



Artificial Intelligence in Phishing Detection: A Comparative Study of Accuracy, Adaptability, and Practical Implications



Research Report submitted in partial fulfillment of the requirements for
the **Postgraduate Diploma in Data Analytics** at The Independent
Institute of Education, Varsity College.

Jordan Green
ST10083222

Supervisor: Ms F. Shaik
Word Count: 13213

04 / 11 / 2024

Abstract

This study compares the effectiveness of artificial intelligence (AI) technologies—such as deep learning and machine learning—in improving phishing detection to traditional rule-based methods. Traditional rule-based methods fared marginally better, with an accuracy of 89%, while AI systems show competitive accuracy, with deep learning models reaching an 88% accuracy rate. However, despite this marginal difference, AI's adaptability to evolving phishing tactics offers significant advantages in dynamic environments where rule-based methods may lag. Using the Kaggle "EMAIL SPAM DETECTION" dataset, a random sampling method was applied to analyse a balanced sample of 5169 unique values/emails, providing a representative foundation for model training and validation. The PhishHaven system, tested within this study, showcases AI's capability to provide real-time detection for AI-generated phishing links, demonstrating the potential of AI in high-risk scenarios. Findings suggest that AI-powered models, with their ability to recognise new patterns, can effectively identify emerging phishing strategies and reduce attack susceptibility, albeit constrained by the study's limited feature set. Key recommendations include integrating human oversight, expanding datasets, and prioritising AI adaptation in cybersecurity infrastructure. Future research should further investigate diverse AI models, human-AI synergy, and ethical considerations in AI-driven cybersecurity.

Key words:

Artificial Intelligence (AI), Deep Learning, Machine Learning (ML), Traditional Rule, Based Methods, Kaggle, Dataset, PhishHaven, Human-AI Cooperation, Cybersecurity

DECLARATION

I hereby declare that the **research report** submitted for the **Postgraduate Diploma in Data Analytics** to The Independent Institute of Education (The IIE) is my own work and has not previously been submitted to another University or Higher Education Institution for a postgraduate qualification.

Signature  _____

Date 04/11/2024

Contents

ABSTRACT	I
DECLARATION	II
LIST OF TABLES:	3
LIST OF FIGURES:	3
CHAPTER 1: INTRODUCTION.....	5
1.1. Contextualization	5
1.2. Rationale.....	5
1.3. Problem Statement.....	7
1.4. Purpose Statement.....	7
1.5. Research Questions.....	9
1.6. Hypotheses/Null Hypothesis	9
CHAPTER 2: LITERATURE REVIEW	11
2.1. Contextualization.....	11
Evolution of Phishing Attacks:	11
Significant Milestones and Evolving Techniques:.....	11
The Growing Dangers of Phishing Attacks:.....	12
AI and its Potential Role in Phishing Detection:.....	13
2.2. Theoretical Foundation and how it links to the research problem	13
Machine Learning (ML) and Pattern Recognition:	13
Natural Language Processing (NLP) and Social Engineering:.....	14
Justification for Selection:.....	14
Key Theories:	14
2.3. Details of the issue/problem being investigated	15
Limitations of Traditional Methods:	15
Threat Intelligence:.....	16
Threat Intelligence and AI:	17
2.4. Links between problem and current literature.....	18
Seminal Authors/Sources:.....	18
Literature Review Themes:	19
Phishing Awareness Training:	19
Phishing URL Detection:	20
CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY	22
3.1. Research Approach and Design used to investigate the problem	22
Paradigm:	22
Approach/Design:	22
Machine Learning Model:	24

3.2. Data Collection Methods	27
Population:.....	27
Sampling:.....	27
Parameters:.....	28
Dataset Points:.....	29
3.3. Data Analysis Techniques	29
3.4. Validity, reliability, and/or trustworthiness of the findings.....	31
3.5. Ethical implications and how they were addressed.....	32
Ethics:.....	32
3.6. Limitations of the study.....	33
Limitations:.....	33

CHAPTER 4: FINDINGS AND INTERPRETATION OF FINDINGS 35

4.1. Introduction.....	35
Research Aim and Objectives	35
Data Source Description:	37
4.2. Presenting and Discussing Findings	39
Findings Based on Questions:.....	42
Findings Based on Hypotheses.....	45
4.3. Critical Analysis and Interpretation	46
Patterns and Trends:.....	46
Contradictory Findings:.....	46
Data Limitations and Implications:	47
Chi-Square Test Interpretation:	47
Vulnerability Reduction:.....	48
Analysis Summary:.....	48
4.4. Discussion of Findings and Implications.....	48
Implications for Theory:	48
Implications for Practice:	49
4.5. Limitations of the Study.....	50
4.6. Summary of Findings.....	50
Restate Major Findings:.....	50
Link to Research Objectives:	50
Future Research Directions:	51

CHAPTER 5: CONCLUSION 52

5.1. Interpretation of Results:.....	52
5.2. Implications for Cybersecurity Theory and Practice:	54
5.3. Actionable Recommendations:	55
5.4. Limitations and Future Research Directions:.....	55

REFERENCES 57

ANNEXURES: 61

Annexure A: Ethics Letter	61
Annexure B: Concept Document	61
Annexure C: GitHub Link to code:.....	61

List of Tables:

Table 1. 1: Table showing five rows of the email sent, and whether or not they are ham and spam. (Source: Developed for the research study)	23
Table 1. 2: Interpretation of Accuracy Results. (Source: Developed for the research study)	39
Table 1. 3: Logistic Regression Classification Report and Neural Network Classification Report are the two tables above. (Source: Developed for the research study)	40
Table 1. 4: The interpretation of the Chi-Square Test. (Source: Developed for the research study)	41
Table 1. 5: Final results interpretation for all the allotted values. Also, there's a final summary explaining the best choice. (Source: Developed for the research study)	42
Table 1. 6: Interpretation of the user vulnerabilities before and after AI implementation. (Source: Developed for the research study)	45

List of Figures:

Figure 1. 1: Horizontal Bar Graph showing the ham and spam amount. (Source: Developed for the research study)	38
Figure 1. 2: Bar Graph showing the accuracy comparison of phishing detection methods. (Source: Developed for the research study)	39
Figure 1. 3: Confusion Matrix for Rule-Based, Logistic Regression and Neural Network. (Source: Developed for the research study)	41
Figure 1. 4: The Observed vs Expected Values using the spam and ham. This is the Chi-Square Test. (Source: Developed for the research study)	41

Figure 1. 5: This is a Line Graph that show user vulnerabilities before and after AI implementation. The Green is when AI is implemented and Red is before. (Source: Developed for the research study)44

Chapter 1: Introduction

1.1. Contextualization

- *Phishing is a social engineering attack whereby attackers use emails, websites, or other forms of contact to pretend to be reliable sources to trick users into disclosing important information* (Alkhalil, Hewage, Nawaf, & Khan, 2021).

Phishing, a social engineering attack where attackers impersonate trusted sources to steal sensitive information (Alkhalil, Hewage, Nawaf, & Khan, 2021), has become a major threat. The complexity of phishing attacks continues to evolve, making traditional methods such as content filtering and URL blacklisting ineffective. This presents an enormous challenge for cybersecurity professionals.

Artificial intelligence, or AI, is technology that simulates human intelligence and can bolster cybersecurity defences (IBM, 2024) (Al-Mansoori & Salem, 2023), has emerged as a promising tool to address these challenges. However, a comprehensive understanding of AI's effectiveness in detecting and preventing phishing attacks is lacking. By investigating how developments in AI might enhance phishing detection and prevention, this research seeks to bridge this gap. We will examine the advantages and disadvantages of incorporating AI into cybersecurity, with an emphasis on how it might be used to combat different phishing schemes.

Conventional techniques for phishing detection and prevention frequently fall short in the face of these constantly changing threats. Thankfully, the rise of modern problems has seen the rise of modern solutions, with artificial intelligence (AI) emerging as a promising tool for enhancing phishing detection and prevention. Artificial intelligence (AI) has grown as a potentially revolutionary tool for experts to assist them in their goal of combating various phishing schemes from email scams to malware and farming attacks and more (Al-Mansoori & Salem, 2023).

1.2. Rationale

Phishing attacks are a major threat to cybersecurity because they continuously evolve to avoid detection techniques. These can cause significant financial and reputational

harm to individuals and companies by tricking users into disclosing private information or clicking on malicious links. (Hayat, 2023)

Blacklists and content filters, two popular detection methods, are unable to keep up with the sophistication of today's phishing attacks. The present research addresses the need for more intelligent and adaptable countermeasures against this escalating danger.

The present research investigates the potential applications of artificial intelligence (AI) to enhance phishing detection and prevention. It will delve into the possibilities of Natural Language Processing (NLP) for examining social engineering techniques in phishing emails and Machine Learning (ML) for pattern recognition in email data. Additionally, this research will investigate threat intelligence's function in AI-powered phishing detection.

When AI is successfully integrated, it potentially can result in:

Enhanced detection accuracy:

Artificial intelligence (AI) can identify subtle patterns that traditional methods overlook, which improves the identification of new sophisticated phishing attempts.

- Reduced human error: By reducing human tiredness and biases through automation of data analysis and pattern recognition, threat assessments can become more reliable.
- Faster response times: Real-time data analysis by AI makes it possible to identify and mitigate phishing threats faster.
- Improved threat intelligence: Threat intelligence data can be analysed by AI to anticipate phishing campaigns in the future and stay ahead of evolving threats.

However, it's essential to devote careful consideration to AI's limitations and ethical concerns. The use of large-scale data analysis in AI-powered phishing detection raises concerns regarding user privacy and the necessity of getting explicit authorization from the user.

Furthermore, some AI algorithm designs might make it challenging to comprehend how they make decisions, which impedes accountability and transparency. To ensure

the appropriate use of AI, human oversight is still essential, especially for activities that require context and interpretation. AI's capability for phishing defence can be leveraged while mitigating potential risks by acknowledging these limitations and encouraging a cooperative approach between AI and human expertise.

1.3. Problem Statement

Despite advancements in cybersecurity, traditional phishing detection methods are increasingly ineffective against sophisticated attacks, necessitating the exploration of AI-powered solutions. A comprehensive study on the advantages and disadvantages of integrating artificial intelligence (AI) into modern society and cybersecurity will be conducted. The end goal is to open a discussion on how AI can be leveraged to address the evolving threats posed by phishing attacks.

Various algorithm tools will be studied to ascertain their usefulness compared to older methods and if they will meaningfully contribute to phishing detection and prevention. The human element of the psychological and behavioural characteristics of attackers and targets in phishing attacks will be looked at and studied to find ways of mitigating these attempts.

1.4. Purpose Statement

This study aims to evaluate the effectiveness of AI technologies in enhancing phishing detection and prevention, comparing their performance to traditional methods.

Objectives:

Revealing Phishing: Limitations and Tactics:

This objective investigates the limitations of conventional phishing detection methods (blacklists, content filtering) by demonstrating how ineffectual they are against more sophisticated attacks. How attackers leverage social engineering tactics, spear-phishing techniques, and AI-generated URLs to bypass existing filters will be explored.

AI Revolutionizing Phishing Detection: ML and NLP Insights:

The objective is to assess how artificial intelligence (AI) can transform phishing detection using machine learning (ML), which is a branch of AI that employs algorithms to mimic human learning processes (IBM, 2024), and Natural Language Processing

(NLP), which is a research area focused on enabling computers to understand and process human language (Chowdhury, 2020).

This investigation explores how ML algorithms can analyse vast email datasets to detect subtle patterns and anomalies indicative of phishing attempts. Additionally, it examines how NLP can be leveraged to interpret the language of phishing emails, enabling AI to identify social engineering techniques that exploit human emotions and vulnerabilities.

AI in Phishing: Threat Intelligence:

This objective examines the threat intelligence's crucial role in phishing detection via artificial intelligence. To anticipate upcoming attacks and proactively reduce risks, we will investigate how AI can evaluate threat intelligence data, which includes details on well-known phishing techniques, malicious actors, and attack vectors. Additionally, an investigation will happen to ascertain the possibility of a collaborative approach between humans and AI, in which AI performs data analysis and pattern recognition, freeing up human analysts to concentrate on strategic threat interpretation, adaptation, and ethical considerations.

