direct comments, this tactic is a more subtly offensive kind of rudeness (Monir Mirhosseini, 2017). Consequences may be implied by offenders without being explicitly stated. "Many individuals who skip this message have experienced issues related to their accounts," for instance, might be a phishing email that subtly implies that noncompliance could have unfavorable consequences.

v. **Sarcasm and Mock politeness:** Ironic or caustic comments are used in this tactic to damage the target's credibility or sense of self (Culpeper, 2005). In their messages, criminals may make fun of victims by using sarcasm. As an example, "Okay, just give me your information." It's an excellent idea. A hostile atmosphere and additional victimization may result from this.

3.5.4 Rationale for using Politeness/Impoliteness theory

Brown and Levinson's (1987) Politeness Theory and further advances in Impoliteness Theory (Culpeper, 1996; Bousfield, 2008) offer a thorough framework for examining how people handle their faces throughout interactions. Because cybercriminal communication frequently depends on either strategic politeness (phishing emails that mimic legitimacy, trust, and respect) or strategic impoliteness (harassment messages that threaten, degrade, or denigrate), these theories are especially appropriate for our topic.

Politeness Theory is modified for phishing to take into consideration deceptive applications of positive or negative politeness techniques, such as deference, courteous requests, or excessive formality, that mask malevolent intent. When it comes to harassment, the Impoliteness Theory is modified to describe hostile behaviors that purposefully harm the victim's face, including insults, derision, or covert animosity (e.g., sarcasm, mockery). Thus, the approach is designed to capture language acts in cybercrime conversation that are both blatantly hostile and deceptively cooperative.

Table 1

Title: Comparing Politeness and Impoliteness Communication Strategies

Communication Strategies	Politeness Theory	Impoliteness Theory
Bald on Record	Direct communication without mitigation.	Directly confront or insult without any politeness.
Positive Politeness/Impoliteness	Enhance the listener's self-esteem and create a sense of camaraderie.	Undermine the listener's self-esteem or social approval.
Negative Politeness/Impoliteness	Respect the listener's autonomy and minimize imposition.	Threatens the listener's autonomy and create discomfort.
Off Record	Indirect communication to avoid confrontation	Indirect statements to avoid direct confrontation.
Sarcasm and Mock Impoliteness		Use sarcasm or mockery to belittle the target.

3.5.5 Integration of the two frameworks

Politeness/Impoliteness theories describe the linguistic realization of cybercrime (how purpose, hostility, or manipulation is transmitted through language), whereas Routine Activity Theory explains the situational dynamics of cybercrime (why and when criminals act). Each framework only addresses one aspect of the issue when used alone. This study bridges linguistic tactics and criminological motivations through their integration. By mapping linguistic behaviors—which are determined by politeness/impoliteness strategies—to the situational conditions of cybercrime, as described in RAT, the integration is accomplished:

Motivated offender → revealed through hostile or manipulative language (e.g., threats, politeness strategies of deception).

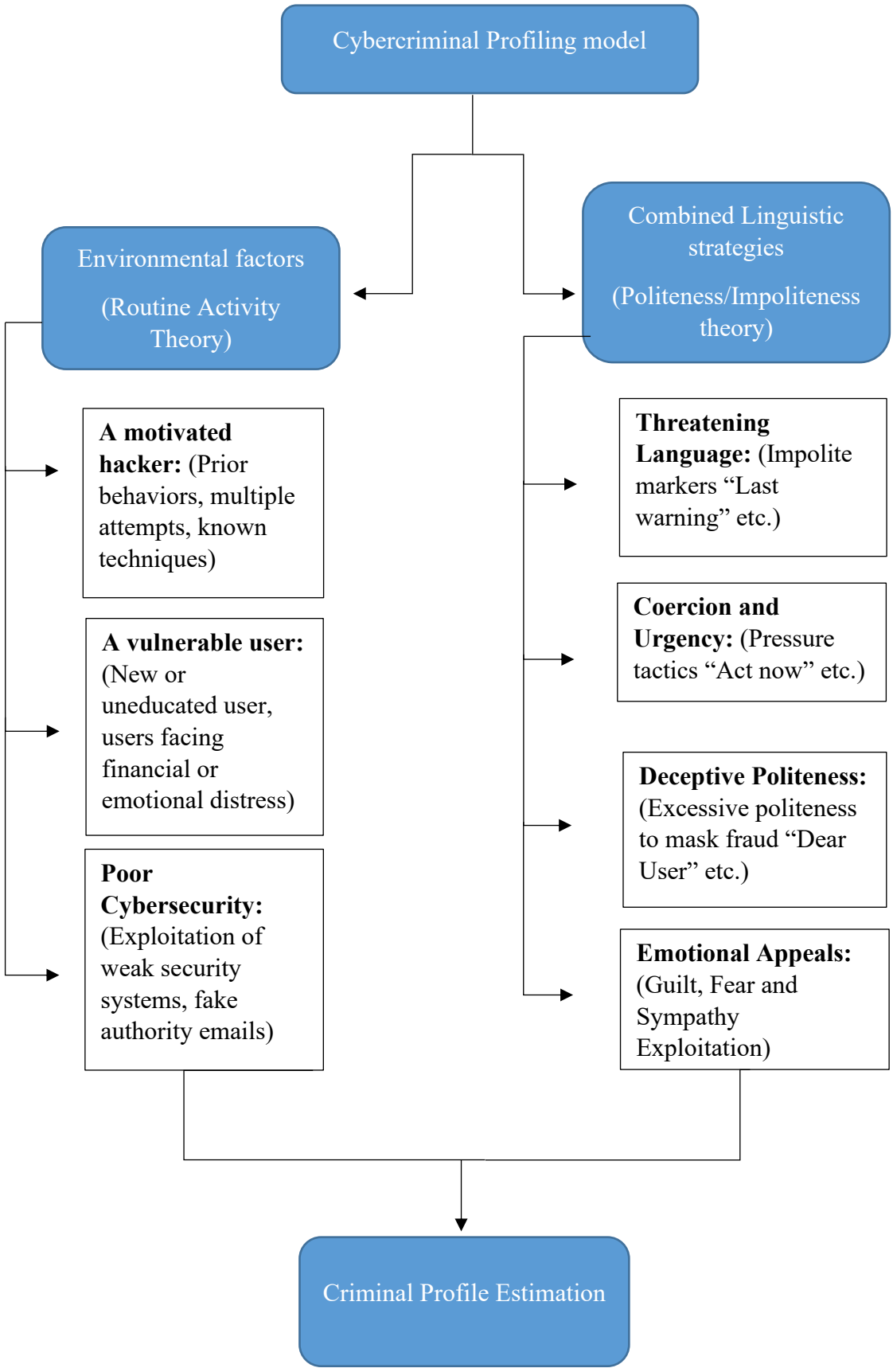
Suitable target → constructed linguistically through appeals to trust, authority, or emotional vulnerability.

Absence of capable guardian → facilitated when linguistic cues bypass detection by automated systems or social moderators.

Therefore, by connecting language usage to the reasons why crimes in digital spaces are successful, this integrated framework makes it possible to profile the behavior of cybercriminals. It offers both analytical depth (through linguistics) and explanatory strength (through criminology), which makes it particularly well-suited to the study's two main goals: identifying cybercrime and deciphering the communication tactics that enable it.

Figure 3

Integrated Framework Design



3.5.7 Significance of Integrated Framework

Through the integration of language theories and RAT, the framework enables a thorough examination of the social and environmental dynamics of cybercrime. It takes into account not just the intentions and actions of criminals, but also the language techniques they use. By identifying the traits of appropriate targets, the framework also helps to explain why some people or communities are more vulnerable to harassment and phishing. Furthermore, this framework can guide the creation of more potent preventative measures and treatments by better understanding how offenders employ politeness and impoliteness. Policymakers and law enforcement organizations can also learn from this framework how crucial it is to profile cybercriminal behavior by taking into account both environmental and linguistic components of cybercrime. Furthermore, this paradigm links psychology, linguistics, and criminology, encouraging multidisciplinary study that may result in novel theoretical advancements. Overall, by including the aforementioned theories, the framework shows that RAT is still relevant in the digital age and broadens its application beyond conventional crime situations.

3.6 Analytical Framework

3.6.1 Framework for Sentiment Analysis

Natural language processing (NLP) technologies are used in sentiment analysis to teach computer programs to comprehend text in a manner akin to that of humans. It is the process of examining digital text to identify if the message is positive, negative, or neutral in terms of emotion (geeks for geeks, 2024). Opinion mining, or sentiment analysis, is another name for this crucial instrument that people use to enhance their goods and services. The goal of applying artificial intelligence (AI)-based sentiment analysis techniques in this research is to obtain objective results while avoiding the personal biases associated with human reviewers. Previously, the analysis has been conducted through natural language toolkit (NLTK) which often uses a rule and lexicon-based approach in which accuracy vary depending upon the complexity of the language as this model struggles with complex sentence structures like sarcasm and irony.

But this research has trained a transformer-based language model Roberta which is more powerful, effective and accurate than all the other models that has been previously used for sentiment analysis due to its ability to comprehend complex patterns from large datasets (geeksforgeeks, 2023). Moreover, it can perform more advanced tasks, like sentiment analysis on different types of text (e.g., news, social media posts, books), and can potentially be fine-tuned for specific tasks. In the present study, it has been fine-tuned for cybercriminal's language for detecting the sentiments and emotions of the cyber criminals.

3.6.1.1 Comparative advantage of using Roberta over traditional models

i. Limitations of traditional approaches:

Before the development of transformer-based models, traditional machine learning techniques like Naive Bayes, Support Vector Machines (SVM), and Logistic Regression were frequently used by the researchers (as had already been discussed in the Literature Review) to tackle cybercrime detection (such as the classification of phishing and cyber harassment), sentiment analysis and emotion detection tasks. Hand-engineered features including part-of-speech tags, keyword frequency counts, and TF-IDF (Term Frequency-Inverse Document Frequency) representations were a major component of these models. However, the ability of such systems to simulate linguistic context and nuance is intrinsically limited. These classifiers frequently misread emotionally or pragmatically rich text because they perceive words as independent units and are unaware of word order or sentence structure. While "you are such a strong girl" has a positive connotation, "you are such a girl" may be used sarcastically or insultingly in cyber harassment circumstances. Because traditional models understand semantics at a surface level, they usually are unable to discern such differences. Furthermore, the more complex linguistic phenomena that are quite common in cyber harassment discourse, such as sarcasm, irony, indirect aggression, or code-switching, are difficult for these models to capture.

ii. Contextual and pragmatic strength of Roberta Model:

Built on a transformer-based architecture, Roberta (A Robustly Optimized BERT Pretraining Approach) gets beyond these restrictions by utilizing contextual embedding and bidirectional encoding. Roberta does not handle words in the absence of context, in contrast to earlier models. Instead, it learns context-sensitive representations of meaning by processing each token in relation to both its previous and succeeding words. Because of its architecture, Roberta is especially well-suited to identify subtle animosity or coded aggressiveness in harassment and deceptive politeness in phishing emails. Roberta can be made to better capture the pragmatic functions and emotional subtexts of language by fine-tuning it on task-specific datasets, as this study did. Because of its versatility, it performs significantly better than static, rule-based classifiers in simulating the linguistically nuanced and strategically misleading communication of real-world cybercriminals.

iii. Elimination of hand-featured engineering process:

Roberta's ability to avoid the manual feature engineering procedure that was previously necessary is another important benefit. Conventional models required handwritten semantic characteristics, sentiment dictionaries, offending keyword lists, or large lexicons. This method was very domain-specific and labor-intensive, which frequently led to models that were not generalizable and could not keep up with the changing strategies of cybercriminals. Roberta, on the other hand, automatically picks up pragmatic, syntactic, and semantic patterns from the training data. This lessens researcher bias in feature selection and increases flexibility in dealing with harmful texts that are structurally identical yet invisible. As a result, the Roberta paradigm improves scalability and enables processing of huge, heterogeneous corpora without requiring frequent feature set updates by hand.

iv. Empirical Superiority in performance:

Roberta's adoption for cybercrime detection is further supported by the performance metrics generated by this study. In evaluation metrics like F1-score, precision, and recall, transformer-based models like Roberta have continuously outperformed conventional methods.

This is especially true when it comes to identifying minority classes like phishing attempts, hate speech, or cyberbullying—domains where previous models have a tendency to perform poorly. The outcomes of this investigation are consistent with these more general conclusions. Even when phishing emails were written in language that appeared to be emotionally neutral or pleasant on the surface, the Roberta model was still able to identify high-confidence negative sentiment. In the same way, the model accurately identified sarcasm, direct and indirect aggressiveness, and biassed expressions in online harassment detection. Even though there were still some misclassifications, especially when there was ironic or manipulative wording, Roberta performed noticeably better than previous machine learning benchmarks documented in the literature on both overt and covert instances of deviant behavior.

v. Flexibility across various domains:

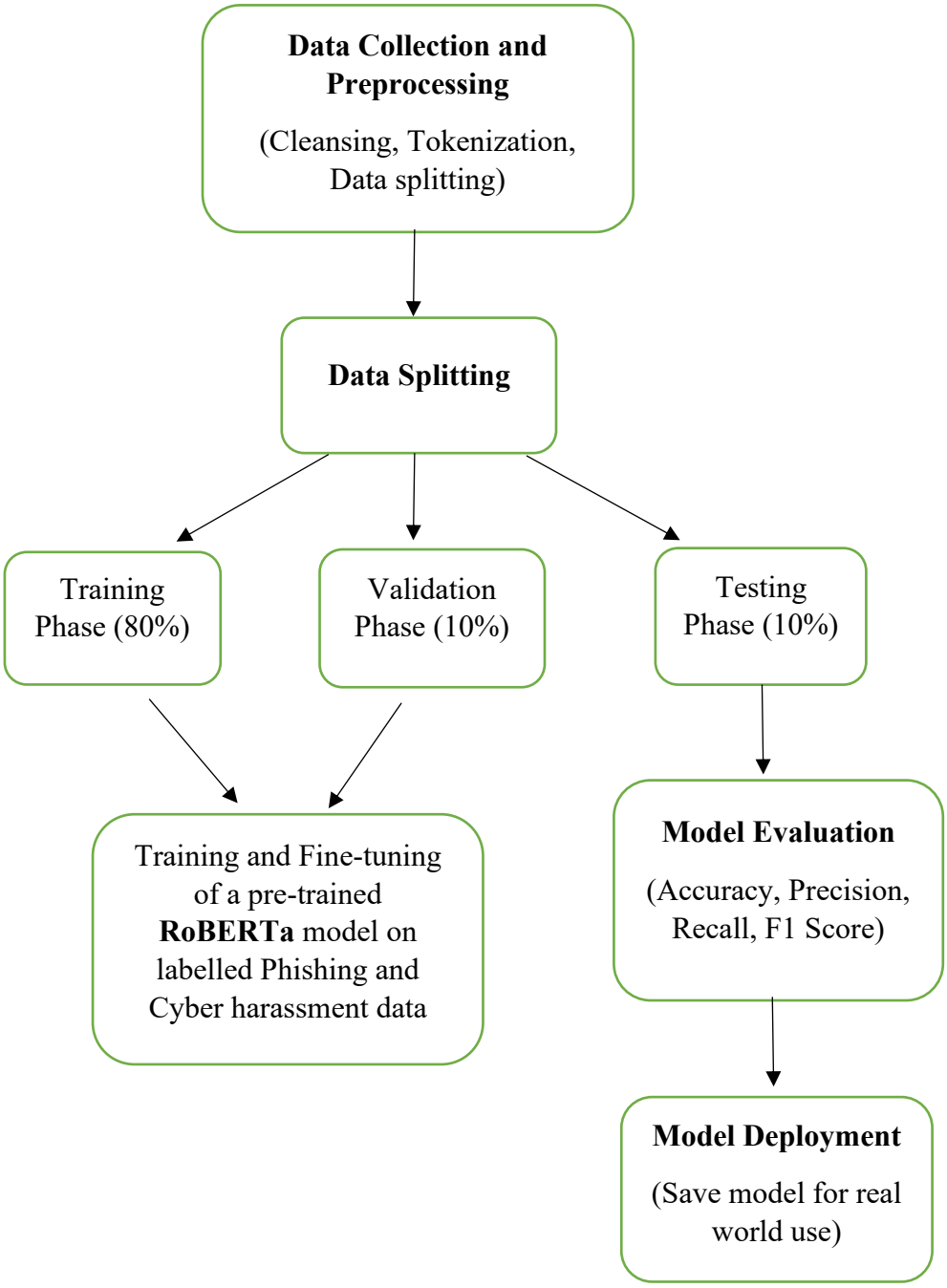
Conventional methods usually required distinct pipelines for phishing detection, hate speech classification, and spam filtering, making them task-specific. Additionally, they showed little ability to transfer knowledge between domains, which is a significant limitation when it comes to behavioral profiling of cybercriminals who use a variety of platforms or strategies. Roberta, on the other hand, enables a unified modelling approach. Roberta was optimized for two different but related objectives in this study: phishing and cyber harassment detection, and it performed well in both situations. The model is a perfect fit for integrated cybercrime detection systems since it showed the capacity to identify cross-domain characteristics like emotional manipulation, dishonest tone, and intent-driven language. By keeping in view, the above-mentioned strengths of Roberta model over traditional machine learning approaches that had been previously used by many researches, this study has adopted it for the analysis of data collected from social-media based communication of two different types of cybercrimes: Phishing and Cyber harassment.

3.6.1.2 Fine Tuning Roberta model

To fine-tune Roberta, this research has taken an already trained Roberta model and then trained it on datasets related to phishing and cyberbullying to make it perform better on detecting the sentiments of cybercriminals.

Figure 4

Steps for fine tuning of RoBERTa model



3.6.1.3 Steps to fine-tune Roberta model

Step 1: Install Dependencies. This step involves installing all the required libraries that will be needed throughout in the model's training. The current research study has installed the appropriate libraries where needed throughout the Roberta base model training procedure. To begin, Scikit-learn has been installed together with transformer datasets.

Step 2: Load and preprocess your dataset. CSV files can be read by deep learning models such as Roberta. Therefore, all of the files were first converted from "excel" to "csv." We have loaded datasets after importing pandas in order to read a CSV file in Python. A Python package called Pandas is used to work with data collections. It provides features for data exploration, cleaning, analysis, and manipulation. Data sets that are disorganized can be cleaned up by pandas to make them more readable and pertinent. Pandas can also remove rows that contain incorrect values, such as empty or NULL values, or are not relevant. We refer to this as data cleaning. The dataset includes text samples (communications, emails, etc.) along with the labels that go with them (e.g., cyberbullying or regular messages, phishing or safe emails). We extracted the "text" and "label" columns once the dataset was loaded. Either print or head commands can be used to verify that data has been loaded correctly and read accurately. This will print a sample of the dataset's initial few entries in the notebook.

Step 3: Tokenization using Roberta Tokenizer. Raw text cannot be directly processed by neural networks. To translate the text into meaningful token IDs which the model can comprehend, tokenization is necessary. Tokenizers for data tokenization are embedded into every NLP model. Roberta Tokenizer is the name of the tokenizer which Roberta model utilized for tokenization purpose. Thus, a code was used to import the Roberta Tokenizer from transformers. It was naturally applied to the load dataset when it was imported from the transformers. Because Roberta stops text if it goes beyond 512 tokens, all text samples were kept at the maximum length of 512.

Step 4: Convert Dataset to PyTorch Format. The Transformers library from Hugging Face supports two deep learning models for deep learning model training. They are TensorFlow and PyTorch. However, PyTorch, which is frequently chosen for fine-tuning transformer models like Roberta, was used in this research study. Built on top of Torch, a Lua-based framework, and Python, PyTorch serves as a deep learning library. For deep learning scholars and developers, it offers dynamic computation graphs, GPU acceleration, and an easy-to-use interface. Because PyTorch uses a "define-by-run" methodology, its computational structures are created dynamically, facilitating improved debugging and model customization. Hugging Face's Trainer API was mainly developed for PyTorch, which makes fine-tuning Roberta lot easier. This is the main justification for choosing PyTorch format. Additionally, the majority of pre-trained Roberta models were initially made available in PyTorch format. This indicates that PyTorch was used to create the initial implementation of Roberta. Therefore, this study has chosen to use PyTorch for fine-tuning Roberta.

Step 5: Splitting of dataset. For the evaluation of the model properly, all the collected Phishing and cyber harassment datasets has been divided into training, validation and testing sets.

i. Training set: In essence, the training set is the part of the dataset which is utilized to train the machine learning model. The Roberta model has been trained using roughly 80% of the dataset. This significant portion of data is used during the training phase in order to educate the model on a large number of datasets and help it grasp the type of data and the results that we want it to produce. Prior to being assigned sentiments (positive, neutral, and negative) in accordance with the data type, the training dataset was first labeled. Datasets were trained in the training phase based on the requirements of the researcher. The Roberta model in this work was trained on two datasets: Phishing and cyber harassment. The datasets were first divided into two groups for Cyber harassment: bullying and non-bullying. Labels for positive, negative, and neutral sentiment were then applied appropriately. After being separated into two groups—safe data and phishing data—the datasets were assigned the proper sentiment labels for phishing.

ii. ***Validation set:*** A part of the dataset intended to assess the model's learning performance during training is known as a validation set. Before final testing, it analyzes model performance, avoids overfitting, and aids in hyper-parameter tuning.

iii. ***Testing set:*** A portion of the dataset put aside for the model's last testing. The testing step made use of the remaining 10% of the dataset. In order to assess the Roberta model's performance during the testing period, unlabeled data was given to it. At the conclusion of the coding process, the model's accuracy was also evaluated to see how well it had been trained.

Step 6: Load pre-trained Roberta Model. A pre-trained Roberta model has been loaded using Hugging Face's Transformers library. The model has already been trained on a variety of general texts, including stories, CC news, and OpenwebText, English Wikipedia, and book corpuses. This project has attempted to use datasets related to cyberbullying and phishing to train a pre-trained model. The model was altered for categorization in order to achieve this.

Step 7: Define Training arguments. Arguments are the values that are passed to the function in question when it is called; they are sometimes abbreviated as **"args"** in Python documentation. The arguments used to train a machine learning model are known as training arguments. Positional arguments and keyword arguments are the two kind of arguments that Python functions can accept. Only keyword arguments, often abbreviated as **"kwargs"**, are accepted by Hugging Face's Training Arguments, and they are passed to them within the Trainer class rather than directly. Keyword arguments are arguments that are equal to the sign (=) and accompanied by a keyword and provided to a given function or method. Since the values have been explicitly assigned, it makes no difference what order a keyword argument is in relation to another keyword argument.

Step 8: Set Hyperparameters for fine tuning Roberta. Configurable factors known as hyperparameters affect a model's learning process during training. Hyperparameters are established prior to training, in contrast to model parameters (such as weights and biases, which the model learns). The following hyperparameters are crucial for adjusting a Roberta model:

i. ***Learning Rate:*** During training, this hyperparameter lets you regulate the pace at which the model updates its weights. It is required because a model might fail to converge when the weights are set too high, and it may learn too slowly if they are set too low. Accordingly, the common learning rate values for Roberta are 2e-5, 3e-5, and 5e-5. The current study's training procedure makes use of 2e-5.

ii. ***Batch Size (per device train batch size):*** This determines the maximum number of samples that can be handled prior to model weight updates. While a smaller batch size necessitates higher-quality training but takes longer, a larger batch size leads to faster training but requires more memory (GPU/TPU). According to the GPU capabilities, the ideal batch sizes include 8, 16, and 32. Batch size 16 has been employed in this model.

iii. ***Number of Epochs:*** An epoch is when the model runs through the whole training dataset. The model views every training sample once throughout each epoch and modifies its weights to enhance predictions. Roberta is often fine-tuned over 3-5 epochs. The epoch number in this model has remained at 3.

iv. ***Weight Decay:*** A regularization technique called weight decay penalizes excessive weights in the model in order to avoid overfitting. Another name for it is L2 regularization. By modifying its weights while training, a model discovers patterns in the data. Overfitting (high efficacy on training data yet poor generalization) can occur when the weights in the model become excessively big due to an over-reliance on specific attributes. Weight decay encourages the model to maintain small weights and improves generalization to unknown data by adding a tiny cost to big weights. Weight decay levels of 0.01 and 0.1 are typical. The value of 0.01 was used in this model. By including a penalty term proportionate to the total of squared weights, weight decay alters the loss function:

$$\text{Loss} = \text{Original Loss} + \lambda \sum w^2$$

Where w is the model weights and λ is the weight decay parameter, which regulates the strength of regularization. An excessively high λ could result in an underfit model whereas an excessively low λ could cause the model to overfit.

v. ***Gradient Accumulation:*** Another hyperparameter that enables training models with longer batch sizes than a device's GPU memory is gradient accumulation, which builds up gradients across a number of steps prior to updating the weights.

For instance, each batch instantly updates the model if batch size = 64 and gradient accumulation is not present. However, batch size 64 can be imitated using collecting gradients if a researcher can fit 16 samples in memory but not 64. If a model is working with higher batch sizes, gradient accumulation is required because it can mimic a larger batch by building up gradients, which increases model stability.

Step 9: Train the Model using Hugging face Trainer API. A code was generated for importing Hugging face Trainer API for finally starting the training process. After five hours, the trained Roberta model was ready to use.

Step 10: Model Evaluation. Evaluation has been done to assess the model on the test set after training is finished. Depending on the type of problem (in this case, sentiment analysis), multiple metrics are employed to evaluate a fine-tuned Roberta model. The accuracy, F1-score, precision, recall, and confusion matrix are the main metrics that have been employed in this study to assess the Roberta model. A model can be evaluated in two ways. There are two ways to achieve it: manually using formulas, or automatically using the Hugging Face API. Hugging Face Trainer's API offers an evaluation library called "evaluate" for these measures, which is used to fine-tune the Roberta model. Calculators for accuracy, precision, recall, F1 score, and confusion metrics are already included in the Hugging Face API. Because it already uses the Hugging Face API for model training, the current paper has chosen to use auto-calculators. After loading each metrics function from "evaluate," the test set was executed.

i. ***Accuracy.*** Accuracy quantifies the percentage of accurately predicted labels among all samples.

$$\text{Accuracy} = TP+TN+FP+FN/TP+TN$$

In this case,

- TP (True Positive) → accurately predicted positive samples
- False Positive (FP) → erroneously anticipated positive samples
- FN (False Negative) → mistakenly predicted negative samples
- TN (True Negative) → properly predicted negative samples

ii. Precision. Precision indicates the proportion of correct projected positives.

$$\text{Precision}=TP+FP/TP$$

iii. Recall. This indicates how many true positives were anticipated accurately.

$$\text{Recall}= TP+FN/TP$$

iv. F1-Score (Harmonic Mean of precision & Recall). This score helps with imbalanced data by balancing accuracy and recall.

$$F1=2\times \text{Precision} + \text{Recall} / \text{Precision} \times \text{Recall}$$

Step 11: Model Deployment. After training, model has been saved and tokenized for later use.

3.6.2 Framework for Linguistic Analysis

Linguistic analysis is an intricate field of study that uses a variety of instruments, processes, and technical terms to investigate language on phonological, morphological, syntactic, semantic, and pragmatic levels (Laghari, 2024). Often called forensic linguistics or cyber profiling, the analysis of language in cyberspace aids investigators in comprehending the motives and operational patterns of cybercriminals and even in identifying possible culprits (CounterCraft). Experts can uncover important hints about an offender's social background, cultural influences, and even geographic location by looking at particular linguistic features like word choices, slang, grammatical errors, and repetitive patterns they use regularly. This information enables them to create strong criminal profiles. The present research study has adapted this strategy in analyzing the language used by cybercriminals in case studies that has been collected in this research to provide valuable insights into the social, educational, cultural and psychological background of the cybercriminals and consequently in determining the overall behavior of these criminals. The results of the study have been summarized to obtain the expected criminal profiles of cybercriminals.

CHAPTER 4
DATA PRESENTATION AND ANALYSIS

4.1 Results Obtained from Sentiment Analysis

Roberta model was trained on large corpus of social media “Cyber harassment” and “Phishing” labelled data separately. The data was classified into three categories “positive, negative and neutral sentiments” for cyber harassment data while Phishing data was assigned only two categories: Positive for Safe emails and Negative for Phishing emails. After the training, the remaining 10% of the data (unlabeled data) obtained from different social media platforms including WhatsApp, Email and Twitter has been used for the testing phase. Both the models have successfully classified the text into their appropriate assigned categories with their confidence scores.

Table 2

Results of the Model Trained for Cyber Harassment Data

Epoch	Training Loss	Validation Loss	Accuracy
1	0.245100	0.410863	0.900000
2	0.153600	0.260466	0.941423
3	0.253900	0.252897	0.941004

As already been discussed in the methodology session, this model has been evaluated on three epochs using training loss, validation loss and accuracy of the trained model as key evaluation metrics. In the first epoch, during training session, a training loss of 0.245100 percent and a relatively high rates of validation loss indicates that model is still learning in learning phase and has not acquired proficiency in generalizing all of the validation set. However, achieving 90% accuracy in the initial stages of training shows a good start. The low rates of training loss and validation loss in the second epoch accompanied with a 94% accuracy indicates that the performance of the training model is learning effectively and is now improving itself in generalizing new and unseen data.

By the time, the model reached its third epoch, the training loss became stabilized but the validation loss has further decreased with a constant accuracy of 94%. It has been concluded from the above results that the model may have reached a plateau and is no longer exhibiting notable gains, as evidenced by the minor change in validation measures between the second and third epochs. This could mean that more training would result in declining returns and increase the danger of overfitting. All things considered, these findings show that the model performed well in a comparatively limited number of epochs, suggesting efficient learning and generalization.

In addition to accuracy, the model has further been evaluated using Precision, Recall, and F1 score evaluation metrics to check how accurate model has been trained in classifying a text across three sentiment categories including Positive, Negative and Neutral. Below are the scores that has been obtained from the evaluation of the trained Roberta model.

Table 3

Evaluation metrics of cyber harassment model

	Precision	**Recall**	**F1-score**	**Support**
Positive	100.00	100.00	100.00	11.00
Negative	100.00	20.00	33.33	5.00
Neutral	69.23	100.00	81.82	9.00
Accuracy	84.00	84.00	84.00	0.84
Macro avg.	89.74	73.33	71.72	25.00
Weighted avg.	88.92	84.00	80.12	25.00

The results obtained from the trained Roberta model on cyber harassment datasets has demonstrated that the model shows an excellent performance in classifying positive sentiments but struggles when it comes to negative sentiments because of the underlying sarcastic tone that usually involves in certain harassment messages. The poor performance of the negative category, as evident by the macro average F1-score of 71.72%, is due to an unweighted average across all the classes.

In the more common Positive and Neutral classes, the model performed well overall, as indicated by the weighted average F1-score of 80.12%, which takes into account the support (number of cases per class). Although the model is successful in detecting positive and neutral feelings, these findings suggest that more negative sentiments are required to increase its efficiency in classifying negative sentiments.

4.1.1.2 Critical Interpretation of Cyber Harassment Model Performance

1. High accuracy but imbalanced class performance. The overall accuracy of the model shows that 84% of the predictions were accurate which means that model performed well at a general level. However, it would be deceptive in this situation to use accuracy as the only evaluation metric because it obscures the uneven performance among sentiment classes. Such differences in class performance requires further investigation given the distorted character of cybercrime speech, where emotionally manipulative language may be nuanced or context-dependent.

2. Performance evaluation of classes.

i. **Positive class: (**Precision = 100%, Recall = 100%, F1 = 100%)

The 'positive' class in the model exhibits perfect classification, indicating that positively framed or emotive language is easily distinguished—probably because of its clear lexical properties and less ambiguous emotional tone.

ii. **Negative class: (**Precision = 100%, Recall = 20%, F1 = 33.33%)

This class is the most problematic. The recall is a startlingly low 20%, which means that even while the model achieves perfect precision which means that when it does predict "negative," it's correct), however, it misses 80% of truly negative instances. This is crucial from a criminolinguistic perspective because, in cases of cyber harassment particularly, negative attitude frequently conveys signs of aggressiveness, threat, or manipulation. The model's applicability in actual detection situations is limited by its inability to appropriately capture them. This deficiency could be caused by overlapping linguistic features between negative and neutral phrases or by implicit hostility (such as indirect threats or sarcasm) that is difficult for surface-level features to capture, even in a transformer model.

iii. Neutral Class: (Precision = 69.23%, Recall = 100%, F1 = 81.82%)

'Neutral' is over predicted by the model, as evidenced by the high recall but low precision. This might indicate a preference for conservative classification, avoiding false positives in "negative" or other high-risk categories. Additionally, it implies that the model might be insensitive to minor emotional or aggressive indications, especially when those indicators are concealed by neutral or formal language, which is a common strategy in sophisticated online harassment or phishing communications.

iv. Macro and Weighted Averages:

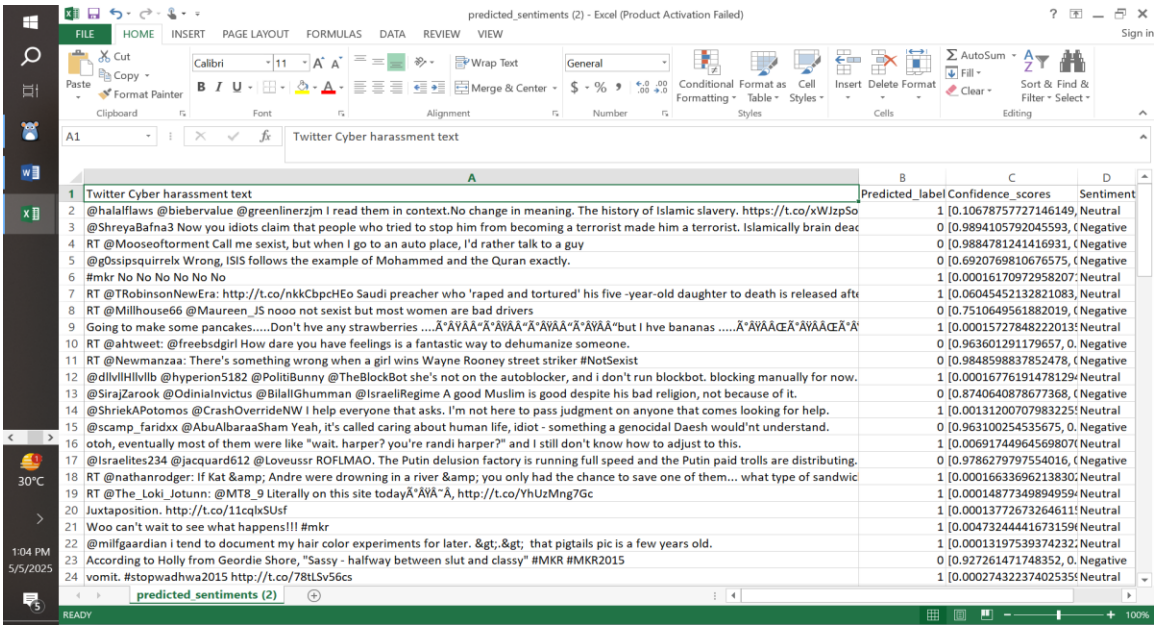
When each class is treated equally, performance significantly declines, as evidenced by the macro-average F1 score (71.72%) and recall (73.33%), which supports uncertainties about class imbalance. The overall average F1-score of 71.72% indicates that the negative category performed poorly, however the weighted averages (F1 = 80.12%) are higher, indicating the influence of the better-performing "positive" and "neutral" classes. This is because all of the classes had an unweighted average. The weighted average F1-score of 80.12%, which accounts for the support, shows that the model did well overall in the more popular Positive and Neutral classes. These results imply that more negative sentiments are needed to improve the model's classification efficiency for negative sentiments, even though it is effective at identifying positive and neutral emotions.

4.1.1.3 Evidence of Model Strengths and Achievements

When the Roberta model is used on the cyber harassment dataset, it successfully captures violent language and contextual emotional undertones, confirming its applicability in identifying abnormal behavior in online conversation. Below are the summarized results of its performance on cyber harassment unlabeled data.

Figure 5

Roberta Model's predicted Cyber harassment text



1. **High confidence in identifying hate speech and direct aggression:**

With very high confidence, a number of tweets were accurately classified as negative sentiment. For Example:

"Now you idiots claim that the person who tried to stop him from becoming a terrorist made him a terrorist..."

This statement is predicted as "negative" by the model with a 0.9894 confidence level. Similarly, another statement **"Call me sexist, but when I go to an auto place, I'd rather talk to a guy"** is predicted "negative" accurately with a confidence score of 0.9884. These instances show how sensitive the model is to aggressiveness and biasness that are linguistically encoded, even in cases where hate speech is not overt. Roberta comprehends contextual language and implied offence, which is crucial in online harassment since hostility is typically indirect, in contrast to traditional ML classifiers that frequently focus on keyword recognition.

2. **Effective Management of Sarcasm and Offensive Humor:**

The model accurate depiction of the statement **"How dare you have feelings. That's a fantastic way to dehumanize someone"** as negative with a confidence score of 0.9630 shows its successful handling of sentences that contain sarcasm or humor. This sentence employs passive-aggression and sarcasm, which are frequent strategies in online abuse. Despite the mocking tone, Roberta correctly deciphers the underlying negative sentiment, demonstrating that the model is not just able to identify explicit emotional markers but also implicit animosity, which is something that conventional sentiment models frequently miss.

3. **Context-aware sentiment prediction:**

A few tweets were categorized as neutral, such as:

"RT @ShriekAPotomos I help everyone that asks. I'm not here to pass judgment…"

"RT @dllvlllHlllvlb... blocking manually for now."

The model effectively avoids over-predicting negativity in factual or non-aggressive utterances, demonstrating its capacity to distinguish between opinion, sarcasm, and actual animosity. Compared to older models, which frequently highlight any strong words as negative without understanding context, this is a significant accomplishment.

4.1.1.4 Results Alignment with Integrated Theoretical Framework

According to the Routine Activity Theory, deviant behavior happens when three essential components come together: a suitable target, a motivated offender, and the lack of a capable guardian. The online spaces where the cyber abuse tweets originated frequently lacked social responsibility, identity verification, and moderation—all of which are prerequisites for digital deviance. This opinion is supported by the Roberta model's findings. Strong negative sentiment, including implicit bias, personal insults, and social antagonism, was detected by the algorithm in tweets. For instance, tweets like "Call me sexist, but" and "You idiots claim that" were tagged as negative with high confidence, suggesting that the offenders were not just voicing their opinions but were using targeted, intentional verbal abuse without any rules or repercussions.

Hence, the model demonstrates how offender motivation is reflected in language behavior and how low-guardianship digital areas serve as fertile ground for harassment—a crucial insight from RAT.

According to Culpeper's Impoliteness Theory, verbal aggression is classified as a strategic face-threatening behavior that is frequently used through assumption, sarcasm, fake politeness, or implied offence. These nuanced linguistic aspects were successfully captured by the Roberta model. Even though they lacked outright threats or explicit profanity, tweets that employed gendered stereotypes, dismissiveness, or mock sincerity were appropriately classified as negative by the cyber harassment model. In contrast to lexical sentiment, this illustrates the model's sensitivity to pragmatic cues of impoliteness. The model's output supports the idea that socially coded or context-dependent language is frequently used in cyberbullying. For example, the model correctly identified negative sentiment in a statement such as "Now you people want to pretend…" because it conveys implicit hostility and group-based othering. According to impoliteness literature, this implies that Roberta has internalized the pragmatic aspects of online abuse.

Sentiment analysis's efficacy in classifying harassment content lends more credence to the notion that emotionally charged language is a sign of abnormal online conduct. The model results serve as a link between the two theories, with Impoliteness Theory decoding the language acts of hostility and Routine Activity Theory addressing the situational conditions for crime:

- Acts of domination, exclusion, or threat are linguistic manifestations of emotionally negative content.
- These behaviors are frequently carried out in situations when social norms are either nonexistent or weak.

The great confidence with which the algorithm was able to identify these behaviors demonstrates that language itself turns into a behavioral signal, reflecting both environmental vulnerability (from the platform context) and intent (from the perpetrator).

Table 4

Results Shown by the Model Trained for Phishing Data

Epoch	Training Loss	Validation Loss	Accuracy
1	0.091700	0.168104	0.965561
2	0.013500	0.083222	0.979420
3	0.052000	0.071800	0.981100

Note. The performance of a Roberta based sentiment analysis model trained on Phishing dataset has been evaluated using the evaluation metrics training loss, validation loss and accuracy over three epochs.

In the first epoch, the model has achieved training and validation loss of "0.0917" and "0.168104" respectively along with a high accuracy rate of 96.55%. This indicates the capability of the trained Roberta model to capture relevant patterns during the learning process. In the second epoch, a significant decrease in training and validation loss with increased accuracy of 97.94% suggests that the model performance has greatly been improved in generalizing unseen data.

By the third epoch, a slight increase in the training loss indicates that the model is adjusting its weights that ultimately contributed to achieve an accuracy of 98.11%. Overall, these findings demonstrate how well the trained Roberta model captures the intricate linguistic patterns for sentiment classification. As seen by the consistent increase in accuracy across epochs and the steady drop in validation loss, the model has obtained a high level of generalization by avoiding overfitting. This makes the model appropriate for real-world applications requiring sentiment analysis of content connected to cybercrime.

Table 5

Evaluation metrics of Phishing model

	Precision	Recall	F1-score	Support
Safe email	100.0	100.0	100.0	17.0
Phishing email	100.0	100.0	100.0	32.0
Accuracy	100.0	100.0	100.0	1.0
Macro avg.	100.0	100.0	100.0	49.0
Weighted avg.	100.0	100.0	100.0	49.0

4.1.2.1 Critical interpretation of the Phishing model performance

1. Excellent Predictive Accuracy:

On the test set, the Roberta model obtains 100% precision, recall, and F1-score for both safe and phishing emails, showing:

- **Precision (100%):** It means that there are no false positives in the test set; every email that is identified as phishing is, in fact, phishing.
- **Recall (100%):** It means that there were no false negatives in the test set; the model correctly identified every phishing email in the test data.
- **F1-score (100%):** It means that recall and accuracy are perfectly balanced.

This implies that the model fits and generalizes the present dataset exceptionally well. It validates that refined transformer models, such as Roberta, are appropriate for phishing detection, since linguistic manipulation frequently include misleading, urgent, and persuasive cues that Roberta may pick up on through context-rich embeddings.

2. **Justification for Roberta's Superiority:**

In contrast to previous models that used keyword-based filtering, SVMs, or Naive Bayes, Roberta gains knowledge of linguistic nuances such as semantic incongruence, scam urgency, and persuasive politeness.