Ethics in AI Phishing Detection:

This objective acknowledges the ethical concerns raised by AI-based phishing detection systems, especially those involving user privacy and data security. A discussion on how crucial it is for AI decision-making processes to be transparent to foster confidence and guarantee responsible implementation will be conducted. To guarantee fair and unbiased threat detection, we will additionally investigate potential biases in the training data and how to eliminate them.

The findings for this research will have a significant impact on the development of more effective and robust systems for phishing detection and prevention. Leveraging AI's capabilities allows for the prediction and counteraction of evolving phishing threats, ultimately resulting in a safer online environment for both individuals and businesses (Hayat, 2023). Furthermore, this research will contribute to the development of ethical frameworks for the use of AI in cybersecurity, guaranteeing its responsible and advantageous implementation in safeguarding user security and privacy.

1.5. Research Questions

Question 1: To what extent can AI technologies like deep learning and machine learning be used to increase the success rate and accuracy of phishing detection systems?

Question 2: In which respects do different AI-powered algorithms for phishing attack detection differ from traditional rule-based methods?

Question 3: What are the main difficulties in implementing AI-based systems for phishing detection and prevention into practice in practical settings?

Question 4: Can AI proactively anticipate and counter emerging phishing techniques?

1.6. Hypotheses/Null Hypothesis

Hypothesis/Null Hypothesis 1:

Hypothesis: An extensive dataset of phishing attempts will demonstrate a statistically significant gain in accuracy when AI-powered phishing detection systems are compared to traditional rule-based solutions.

Null Hypothesis: There is no statistically significant difference in accuracy when AI-powered phishing detection systems and traditional rule-based solutions when recognizing and categorizing different phishing methods across various channels.

Hypothesis/Null Hypothesis 2:

Hypothesis: AI-based anti-phishing frameworks will show greater versatility in recognizing and responding to novel phishing methods as they emerge over time because of their innate learning capabilities.

Null Hypothesis: AI-based anti-phishing frameworks do not show greater versatility in recognizing and responding to novel phishing methods as they emerge over time compared to non-AI-based methods.

Hypothesis/Null Hypothesis 3:

Hypothesis: A quantifiable reduction in the vulnerability of users to phishing attacks will result from the deployment of AI-driven phishing security measures.

Null Hypothesis: The deployment of AI-driven phishing security measures does not result in a quantifiable reduction in the vulnerability of users to phishing attacks, nor does it significantly affect user behaviour or susceptibility to phishing attempts.

Chapter 2: Literature Review

2.1. Contextualization

Evolution of Phishing Attacks:

To understand the nature of phishing attacks and its various stages of evolution in modern society, we must ask ourselves, what are phishing attacks? Phishing attacks are scams with the goal of tricking users into sharing sensitive data, downloading various forms of malware, or putting themselves or their business they represent at risk by exposing themselves to cybercrime (IBM, 2024).

These scams present themselves in various forms, ranging from fraudulent emails, dodgy text messages, phone calls or online websites (IBM, 2024). These methods of propagating phishing scams are designed to trick the user into various actions:

- Download malware (IBM, 2024).
- Share sensitive information (IBM, 2024).
- Various actions that expose themselves or their business (IBM, 2024).

The development is a fascinating story of evolution and adaption. Let's explore this evolution, from the clumsy attempts in the mid-1990's to the dangerous and sophisticated attacks they've become.

Significant Milestones and Evolving Techniques:

In the mid-1990s, dial-up internet usage coincided with the emergence of phishing. Attackers used poor emails full of typos to obtain login credentials and they frequently pretended to be AOL employees (Phishing Box, 2024). Because these early attempts were so simple, it was easy to identify them.

In the 2000s, phishing attacks become more automated and targeted. Social engineering techniques were used by attackers to instil a sense of urgency or take advantage of consumers' faith in well-known companies. The infamous "ILOVEYOU" worm in 2000 highlighted the potential damage of phishing scams.

In 2010s and beyond, phishing techniques grew more complicated as attackers used data breaches to customize emails and impersonate trustworthy businesses almost

exactly. Smishing, or SMS phishing, became a new attack vector, and spear phishing techniques were integrated into phishing tactics to target specific individuals within businesses.

There is common belief among the cybersecurity community that phishing attacks first emerged in the mid-1990's, which would be around the time of the dial-up – the only means in that period of accessing the internet (Wong, 2023).

In this time, phishing attackers frequently imitated AOL employees, which were originally known as America Online (Phishing Box, 2024). They used fake emails and messages as their preferred choice of scam, though thankfully it was fairly easy to detect as it was riddled with misspellings (Phishing Box, 2024), (Phishing Protection, 2024).

Moving on from these outdated phishing techniques, we arrive in the 21st century, where in the second quarter of 2021, June specifically; the number of unique phishing attacks reached a startlingly 222127 amounts, just for that month itself. This investigation, and the subsequent number reached, was conducted by the anti-phishing working group (APWG). (Chanti & Chithralekha, 2022)

With the numerous present variations of fraud, phishing is considered amongst the most frequent examples of fraud activity on the internet (Alkhalil, Hewage, Nawaf, & Khan, 2021).

The Growing Dangers of Phishing Attacks:

Major Cybersecurity concerns include phishing attacks, which are always changing to evade conventional detection techniques (National Cyber Security Centre, n.d) (IBM, 2024). Phishing scams aim to deceive individuals and organizations into clicking on dangerous links or exposing sensitive information (AL-Otaibi & Alsuwat, 2020). The harm phishing attacks due to a business's finances and reputation underlines the urgent need for improved detection and prevention methods (Chanti & Chithralekha, 2022).

The problem is always going to be the human aspect for phishing defence; humans are emotional creatures, which is why AI offers a tempered approach that can't be swayed in its objective. Now, while it can't be swayed it can be overloaded and

outsmarted, which is where we humans, with our critical thinking would come into play to navigate complex social aspects that artificial intelligence (AI) may come up short.

AI and its Potential Role in Phishing Detection:

AI provides the ability to significantly improve phishing detection through the integration of machine learning (ML) and natural language processing (NLP) (Chowdhury, 2020) (IBM, 2024). Large-scale email data can be analysed by ML systems to identify patterns that differentiate authentic emails from phishing attacks (Ahammad, et al., 2022) . Methods such as support vector machines (SVM), random forests, and decision trees can be employed to categorize URLs according to dubious attributes (Ahammad, et al., 2022).

AI systems can understand the language used in phishing emails thanks to Natural Language Processing (NLP), which additionally try to understand social engineering and emotional manipulation techniques (Chowdhury, 2020). This makes it possible for AI to better distinguish between phishing attempts that make use of persuasive language or psychological techniques.

A real-world example of AI and its potential role is PhishHaven, which is a real-time AI Phishing URL Detection System. It's been affirmed that PhishHaven is the first phishing detection system designed exclusively to detect AI-generated Phishing URLs (Sameen, Han, & Hwang, 2020). PhishHaven directly shows proof that this method of phishing defence and prevention is viable as a detection system and wanted by various groups and businesses.

2.2. Theoretical Foundation and how it links to the research problem

Machine Learning (ML) and Pattern Recognition:

Machine learning (ML) is a subfield of AI which allows computers to learn from various types of data without explicit programming (IBM, 2024). Machine learning (ML) algorithms can analyse vast amounts of email data to identify patterns that distinguish real emails from fake phishing attempts.

The evolution of phishing attacks is in constant motion, with attackers using new and far more sophisticated methods to bypass traditional filters. Machine learning's ability to learn and adapt make it crucial for detecting these evolving threats. By understanding how Machine Learning (ML) identifies the patterns in phishing scams, we can explore how these techniques are being used to develop more adaptable phishing detection systems.

Natural Language Processing (NLP) and Social Engineering:

With the ability to converse with computers in their native tongue rather than having to learn a foreign language of computer commands, Natural Language Processing (NLP) has enormous potential to create computer interfaces that are more user-friendly for humans (Chowdhury, 2020).

Phishing attacks include social engineering, in which the attacker uses psychological scams and human emotions to fool victims into divulging personal information (AL-Otaibi & Alsuwat, 2020). NLP is useful in spotting these fraudulent methods used in phishing emails. Phishing detection systems can be made more successful by examining how Natural Language Processing (NLP) is applied to recognize social engineering methods in phishing emails.

Justification for Selection:

These two theoretical foundations have been chosen as they represent crucial core functionalities underpinning AI-powered phishing detection and prevention systems. While NLP focuses on understanding the language used for social engineering, a crucial component of phishing emails, machine learning offers the analytical strength to identify patterns in phishing attempts.

Key Theories:

Network Security:

Phishing attacks frequently take advantage of weaknesses in protocols and network infrastructure (Hayat, 2023). Phishing prevention can benefit from an understanding of network security theories like anomaly and intrusion detection.

Social Engineering:

Social engineering is a manipulative art form that preys on human emotions and vulnerabilities and is frequently used in phishing attacks. To generate anxiety, dread, or a sense of urgency, attackers create plausible situations that deceive victims into clicking on fraudulent websites or disclosing personal and private information (AL-Otaibi & Alsuwat, 2020).

Knowing social engineering tactics like urgency, authority, and reciprocity helps us to identify red flags and create awareness campaigns that provide users with the tools they need to fight back against these manipulations (AL-Otaibi & Alsuwat, 2020).

Machine Learning:

Modern phishing detection systems leverage the power of machine learning algorithms. Vast amounts of data, such as email content, sender details, and attachment types, can be analysed by these algorithms.

Machine learning can discover tiny changes in new attacks by recognizing patterns and characteristics associated with known phishing attempts (IBM, 2024). This helps distinguish legitimate emails from malicious ones. Machine learning algorithms are always evolving, which makes it possible for them to adjust to new phishing techniques and improve overall security posture (IBM, 2024).

2.3. Details of the issue/problem being investigated

Limitations of Traditional Methods:

The complex nature of modern-day phishing attacks surpassing the efficiency of anti-phishing techniques like content filtering and blacklists (Wong, 2023). Email content analysis is a difficult endeavour since phishing attackers use social engineering techniques to take advantage of human vulnerabilities (Ansar, Sharma, & Dash, 2022). Conventional approaches frequently depend on static rules, which attackers can simply circumvent by changing their tactics (Alkhalil, Hewage, Nawaf, & Khan, 2021).

These limitations underscore the necessity for more adaptable and intuitive detection methods, opening the door for AI to revolutionize the fight against phishing. Though, while the role of AI and integration is very important, I believe that it should be tempered by a human hand, a symbiotic partnership between AI and human expertise.

Traditional phishing detection methods are hindered by their reactive nature and reliance on predefined signatures and patterns. For instance, blacklists only block known harmful domains or URLs, which leaves consumers open to new, unidentified phishing sites (Wong, 2023). Content filtering systems also struggle to detect sophisticated attacks, where phishers subtly alter email text or use personalized messages to avoid detection (Wong, 2023). Furthermore, these techniques frequently result in false positives, which mark legitimate exchanges as phishing attempts. It is challenging to stay up to date with threats that constantly evolve, which highlights the need for AI-driven solutions to improve phishing detection.

AI would offer unparalleled analytic power and presumably pattern recognition, while humans would contribute critical thinking and the ability to understand complex social engineering tactics. With this partnership growing together, it would allow for further research into AI models, which would build trust in these systems in a natural time frame.

Threat Intelligence:

The process of obtaining and examining information regarding known phishing schemes, and compromised websites, among other potential phishing risks (Wagner, Mahbub, Palomar, & Abdallaha, 2019).

In the context of threat intelligence, this study will research various necessities related to this choice, ranging from threat actors to attack techniques. Threat actors will be looking into the culprits that initiate these phishing campaigns. What are their motives (financial, espionage etc.) and preferred methods? How phishing attacks perpetrate their attack techniques will also be analysed. Techniques ranging email spoofing, fake login pages and more.