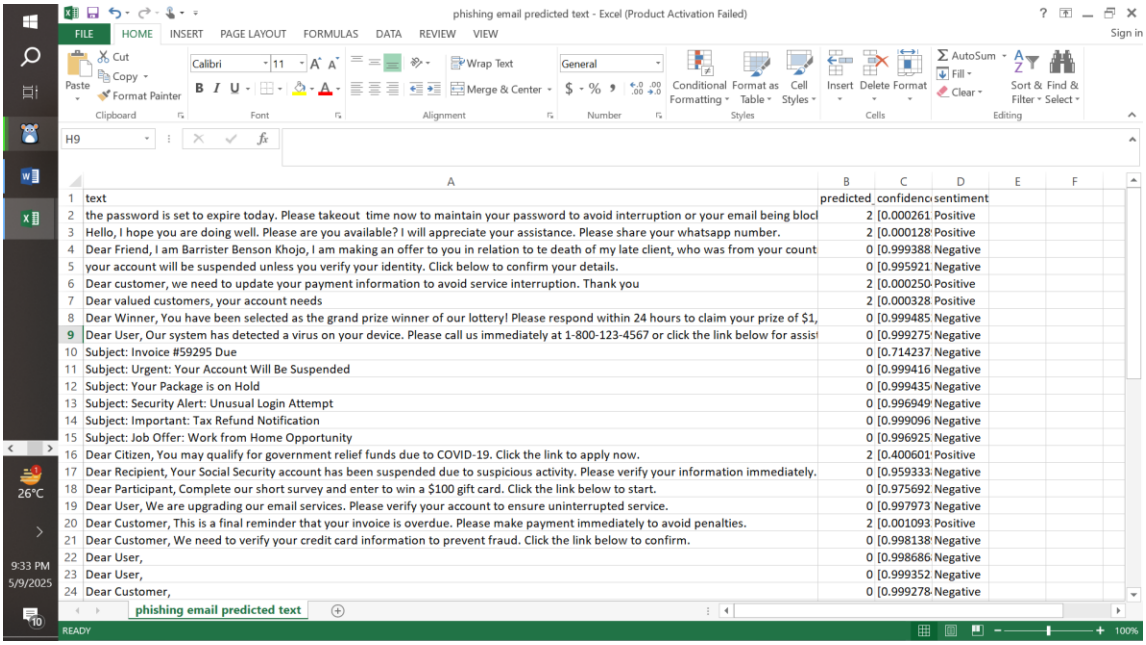
Moreover, it carries out context-aware classification, which improves its ability to detect complex phishing schemes that don't depend on obvious clues.

3. Context-aware sentiment classification:

Phishing alerts such as "Dear User, we detected a virus" and "Your account will be suspended" were appropriately identified by the model. On the other hand, sentences that were informal and non-threatening, such as "Going to make some pancakes," were appropriately evaluated as safe. This shows that the model is actually comprehending context and tone, something that previous models have continually failed to achieve, rather than just pattern-matching spam sentences.

Figure 6

Roberta Model's predicted Phishing text



	text	predicted_	confidenc	sentiment
2	the password is set to expire today. Please takeout time now to maintain your password to avoid interruption or your email being bloc	2	[0.000261	Positive
3	Hello, I hope you are doing well. Please are you available? I will appreciate your assistance. Please share your whatsapp number.	2	[0.000128	Positive
4	Dear Friend, I am Barrister Benson Khojo, I am making an offer to you in relation to te death of my late client, who was from your count	0	[0.999388	Negative
5	your account will be suspended unless you verify your identity. Click below to confirm your details.	0	[0.995921	Negative
6	Dear customer, we need to update your payment information to avoid service interruption. Thank you	2	[0.000250	Positive
7	Dear valued customers, your account needs	2	[0.000328	Positive
8	Dear Winner, You have been selected as the grand prize winner of our lottery! Please respond within 24 hours to claim your prize of $1,	0	[0.999485	Negative
9	Dear User, Our system has detected a virus on your device. Please call us immediately at 1-800-123-4567 or click the link below for assis	0	[0.999275	Negative
10	Subject: Invoice #59295 Due	0	[0.714237	Negative
11	Subject: Urgent: Your Account Will Be Suspended	0	[0.999416	Negative
12	Subject: Your Package is on Hold	0	[0.999435	Negative
13	Subject: Security Alert: Unusual Login Attempt	0	[0.996949	Negative
14	Subject: Important: Tax Refund Notification	0	[0.999096	Negative
15	Subject: Job Offer: Work from Home Opportunity	0	[0.996925	Negative
16	Dear Citizen, You may qualify for government relief funds due to COVID-19. Click the link to apply now.	2	[0.400601	Positive
17	Dear Recipient, Your Social Security account has been suspended due to suspicious activity. Please verify your information immediately.	0	[0.959333	Negative
18	Dear Participant, Complete our short survey and enter to win a $100 gift card. Click the link below to start.	0	[0.975692	Negative
19	Dear User, We are upgrading our email services. Please verify your account to ensure uninterrupted service.	0	[0.997973	Negative
20	Dear Customer, This is a final reminder that your invoice is overdue. Please make payment immediately to avoid penalties.	2	[0.001093	Positive
21	Dear Customer, We need to verify your credit card information to prevent fraud. Click the link below to confirm.	0	[0.998138	Negative
22	Dear User,	0	[0.998686	Negative
23	Dear User,	0	[0.999352	Negative
24	Dear Customer,	0	[0.999278	Negative

4. Critical Evaluation of the Findings:

The results obtained from the test data could not be taken as a definitive proof of model superiority. Even though they are ideal, perfect grades should be carefully considered because overfitting occurs when the training and test data have structural or lexical similarity. Moreover, it is possible that the dataset's phishing examples don't accurately capture the variety and intricacy of actual attacks.

Tests of the model's resilience to adversarial inputs or more ambiguous messages that combine malicious and safe intent have not yet been conducted. Since the style, sophistication, and social engineering techniques of phishing emails vary widely in practice, this reflection is crucial. The efficiency of the model on the provided dataset is thus confirmed by the current results, but more external validation is required.

5. Results Alignment with Integrated Theoretical Framework:

The comprehensive theoretical framework used in this study is highly consistent with these findings. The model's effectiveness indicates that hackers' (motivated offenders') discourse follows recognizable patterns that may be computationally modelled, according to Routine Activity Theory (RAT). The Roberta model's detection of phishing communications consistently reflects the strategies of motivated criminals, including deceit, urgency, and social engineering. Phrases such as "Your account will be suspended..." or "Claim your prize now..." demonstrate efforts to take advantage of victims' weaknesses when they are engaging in routine online activity. RAT's presumption that criminal behavior is opportunistic and linguistically strategic is supported by Roberta's capacity to contextually identify manipulative intent in these messages.

Understanding cybercriminal behavior through linguistic cues depends on the model's capacity to distinguish between manipulative politeness (e.g., "Dear Customer, your account will be closed...") and real politeness. Politeness Theory also clarifies how phishing actors conceal their intentions by using politeness at the surface level (e.g., "Dear Customer") and concealing coercive or threatening content beneath. The linguistic duality of phishing texts—where face-saving techniques are employed to reduce suspicion and evade detection—is highlighted by the model's ability to recognize negative emotion inside polite framing. It supports the idea that language serves as a tool for carrying out deviant activities in addition to being a medium of communication.

6. Comparative Advantage over Conventional Models:

These results further support the choice to abandon conventional machine learning models, such as SVM and Naive Bayes, which usually rely on bag-of-words assumptions and static features.

In contrast to other methods, Roberta uses deep contextual embeddings to identify intention and sentiment in larger discourse patterns. In cybercrime research, where language is frequently designed to evade keyword filters and take advantage of user psychology, this capacity is especially helpful. To conclude, the Roberta model's superiority over conventional machine learning models for cybercrime analysis is empirically validated by the study's results.

Roberta showed great confidence in detecting coercive and frightening language in phishing emails, even when the information was presented in a formal or courteous manner. This is an area where traditional models frequently fall short. In a similar vein, the model demonstrated its capacity to comprehend context and pragmatic complexity in cyber harassment detection by successfully identifying indirect aggressiveness, sarcasm, and coded hostility. Furthermore, by including confidence scores, which are usually lacking in earlier classification techniques, Roberta provides interpretive granularity that enables researchers and security systems to evaluate prediction dependability. The model is positioned as a very powerful tool for integrated behavioral profiling in cybercrime situations due to its cross-domain adaptation, which further highlights its resilience and usefulness.

7. Limitations of the Roberta Model:

The results obtained from sentiment analysis of cyber harassment messages has demonstrated that although the model shows an excellent performance in classifying positive and neutral sentiments but struggles when it comes to negative sentiments because of the underlying sarcastic tone that usually involves in certain harassment messages. These findings highlight the importance of combining discourse-level linguistic interpretation with sentiment analysis. By combining the theories of politeness and impoliteness, this study shows that hackers frequently use performative politeness as a deception to avoid being discovered by sentiment-based classifiers alone. Hence, it has been observed from the results that sentiment analysis alone is not sufficient to explain the phenomenon under study. Therefore, a combined analytical framework that involves both sentiment and linguistic analysis is a requirement that has already been met by this study.

4.2 Results obtained from Linguistic Analysis of Case Studies

4.2.1 Case Study # 1

1. **Use of Politeness Strategies:**

Replacing the victim's name with a generic "Dear User" is a clear indication that the sender isn't real. They usually address you as "Dear Jane Smith," so that gives away whether you are talking to an automated service or a person.

2. **Establishing a Sense of Importance:**

The term "valued journalist" was designed to give flattery to the take so that the journalist feels very important and appreciated. "Your account is essential for our community" is also designed to add responsibility to the recipient. Its supper tries to make the taker feel trusted and encouraged. They are been told that their account is vital in the community. This suggestion adds a lot of seriousness to the email. Pouring praise reduces the recipient's defenses and makes them more likely to comply.

3. **Knowing Target Credentials:**

By tagging them as a journalist, the email tries to give a predefined sense of targeting. This means that the person who has composed the message intends this exclusively and so is likely to be known at a far greater level than a random sender at some other company.

4. **Fake Verification Links:**

'Click on the link below' sounds suspicious. Authentic companies do not request that you click on any links found in emails that you did not specifically request.

5. **Hints of Phishing:**

The email tries to prompt action without a second thought by creating urgency (immediately, failure to comply may result in suspension), which is commonly used in phishing attempts.

6. **Requesting Information that is Out of the Ordinary:**

In attempting to fish for sensitive information like phone number or account details, the scammers will usually ask these questions.

7. **Basic Literacy and Writing Skills:**

Phishing emails are sent with many grammatical mistakes or unprofessional phrasing. Although this message seems fairly neat and structured, suggesting that the scammer has basic literacy and writing skills, his ability to construct an email also demonstrates a particular level of education and a focus on writing and communication. This suggests that he has obtained at least secondary education or its equivalent.

8. **Cultural Knowledge:**

The scammer probably possesses some level of social competence which allows them to emotionally manipulate the victims with ease. So, it demonstrates an experience on human behavior and exploiting vulnerabilities. The expression “valued journalist” alongside references to community shows that there are certain phraseologies which suggest western culture, especially around professional and social belonging. This indicates that the origin of the scammer’s values displays those.

9. **Language Skills:**

The polished language indicates that the scammer is proficient in English whether structurally ascribed to being a native or having been in English speaking countries. The scammers’ level of comfort with online environments and subtleties of online language reveals that they are also likely to be somewhat technology-literate as per the standards of the society, a trait that in many cases is linked with education and access to resources.

10. **Psychological Health:**

The scammer is likely to be manipulating in nature and could carry narcissistic traits as they put their benefit over the well-being of others. Second, by engaging in illegal behavior such as phishing the scammer has taken a risk and taking risk may, in turn, be associated with a variety of underlying psychological variables, such as thrill-seeking or impulsiveness.

4.2.1.1. Integrated Analysis:

The above case study clearly demonstrates how the convergence of a motivated offender "a cybercriminal", a vulnerable user who is "a verified journalist" in this case and poor cybersecurity measures results in a cybercrime. The scammer's demonstration of knowing the identity of the user by mentioning his or her professional status make the scam seems more reliable. The urgency and immediate tone of the message lead the victim's anxiety to act in a spur of moment in order to save their account from losing the verified status, hence, making the victim a suitable target. Moreover, the victim's trust on the reliability of "Twitter" and lack of awareness about such spam emails has resulted in the cybercrime.

4.2.1.2 Estimated Criminal Profile

All in all, this phishing email has provided useful insights into the social, cultural, educational, and psychological context of the scammer. Apparently, the scammer can construct a sentence and elicit an emotion which demonstrates that he or she is probably educated and socially-savvy to some degree, yet being so cold-hearted and morally bankrupt to exploit a person indicates some serious psychological issues.

4.2.2 Case Study # 2

1. Use of Politeness:

The message starts with an informal greeting which indicated a sense of friendliness and positivity. Cybercriminals usually use this tactic in their conversations to build a rapport with the victims in order to lure them into their sweet talk. The use of "Please" and "Thanks" also indicates positive politeness and may be seen as a cultural norm that values "respect" and "kindness" usually found in eastern culture.

2. Tone of the Message:

The message contains a formal introduction which communicates a sense of professionalism. It also signals towards a certain level of educational background; however, it is a general way of introduction across various contexts.

3. Requests for Information:

It is unusual for a courier company to ask for biometric identification. This unprofessional way of asking for personal information makes it further suspicious and unauthentic. Moreover, by calling it a standard procedure scammer has tried to normalize this unusual request.

4. Urgency and Incentive:

The urgent nature of the message compelled the victim to act hastily without confirming the authenticity of it. In addition to this, the incentive of "having the parcel delivered right away" further lure the victim, who is waiting for the parcel eagerly, to comply with these unusual requests of information.

5. Use of Manipulative Language:

In order to trick the victim, scammer has mentioned the phrase "to ensure safe delivery deliberately in the text to create a fake sense of security in the mind of the victim. Likewise, the use of the term "cooperation" further implies compliance to the request.

6. Hints of Phishing:

Given that the message is a formal message sent by a courier company, the informal greeting in the start of the message looks awkward. Moreover, the request for a thumbprint also seems out of place and unprofessional as legitimate companies usually do not ask for such type of sensitive information.

4.2.2.1 Integrated Analysis

This cybercrime occurs through a combination of a motivated offender (an impersonating courier person), a suitable target (a person who is eagerly waiting for his delivery) and insufficient security measures for identifying a phishing request. The urgent and polite nature of message has made the victim vulnerable to this scam. Moreover, the reassurance of authenticity with the addition incentive of having the parcel delivered on time has increased the likelihood of a victim to accept the message as authentic without deep consideration.

Furthermore, the lack of any security measures to identify a text message as potentially a scam has further accelerated the process of such types of Phishing attempts.

4.2.2.2 Estimated Criminal Profile

By keeping in view all the above linguistic cues, it is not possible to generalize the cultural and educational background of the scammer. The level of formality politeness strategies that has been used in the above settings are a basic knowledge that many individuals possess without having any type of education. However, the absence of a courier company name and unusual requests of information are the strong indicators of a potential phishing attempt.

4.2.3 Case study # 3

1. Positive Politeness:

This is a common tactic used by the scammers to call the victims with alluring salutations like “Dear friend” to develop a friendly connection with the victim and make them comply to their requests readily and easily.

2. Manipulating Emotions:

By mentioning a deceased client and unsuccessful attempts of finding his relations, the scammer has tried to evoke an emotional response from the victim that is to comply with his request out of sympathy.

3. Dubious Credentials:

The scammer has pretended to be a barrister in the text message, which may be a tactic to let the victim believe in his ridiculous talk. Scammers typically use fake and reputable identities in phishing scams to appear to be more real. The term Barrister is typically used in British legal contexts so there can be a possibility that the scammer has a UK origin or maybe he belongs to a country which follows British English in academic settings.

4. Writing Errors:

Although the message implies a formal tone, yet the presence of grammatical errors like “te” and “yo” and the use of “martial” instead of “marital” indicates a lack of basic writing skills in English language. Grammatical errors are common in scam messages. Errors that have been spotted in this scam message suggests that writer is from a not a native speaker of English.

5. Asking for a list of information:

It is quite suspicious that the scammer has asked for a whole list of personal information including mobile number, age, account details etc. This request for a whole set of personal information is out of ordinary. A person who claimed to be a barrister must be aware of the fact that asking for a whole lot of personal information is insane. This attempt has further make the professional status of the sender suspicious.

6. Cultural Background:

It is a popular opinion that West Africa has a history of such fraudulent Phishing scams which are usually based on a deceased client leaving huge amount of money behind. Taking this point of view, it can be assumed that scammer may be of a West African origin or may have some connection with West African communities.

4.2.3.1 Integrated Analysis

The combined analysis of all the cues from the above spam email confirms the presence of a motivated offender in the form of an impersonating barrister, a suitable target who is facing financial pressures and seeking a financial opportunity to prosper and lack of the availability of appropriate safety procedures which collectively makes an environment conductive to cybercrime. Through the use of both politeness (friendly tone) and impoliteness strategies (unusual requests) offender has made the victim comply with his requests. The emotional and opportunist tone of the email puts psychological pressure on the victim to agree with the offender’s offer.

4.2.3.2 Estimated Criminal Profile

The combination of the certain words and phrases such as those used in the email implies a criminal intention when seen through the eyes of a forensic linguist. Although the use of the word barrister implies a British influence, yet, the prominent grammatical errors indicate lack of proficiency in English language, hinting at the fact that scammer might be from a regional background where English language is not strictly followed and used.

4.2.4 Case Study # 4

1. Casual Language:

Like all other Phishing scams previously analyzed in this research study, this message also starts with a friendly gesture. It means, this is a common tactic of phishers to use polite language to lure the victims into falling for their sweet lies. However, the use of an Islamic way of greeting indicates a possible regional or cultural influence. Either the scammer is a Muslim or he or she is living in a Muslim community.

2. Emotional Appeal:

The use of emotional language is another significant tactic seen in the phishing attempts. By using the phrases like “we, poor people”, “can’t afford to lose money” or “you can help us” sets an emotional tone. Innocent people who feel sorry for poor people are more likely to fall for this tactic.

3. Urgent Matter:

Urgency is a common way of letting the victims act without thinking. Scammers frequently combine emotional appeal with urgency to deceive innocent people first by taking sympathy and then by making them act irrationally, giving them no time to think about the authenticity of the message.

4. Hints of Regional Background:

Using the name of the mobile payment app that is currently working in Pakistan confirms his nationality.

Moreover, the hint of belonging to a small town in a message suggests that the offender is operating from a small town in Pakistan, which must be of low socio-economic status as previously observed in many phishing cases reported.

5. Requesting for an OTP (One time password):

Asking for an OTP is also a suspicious activity which confirms the fraudulent nature of this message. Professional companies usually do not ask for such information and if they want such type of sensitive information, they usually do not ask via unsolicited messages.

6. Writing Style:

The excessive use of an exclamation mark seems to be habitual nature of this scammer. Scammer might use them intentionally to deceive the cyber experts by adapting the writing style of another person who can serve as potential suspect.

4.2.4.1 Integrated Analysis

By applying the RAT theory, it can be stated that an environment conductive to exploitation has been created through the combination of three elements: a person disguising as an Easypaisa representative (motivated offender), a person who is well-aware of the digital transactions occurred via these mobile payment applications (suitable target) and ignorance about the existence of such scams because of the weak cyber security organizations. The use of politeness strategies and emotional blackmailing allowed the offender to build rapport with the victim and make them more likely to become their prey. The lack of awareness about such scams may give offender the confidence to commit such crimes and get succeeded.

4.2.4.2 Estimated Criminal Profile

All in all, the text involves multiple indicators of potential criminal intent, predominantly when seen in the context of a financial scam. These indicators include urgency, requests for sensitive information, and emotional manipulation which suggests an increased likelihood of fraudulent behavior.

The language used in the message and the repetitive references to culture followed by Asian countries indicates a South Asian context, specifically linked to Pakistan, where Easypaisa operates.

4.2.5 Case Study # 5:

1. Urgent Nature:

Phishers always made use of urgency in carrying out a successful phishing attempt. Without this tactic of urgency, they may likely fail to commit a crime because all irrational decisions are made spontaneously. If the criminals give plenty of time to the victims to think, victims can involve police in the matter, hence, there is an increased likelihood of getting caught. In this phishing email, offenders have made use of urgency tactic by giving a time duration of 2-3 days.

2. Financial Scam:

As offenders are aware of the fact that people are more prone towards a financial incentive, therefore, these phishing attempts mostly involve some kind of financial benefit like this scam. The offer of a tax refund of a huge amount “52,088.50 Pakistani Rupee” make a person dealing with financial crisis a suitable target.

3. Link to External Site:

The use of a malicious link which directed victims to fake websites is another sign of a phishing scam. The inclusion of a link (http://www.fbr.gov.pk/) without context confirms its fraudulent nature.

4. Ambiguous Language:

The use of the phrase “there could be variety of reasons for delays” shows vagueness. The official government websites usually do not use such vague language. Moreover, the absence of receiver’s details also makes the email seems more suspicious.

5. Cultural References:

“Pakistani rupee” and “Federal Board of Revenue” are clear indicators of Pakistani origin, hinting at the operational base of the scammer.

6. Formal Tone:

Although the scammer has tried hard to make the email look formal yet he failed at formatting. The absence of bullet points and structured instructions has revealed the fake tone of the message.

7. Coercive Requests:

The mentioning of potential delays due to invalid records can be seen as a scammer tactic to get valid information by intimidating victims to delay the refunds for not complying with the request.

4.2.5.1 Integrated Analysis

This phishing attempt is made possible through the convergence of a cybercriminal seeking sensitive information using fake governmental websites, a victim who is in search for a financial opportunity to come and unavailability of security that allows the users to detect the difference between a potential scam and reality. Through the use of a fake official website, this email has tried to exploit the trust people have on official websites. By playing with the fear of losing a financial opportunity, the scammer has led victims to fulfill his demands within the given deadline, ultimately pressurizing them to act immediately without giving a second thought to it.

4.2.5.2 Estimated Criminal Profile

Urgency, absence of sender details, unstructured data are the powerful indicators of a phishing scam. The language usage and reference to Pakistani culture gives the clear indication of a Pakistani origin. Moreover, the construction of correct formal sentences indicates a certain level of educational background in legal contexts.

4.2.6 Case Study # 6

1. Informal Tone:

By direct addressing the victim by her name, the harasser has tried to convey a false sense of intimate connection between them.

The use of word “dear” creates an illusion that the harasser still wants to stir the victim’s emotions by making the threat feel more personal.

2. **Use of threats:**

Using threatening language is a common tactic of harassers to harass people. People in fear of getting their personal pictures or data leaked usually do not report such cases. In this case study, the harasser is threatening his ex-wife to post his photos from the intimate moments they have enjoyed together if he filed for a divorce from him.

3. **Ultimatums:**

The harasser has given an ultimatum of 6 days to the victim to think about the matter seriously and to reconsider her decision of taking divorce or else. This tactic has been used to put psychological pressures on the victim so that he or she can act irrationally and more probably in the favor of the harasser to avoid public humiliation.

4. **Emotional Blackmailing:**

The statement that “either I will take my life or yours” serves an emotional purpose. This is meant to make the victim feel emotionally damaged by thinking of the fact that someone will take his life because of her.

5. **Cultural Dynamics:**

The use of informal and direct tone in familial relationships conveys that the speaker belongs to South Asian culture as such practices are common in this culture. Moreover, the over-emphasis on exposing the private photos further confirms South Asian context in which family honor and reputation is considered a matter of life and death. The harasser is well-aware of the vulnerability of the victim that is his family honor and he is taking full advantage of it.

6. **Educational Background:**

The reference to a social media platform “Facebook” to expose photos indicates certain of level of technological awareness, yet the inappropriate use of such platform suggests a lack of understanding of the ethical boundaries.

7. **Mental Well-being:**

The aggressive and manipulative language that has been used throughout the message exhibits that harasser has anti-social and narcissist traits. Moreover, the threats of giving or taking lives indicates that harasser is mentally sick. This means that he has developed suicidal tendencies and can resort to extreme measures if his demands are not fulfilled. These threats are meant to instill fear in the victim, which is a common behavior of abuser in abusive relationships. Furthermore, the list of demands which the harasser wants to be fulfilled in a fixed period of time suggests impulsive behavior where the sender does not consider the consequences of his actions or behavior.

4.1.6.1 Integrated Analysis

The message illustrates the creation of an environment suitable for intimidation through the integration of a motivated offender (a harasser seeking revenge), a suitable target (an emotionally vulnerable victim) and poor cybersecurity which is often associated with cyber harassment cases. The use of manipulative as well as threatening language by the offender made the emotionally vulnerable victim a suitable target for this crime. The language creates an atmosphere of psychological pressure which can lead victim to take extreme measures such as a suicide attempt.

4.1.6.2 Estimated Criminal Profile

The text message contains numerous linguistic signs that can be enough to prove it to be a potential cybercrime. The reference to the social media platform indicates a certain level of technological education. Culturally, the harasser seems to be someone belonging to South Asian context where family honor is considered a valuable asset and must be protected by any means. Psychologically, he seems to be mentally sick and have extreme violent tendencies which can result in harming himself or others.

4.2.7 Case study # 7

1. Aggressive Tone:

The harassers have started the message with the demeaning language and has maintained it throughout the message to assert dominance and to exhibit control over the victim. Instead of addressing the person with a title or name, the use of a second person pronoun “You” indicates a lack of respect for the victim. It can be seen as a conscious attempt to reduce the person to the level of an object which suggests an abusive mindset.

2. Use of Slangs:

The use of informal words like “lil girl, BS, hoe” indicated a communication style that is common among youth now-a-days. It can be a useful hint about the age group of the sender. Moreover, it also communicates a lower educational background of the sender.

3. Direct Insults:

The use of the phrase "u are just a lil girl who don’t know her place" is an intentional attempt to snub the victim to make her lose his confidence.

4. Threatening Language:

The mention of threats time and again in the whole conversation is a tactic to instill fear in the mind of the victim. The threat of public humiliation by exposing sensitive information is another common tactic used in harassment messages like this one. The confrontational phrase "Think again" heightens intimidation and indicates aggressive intent. The phrase "wait and watch" implies already planned actions that will be taken against the victim if victim fails to comply with the harasser’s request. Last, the threat of ruining the reputation of the victim indicates a desire to damage the victim's social standing, a common motive seen in many harassment cases.

5. Anti-feminist:

The dismissive reference to feminism in the message indicates a cultural resistance towards women’s rights which can be an indicative of anti-feminist mindset or relation to a community having such attitudes towards feminism.

6. Power Dynamics:

The use of derogatory and threatening statements is indeed an attempt to exert dominance and control over the victim. The use of word “us” in the phrase “stop acting like u better than us” indicates a collective identity. By using this speaker has tried to position himself and his group against the victim and possibly his group, creating an in-group vs out-group power dynamic. By forcing the victim to “stop acting” shows that the harasser think that he has the power to dictate others, which ultimately showcasing his desire to have control and dominance.

7. Digital Literacy:

Through the demonstration of an understanding of digital platforms and image manipulation technique, harasser has conveyed that he is technologically literate. This technological literary, however, doesn’t indicate an educational background as it can be obtained with the mere engagement with contemporary digital trends without having any formal education.

4.1.7.1 Integrated Analysis

A person or possibly a group seeking revenge is the motivation behind this cybercrime and an outgroup person who is outspoken and confident is the ultimate target of the harasser. Together with the benefit of anonymity in the online communications creates an environment suitable to conduct a successful crime. The harasser’s use of threatening and coercive language to threaten the victim to remain silent was further aggravated by the insufficient reporting mechanisms. Consequently, the language creates an unusual pressure on the victim to comply with the requests or to face the consequences.

4.1.7.2 Estimated Criminal Profile

In conclusion, by keeping in view the informal and aggressive tone we can conclude that the harasser is of limited educational background. However, there are some evidences of digital literacy found in the text which indicates a certain level of engagement with the digital culture.

The distinctive attitude towards women's rights indicates anti-feminist mindset which can be seen a cultivation of a regional or cultural background which prioritizes male dominance.

4.2.8 Case Study # 8

1. Tone of the Message:

The message opens up with a complimentary note which communicates a sense of goodwill and creates avenues for further communication. Flattery is another most useful and powerful tactic used in harassment messages as females are more prone to such sweet talk.

2. Positive Critique:

The suggestion of "lighten up" and "smile more" are the indicators of a possible critique though framed as a friendly advice. This may be seen as patronizing attitude often observed in workplace environment as it implies that the person appearance is inappropriate and needs attention.

3. Initiative for Social Interaction:

The offer to grab a coffee sometime in the very first meeting suggests the speaker's desire to build rapport with the hearer. But the critique in the previous line also suggests that she should introduce some changes into his appearance to look more attractive and likeable.

4. Gender Dynamics:

The message reflects a common gender dynamic where a male colleague is suggesting a female colleague to change his appearance to look more presentable. This can be perceived as reinforcing gendered stereotypes that states women should be presentable in professional settings.

5. Blurring the Boundaries:

The use of the phrase “unwind a little” seems a bit awkward in a professional setting. It indicates that the person is deliberately trying to cross the line depending on the relationship between the speaker and the hearer.

6. Politeness Strategies:

The speaker has tried to maintain a positive face through the complimentary opening of the message. By using the phrase “just don’t take it in the wrong way”, he has tried to mitigate the effect of the critique he made, which is a useful politeness strategy often used in conversations.

7. Impoliteness Strategies:

Although presented as a suggestion, the implication that a person is "uptight" may be interpreted as an implicit criticism, thus undermining the teacher's authority as a professional. It can come across as patronizing to tell someone to smile more or how they should act, particularly if they didn't ask for the input.

4.1.8.1 Integrated Analysis

A motivated sender (the underlying motive is to engage with the female teacher), a suitable target (a female teacher who looks a bit disturbed and serious), and a lack of capable guardianship (the absence of a supportive network to report workplace harassment) can all combine to produce an environment in which uninvited advice can be given with a pleasant appearance but still have the implications of power dynamics and criticism. The message demonstrates the complex nature of interpersonal communication in work environments, where compliments can coexist with implicit criticism and gender dynamics can be affected by the interpretation of intent behind the message.