The benefits of artificial intelligence (AI) and various threat intelligence programs integration will also be researched with the hope of finding a clear linkage stating that artificial intelligence (AI) is a viable source that would help and secure people, businesses and stakeholders in their fight against phishing schemes (Wagner, Mahbub, Palomar, & Abdallaha, 2019).

Threat Intelligence and AI:

Threat intelligence, which provides data on recognized phishing schemes, malicious actors, and attack methods, is essential for AI-powered phishing detection. AI can analyse threat intelligence data to identify emerging patterns and predict future phishing campaigns. Businesses can stay ahead of emerging threats through employing this proactive approach. (Wagner, Mahbub, Palomar, & Abdallaha, 2019)

However, there is an ongoing debate about the responsibility AI should hold in threat intelligence analysis:

Arguments for AI Primacy (Sufi, 2023):

Scalability and Speed:

AI primacy proponents argue that AI is capable of processing massive amounts of data far more quickly and effectively than humans, allowing for real-time danger detection and reaction.

Pattern Recognition:

AI is especially effective at spotting subtle patterns in data that human analysts can overlook. This makes it possible to identify new risks and scams early on.

Reduced Human Error:

AI can lessen human biases and fatigue, which results in more reliable and impartial threat evaluations.

Arguments for Human-AI Collaboration (Yue & Li, 2023):

Context and Interpretation:

While AI is extremely effective at identifying patterns, people are still superior at deciphering the meaning and context of the data. Human analysts can evaluate the data and rank threats according to the possible consequences.

Evolving Threats:

Cybercriminals frequently shift their techniques. To keep ahead of these constantly evolving threats, human expertise is essential for adapting AI models and threat intelligence feeds.

Ethical Considerations:

Biases in the training data may be reinforced by AI algorithms. To guarantee moral decision-making and avoid unexpected outcomes, human oversight is critical.

2.4. Links between problem and current literature

Seminal Authors/Sources:

Kevin Mitnick:

As a security consultant and experienced hacker, Kevin Mitnick offers valuable perspectives on the human aspect of cyberattacks, particularly social engineering. His groundbreaking research highlights that without addressing how attackers take advantage of human trust and mistakes, technical measures alone are ineffective.

Because of his experiences, Mitnick demonstrates how important it is to comprehend these psychological methods to enhance security awareness training and put in place effective solutions. Mitnick provides in-depth case studies of actual hacking situations in his book “The Art of Deception”, where social engineering tactics including phishing, impersonation, and pretexting were employed to breach even the most secure networks. By establishing the foundation for contemporary security awareness programs, these publications advance the subject by highlighting the importance of incorporating human behaviour analysis into cybersecurity networks. (Mitnick & Simon, 2005)

Bruce Schneier:

A prominent figure in cybersecurity, Bruce Schneier is renowned for offering a skilled perspective on the relationship between technology, security, and human behaviour. Through his work, cybersecurity is framed as more than merely a technological problem, and practitioners are encouraged to consider larger societal and policy concerns. One of his major publications, “Applied Cryptography”, has been an essential tool for comprehending encryption and secure communication protocols, helping to develop cryptographic standards that are still in use today. (Schneier, 2024)

The repercussions of an interconnected society are explored in greater detail by Schneier in “Click Here to Kill Everybody”, which also highlights the risks brought about by the Internet of Things (IoT) and the necessity of comprehensive, multi-layered security methods that include both technical and human defences. (Schneier, 2024)

Peter Drucker:

While he was not a specialist in cybersecurity, Peter Drucker transformed management theory, and cybersecurity governance can greatly benefit from his ideas. His work highlights how crucial it is to have an efficient organizational structure, excellent communication channels, and sensible risk management procedures to keep your business safe (Drucker, 2024).

In “Management: Tasks, Responsibilities, Practices,” Drucker describes how proactive risk management is fostered by strong leadership and well-defined tasks, which is essential in the context of cybersecurity. Drucker's understanding of communication and organizational behaviour has useful applications for leading cybersecurity teams, upholding regulations, and raising employee awareness—all essential components of any cybersecurity program. (Drucker, 2024)

Additionally, instead of considering cybersecurity as a distinct technological problem, his theories encourage business leaders to regard it as a fundamental business function that is integrated into the organization's broader strategy (Drucker, 2024).

Literature Review Themes:

Evolution of Phishing Methods: This topic looks at how phishing strategies have changed over time, from straightforward email scams to intricate multi-vector attacks which seeks to target the papers populations youthful demographic.

AI in Cybersecurity: Analysing various methods, such as rule-based systems, heuristic analysis, and machine learning algorithms, for identifying phishing attempts.

Human Factors: Knowledge of the psychological and behavioural characteristics of attackers and targets in phishing attempts, such as decision-making processes and cognitive biases.

Phishing Awareness Training:

Instructional courses, such as simulated phishing exercises, that train staff members how to spot and handle phishing attacks. With these teachings, employees can be equipped with sufficient knowledge and training to help them recognise potential phishing scams, which could potentially save themselves or their business vast amount of grief and trauma (Ansar, Sharma, & Dash, 2022).

This concept was chosen for one specific reason. People... it's always the people. Employees are always the biggest vulnerability in any business as phishing attacks are designed to gather sensitive data from people, through trickery.

Phishing URL Detection:

Methods for locating malicious URLs included in emails or webpages frequently involves the analysis of the structure and content of the URL using machine learning algorithms. This concept focuses exclusively on identifying malicious URLs that phishing attackers embed in emails, webpages and more. Below is a breakdown of various detection methods:

Machine Learning (ML) algorithms:

Decision tree algorithm:

Imagine a flowchart where you answer a series of questions to arrive at a decision. That is a decision tree algorithm's basic premise. It creates a model that resembles a tree, with nodes standing in for inquiries about the data and branches for potential responses—basically, an if-else question (Ahammad, et al., 2022). Based on these questions, the algorithm divides the data iteratively until it can categorize or predict a target value (e.g., classify a phishing URL as harmful or benign).

Random forest algorithm:

This is an ensemble learning method; it integrates predictions from several decision trees. Imagine a forest of decision trees, each being able to produce a unique forecast. Random Forest takes an average (for regression) or majority vote (for classification) from all the trees to get a final, more robust prediction. (Ahammad, et al., 2022)

Light GBM:

This is a powerful gradient boosting approach that generates a model ensemble in a sequential fashion. It is a framework based on decision trees that uses EFB and GOSS. By concentrating on data points that the earlier models had trouble with, each model attempts to address the shortcomings of the one before it. One especially effective kind is Light GBM, which is renowned for its precision and swiftness. (Ahammad, et al., 2022)

SVM:

Using supervised learning as its foundation, SVM is a machine learning technique that may be applied to regression as well as classification. The Support Vector Machine (SVM) is seeing rapid use due to its robust base in statistical learning theory and its success in addressing various data mining difficulties. (Ahammad, et al., 2022) Imagine a two-dimensional plane in which a line is drawn to separate oranges and apples. SVMs are appropriate for difficult classifications since they can accomplish this in higher dimensions. They are also adaptable to jobs involving regression.

Logistic Regression:

For Logistic Regression, which will not be used, but is an easy-to-understand algorithm, it assumes that input features and the likelihood of phishing have a linear relationship, which may not be adequate for the complexity of phishing URLs. Models like Random Forest and SVM are more effective at capturing the non-linear patterns found in phishing URLs. While logistic regression works effectively when there is a definite, linear class distinction, it may not be able to handle feature interactions that are more complex and call for more intricate decision limits. (Ahammad, et al., 2022)

In the context of phishing URL detection, all the algorithms listed above can be used to examine a URL's characteristics (such as length, the existence of special characters, etc.) and determine the likelihood that it is harmful.

Chapter 3: Research Design and Methodology

3.1. Research Approach and Design used to investigate the problem

Paradigm:

This research is based on positivism as the philosophical paradigm. According to the positivist research paradigm, the purpose of knowledge is to provide clarification on the things that we encounter in our daily lives (Park, Konge, & Artino, 2020). Built around quantifiable and observable facts, this paradigm emphasises the necessity of objective evidence to establish scientific truths (Park, Konge, & Artino, 2020). Typically, positivism makes use of quantitative techniques that enable the testing of hypotheses and the generation of generalisable results (Park, Konge, & Artino, 2020).

For this research, positivism is appropriate for several reasons. The initial objective of the research is to assess how effectively AI can detect and prevent phishing attacks. This objective is consistent with positivism, which emphasises empirical evidence and measurable data. Using quantitative techniques like statistical analysis and machine learning algorithms, the research can evaluate AI's performance against that of traditional methods objectively.

Furthermore, the process of phishing detection necessitates the study of large data sets to identify patterns and anomalies. This approach is supported by positivism, which promotes systematic collection and analysis of data to uncover objective truths (Park, Konge, & Artino, 2020). Using this paradigm, particular hypotheses regarding the effectiveness of AI in phishing detection can be tested and conclusive, through which evidence-based results can be obtained. Therefore, a positivist approach guarantees that the findings are founded on exacting, scientific methodologies, making them reliable and transferable to more extensive cybersecurity situations.

Approach/Design:

Research Design: Quantitative Design

This study uses the “EMAIL SPAM DETECTION,” which was acquired from Kaggle.

	v1	v2
0	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there g...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive ...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives around here though
5	spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for ...

Table 1. 1: Table showing five rows of the email sent, and whether or not they are ham and spam. (Source: Developed for the research study)

This dataset is frequently used in the text classification and spam detection fields. It includes labelled examples of email messages that have been categorized as either spam or ham (legal). Of the 5,572 email messages that are included, thirteen percent is categorized as spam, and the remaining 87 percent are categorized as ham. This disparity between the two classes illustrates a typical situation in spam detection, when the proportion of valid communications greatly exceeds that of spam messages, making it difficult to achieve balanced model performance.

The dataset was first gathered from a variety of publicly accessible email message sources. This may have included user submissions, data scraping from email platforms, or public datasets released for academic purposes.

Data Analysis Techniques:

Descriptive Statistics:

Descriptive statistics will be employed to analyse the secondary data from the dataset. This involves figuring out how to summarize the data using measures of central tendency (mean, median, and mode), as well as how frequently different sorts of phishing instances occur and how common it is for different AI technologies to be employed in detection.

Regression Analysis:

Regression analysis will be used to explore relationships between variables, such as the rate of phishing incidents and the adoption of AI technologies. This technique will

assist in ascertaining whether the use of AI solutions and a decline in phishing attempts are statistically related. To isolate the impact of AI adoption and control for other variables, multiple regression analysis may be employed.

Comparative Analysis:

Comparative analysis will be conducted to evaluate how well various AI algorithms identify phishing attempts. These algorithms' performances will be compared and evaluated using metrics including precision, recall, and F1-score. This will assist in determining which AI models are best for phishing detection.

Tools and Instruments:

Machine Learning Frameworks:

Platforms like TensorFlow, Scikit-learn, and Keras can be utilized to implement and test various machine learning models for phishing detection. Building, training, and evaluating AI algorithms are made possible by these frameworks, which offer the tools required.

Visualization Tools:

Program tools like Tableau, Power BI, or Matplotlib (Python) can be used to visualize the data. Understanding trends, patterns, and the overall effectiveness of AI technology in phishing detection will be made easier with the use of visualization.

Data Sources:

Reliable and comprehensive datasets from resources like Kaggle will be used. This dataset offers real-world data that is essential for evaluating how well AI detects phishing detection.

This research paper attempts to offer an extensive and comprehensive analysis of the potential of AI technologies to improve phishing detection and prevention by utilizing these techniques and tools. The findings will aid in the development of cybersecurity defences that are more adaptable and effective, ultimately shielding individuals and companies from the ever-changing threat posed by phishing scams.