4.1.8.2 Estimated Criminal Profile

Overall, the tone of the message looks friendly and polite, yet the criticism which has been made in disguise of friendly advice implies a sense of judgmental and condescending tone. The profession of the person is quite obvious and hence indicates a good educational background.

The suggestive tone of the message also reflects the authoritative and dominant nature of the criminal which can be a perceived as a possible influence of male dominating culture.

4.2.9 Case Study # 9

1. Use of Rhetorical Questions:

The use of rhetorical questions throughout the message implies that the harassing is not seeking the answers but instead he wants to provoke a reaction. The context explains that the harasser has feel insulted on being rejected by a girl which has aroused the emotions of rage in him.

2. Derogatory Language:

The use of slangs and colloquial expressions indicates a low educational background. Moreover, the use of the words “cheap girl” suggests a misogynistic viewpoint which implies connection to a culture that devalues women.

3. Threatening Language:

By using repetitive phrase structures like “you think I am scared of law” or “You think you are too good for me” or “you think you can stop me” etc. suggests that the offender is trying to reinforce his authority and control over the victim and wanted to threaten the victim so that she shows compliance with his requests.

4. Power Dynamics:

In order to maintain control and influence, the harasser takes use of the victim's previous behavior, such as exchanging images. This illustrates a strong power dynamic in which the harasser feels entitled to set terms since they believe they own the personal data.

5. Deceptive Politeness:

The use of emojis and informal tone implies a false sense of friendliness and can be seen as a tactic to lower recipient’s defensive behavior.

4.1.9.1 Integrated Analysis

It is evident that the harasser is driven by the desire to exert control and dominance on the victim. The harasser has made use of manipulative and threatening language to induce fear and obedience in the victim's mind. On the other hand, recipient seems to be someone who has shared personal pictures with the harasser when they were in a relationship, making them vulnerable to the harassment. Lack of any support mechanism like family, victim support networks etc. may made the victim further vulnerable to this type of harassment. Sometimes, even in the presence of such support systems victims do not report such crimes and get exploited at the hands of harasser because of the fear of public humiliation which gives offender more confidence to conduct such crimes in future. Overall, as RAT theory suggests, that combination of these three factors a motivated offender, a suitable target and lack of security results in a cybercrime.

4.1.9.2 Estimated Criminal Profile

The dynamic depicted in the text is extremely distressing and is marked by aggressiveness, manipulation, and a flagrant disrespect for the victim's autonomy and welfare. The harassment's language suggests criminal intent, a possible lack of education, and serious psychological problems. This analysis emphasizes how serious these kinds of communications are and how appropriate action is required to stop harassment and safeguard victims.

4.2.10 Case Study # 10

1. Threatening Behavior:

The use of threatening language is the common tactic used in the harassment cases specifically the threat of leaking personal information. This case is also an embodiment of such scams. The boy's family is attempting to intimidate the lady's father into accepting the proposal otherwise they will leak the photos of the girl shared at the time when the proposal came.

2. Vulnerability of the Victim:

The emotional and social implications associated with a proposal's rejection made the lady and his family vulnerable to this harassment.

3. Formal and Respectful Tone:

The dialogue starts with a respectful tone acknowledging the importance of discussion. Phrases like "we appreciate your interest" and "thank you for your time" reflects the use of positive politeness strategies and formal communication styles common in South Asian cultures, highlighting respect and honor.

4. Family Dynamics:

The dialogue reflects a regional or cultural background where parents hold authoritative power in decision making process. This implies that the parties involved must belong to a cultural context where familial consent is important in personal relationships.

5. Sophisticated Language:

The use of structured and sophisticated language suggests a certain level of educational background of both the families involved.

6. Control and Coercion:

The dialogue reflects a power imbalance where boy's family is trying to exert control and dominance over the lady's family, a power dynamic usually seen in South Asian context.

7. Justification and Denial:

The boy's family highlights both possibility ("Aisha could have a good life with him") and transformation ("He has changed") throughout the dialogue. Their inability to resolve their son's past with their aspirations for his future is a sign of cognitive dissonance, which can be a sign of denial or an unwillingness to accept truth.

8. Emotional Manipulation:

The use of the phrase “my son truly cares for Ayesha” suggests victim feelings for the girl as well as a strong desire for familial connection.

4.1.10.1 Integrated Analysis

The dialogue depicts a situation in which the lady's family upholds a position of integrity and respectability while the boy's family tries to use threats to pressure them. The exchange demonstrates the emotional costs associated with marriage proposals by highlighting the conflict between respect, familial duty, and the coercive methods used by the boy's family.

4.1.10.2 Estimated Criminal Profile

The dialogue reveals useful insights about the social and educational background of the families involved. The formal and structured way of using language indicates higher level of educational background. In addition to this, the dialogue also gives hint about the possible south Asian context and a middle-class family. Moreover, manipulative and threatening language tells us about the psychological heath of the boy’s family, suggesting insecurity and possible psychological distress associated with proposal rejection.

4.3 Combined Results from both Sentiment and Linguistic analysis

It has been observed from the results obtained from sentiment analysis that although training these transformer-based machine learning models has made it easier for internet users, cybersecurity experts, and law enforcement agencies to detect a message as a potential phishing or cyber harassment scam by classifying a text into positive, negative and neutral sentiments within a limited amount of time. It has also been observed during the training phase that these models also possess the capabilities to even detect the emotions behind the language used by cybercriminals and classify them into particular emotions by simply fine-tuning them on labeled datasets. However, they cannot give us stylometric information about a cybercriminal which sometimes plays a very crucial role in criminal profiling.

Hence, for a comprehensive analysis, linguistic cues (that are often expressed by cybercriminals in their speech) must go through deep linguistic analysis for building an overall understanding of a cybercriminal. By keeping in view these limitations associated with using only sentiment analysis, this research has also incorporated linguistic analysis in it.

The results obtained from linguistic analysis has enhanced the understanding of this research study further, by giving detailed information about the structure, patterns and features of language use. By integrating sentiment analysis with linguistic analysis, this research study has provided a deeper understanding of not only what has been said (literal meaning) but also how it is said (implied meaning). The combined results obtained by integrating both sentiment and linguistic analysis enhances the accuracy of cybercriminal profiling by linking sentiments with linguistic choices, revealing underlying intentions, behaviors, motivations and even the psychological state of the cybercriminals.

4.4 Results Obtained from Open-ended Questionnaires

Open-ended questionnaire has been devised in this research to get further validation for this research study. This questionnaire consists of three sections. First section deals with the personal information necessary for the selection of a suitable candidate for this research. For example, this research study has taken only those responses given by advocates who have some experience in handling cybercrime cases in courts. Second section has been devised to get information about the background knowledge of these advocates about legal aspects of cybercrime. Third section contains eight questions regarding the effectiveness of this research study and its implementation in the existing cybersecurity framework and practices. Following is the summarized view of results obtained from the responses collected from this open-ended questionnaire. Major themes of this questionnaire are as follows:

1. ***Significance of using Linguistic Evidence in Cybercrimes from a Legal Standpoint:***

3 out of 5 candidates have firm believe that linguistic evidence is crucial when dealing with cases related to cybercrimes as it can give valuable insights into the social, educational, cultural and even psychological background of the criminal which in turn helps lawyers to form a powerful criminal profile and to cut down a huge list of suspects to a few potential suspects having all the characteristics mentioned in the profile. On the other hand, 2 candidates although agreed with the growing importance of using linguistic evidence in cybercrimes, yet, they are of the view that it cannot be considered as a primary and necessary source of information.

2. ***Effectiveness of Language Analysis in Detecting Cybercriminal's Behavior:***

All of the participants agree with the fact that language a person use can provide useful insights about a cybercriminal's behavior. Every person carries a unique linguistic mark with them which tells us a lot about the background of the person. Language patterns and choice of specific words vary from person to person and can provide a strong link between suspects and criminal activity. Courts can use this kind of evidence in analyzing cybercrime cases if supported by a forensic linguistic expert. One of the participants stated his response in these words "language analysis is a valuable tool in detecting criminal behavior, contributing to criminal profiling and investigations. It helps identify potential suspects by analyzing specific linguistic patterns used in cybercriminal communications, ultimately revealing about the criminal's background, motives, and affiliations".

3. ***Integration of Language Analysis into Existing Cybersecurity Frameworks and Practices:***

This question was set in relation to the response from above question. It states that, "if language analysis is effective, then should it be integrated in the existing cybersecurity frameworks and practices and how?" All the participants except one are in the favor of integrating linguistic analysis into existing cybersecurity frameworks and practices. They believe that linguistic analysis could be effectively integrated into existing cyber security frameworks by working in collaboration with forensic linguists.

Moreover, training programs can be introduced to educate cyber security professionals on recognizing linguistic evidence. In addition to this, awareness programs can be created for users to recognize suspicious language in communication. Integrating these practices into existing framework will not only improves vigilance of cybersecurity networks but it will also strengthen defenses against cybercrime by providing a deeper understanding of criminal behavior. One of the participants who is against this integration practice states that it is not legally effective therefore it is difficult to use in current practices.

4. ***Combination of Environmental and Linguistic Factors in Detecting Behavior of Cybercriminals:***

All of the participants agreed that both environmental and linguistic factors play a vital role in detecting the behavior of a cybercriminal. More or less, they are of the same opinion that unique linguistic patterns help identify intent, authorship, and emotional state of the cybercriminal while the environmental factors help establish motive, opportunity, and digital traces. Linguistic and environmental factors alone cannot be sufficient for detecting a person as a potential suspect; however, a combined approach centered on both linguistic and environmental factors can be proven beneficial for determining a potential suspect as has been used by this research study.

5. ***Robustness of Roberta Model Used in this Research Study for Sentiment Analysis:***

Two out of the five participants do not have sufficient knowledge about these machine learning models. One of the participants agreed to its robustness but didn't provide any explanation. One of them responded saying "as the model is showing 94 percent accuracy, I think it's enough to measure its robustness". Another participant submit his response in these words "The Roberta model, being transformer-based, excels at understanding context and language nuances. This makes it highly suitable for sentiment analysis tasks. Its deep contextual learning strengthens detection accuracy. Thus, the model appears robust for the study's objectives".

6. Limitations in Using Linguistic Data for Profiling Cybercriminal Behavior:

3 out of 5 participants replied that they do not see any such limitation in using linguistic data for profiling cybercriminal behavior. While the remaining two participants stated several limitations. One of them is the adaptive nature of cybercriminals. They may adapt the linguistic style of other criminals to deceive linguistic experts. Moreover, with the evolving nature of online communication, a cybercriminal language may also evolve with the passage of time. Hence, it will be difficult to detect repeated offenders using only the linguistic evidence.

7. Challenges in Interpreting Linguistic Data:

Some of the participants are of the view that relying solely on linguistic cues can introduce biases and misinterpretations in language analysis due to cultural differences. For example, an idea or an activity can be seen as offensive in one culture but it may seem normal to the people of other cultures. Similarly, situational factors can also alter the meaning of language used. Criminals may also adapt their language to evade profiling. Thus, one should not be proven guilty or innocent just on the basis of linguistic evidence. One of the participants put this in these terms "By keeping in view these challenges we should use linguistic analysis for scrutinizing a big criminal list but not solely depend on it for final judgment". Another participant also gave the solution of combining the linguistic analysis with other analytical methods to gain a deeper understanding of all the cues available.

8. Recommendations for this Research Study or Future Studies:

Three out of five candidates provided recommendations, one for the current study and two for the future studies. One of the participants said that this research study should have analyzed emotions with sentiment analysis to make it more commendable. Two recommendations for future studies include: conducting longitudinal studies to keep track of the changes in linguistic patterns over time and catering other types of cybercrimes as well to expand applicability and strengthen findings.

4.5 Discussion

This study aimed to explore the linguistic patterns that contributes to detect cybercriminal behavior in online communications containing cyber harassment and Phishing scams. The results of the study revealed that cybercriminals frequently make use of certain specific linguistic markers such as threatening language, aggressive tone, urgency, emotional appeals and deceptive politeness etc., in their online communications as evident by the case studies analyzed in this research study. It has been observed that the main purpose of using these linguistic strategies is to exert dominance and control, to intimidate the victims so that they act without giving a second thought to their actions and to make them obedient to their insane requests.

Typically, cybercriminals involved in phishing have seen to use more polite language in an attempt to establish a friendly connection between them and the victims. Therefore, most of the Phishing scams includes friendly greetings, informal tone and polite language. Flattery and polite language are the powerful weapons of Phishers for making the victims complying with their requests. This tactic is often used by scammers to lure the innocent victims into their sweet talk by building a fake rapport with them to make them feel safe around them. Another common linguistic strategy found in almost all phishing scams that have been analyzed in this research study is urgent nature of the message. Scammers often use time pressures in their phishing attempts to put pressure on the victims to act in haste so that they don't have enough time to think of the message as a potential phishing scam and report it in the meantime. Moreover, phishing scams also involves financial incentives (such as tax refund, winning exclusive prizes for sharing a fake verification link etc.) to play with the feelings of the poor and innocent people who are in search for a financial opportunity to prosper in their lives, hence making them the suitable target for the scam. Furthermore, lack of proper cyber security systems to detect a message for its possibility of being a potential phishing message further aggravates the situation.

On the other hand, harassers often use impolite markers in their communications to intimidate the victims and to reinforce their authority to exert power and influence over the victims so that they show compliance with their requests.

The use of threatening language is the common tactic used in harassment cases specifically the threat of leaking personal information such as private photos or videos. Victims, in fear of getting their private photos leaked, become vulnerable to such cyber harassment scams and usually do not report such cases in fear of public humiliation. In addition to this, emotional manipulation has proven to be another powerful weapon in the exploitation of weak and innocent people. Harassers often make use of this method by threatening the victims to harm someone which are very near and dear to them in order to ensure compliance to their unusual requests. Unlike Phishing attempts, cybercriminals have seemed to use slang and colloquial expressions in their conversations with the target which indicates a possibility that cyber harassers usually have low educational background.

The use of derogatory language towards women also suggests that they belong to a cultural background that devalues women and women's rights. Furthermore, the psychological health of the harassers can be predicted through their aggressive tone and anti-social traits and narcissist traits frequently observed in the analysis of cyber harassment cases.

In addition to exploring linguistic cues, this research also aimed at training a transformer-based model Roberta for the analysis of the sentiments and emotions expressed by cybercriminals in online communications. The results of the study showed that the Roberta model trained for cyber harassment and Phishing data has achieved 94.1% and 98% accuracy respectively in assigning appropriate sentiments to the texts given to it for predictions in testing phase.

This means that sentiment analysis of online communications can significantly enhance the ability to predict and identify potential cybercriminal activities, particularly phishing and cyber harassment. The model ability to understand contextual information as well as emotional tone allowed it to identify malicious intent in cybercriminal behaviors better than all the traditional keyword-based systems exists before it. For example, phishing scams mostly use polite and manipulative language such as informal tone, urgency, and

emotional appeals to pressurize the victims to take immediate actions without taking into account the consequences of those actions.

A Roberta based sentiment model can detect these negative and deceptive tones in real world cyber harassment and Phishing cases. Likewise, in instances of online harassment, the model has the ability to identify communications containing offensive, threatening, and aggressive language, even if they are subtle or indirect. By examining the sentiment patterns in online communications over time, cybersecurity detection systems can identify changes toward abusive or coercive communication and can construct cybercriminals behavioral profiles easily. Because of these trained machine learning transformer-based models, cybersecurity systems can react proactively and give priority to high-risk connections. Roberta's integration of sentiment analysis into pre-existing threat detection frameworks can improve both the detection of malicious content and the comprehension of its underlying intent, making it an effective tool for reducing dangers on digital platforms.

Specifically, the study examines how social dynamics and online community standards influence cybercriminal behavior in the areas of online harassment and phishing. Using a hybrid methodology that combines sentiment analysis with linguistic profiling, the study finds important emotional and linguistic patterns that show how criminals negotiate and take advantage of the socio-discursive context of digital platforms. Brown and Levinson's theory of impoliteness is consistent with the frequent infractions of politeness norms found in cyber harassment messages, including the use of imperatives, threats, and face-threatening acts. According to these linguistic cues, societal norms may have been purposefully broken in order to establish authority or emotionally control the victim.

Additionally, a pattern of emotional manipulation may be seen in sentiment analysis of both harassing and phishing texts. This pattern ranges from extreme negativity in harassment (such as insults and anger) to artificially optimistic tones in phishing (such as rewards or support). The norms of the online communities where these discussions take place frequently influence these sentiment-driven strategies. Cybercriminal activities are not only accepted but occasionally encouraged by group dynamics in permissive or unmoderated digital venues where aggressive talk and toxic behaviors are normalized. Due

to the decreased likelihood of social pushback or punishment, this normalization effect emphasizes how community standards encourage recurring offences.

Hence, by integrating computational and manual linguistic approaches, this study offers empirical evidence of how cybercriminals reject or manipulate online community norms. This allows for a more nuanced understanding of the socio-pragmatic factors that facilitate digital crimes.

CHAPTER 5

CONCLUSION

The rapid evolution of technology has greatly changed communication patterns, which not only results in opening up new ways for engaging in online communications but also gave rise to new types of cybercrimes. This research has enriched our knowledge of the linguistic cues and emotional tones used by scammers in Phishing and Cyber harassment frauds. The research contributes to the multidisciplinary understanding of cybercriminal behavior by offering an integrated framework for analyzing and categorizing cyber risks through the integration of environmental and linguistic factors that contributes to the creation of an environment conductive to cybercrime. This research is based on a central idea that language is not just a means of communication, but it can be used to control emotions and behaviors of cybercriminals. The results of the study revealed that in order to make their messages seem more convincing and authentic, scammers frequently make use of particular lexical choices, emotional tones, and grammatical patterns such as using politeness in combination with urgent appeals to increase the chances that their phishing attempts would be successful. This sophisticated knowledge of language tactics emphasizes how crucial language is in conducting a successful cybercrime.

The analytical approach (which involves the fine-tuning of the Roberta model on phishing and cyber harassment datasets) used in this study is another important development in the field of cybersecurity analysis. This research study further expands the scope of linguistic analysis by using an advanced machine learning model “Roberta” which enables a more refined understanding of the linguistic strategies used by cybercriminals in their online communications with the victims. Roberta, a transformer-based model, is well-known for its robust performance on tasks involving natural language processing such as sentiment analysis and emotion detection. It is particularly good at identifying semantic relations and contextual subtleties in a text. By categorizing and evaluating phishing attempts and cyber harassment messages accurately to some extent, the fine-tuned Roberta model can help create detection mechanisms that can be more successful. In addition to this, Roberta model allows for the faster detection of possible threats which helps enterprises to react quickly and reduce risks.

Moreover, this research has offered useful insights into the psychological and emotional aspects of cybercriminal behavior by using the data collected from various social media sites. This innovative approach of Roberta model has clearly demonstrated how advanced NLP models can evaluate complex linguistic features and enable a deeper understanding about how language used in online communications conveys criminal intent.

In addition to this, the results of the study gave us important information about the ways cybercriminals behave in online communications related to phishing and cyber harassment scams. The use of language strategically by cybercriminals is highlighted by the identification of speech involving threatening and manipulative language, unusual requests and demands and justifications for their actions. These linguistic cues further reveal the psychological drives and techniques often used by the scammers to further their objectives. For example, threatening language can be used by cybercriminals to frighten their victims, urgent nature of these messages signals towards the necessity of an immediate action by the victims without careful consideration, use of polite strategies often accompanied with informal tone suggests the establishment of a fake sense of connection with the victims to make them feel secure, and excuses like justifications or emotional appeals can give the situation a false sense of legitimacy and authenticity. Having an understanding of these trivial clues that have the ability to measure the malicious intent behind a text message is necessary for cybersecurity experts to detect these evils.

Furthermore, by improving threat detection mechanisms with the help of this study's insights, cybersecurity experts will be better equipped with the desired knowledge to recognize and address online threats. Cyber analysts will be able to effectively allocate resources and prioritize high risk cases by including linguistic and psychology markers in their analysis. This pre-emptive strategy will eventually help to create a safer online environment in addition to increasing overall efficiency of cybersecurity measures.

Moreover, this research study also provides useful insights to the guardian of virtual communities who are responsible for creating a safer online environment for its users. By getting useful information about the emotional tones and language clues that indicate cyber threats, moderators can use more successful content moderation techniques.

The study's insights enable moderators to take preventive steps against phishing attempts and cyber abuse by eliminating damaging content, creating awareness programs, and promoting a culture that prioritizes accountability among the members of community, ultimately making the internet a safer place for all users. The study's findings can also be used by law enforcement agencies to detect and capture cybercriminals more effectively and easily by integrating language profiling into their investigation processes. Investigators can gain essential details that can guide their strategies by analyzing recurrent communication methods and emotional manipulation techniques. If law enforcement training programs integrate these findings into their investigation procedures, law enforcement officers may be able to recognize and handle cyber threats. Law enforcement must use innovative techniques like linguistic analysis into their operations to stay ahead of emerging threats in the constantly changing realm of cybercrime.

Furthermore, this study has made significant contributions to the fields of criminology, cybersecurity, forensic and computational linguistics by using a multidisciplinary approach. This study has built a thorough understanding of the behaviors and intentions of cybercriminals by taking interdisciplinary insights from all these academic disciplines into account. Future research can be directed by using interdisciplinary insights from other academic disciplines to further expand the understanding of these concepts. This research has set the stage for all the further research that wish to further investigate the linguistic and psychological aspects of cybercrime, hence opening the doors to the development of advanced methodologies and frameworks for the investigation and prevention purposes. As this research has kept delimited to the two types of cybercrime: phishing and online harassment, future studies can involve other types of cybercrimes into account to look for the similarity and differences in using particular linguistic patterns in other types of cybercrimes such as hacking, identity theft and ransomware etc. As technology is evolving day by day, cybercriminals has also adopted new and better ways of conducting cybercrimes. Therefore, continuous investigation in the realm of cybercrime is necessary to keep track with these changing forms of conducting cybercrimes. This requires adaptation of the new and advanced threat detection models and investigative techniques which can react to the changing circumstances more effectively.

For this to happen, just like cybercriminals the sector of cybersecurity should also equip themselves with the necessary training to keep pace with the rapidly changing trends in cybercrime. The results of this study highlight the importance of continuous research and development in the field of cybersecurity criminology and forensic linguistics in order to prepare themselves ahead of new dangers. This battle against cybercrimes can be fought effectively only when linguistic scholars, law enforcement officials and cybersecurity experts will work in collaboration.

In a nutshell, this research has made significant contribution to our knowledge of linguistic and emotional behaviors of cybercriminals which in turn will help the law enforcement officials to proactively address cybercrimes, especially phishing and online harassment. A multidisciplinary approach used by this study was necessary for deeply analyzing the linguistic and psychological aspects of cybercrime by taking into account interdisciplinary insights from the disciplines of criminology, cyber psychology, forensic profiling and computational linguistics. This research emphasizes the importance of language and explored the ways in which it can be manipulated by cybercriminals for serving evil purposes. By taking useful insights from this research, future studies can develop more advanced threat detection models to further improve cybersecurity. Ultimately, the results of this study have helped in achieving a more general objective of making the internet a safer place for all its users, eventually promoting security and trust in a world that is becoming more interconnected by the day. To conclude, it can be stated that there is an urgent and dire need of continuous research and innovation in the realm of cybercrime for keeping pace with the rapidly developing digital world and continuously evolving nature of cybercrimes. By utilizing the useful insights from this research and future research that will be conducted in this context, a secure and more resilient digital environment can be created to promote users' security and trust in an increasingly complex digital world.

5.1 Limitations of the Study

This research study is subject to various limitations.

Firstly, it involves a small sample size for case studies which although can be sufficient for gaining initial insights into the matter but not enough to generalize the results to varied demographics of the larger communities. Secondly, the choice of particular social media platforms for data collection can restrict the application of the findings on the cybercriminal communications that occurred via other online platforms. Moreover, the victims of cyber harassment, in fear of public humiliation by getting their identities leaked, usually do not report these crimes and do not allow their data to be a part of any research study, which limits the study to a small sample size. Last but not least, manual interpretation of linguistic cues, although involved due care, may add biasness to the results and the conclusions reached. Recognizing these limitations makes it easier to comprehend the study's findings and identify areas that require further research to overcome these obstacles.

5.2 Recommendations for Future Research

To address the limitations of this study and to provide more comprehensive understanding of cybercriminal behavior, the future research might involve broader populations and cultural contexts by increasing the sample size to ensure diversity. Moreover, longitudinal studies can be conducted to track the evolving nature of language as well as cybercrimes in the rapidly advancing world of technology. In addition to this, future research can look into a wide variety of online and offline platforms that has not been explored in this research study in order to confirm the similarity or differentiation of the results in other settings. By keeping in view the ethical considerations, future research should look for the methods that result in collecting rich qualitative data while still maintaining user privacy and consent. Like this study, future studies can be conducted by collaborating with other disciplines for useful interdisciplinary insights to gain a more comprehensive understanding of cybercriminal behavior. Eventually, creating powerful threat detection frameworks in real time that involve linguistic and emotional indicators will be essential for the creation of proactive cybersecurity tactics. Future research can greatly aid in this continuous effort to minimize the complexity of cybercrimes by tackling the above-mentioned themes.

REFERENCES

Adha, A. (2020). Linguistic Based Cues In Detecting Deception In Indonesian Language Use. *Argumentum 16*. https://doi.org/DOI: 10.34103/ARGUMENTUM/2020/2

Aimanshahbaz. (2022, December 17). medium. https://medium.com/@aimanshahbaz34/the-speech-act-theory-by-austin-44aa586388b0

all, J. M. (2024, August 9). Psychological Profiling in Cybersecurity: A Look at LLMs and Psycholinguistic Features.

Ameiruel Azwan Ab Aziz, N. A. (2023). Linguistic Cues of Deception in Malaysian Online Investment Scams'. *Journal of Language Studies, 23 (4)*. https://doi.org/ http://doi.org/10.17576/gema-2023-2304-09

Aseel Addawood, A. B. (2019). Linguistic Cues to Deception: Identifying Political Trolls on Social Media. *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019).* https://doi.org/https://doi.org/10.1609/icwsm.v13i01.3205

Austin, J. L. (1962). *How To Do Things With Words.* Clarendon Press.

Aydınoğlu, N. (2013). Politeness and impoliteness strategies: an analysis of gender differences in Geralyn l. Horton's plays . *Procedia - Social and Behavioral Sciences 83 (2013) 473 – 482* . https://doi.org/10.1016/j.sbspro.2013.06.093

Azianura Hani Shaari, M. M. (2019). Online-Dating Romance Scam in Malaysia: An Analysis of Online Conversations between Scammers and Victims. *Journal of Language Studies, 19(1)*. https://doi.org/ http://doi.org/10.17576/gema-2019-1901-06

Bart Desmet, V. H. (2013). Emotion detection in suicide notes. *Researchgate*. https://doi.org/10.1016/j.eswa.2013.05.050

Bayan B. Rababa'h, G. R. (2021). The Impact of Culture and Gender on Impoliteness Strategies in Jordanian and American TV Sitcoms. *Academy Publication*(Vol. 11 No. 2 (2021)). https://doi.org/https://doi.org/10.17507/tpls.1102.06

Bjarke Felbo, A. M. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Reserchgate*. https://doi.org/DOI:10.18653/v1/D17-1169

Cybersecurity ASEE. (2025, July 14). Phishing attacks: Recognizing & preventing phishing fraud. Retrieved October 1, 2025, from https://cybersecurity.asee.io/blog/authentication/phishing-attacks-how-to-recognize-phishing-scams/

cohen, L. a. (1979). Social Chnage and Crime rate trends: A routine activity approach. *American Sociological Review*.

Cohen, L. E. (1981). Social Inequality and Predatory Criminal Victimization: An Exposition and Test of a Formal Theory. *American sociological review, 46(5), 505-524.* https://doi.org/https://doi.org/10.2307/2094935

CounterCraft, R. B. (n.d.). *Inside the mind of the enemy: A guide to profiling cybercriminals.*

Culpeper, J. (2005). Impoliteness and Entertainment in the Television Quiz. *Journal of Politeness Research Language Behaviour Culture*. https://doi.org/10.1515/jplr.2005.1.1.35

Dawn News. (2019, April 6). *Language is more than a means of communication*. Dawn.

Fan Zhang, H. X. (2016). Grasp the implicit features: Hierarchical emotion classification based on topic model and SVM. *Researchgate*. https://doi.org/10.1109/IJCNN.2016.7727661

Genao, D. R. (2021). Identification of Fraudulent Financial Statements: The Detection of Deception and Collusion in Earnings Calls. *ProQuest*. https://doi.org/https://www.proquest.com/docview/2551555139?pqorigsite=gscholar&fromopenview

Goffman, E. (1967). *Interaction Ritual: Essays on Face to Face Behavior.* NewYork: Anchor Books.