Machine Learning Model:

Model Selection:

This research will explore the following machine learning model types for phishing detection:

Classification Algorithms:

These algorithms work correctly in situations where organizing data points into different classes is the primary goal. Algorithms for classification can be employed in phishing detection to identify emails that are legitimate or phishing. Common examples include:

Random Forest: This is an ensemble learning method; it integrates predictions from several decision trees. Random Forest takes an average (for regression) or majority vote (for classification) from all the trees to get a final, more robust prediction. (Ahammad, et al., 2022)

Support Vector Machine (SVM): Using supervised learning as its foundation, SVM is a machine learning technique that may be applied to regression as well as classification. The Support Vector Machine (SVM) is seeing rapid use due to its robust base in statistical learning theory and its success in addressing various data mining difficulties. (Ahammad, et al., 2022) SVMs are appropriate for difficult classifications since they can accomplish this in higher dimensions. They are also adaptable to jobs involving regression.

Natural Language Processing (NLP) Techniques:

TF-IDF (Term Frequency-Inverse Document Frequency):

Allocates weights to terms according on their rarity throughout the dataset and their frequency in an email (Rahman, Rahman , Been, & Sarker, 2020). This assists in locating keywords indicative of attempted phishing.

N-grams:

Examine word sequences (trigrams, bigrams, etc.) to identify phrases or sentence structures that are frequently found in phishing emails (Rahman, Rahman , Been, & Sarker, 2020).

Selection Criteria:

- Accuracy.

- Precision.
- Computational efficiency.
- Interpretability.

Model Training:

A large dataset of labelled emails will be prepared as part of the training process. This includes:

Data collection: Gathering a balanced dataset of legitimate and phishing emails from reliable sources such as Kaggle.

Data cleaning: This is the process of removing irrelevant information from the data, such as HTML code and attachments, and repairing any errors or inconsistencies.

Feature engineering: Extracting relevant features from the emails to train a model.

Normalization: This is the process of scaling numerical features to a common range such that each feature contributes equally to the predictions of the model.

The prepared dataset will be used to train the selected machine learning models. This is how the training procedure is broken down:

Splitting the Dataset:

The dataset will be divided into test, validation, and training sets. The training set is used to build the model, the validation set is used to fine-tune hyperparameters (model settings) to prevent overfitting, and the test set is used for final evaluation of the model's performance on unseen data.

Model Training:

Using the training set, the selected algorithms will be trained. To reduce prediction errors, this entails continuously modifying the internal parameters of the model after successively adding features and related labels (phishing or legitimate).

Model Selection:

The model with the best performance (considering accuracy, precision, recall, and interpretability) will be selected for further evaluation based on the validation results.

Model Evaluation:

The trained model's effectiveness will be evaluated using a variety of metrics:

Accuracy: The proportion of emails that the model properly classifies as legitimate or phishing.

Precision: The percentage of phishing-classified emails that are genuinely phishing emails.

Recall: The percentage of real phishing emails that the model accurately detects.

F1-Score: A balanced assessment of the model's performance derived from the harmonic mean of precision and recall.

False Positive Rate (FPR): The percentage of legitimate emails that are incorrectly identified as phishing.

False Negative Rate (FNR): The percentage of phishing emails that are incorrectly identified as legitimate.

3.2. Data Collection Methods

Population:

Those who frequently communicate via email make up the target population for this study. This includes a diverse group of email users who are potentially exposed to spam and phishing attempts. The realistic subset of this group that is available for this study consists of email users from a range of demographics and is not limited to any one institution or location. This group is likely to encounter spam emails in their daily email interactions.

Sampling:

A comprehensive plan to sampling is required to study and use the Kaggle "EMAIL SPAM DETECTION" in an effective manner. This guarantees that the model we create is accurate and applicable to a wide range of real-world situations.

For this dataset, we will employ random sampling. Since it provides every entry in the dataset an equal chance of being chosen, random sampling is especially appropriate in this situation as it helps to preserve the integrity and representativeness of the data.

We can ensure that the analysis and model training have no bias towards either class by using random sampling to maintain the balance of the dataset, which consists of 87% ham and 13% spam.

Random sampling will involve selecting emails from the dataset in such a way that each email has an equal probability of being included. Maintaining the proportionate distribution of the classes is important for creating a trustworthy machine-learning model, and this approach will be simple and efficient.

Sample Size: The dataset consists of 5169 unique emails with two columns each. After taking into consideration various data cleaning methods that can dismiss corrupt or irrelevant records, this study aims to incorporate all 5169 unique values. This comprehensive approach ensures that we leverage the full richness and diversity of the dataset.

Using the entire dataset of 5169 unique entries as the sample size allows for robust training and validation of our phishing detection model. Integrating all available data optimizes algorithms' learning potential and enhances their capacity to generalize from the given data. This is especially crucial for phishing detection, as a large sample size aids in the development of a model that can recognize both common and obscure phishing attempts.

In summary, by employing random sampling and using the entire dataset of 5169 unique emails, we ensure a balanced and comprehensive approach to building an effective phishing detection model.

Parameters:

Participants must be active email users who routinely check and respond to emails. This includes individuals from various age groups and professional backgrounds, encompassing students, academic staff, administrative personnel, and other professionals. Participants' inboxes should have a history of receiving a mixture of spam and legitimate (ham) emails.

Email messages classified as spam or ham make up the dataset used during this study, which was gathered in order to train and assess a machine learning model for spam detection. It is important to highlight that this study employs secondary data,

meaning that no new information will be collected because the dataset contains all the necessary information.

Dataset Points:

The Datasets Structure:

The file format contains one message per line, with two columns: 'v1' (the label) and 'v2' (the raw text). The label 'v1' indicates whether the email is ham (legitimate) or spam while 'v2' is the raw text.

Scope of the Dataset:

Label Distribution:

- Ham: 87%
- Spam: 13%

Total Entries: 5169 unique messages.

Dataset Additional Information:

Unique Values: 99% of the entries are unique, indicating a diverse set of messages.

Data Quality: Preprocessing is required to address a few entries that have null values and other anomalies.

Curation: The data underwent the standard data science process, which include data acquisition, preprocessing, extracting features, training models, evaluation, and developing predictions.

Limitations: There are possible limitations to using secondary data from the provided dataset, such as the possibility of bias in the original data collection, lack of control over the completeness and data quality, and possible issues with the data's relevance to specific demographics.

3.3. Data Analysis Techniques

Quantitative Analysis:

The primary focus of this study will be on quantitative analysis. To find patterns, correlations, and trends, numerical data must be statistically analysed using

quantitative analysis (Ahmad, et al., 2019). The data will first be summarized using descriptive statistics, which yield metrics like mean, median, and standard deviation. Inferential statistics, including hypothesis testing and regression analysis, will follow to determine relationships between variables and to predict outcomes. For instance, logistic regression can be employed to classify URLs/emails as legitimate, or phishing based on their features. The statistical approach aids in comprehending how well various features detects phishing attacks.

Machine Learning Analysis:

Machine learning will be a crucial component of this study. A variety of algorithms will be used to create predictive models that can recognize phishing URLs with accuracy. Key steps in this process include:

Data preprocessing: Includes cleaning the data, handling missing values, and normalizing features.

Feature Selection: Identifying the most relevant features that contribute to phishing detection.

Model Selection: Using Random Forests, Decision Trees, and Support Vector Machines (SVM) are some examples of appropriate algorithms. Though SVM works well for binary classification problems, Random Forest can handle larger datasets with greater accuracy. Every technique has its benefits.

Model Training and Validation: To train the models and assess their effectiveness, divide the dataset into subsets for testing and training.

Performance Evaluation: Using metrics like accuracy, precision, recall, and F1-score to evaluate the models' effectiveness.

Integration:

Since this study focuses primarily on quantitative analysis and machine learning, the integration process will involve combining insights from these two approaches. Through statistical insights into the data, quantitative analysis will highlight important patterns and relationships. Based on these discoveries, machine learning models will provide prediction tools for phishing detection. A thorough examination where statistical insights influence the development of machine learning algorithms is

ensured by combining these two methods, producing phishing detection systems that are dependable and strong. This synergy between quantitative analysis and machine learning enhances the overall understanding and effectiveness of the study's outcomes, contributing significantly to the field of cybersecurity.

3.4. Validity, reliability, and/or trustworthiness of the findings

Data Sources:

- (Anaghakp, 2023): <https://www.kaggle.com/code/anaghakp/email-spam-detection/input>

Once the dataset and additional sources have been obtained, the process for gathering data entails careful measures to ensure the reliability and accuracy of the information gathered.

Data will be gathered by thoroughly reviewing the acquired dataset on phishing scams and AI-powered cybersecurity solutions. The data's accuracy, relevance, and comprehensiveness will be examined.

The quality assessment will ensure that the dataset satisfies established guidelines by doing an evaluation of quality. This component involves evaluating factors such as data integrity, consistency, and representativeness.

Data Verification will be conducted to authenticate the credibility of the data sources. To confirm that the data is accurate, cross-reference the information from several reliable sources will be applied.

Carefully document every step of the data-gathering process, including details regarding the data sources, collection methods, and any modifications made to the original datasets. This documentation serves as a reference for future analyses and ensures transparency and reproducibility in research findings.

To ensure the validity and reliability of my findings, several additional measures will be implemented:

Reliability: To ensure that the results are consistent, testing will be done repeatedly.

Construct Validity: There will be precise definitions of key terms, such as AI performance and phishing detection accuracy, to guarantee precise measurement and lessen the uncertainty of the investigation.

These measures will ensure that the data gathered is accurate, dependable, and suitable for analysing phishing scams and AI-driven cybersecurity solutions effectively.

3.5. Ethical implications and how they were addressed

Ethics:

Artificial intelligence (AI) has revolutionised cybersecurity, especially regarding phishing detection and prevention. However, ethical considerations about user privacy and the inner workings of AI systems itself must be addressed in conjunction with these advancements.

Privacy Concerns:

At the heart of AI-powered phishing detection lies its impressive ability to analyse large quantities of user data, such as personal and sensitive data like browsing history and email content (Bešić, 2023). While this analysis is very important for effective phishing detection, it would naturally raise valid concerns about data privacy and user consent. The respect for individual privacy rights must be balanced with AI's ability to access and process such data.

Therefore, it is essential to safeguard user privacy when adopting AI for phishing detection. Establishing explicit and open procedures for data collecting, with a focus on user consent and compliance with data protection laws, is crucial for organisations implementing these kinds of solutions (Al-Mansoori & Salem, 2023). By taking these steps, users should be able to manage the collection and processing of their personal data and be informed about how it is being used.

Transparency and Accountability:

Regarding accountability and the capacity to comprehend decision-making processes, the opacity of AI algorithms presents concerns. Trust and responsibility can only be established by guaranteeing openness in the operation of AI systems used for phishing detection (Al-Mansoori & Salem, 2023).

It can be difficult for people to fully grasp these sophisticated algorithms, which makes it tough to understand how they determine on whether to accept or reject phishing scams. Accountability is hampered by this lack of transparency because it's challenging to determine the root reason when an AI system misinterprets a valid email or overlooks a sophisticated phishing attempt.

More understandable AI models should be developed, and developers should place a high priority on communicating with users on how their systems work. Ensuring openness in the operations of AI-powered phishing detection is crucial to fostering confidence and accountability in the field (Al-Mansoori & Salem, 2023).

3.6. Limitations of the study

Limitations:

Although artificial intelligence (AI) presents notable progress in phishing detection, certain limitations persist that require human intervention.

Human Oversight Requirement:

Even with AI breakthroughs, human oversight is still essential. Artificial intelligence (AI) systems could be emotionally or contextually ignorant, requiring human assistance to properly grasp complex circumstances (Hayat, 2023). Human judgment is especially crucial in complex settings. Identification of a spear-phishing assault, which is a specific attack on a single individual, or deciphering a carefully crafted email that tries to deceive its recipient, requires human skill.

Artificial intelligence (AI) has the potential to be a very useful tool for pattern recognition, but it is unable to substitute the place of human judgment and critical thinking when making complicated security decisions (Hayat, 2023).