GeeksforGeeks. (2025, July 23). *Overview of RoBERTa model*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/overview-of-roberta-model/

GeeksforGeeks. (2025, July 12). *What is Sentiment Analysis?* GeeksforGeeks. https://www.geeksforgeeks.org/what-is-sentiment-analysis/

Hulatt, L. (2022). Robert K. Merton. studysmarter. https://www.studysmarter.co.uk/explanations/social-studies/famous-sociologists/robert-k-merton/

Hafeez, E. (2014). researchgate. https://www.researchgate.net/publication/270342188_Cyber_Harassment_its_implications_on_youth_in_Pakistan

Hafeez, E. (2014). Cyber Harassment & its implications on youth in Pakistan. *researchgate*. https://doi.org/https://www.researchgate.net/publication/270342188_Cyber_Harassment_its_implications_on_youth_in_Pakistan

Hobashia Saleem, J. J. (2022). Cyber Crimes Trends in Pakistan: Analyzing the Legal Framework and Enforcement Challenges. *researchgate*. https://doi.org/10.62585/slpr.v1i1.21

Inc. (2021, January 5). *Computer crimes*. Inc.com. https://www.inc.com/encyclopedia/computer-crimes.html

Ilyasova, R. S. (2018). *Language Personality In The Context Of Interaction Of Language And Culture.* https://doi.org/10.23947/2414-1143-2018-16-4-32-36

Jibran Jamshed, W. R. (2022). Critical Analysis of Cybercrimes in Pakistan: Legislative Measures and reforms. *International Journal of Business and Economic Affairs (IJBEA)*. https://doi.org/DOI: 10.24088/IJBEA-2022-71002

Jonathan Culpeper. (1996). Towards an anatomy of impoliteness. *journal of pragmatics*, Pages 349-367. https://doi.org/https://doi.org/10.1016/0378-2166(95)00014-3.

Kate Brush, M. C. (2024, January). Techtarget. https://www.techtarget.com

Laghari, R. (2024, september 29). Medium. https://medium.com/@riazleghari/linguistic-analysis-67cae4052118

Levinson, P. B. (1987). *Politeness: Some universals in language use.* Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511813085

Li, X. (2017, FEBRUARY 19). A Review of Motivations of Illegal Cyber Activities.

Luo, W., & Lee, Y. (2020, May). Unveiling the cloak of deviance: Linguistic cues for psychological processes in fake online reviews. *International Journal of Hospitality Management, 87,* 102464. https://doi.org/10.1016/j.ijhm.2019.102464

Maryam Hasan, E. R. (2018). Automatic emotion detection in text streams by analyzing Twitter data. *Researchgate*. https://doi.org/10.1007/s41060-018-0096-z

Merton, R. K. (1968). *Social theory and social structure.* The Free press.

miro` llinares, F. (2014). *The Encyclopedia of Theoretical Criminology Online (pp.1-7).* Blackwell Publishing Ltd. https://doi.org/10.1002/9781118517390/wbetc198

Monir Mirhosseini, M. M. (2017). Impoliteness Strategies Based on Culpeperâ€™s Model: An Analysis of Gender Differences between Two Characters in the movie Mother. *Journal of Applied Linguistics and Language Research*.

Nickerson. (2022). Simple Psychology. www.simplypsychology.org/routine-activities-theory.html

Proofpoint. (2025, August 19). *What is cyber crime? – Types, impact, prevention*. Proofpoint. https://www.proofpoint.com/us/threat-reference/cyber-crime

Perera, A. (2024, Feburary 13). *simple psychology.* https://www.simplypsychology.org/rational-choice-theory-of-criminology.html

R Jayakrishnan, G. N. (2018, january 6). Multi-Class Emotion Detection and Annotation in Malayalam Novels. *Researchgate*. https://doi.org/DOI:10.1109/ICCCI.2018.8441492

Roderick S. Graham, '. K. (2019). *Cybercrime an digital deviance (1st Edition).* Routledge. https://doi.org/https://doi.org/10.4324/9781351238090

Rudianto, K. a. (2023). An Analysis of Impoliteness Strategies in "Can You Ever Forgive Me?" Movie. *Journal of Language Teaching and Learning, Linguistics and Literature*. https://doi.org/10.24256/ideas.v11i1.3877

Saleh Mohamed, L. F. (2024). Social Media and Misinformation: Amplifying Discrimination and Violence. *researchgate*.

Saleh Mohamed, L. F. (2024). Social Media and Misinformation: Amplifying Discrimination and Violence. *Social Media and Technology*.

Shmyla Khan, S. K. (2019, January). DigitalRightsFoundation. https://digitalrightsfoundation.pk/wp-content/uploads/2019/01/Research-Work.pdf

Simister, N. (2017). intrac for civil society. https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Basic-tools-for-data-collection.pdf

Smith, R. S. (2024). *Cybercrime and Digital Deviance: Second Edition.* Routledge. https://doi.org/https://doi.org/10.4324/9781003283256

Sonja Gievska, K. K. (2015). A Hybrid Approach for Emotion Detection in Support of Affective Interaction. *Researchgate*. https://doi.org/10.1109/ICDMW.2014.130

Stacey. (2015, October 25). Language Bird. https://www.languagebird.com/language-is-a-lot-more-than-just-communication/

StudySmarter UK. (n.d.). Explanations: Politeness theory. StudySmarter UK. https://www.studysmarter.co.uk/explanations/english/pragmatics/politeness-theory/

StudySmarter UK. (2023, November 30). Explanations: Criminal profiling. StudySmarter UK. https://www.studysmarter.co.uk/explanations/psychology/forensic-psychology/criminal-profiling/

The Nation. (2022, July 6). The biggest cyber threats faced by small businesses in Pakistan. The Nation. https://www.nation.com.pk/06-Jul-2022/the-biggest-cyber-threats-faced-by-small-businesses-in-pakistan

team, d. e. (2023, february). Dovetail. https://dovetail.com/research/purposive-sampling/#:~:text=Purposive%20sampling%20is%20a%20technique,judgmental%20sampling%20or%20selective%20sampling.

Thompson, K. (2023, November 16). revisesociology. https://revisesociology.com/2016/04/16/mertons-strain-theory-deviance/

Umema Hani, K. K. (2024). Psychological profiling of hackers via machine learning toward sustainable cybersecurity. *frontiers*. https://doi.org/10.3389/fcomp.2024.1381351

UniversalClass. (n.d.). *The politeness theory: A guide for everyone*. UniversalClass.com. https://www.universalclass.com/articles/business/communication-studies/politeness-theory.html

Voilino, B. (2023, January). CNBC. https://www.cnbc.com/2023/01/07/phishing-attacks-are-increasing-and-getting-more-sophisticated.html

Windy Amelia, N. U. (2016). Dominant emotion recognition in short story using keyword spotting technique and learning-based method. *Researchgate*. https://doi.org/10.1109/ICAICTA.2016.7803131

Zafarani, X. Z. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys (CSUR)*. https://doi.org/https://doi.org/10.1145/3395046

Zainab Alkhalil, I. K. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *frontiers, 3*. https://doi.org/https://doi.org/10.3389/fcomp.2021.563060

ANNEXURE (A & B)

A) Data collected for training Roberta model:
Because of its big size, the coding scheme, labelled and unlabeled data that has been used for the training of both Roberta models has been saved in separate word and excel files respectively which will be attached with the thesis when asked.

B) Data collected for Case Studies:

Case Study # 1:

"Dear User,

Congratulations on your recent verification! As a valued journalist, your account is crucial to our community. To ensure the security and continued verification of your account, we need you to follow the instructions below immediately. You need to follow three simple steps to get your verified status:

1. Confirm Your Identity by Clicking the link below to verify your identity and maintain your blue tick status: Verify Your Account

2. Update Your Account Information including your phone number and any additional details requested.

3. Change your password to protect against unauthorized access.

Failure to comply with any of these instructions may result in the temporary suspension of your verified status. Thank you for your immediate consideration to this matter. We appreciate your contribution to the Twitter community!

Best regards,

Twitter Support Team"

Case Study # 2:

"**Hi** [Victim's Name], **this is Ali** from **Courier Company**. We have a delivery scheduled for you today, but we need to verify your identity first. Can you **please** provide your **thumbprint** to the rider for our records? It's just a **standard procedure** to **ensure safe delivery**.

If you could do this **quickly**, I can have your package **delivered right away**! **Thanks** for your **cooperation**!

Case Study # 3:

Dear Friend, I am **Barrister** Benson Khojo, I am making an offer to you in relation to **te death of my late client**, who was from your country before his death, leaving some **huge amount of money** in the bank, after **unsuccessful attempts to find his relations** here, i decided to contact **yo**. For more details and information, please send to me the following information if you are interested. Your full name; Age; **martial** status; telephone number; private email; residence; office address.

Case Study # 4:

"**Hello** Amar **brother**, this is Rizwan from **Easypaisa**. Listen, there's a **problem**! A shopkeeper **mistakenly sent money** to your number. An **OTP** will come to you shortly; **please share it with me** so we can resolve this issue. I'm from a **small town** and **can't afford to lose this money**. **Your help is really important!** Thank you!

Case Study # 5:

After the last annual calculations of your fiscal activity, we have determined that you are eligible to receive a tax refund of 52,088.50 **Pakistani Rupee**. **Please submit the tax refund request** and allow us **2-3 days** in order to process it.

Click the link below to submit your tax refund request.

http://www.fbr.gov.pk/

Note: A refund can be delayed **a variety of reasons**, for example submitting invalid records or applying after deadline.

Best Regards

Federal Board of Revenue Government of Pakistan

Copyright…All rights reserved

Case Study # 6:

Yes dear (victim name), I also want you to take divorce from me and **I will post the bed video on Facebook** so that people can see you naked and you cannot face anyone. Shakoor, **I will give all your girl photos on the internet** by **this evening**. Finish the work on any matter with me, **either take a divorce from me and return the belongings**, or come. **You have 6 days**; I will come to your house again. Now I will either **give my life or take yours**."

Case Study # 7:

Hey you, you think you can just keep spouting ur **feminist BS** and not get consequences? u just a **lil girl** who don't know her place. got your pics, just few clicks, and I can totally make them hot. Stop acting like u better than us. keep talking, I'll make sure you regret it. u think ur safe behind ur screen? **Think again**. 😈 Just **wait and watch** hoe em going to **ruin ur repo**. u'll wish u never opened ur mouth about women's rights. better **shut up** or I'll make ur **new pics go viral**🔥

Case Study # 8:

Hi [Female Teacher's Name],

I just wanted to say that I think you did a fantastic job on that last project. You really know how to engage the students. But honestly, I think you should lighten up a bit. Everyone talks about how serious you are all the time. Maybe if you smiled more, people wouldn't think you're so uptight. Let's grab coffee sometime. It'll be good for you to unwind a little. Just don't take it the wrong way; I'm just trying to help.

Looking forward to your reply.

[Male Teacher's Name]

Case Study # 9:

Harasser: "Hey, why did you ignore my proposal? **You think you're too good for me?** 😉"

Harasser: "Married or not, doesn't matter. You shared your pics with me. **I have everything**. Don't think you can just get away with this. 😈"

Harasser: **"You better start talking to me again. Or else…"**

Harasser: "I have made fake account on your name and you will soon see your naked pics everywhere. Let's see how your husband feels about that! 😂"

Harasser: **"Illegal? Who cares! You think I'm scared of the law?** Just admit you want me, and I might let it slide. Otherwise, get ready for some fun! 😜"

Harasser: "Go ahead! I'll just make another account. **You think you can stop me?** You're just a **cheap girl** who played with my feelings. Now **face the consequences!** 😉"

Harasser: "Begging won't help you! Just remember, **I have all the power here**. You better **keep quiet**, or I'll **ruin your life**! 💀"

Case Study # 10:

Lady's Father: "Hello, this is **Mr. Khan**. We wanted to discuss the proposal for my daughter, **Aisha**."

Boy's Family: "Hi, Mr. Khan. Yes, we were hoping to hear from you. Have you made a decision?"

Lady's Father: "We appreciate your interest, but after some consideration and further inquiries, we have decided to decline the proposal."

Boy's Family: **"Oh? May I ask why?** We thought everything was going well."

Lady's Father: "We learned some concerning information about your son. It seems he has a history that we were unaware of."

Boy's Family: "What do you mean? Is this about his previous marriage?"

Lady's Father: "Yes, we found out that he was divorced due to extramarital affairs. This raises serious concerns for us regarding his character."

Boy's Family: **"That's in the past! He has changed**. We thought you would see his potential and how much he has grown since then."

Lady's Father: "While we appreciate that, we cannot overlook the implications of such behavior. We believe it's in Aisha's best interest to move on."

Boy's Family: "I understand, but I think you're making a mistake. **My son truly cares for Aisha**."

Lady's Father: "We respect your opinion, but our decision is final."

Boy's Family: "Before you finalize anything, I need to warn you. There are pictures that were shared during the proposal process. **If this goes public, it could be damaging**."

Lady's Father: "Are you threatening us? This is highly inappropriate."

Boy's Family: "Not a threat, just a reality. If you're rejecting the proposal, **you might want to reconsider**. Those pictures could create quite a stir."

Lady's Father: "This is not how a respectable family should conduct themselves. We will not be intimidated into changing our decision."

Boy's Family: **"Just think about it. It's not too late to reconsider**. Aisha could have a good life with him."

Lady's Father: "Aisha deserves a partner who respects her and her family. We will not be swayed by threats. Our decision stands."

Boy's Family: "Fine, but remember, you can't control what happens next."

Lady's Father: "We will handle this matter appropriately. **Thank you** for your time."

ANNEXURE (C):

Open-Ended Questionnaire:

Title:

Linguistic cues in cybercrime: profiling criminal behavior through language analysis:

Description:

This interview questionnaire is meant for a research study that aims to analyze the language of cybercriminals through deep linguistic analysis that will help lawyers, law enforcement officials and cybersecurity experts to have deeper insights into the social, educational or cultural background of the criminal as well as psychological behavior of him or her. This research study also aims to analyze the sentiments of the cybercriminals through the language they used to gain an understanding of the emotional well-being of the cybercriminals.

1) **Personal Information:**

Name:

Designation:

Organization name:

Email Address:

---

2) **Background knowledge of the subject under discussion:**

1. How familiar are you with the topic of this research?

2. Can you briefly describe your experience in cybersecurity or legal aspects of cybercrime? In your experience, have you encountered cases where linguistic analysis provided significant insights into criminal intent?

3) **Questions related to research:**

Q1: From a legal standpoint, how important is linguistic evidence in prosecuting cybercrime cases?

Q2: In your opinion, are both environmental and linguistic factors important in detecting a cybercriminal behavior?

Q3: In the light of the above research, how effective do you find language analysis in detecting criminal behavior?

Q4: If effective, how could it be integrated into existing cybersecurity frameworks or practices?

Q5. Do you think the integrated cybercriminal profiling model and sentiment analysis model Roberta that has been used for this research study is robust enough?

Q6: What limitations do you see in using linguistic cues for profiling cybercriminal behavior?

Q7: Are there challenges in interpreting linguistic data that you believe need to be addressed in the research?

Q8: Do you have any recommendations for improving the study or additional resources to consider?

----------The End----------