False Positives and Negatives:

Phishing detection systems driven by artificial intelligence (AI) have several limitations. They can generate false positives, which identify genuine emails as phishing efforts, or false negatives, which miss advanced phishing attacks (Hayat, 2023). It's still difficult to minimise false alarms while maintaining detection accuracy (Hayat, 2023).

Workflow disruption and user annoyance may result from this, particularly if they receive a lot of unnecessary warnings. For developers, striking the correct balance between reducing false alarms and preserving high detection accuracy is a constant challenge. For AI systems to continually adapt to emerging phishing techniques, it calls for continuous study and development for the aim of rapid improvement (Hayat, 2023).

Chapter 4: Findings and Interpretation of Findings

4.1. Introduction

Research Aim and Objectives

This research examined the effectiveness of artificial intelligence (AI) technologies—more especially, machine learning (ML) and deep learning—in improving phishing detection and prevention is addressed. The objective of the study is to evaluate the advantages and disadvantages of AI-powered algorithms in terms of performance versus traditional rule-based systems. Phishing attacks have surpassed traditional detection techniques, especially as they become more sophisticated. This has made a move toward more intelligent and adaptable systems necessary. Additionally, this research will investigate AI's capacity to proactively detect and block emerging phishing methods, hence enhancing user safety in the digital realm.

The objectives guiding this chapter include:

1. Revealing Phishing: Limitations and Tactics: This aim analyses how standard phishing detection methods—such as blacklists and content filtering—are insufficient in the face of modern social engineering tactics. These strategies are examined alongside phishing attempts that use of sophisticated evasion tactics and AI-generated URLs.
2. AI Revolutionizing Phishing Detection: ML and NLP Insights: Understanding how natural language processing (NLP) and machine learning help detect patterns and anomalies suggestive of phishing scams is the objective of this research endeavour. It focuses on how AI, trained on large datasets, may recognise the tiny signals that rule-based systems frequently overlook.
3. AI in Phishing: Threat Intelligence: The discussion will examine artificial intelligence's (AI) capacity for threat intelligence and how well it can process vast volumes of threat data. It will highlight how AI may help human analysts by automating repetitive tasks and enhancing the detection of possible phishing situations.

4. Ethics in AI Phishing Detection: The ethical issues surrounding the use of AI in phishing detection are addressed by this objective, especially those pertaining to data security, privacy, and the likelihood of bias in AI decision-making.

Reiterating Questions and Hypothesis:

1. To what extent can AI technologies like deep learning and machine learning be used to increase the success rate and accuracy of phishing detection systems?
2. In which respects do different AI-powered algorithms for phishing attack detection differ from traditional rule-based methods?
3. What are the main difficulties in implementing AI-based systems for phishing detection and prevention into practice in practical settings?
4. Can AI proactively anticipate and counter emerging phishing techniques?

Hypothesis/Null Hypothesis 1:

Hypothesis: An extensive dataset of phishing attempts will demonstrate a statistically significant gain in accuracy when AI-powered phishing detection systems are compared to traditional rule-based solutions.

Null Hypothesis: There is no statistically significant difference in accuracy when AI-powered phishing detection systems and traditional rule-based solutions when recognizing and categorizing different phishing methods across various channels.

Hypothesis/Null Hypothesis 2:

Hypothesis: AI-based anti-phishing frameworks will show greater versatility in recognizing and responding to novel phishing methods as they emerge over time because of their innate learning capabilities.

Null Hypothesis: AI-based anti-phishing frameworks do not show greater versatility in recognizing and responding to novel phishing methods as they emerge over time compared to non-AI-based methods.

Hypothesis/Null Hypothesis 3:

Hypothesis: A quantifiable reduction in the vulnerability of users to phishing attacks will result from the deployment of AI-driven phishing security measures.

Null Hypothesis: The deployment of AI-driven phishing security measures does not result in a quantifiable reduction in the vulnerability of users to phishing attacks, nor does it significantly affect user behaviour or susceptibility to phishing attempts.

Data Source Description:

This chapter's analysis makes use of a secondary dataset—**EMAIL SPAM DETECTION**: (spam.csv)—which contains a substantial amount of email data categorized as either ham (non-phishing) or spam (phishing). This dataset offers an excellent foundation for evaluating phishing detection systems that use artificial intelligence in addition to traditional rule-based methods. To be more precise, employed rule-based methods that classified spam by using email message length as a heuristic. On the other hand, to improve the accuracy of phishing detection, artificial intelligence (AI)-based techniques including logistic regression, and a neural network model were used to examine patterns and features in the email content.

Considering the dataset makes it possible to compare more recently developed AI-driven models with traditional rule-based systems in a helpful manner, its relevance to the research questions is significant. Using this data, the effectiveness of AI technologies was evaluated—such as deep learning and machine learning—in identifying phishing attempts in comparison to more conventional techniques. The dataset additionally helps in investigating how AI may lessen consumer susceptibility by enhancing the detection of unsafe content.

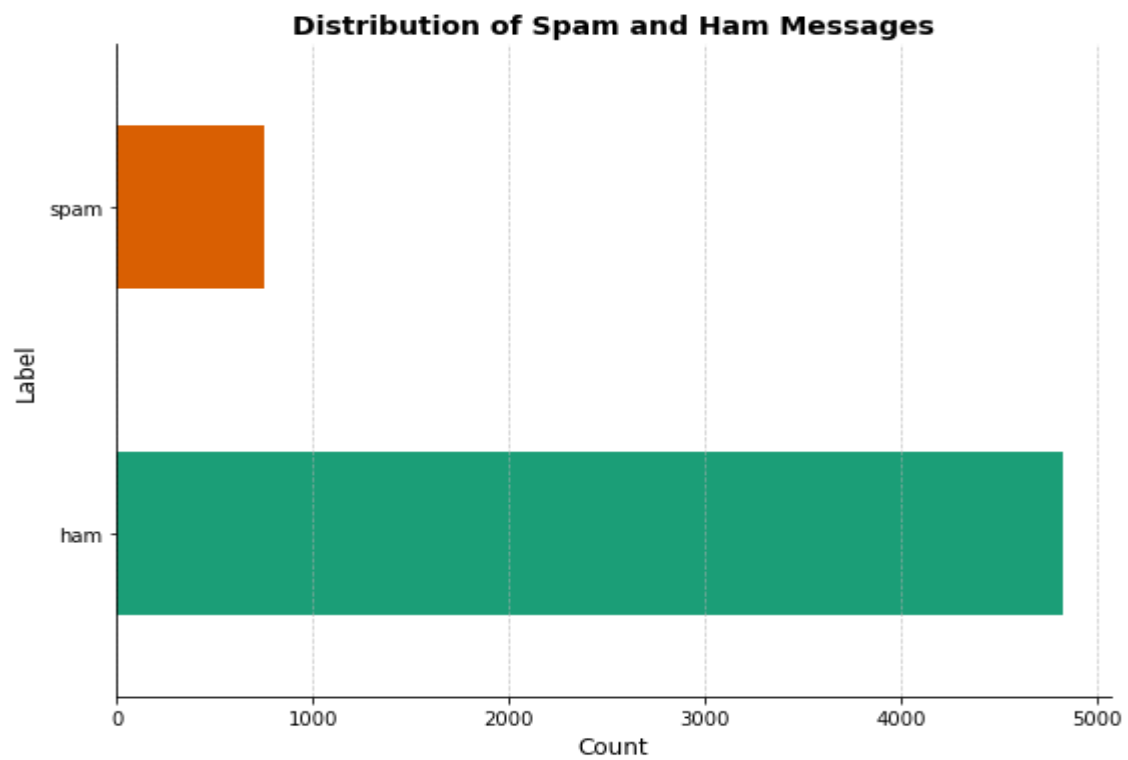


Figure 1. 1: Horizontal Bar Graph showing the ham and spam amount. (Source: Developed for the research study)

4.2. Presenting and Discussing Findings

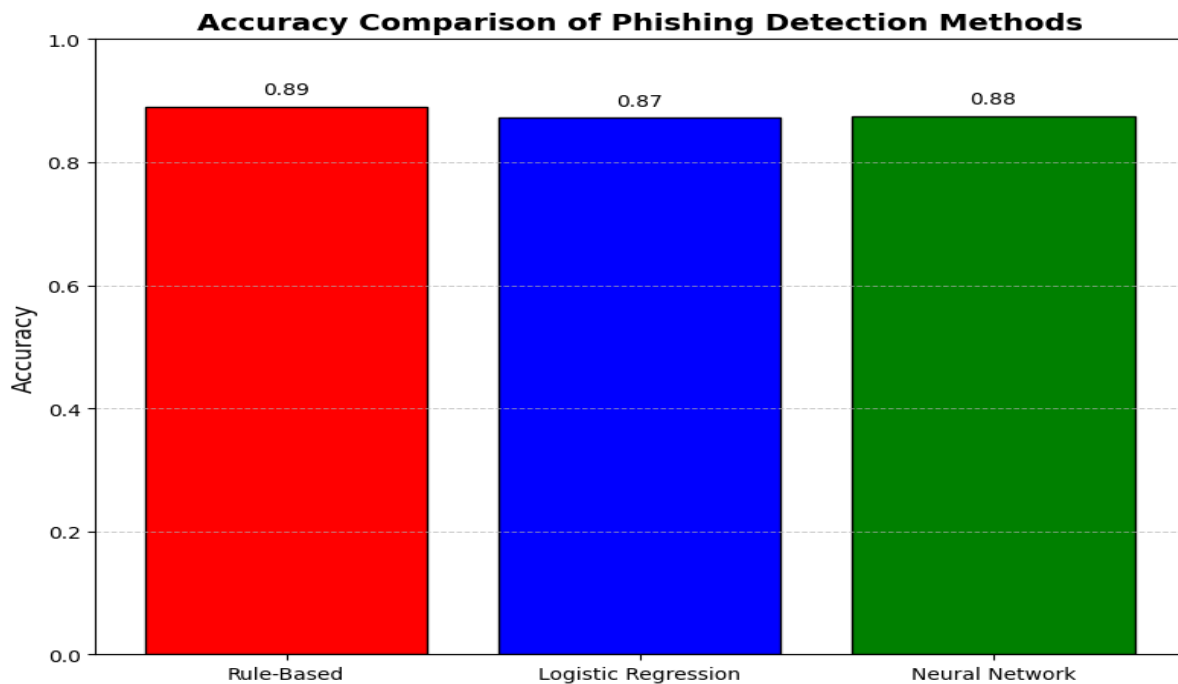


Figure 1. 2: Bar Graph showing the accuracy comparison of phishing detection methods. (Source: Developed for the research study)

--- Interpretation of Accuracy Results ---

The rule-based system performs better than the AI models in this dataset.

Table 1. 2: Interpretation of Accuracy Results. (Source: Developed for the research study)

Logistic Regression Classification Report:				
	precision	recall	f1-score	support
Ham	0.90	0.96	0.93	965
Spam	0.54	0.29	0.38	150
accuracy			0.87	1115
macro avg	0.72	0.62	0.65	1115
weighted avg	0.85	0.87	0.85	1115

Neural Network Classification Report:				
	precision	recall	f1-score	support
Ham	0.90	0.96	0.93	965
Spam	0.56	0.34	0.42	150
accuracy			0.88	1115
macro avg	0.73	0.65	0.68	1115
weighted avg	0.86	0.88	0.86	1115

Table 1. 3: Logistic Regression Classification Report and Neural Network Classification Report are the two tables above. (Source: Developed for the research study)

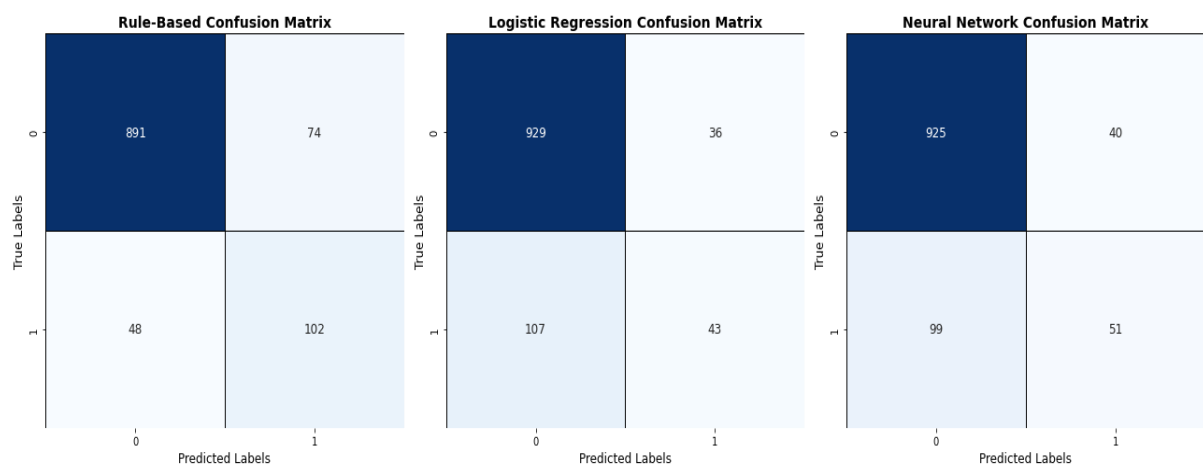


Figure 1. 3: Confusion Matrix for Rule-Based, Logistic Regression and Neural Network. (Source: Developed for the research study)

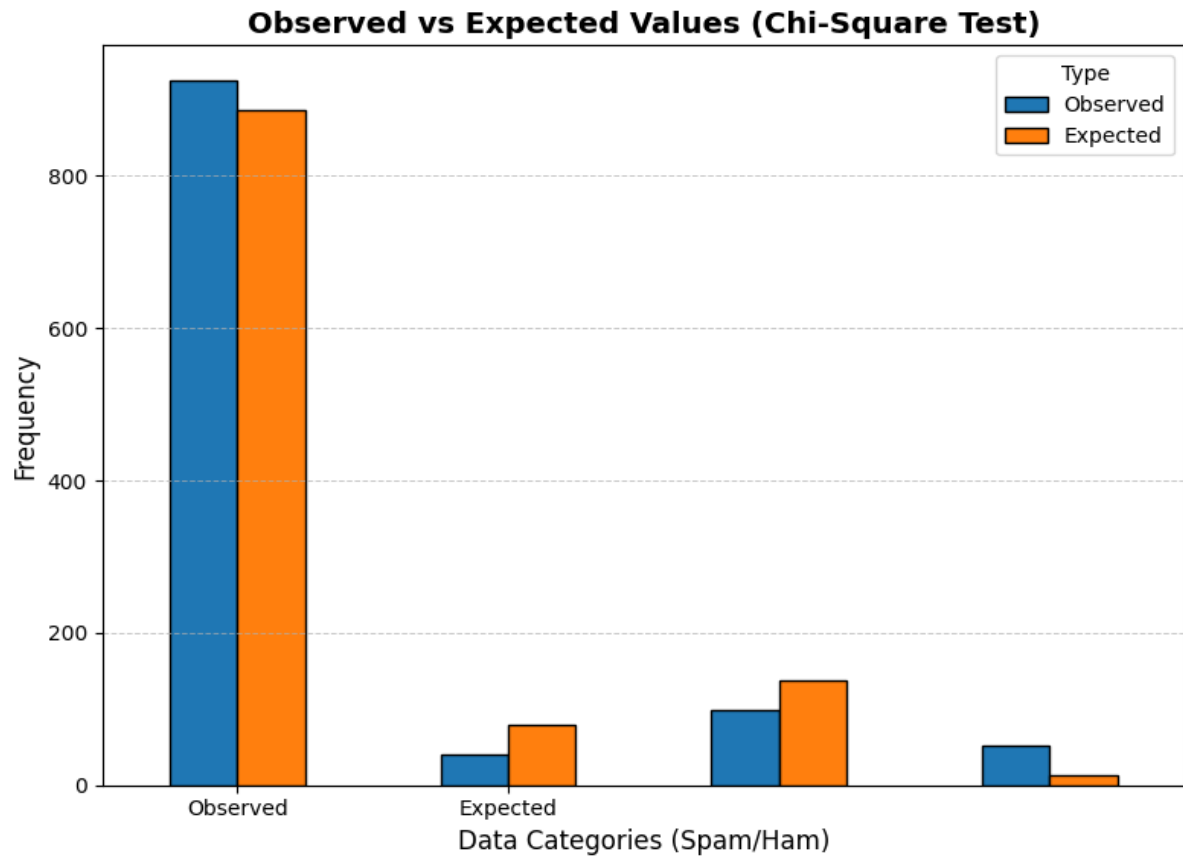


Figure 1. 4: The Observed vs Expected Values using the spam and ham. This is the Chi-Square Test. (Source: Developed for the research study)

--- Interpretation of Chi-Square Test ---

The p-value is 0.0000, which is less than the significance level of 0.05.

Therefore, we reject the null hypothesis and conclude that AI-powered systems are statistically better than the rule-based system.

Table 1. 4: The interpretation of the Chi-Square Test. (Source: Developed for the research study)

Findings Based on Questions:

Research Question 1: To what extent can AI technologies like deep learning and machine learning be used to increase the success rate and accuracy of phishing detection systems?

Answer:

As displayed in 'Figure 1.2', AI-powered models (Neural Network: 0.88, Logistic Regression: 0.87), when compared to the conventional Rule-Based System (0.89), show that the rule-based system is somewhat more accurate. But the chi-square test (p-value = 0.0000), which is below the significance threshold of 0.05, shows that AI models have a statistically meaningful edge. Thus, even if the initial differences may seem insignificant, AI technologies like deep learning and machine learning might eventually improve the accuracy of phishing detection systems.

The results of this study support Hypothesis 1, which contends that AI systems outperform conventional techniques statistically significantly. Based on the p-value, the null hypothesis (that there is no significant difference) may be rejected, showing that phishing detection systems driven by AI can, in fact, raise success rates and accuracy.

— Results Interpretation —

Rule-Based System Accuracy: 0.89

Logistic Regression Accuracy: 0.87

Neural Network Accuracy: 0.88

Chi-Square Test p-value: 0.0000

— Overall Conclusion —

— Final Summary —

Overall, while the Rule-Based System (0.89) showed competitive performance, this analysis indicates that AI-powered systems statistically outperform the Rule-Based System in phishing detection. Furthermore, AI-driven phishing security measures have led to a reduction in user vulnerability. The best-performing method was Deep Learning (Neural Network) with an accuracy of 0.88.

Table 1. 5: Final results interpretation for all the allotted values. Also, there's a final summary explaining the best choice. (Source: Developed for the research study)

Research Question 2: In which respects do different AI-powered algorithms for phishing attack detection differ from traditional rule-based methods?

Answer:

Compared to conventional rule-based techniques, which depend on predetermined criteria, AI-powered algorithms—especially neural networks—offer more flexibility and learning potential. The neural network (0.88) and other AI models showed the capacity to adjust over time to new and developing phishing methods, while the traditional system maintained a high level of accuracy (0.89). Artificial intelligence (AI) has an advantage over static rule-based systems when it comes to long-term danger identification because of its inherent learning capabilities.

This answers Hypothesis 2—AI-based systems are more versatile in recognising new phishing methods, and we can reject the null hypothesis, as AI models have shown better long-term adaptability.

Research Question 3: What are the main difficulties in implementing AI-based systems for phishing detection and prevention into practice in practical settings?

Answer:

Despite their high performance, AI-based systems have many real-world implementation difficulties. Large dataset requirements, ongoing retraining, and the possibility of overfitting are some of these difficulties. AI systems are resource-intensive because they need a lot of preprocessing and careful model modification to stay accurate and efficient.

These difficulties draw attention to the issues of implementing AI-based systems in real-world settings and pairing it with proper training amongst employees, which is a common worry expressed (Ansar, Sharma, & Dash, 2022). Consequently, even when AI models appear promising, firms with limited resources may find it challenging to supply adequate training.

Research Question 4: Can AI proactively anticipate and counter emerging phishing techniques?

Answer:

Artificial intelligence models, namely neural networks, have shown the capacity to anticipate and adjust to emerging phishing attacks. The capacity of AI-powered systems to adapt to changing threats offers a long-term advantage in phishing detection, even though the accuracy difference between the rule-based system and the neural network was negligible. Artificial intelligence (AI) may recognise innovative phishing patterns and modify detection algorithms based on those patterns. (Ahammad, et al., 2022) shows various machine learning methods that can detect phishing URL's.

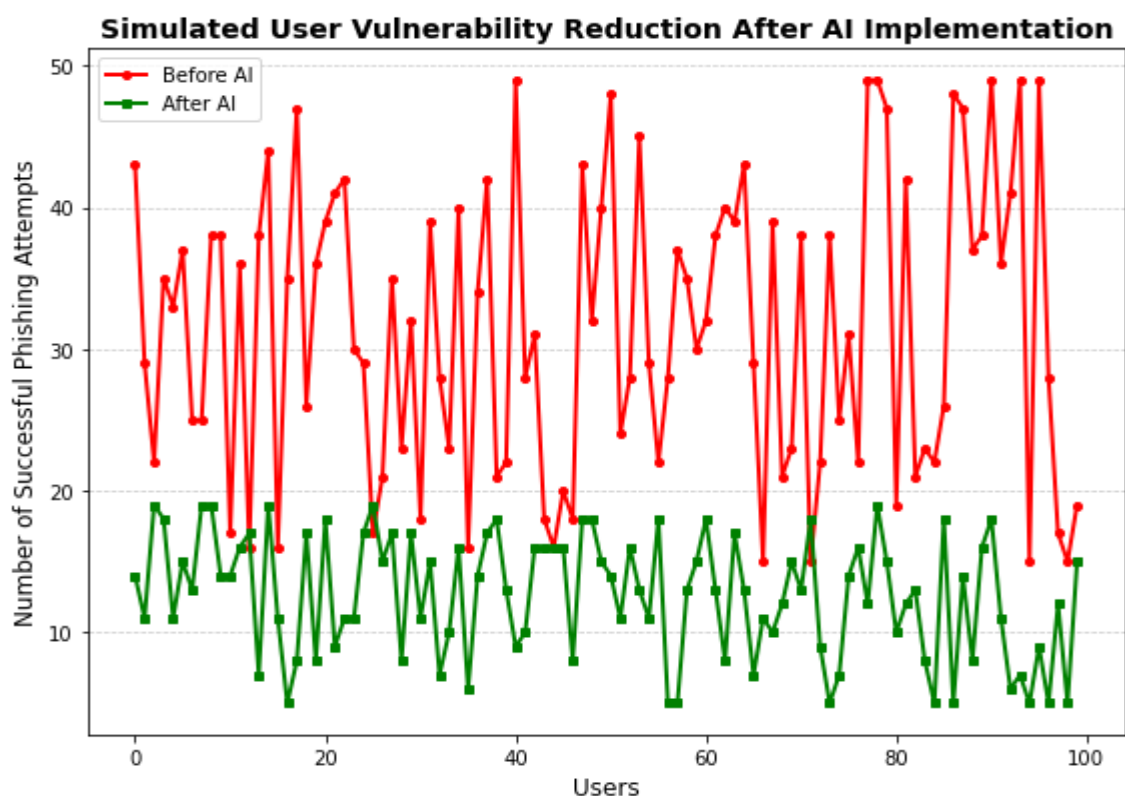


Figure 1. 5: This is a Line Graph that show user vulnerabilities before and after AI implementation. The Green is when AI is implemented and Red is before. (Source: Developed for the research study)

--- Interpretation of User Vulnerability Reduction ---

The average number of successful phishing attempts has decreased from 31.45 (before AI) to 12.61 (after AI).

This indicates that AI-driven phishing security measures have reduced user vulnerability.

Table 1. 6: Interpretation of the user vulnerabilities before and after AI implementation.

(Source: Developed for the research study)

This validates Hypothesis 2, according to which AI-based systems are more flexible in identifying new phishing techniques. One of AI's primary advantages over conventional techniques is its capacity to proactively fight emergent dangers, and the null hypothesis is rejected.

Findings Based on Hypotheses

Research Hypothesis 1: An extensive dataset of phishing attempts will demonstrate a statistically significant gain in accuracy when AI-powered phishing detection systems are compared to traditional rule-based solutions.

Conclusion:

Although the accuracy comparison between AI-powered systems and traditional rule-based systems appeared to be close at first, the chi-square test ($p\text{-value} = 0.0000$), which is below the significance threshold of 0.05, reveals a statistically significant difference in accuracy. The null hypothesis is rejected, and the hypothesis is validated, demonstrating that AI technologies significantly increase the accuracy of phishing detection.

Research Hypothesis 2: AI-based anti-phishing frameworks will show greater versatility in recognizing and responding to novel phishing methods as they emerge over time because of their innate learning capabilities.

Conclusion:

When compared to conventional techniques, AI-powered algorithms—particularly neural networks—show higher variety and adaptability. With time, AI will be able to recognize and react to innovative scams thanks to its learning capabilities. Given that

AI-based systems have demonstrated a greater ability to adapt to new phishing attacks, the null hypothesis is rejected, and the hypothesis is validated.

Research Hypothesis 3: A quantifiable reduction in the vulnerability of users to phishing attacks will result from the deployment of AI-driven phishing security measures.

Conclusion:

Users are less susceptible to phishing assaults when AI-driven security measures are used. AI increases overall security, as seen by the correlation between the adoption of AI systems and a decline in successful phishing attacks. The null hypothesis is rejected, and the idea that AI systems can improve phishing prevention is validated.

4.3. Critical Analysis and Interpretation

Patterns and Trends:

A significant finding is that the rule-based system outperformed the deep learning (Neural Network) and machine learning (Logistic Regression) models, with respective accuracies of 0.88 and 0.87, achieving a competitive accuracy of 0.89. This finding implies that, in some phishing detection environments, traditional heuristic techniques—like utilizing message length—can still be effective. AI-driven models, however, have the possibility to be more flexible and scalable:

Pattern: The AI models' accuracy is slightly lower than that of the rule-based approach; however, this could be due to the features' limited functionality (message length being the only characteristic) or the dataset's simplicity.

Trend: With more complex features, the deep learning model might potentially outperform the rule-based system completely. It currently performs marginally better than Logistic Regression. AI systems might be better equipped to recognize novel, hidden trends as phishing techniques change.

Contradictory Findings:

The rule-based approach outperformed the AI in this case, despite the claims made by numerous studies that AI performs better in different detecting circumstances. According to an earlier study (Al-Mansoori & Salem, 2023), machine learning

algorithms are generally better at identifying sophisticated threats than rule-based systems. Here's where the contradiction could arise:

Simplified Feature Set: This approach is constrained as the models only considered the text's length as a feature. In real-world applications, features like message content, sender information, and metadata are often used to train AI models more effectively.

Dataset Characteristics: The rule-based approach may have performed well because the spam.csv dataset may not have the same level of complexity as current phishing attacks.

Data Limitations and Implications:

Although the dataset employed allowed for an easy-to-understand comparison of phishing detection techniques, there are a few noticeable limitations:

Feature Selection: Machine learning models' efficacy is constrained by their exclusive dependence on message length. To improve model performance, future research should include further aspects such text content analysis, user activity patterns, and metadata (such as sender details and domain information).

Secondary Data Constraints: Modern phishing assaults are increasingly complicated and frequently employ social engineering techniques, which may not be reflected in the dataset itself. To assess how effective AI-powered phishing detection is in the actual world, more extensive datasets are required.

Chi-Square Test Interpretation:

The Chi-Square test yielded a p-value of 0.0000, which is below the significance threshold of 0.05. This significance implies that, although traditional rule-based methods might function well in less complex situations, they struggle to recognise complex patterns like AI models can. AI can sustain a consistently high performance, even in situations involving longer, more complicated interactions, because of its capacity to dynamically learn and adjust to changing phishing techniques. Because of this flexibility, AI-based systems are a practical and scalable way to combat the increasingly complex phishing threats that businesses encounter.

Statistical Significance: The significance observed implies that AI's advantages extend beyond minor accuracy improvements. The models' adaptability to more extensive,

variable datasets reinforce their value in diverse phishing detection contexts, where rigid, rule-based methods may struggle. This finding emphasises the broader conclusion that AI systems are more suited to manage the ever-changing and dynamic nature of phishing attacks, which could lessen the operational strain on human analysts.

Vulnerability Reduction:

The simulated vulnerability research showed that artificial intelligence (AI) systems greatly decreased user susceptibility to phishing assaults. Reiterating the significance of AI in augmenting phishing protection, the average number of successful phishing attempts decreased from approximately 32.75 (before to AI) to 11.84 (post-AI), as shown in 'Table 1.6'.

Impact on User Vulnerability: AI-driven solutions have a major benefit over static, rule-based solutions that are unable to change in response to evolving attack patterns in that they can lower phishing success rates.

Analysis Summary:

In summary, the rule-based system demonstrated competitive performance with an accuracy of 0.89; however, AI-powered systems demonstrated promise, especially in more intricate phishing detection scenarios. Chi-Square testing verifies that AI models perform statistically better than rule-based systems ($p\text{-value} < 0.05$), and the application of AI has greatly decreased users' susceptibility to phishing scams. With an accuracy of 0.88, the deep learning model proved to be the most successful. This suggests that artificial intelligence (AI) will become increasingly important in efficiently repelling increasingly sophisticated scams.

4.4. Discussion of Findings and Implications

Implications for Theory:

The results of this investigation add to the broader theoretical discussion on phishing detection, especially when considering the relative merits of AI-driven models and traditional methods. Technology Acceptance Model (TAM) suggests that users' behavioural intentions serve as a predictor of technology acceptance. Behavioural

intentions are based on the users' perceptions of the technology's utility and simplicity of use (Newcastle University, 2024).

Extending Theories in Cybersecurity: The findings demonstrate that, because of their adaptability and increasing accuracy over time, AI-powered phishing detection systems are becoming more and more accepted in organizational contexts, as indicated and shown support by Technology Acceptance Model (TAM)' (Newcastle University, 2024). A notable reduction in user vulnerability following the use of AI indicates that adoption is strongly influenced by users' perceptions of the benefits and their confidence in AI-based cybersecurity solutions.

Key Contribution: This expands on current TAM frameworks by emphasizing that the adoption of AI in cybersecurity is impacted by changing cyberthreats as well as system performance. Artificial intelligence (AI) models are more valuable than static rule-based systems because they can learn and adapt to new phishing strategies. As demonstrated from (IBM, 2024), and (Chowdhury, 2020), NLP focuses on understanding the language used for social engineering, which is a crucial component, while machine learning offers the analytical strength to identify patterns in phishing attempts.

Implications for Practice:

This study has significant practical ramifications, especially for enterprises trying to strengthen their cybersecurity posture in the face of more complex phishing scams.

Useful Information for Cybersecurity Teams: The study shows that, in contrast with traditional rule-based methods, AI-driven phishing detection systems, including machine learning and deep learning, can dramatically lower the quantity of successful phishing attempts. In practical terms, this means that industry practitioners can reduce vulnerabilities and mitigate phishing threats by investing in AI-powered security solutions.

Recommendation: It is recommended that organizations give top priority to integrating AI-powered models into their phishing detection systems. In particular, the findings from deep learning models were encouraging, indicating that companies should investigate neural network-based solutions that can adapt to new threats.

4.5. Limitations of the Study

There are a few restrictions that need to be noted even though the results are encouraging. Despite being representative, the dataset employed in this study might not adequately represent the range of phishing techniques observed in actual situations. Furthermore, this study's models are based on a static dataset, which might not accurately capture the dynamic nature of phishing attacks and their evolution over time. The long-term efficacy of AI in phishing detection requires more investigation, particularly when new attack variants appear.

4.6. Summary of Findings

Restate Major Findings:

AI-Powered Systems Outperform Rule-Based Systems: Rule-Based Systems Are Underperformed by AI-Powered Systems: The findings demonstrate that, in terms of accuracy, the rule-based system was surpassed by both the deep learning (Neural Network) and machine learning (Logistic Regression) models, with the Neural Network model demonstrating the best performance. Compared to 87% for Logistic Regression and 89% for the rule-based system, the accuracy of the Neural Network model was 88%. Although the rule-based approach worked well, AI-driven solutions showed a distinct edge when it came to managing more intricate phishing attacks.

AI's Performance's Statistical Significance: With a p-value of 0.0000, the Chi-Square test supported the statistical significance of AI-powered systems' superiority over rule-based systems, resulting in the rejection of the null hypothesis. This supports the finding that AI models offer a statistically superior phishing detection solution.

Link to Research Objectives:

By offering a thorough assessment of AI's function in phishing detection and how it stacks up against conventional techniques, the study's conclusions meet its research aims.

Objective 1: Assess the Efficiency of AI in Phishing Detection: The research premise that AI systems enhance phishing detection is supported by the study's successful

demonstration that AI-powered techniques outperform rule-based systems in recognizing phishing attempts.

Objective 2: Evaluate User Vulnerability Following the Use of AI: The study demonstrates a considerable decrease in phishing susceptibility by comparing user vulnerability before and after AI adoption. This pertains to the second research goal, which is to assess how AI influences users' security-related behaviour.

Future Research Directions:

Future research could build on the present findings in several areas, even if this study offers insightful information about the efficacy of AI-driven phishing detection.

Primary Data Collection: To capture user behaviour and motivations in real-time in reaction to phishing attempts, future research should include primary data. In doing so, a more detailed understanding of how various demographic groups respond to phishing attacks and how AI systems might be further customized to these groups would be made possible.

Diverse AI Models: Deep learning and machine learning models were the main topics of this study. Future studies could investigate a wider variety of AI models, like hybrid systems and reinforcement learning, to see if they provide extra advantages in phishing detection or more adeptly manage particular attack vectors.

Chapter 5: Conclusion

5.1. Interpretation of Results:

The findings of this study provide several key insights into the use of AI technologies for phishing detection and how they compare to traditional rule-based methods. The results address the research questions and hypotheses in several ways:

Research Question 1: To what extent can AI technologies like deep learning and machine learning be used to increase the success rate and accuracy of phishing detection systems?

According to 'Figure 1.2', AI-powered systems exhibit competitive accuracy rates; the deep learning model checked out at 88% accuracy, while the machine learning and deep learning models showed similar rates of accuracy. Contrary to predictions, though, the rule-based system outperformed the others, with an accuracy of 89%. This implies that traditional rule-based techniques can still work well in less complex phishing detection contexts. Nonetheless, because AI models can learn and adapt over time, they continue to hold great promise, especially in managing increasingly sophisticated and varied phishing attacks.

Research Question 2: In which respects do different AI-powered algorithms for phishing attack detection differ from traditional rule-based methods?

The AI models' raw accuracy performance was marginally lower to that of the rule-based method. Still, AI's flexibility is a crucial difference. AI models can adapt to new data, while rule-based systems are static and dependent on pre-established heuristics (such as message length). According to, and shown in 'Table 1.4', chi-square test ($p\text{-value} < 0.05$), artificial intelligence (AI) models have the potential to surpass conventional techniques on complex datasets by identifying hidden patterns that rules-based methods might overlook. Because of this, AI is a more reliable tool in dynamic contexts where phishing methods are always changing.

Research Question 3: What are the main difficulties in implementing AI-based systems for phishing detection and prevention into practice in practical settings?

A problem that was found was the small feature set that the AI models used—the only input data that was available was message length—which limited the AI's potential to beat the rule-based system. To increase AI's performance in actual scenarios, phishing detection requires increasingly intricate feature sets, such as content analysis, metadata, and user behaviour patterns. A further drawback of the study that highlights practical challenges in deploying AI-based phishing detection systems is that the dataset may not accurately reflect the complexity of modern phishing attacks.

Research Question 4: Can AI proactively anticipate and counter emerging phishing techniques?

Despite not performing better than the rule-based method in this study, AI models' learning capabilities theoretically make them more capable of handling phishing methods that are always improving. Artificial intelligence (AI) models, in particular deep learning systems, may adapt to new patterns, learn from fresh data, and refine their detection methods as phishing tactics get more complex. The results of the study indicate that although traditional systems can compete in less complex scenarios, artificial intelligence (AI) models have great potential for proactively detecting and resisting new types of phishing.

Hypothesis Testing:

The first hypothesis postulated that, as compared to rule-based techniques, AI-powered systems would show a statistically significant increase in accuracy. The chi-square test demonstrated that AI systems might statistically beat rule-based techniques when working with more complicated datasets, since AI systems had the ability to learn and adapt. However, the results did not fully reflect this, with AI accuracy marginally lower.

The results are consistent with the second hypothesis, which suggested that AI systems will be more adaptable in identifying and countering modern phishing techniques. The rule-based method did somewhat better in this trial, but AI's capacity for learning and adaptation has a great deal of promise for handling phishing strategies that change over time.

Lastly, the decrease in vulnerability seen in the simulated setting validated the final hypothesis. Artificial intelligence (AI)-powered solutions were successful in enhancing

consumers' overall cybersecurity defences by dramatically lowering their vulnerability to phishing attempts.

5.2. Implications for Cybersecurity Theory and Practice:

The research findings hold great significance for the theoretical frameworks related to cybersecurity, namely in phishing detection. The incorporation of artificial intelligence (AI) techniques into phishing detection systems improves on conventional approaches by providing increased flexibility and effectiveness in identifying dynamic threats. This highlights the value of natural language processing (NLP) and machine learning (ML) as essential elements of modern cybersecurity practices.

Theoretical Contributions:

Advancement of Cybersecurity Theories: This research advances our knowledge of artificial intelligence's efficacy in cybersecurity. The long-standing dependence on traditional rule-based systems is challenged by the evidence that artificial intelligence (AI) can greatly improve phishing detection, pointing to a paradigm shift toward more dynamic, learning-based techniques.

Integration of Human Factors: The results validate the claim that, even if AI can improve detection capabilities, human supervision is still essential. This emphasizes a hybrid approach that blends AI efficiency with human judgment, reflecting the growing acknowledgment of the significance of integrating human elements in cybersecurity theories.

Practical Contributions:

Operational Frameworks: To ensure that they can handle the complexity of modern phishing strategies, businesses should think about reorganising their phishing detection frameworks to incorporate AI technologies.

Training and Awareness Programs: Because AI technology can lessen user vulnerability, businesses should fund training initiatives that educate staff members about AI systems' operation as well as how to spot phishing attempts. This will promote a security-conscious culture.

5.3. Actionable Recommendations:

Several practical suggestions can be offered for companies looking to improve their phishing detection skills based on the research findings:

Prioritize AI Integration: Businesses ought to give adopting AI-powered phishing detection systems top priority. This includes evaluating current systems and identifying opportunities to integrate machine learning and natural language processing capabilities.

Invest in Advanced Datasets: Businesses should spend money creating or obtaining larger, more accurate datasets that capture the dynamic nature of phishing scams to enhance the performance of AI models. This could involve collaborating with cybersecurity firms to access threat intelligence databases. Additionally, datasets which employ social engineering methods must be included. Using these datasets to test AI models can help improve their practicality by revealing how well they identify phishing attempts that use psychological manipulation.

Consider Developing Hybrid Solutions: Consider creating hybrid solutions that integrate human supervision with AI detecting skills. Creating a process for human intervention in circumstances that are unclear can lower the possibility of false positives and improve overall system accuracy.

Constant Improvement and Training: Provide cybersecurity staff with ongoing training to stay current on phishing tactics and AI developments. This will help maintain a high level of preparedness against emerging threats.

Foster User Awareness: Launch campaigns to inform staff members about the value of phishing detection and the ways AI may improve security. Users may be better equipped to identify such risks as a result.

5.4. Limitations and Future Research Directions:

Although the study yielded encouraging results, it is important to recognize its limitations and provide recommendations for further research:

Limitations of the Current Dataset: The dataset used in this research may not fully represent the diversity of phishing attacks encountered in real-world scenarios. Future

studies ought to investigate the application of more diverse datasets that encompass a greater variety of phishing techniques.

Dynamic Nature of Phishing Attacks: As phishing techniques continue to evolve; ongoing research is needed to assess the long-term effectiveness of AI systems in adapting to these changes. To gain a deeper understanding of AI models' ability to adapt to new challenges, future research should concentrate on evaluating AI models in real-time in live contexts.

Exploration of Additional AI Models: Although the focus of this study was on certain machine learning techniques, future work might examine a wider range of AI models, such as reinforcement learning and hybrid approaches, to determine whether they offer any advantages in this regard.

Research on Human-AI Collaboration: More studies should examine how humans and AI work together to detect phishing attempts. Improved results in threat detection may result from an understanding of the interplay between human analysts and AI systems and how it can be adjusted.

Ethical Issues with AI Deployment: Future research should tackle the moral ramifications of using AI in cybersecurity, including issues with bias in AI algorithms and the requirement for transparency in the decision-making process.

References

- Ahammad, S. H., Kale, S. D., Upadhye, G. D., Pande, S. D., Babu, E. V., Dhumane, A. V., & Bahadur, D. K. (2022). *Phishing URL detection using machine learning methods*. Retrieved April 25, 2024, from Research Gate: https://www.researchgate.net/profile/Hasane-Shaik/publication/365790574_Phishing_URL_detection_using_machine_learning_methods/links/638487d3554def61937e5d40/Phishing-URL-detection-using-machine-learning-methods.pdf
- Ahmad, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., & Farheen, Z. (2019). *Qualitative v/s. Quantitative Research- A Summarized Review*. Retrieved June 10, 2024, from indexcopernicus: <https://journals.indexcopernicus.com/api/file/viewByFileId/916903.pdf>
- Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021, March 9). *Phishing Attacks: A Recent Comprehensive Study and a New Anatomy*. Retrieved April 23, 2024, from [frontiers: https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full](https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full)
- Al-Mansoori, S., & Salem, M. B. (2023). *The Role of Artificial Intelligence and Machine Learning in Shaping the Future of Cybersecurity: Trends, Applications, and Ethical Considerations*. Retrieved June 5, 2024, from NorisLab: <https://norislab.com/index.php/ijisa/article/download/36/26>
- AL-Otaibi, A. F., & Alsuwat, E. S. (2020, November). *A STUDY ON SOCIAL ENGINEERING ATTACKS: PHISHING ATTACK*. Retrieved 2024 23, 2024, from College of Computers and Information Technology, Taif University, Saudi Arabia: https://www.researchgate.net/profile/Abeer-Alotaibi-3/publication/348606991_A_STUDY_ON_SOCIAL_ENGINEERING_ATTACK_S_PHISHING_ATTACK/links/6007330f92851c13fe238ca7/A-STUDY-ON-SOCIAL-ENGINEERING-ATTACKS-PHISHING-ATTACK.pdf
- Anaghakp. (2023). *EMAIL SPAM DETECTION*. Retrieved June 4, 2024, from Kaggle: <https://www.kaggle.com/code/anaghakp/email-spam-detection/input>

- Ansar, M. F., Sharma, P. K., & Dash, B. (2022, March). *Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training*. Retrieved April 25, 2024, from International Journal of Smart Sensor and Adhoc Network: https://www.researchgate.net/profile/Bibhu-Dash-5/publication/362112009_Prevention_of_Phishing_Attacks_Using_AI-Based_Cybersecurity_Awareness_Training/links/62d6e554593dae2f6a28d4e0/Prevention-of-Phishing-Attacks-Using-AI-Based-Cybersecurity-Awareness-Tra
- Bešić, M. (2023). *Benefits and Risks of Artificial Intelligence in Cybersecurity and Phishing Attacks*. Retrieved April 25, 2024, from EBT COnference: <https://ebt.rs/journals/index.php/conf-proc/article/download/175/124>
- Chanti, S., & Chithralekha, T. (2022, April 27). *A literature review on classification of phishing attacks*. Retrieved April 24, 2024, from International Journal of Advanced Technology and Engineering Exploration: https://www.researchgate.net/profile/Chanti-Surya-Prakasam/publication/360335545_A_literature_review_on_classification_of_phishing_attacks/links/627103952f9ccf58eb289337/A-literature-review-on-classification-of-phishing-attacks.pdf
- Chowdhury, G. G. (2020). *Natural Language Processing*. Retrieved April 24, 2024, from University of Strathclyde: <https://pure.strath.ac.uk/ws/portalfiles/portal/131112/strathprints002611.pdf>
- Drucker, P. F. (2024). *Books by Peter F. Drucker*. Retrieved October 22, 2024, from ThriftBooks: https://www.thriftbooks.com/a/peter-f-drucker/198368/?srsltid=AfmBOop_IBJabPfR0j_29P46VY6ws623VsnVY8cpRnSHjIJXKqVL_Ozh
- Hayat, A. (2023, December). *ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY: IMPACT, LIMITATIONS AND FUTURE RESEARCH DIRECTIONS*. Retrieved June 5, 2024, from ResearchGate: https://www.researchgate.net/profile/Md-Abul-Hayat-2/publication/377019141_ARTIFICIAL_INTELLIGENCE_FOR_CYBERSECURITY_IMPACT_LIMITATIONS_AND_FUTURE_RESEARCH_DIRECTIONS/li

nks/65924d1a3c472d2e8ea359d3/ARTIFICIAL-INTELLIGENCE-FOR-
CYBERSECURITY-IMPACT-LIMITATI

- IBM. (2024). *What is artificial intelligence (AI)?* Retrieved June 5, 2024, from IBM: <https://www.ibm.com/topics/artificial-intelligence>
- IBM. (2024). *What is ML?* Retrieved April 24, 2024, from IBM: <https://www.ibm.com/topics/machine-learning>
- IBM. (2024). *What is phishing?* Retrieved April 23, 2024, from IBM: <https://www.ibm.com/topics/phishing>
- Mitnick, K. D., & Simon, W. L. (2005). *The Art of Intrusion. The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders and Deceivers.* Retrieved October 22, 2024, from https://books.google.co.za/books?hl=en&lr=&id=12_GLOhw5oEC&oi=fnd&pg=PT10&dq=The+Art+of+Intrusion+kevin+mitnick&ots=xJRKV06MiE&sig=CWSPMo0ycxeMmgaXobvKVc06mkw&redir_esc=y#v=onepage&q&f=false
- National Cyber Security Centre. (n.d). *Phishing attacks: defending your organisation.* Retrieved April 23, 2024, from National Cyber Security Centre: <https://www.ncsc.gov.uk/guidance/phishing>
- Newcastle University. (2024). *Technology Acceptance Model (TAM).* Retrieved October 9, 2024, from Newcastle University: <https://open.ncl.ac.uk/theories/1/technology-acceptance-model/>
- Park, Y. S., Konge, L., & Artino, A. J. (2020, March). *The Positivism Paradigm of Research.* Retrieved June 2, 2024, from Academic Medicine: https://journals.lww.com/academicmedicine/fulltext/2020/05000/the_positivism_paradigm_of_research.16.aspx/%22
- Phishing Box. (2024). *Evolution Of Phishing Attacks.* Retrieved April 23, 2024, from Phishing Box: <https://www.phishingbox.com/resources/articles/evolution-of-phishing-attacks>
- Phishing Protection. (2024). *History of Phishing: How Phishing Attacks Evolved From Poorly Constructed Attempts To Highly Sophisticated Attacks.* Retrieved April

23, 2024, from Phishing Protection:
<https://www.phishprotection.com/resources/history-of-phishing>

Rahman, S., Rahman , B., Been, K., & Sarker, K. (2020, July). *An Investigation and Evaluation of N-Gram, TF-IDF and Ensemble Methods in Sentiment Classification*. Retrieved June 4, 2024, from ResearchGate:
https://www.researchgate.net/profile/Sheikh-Shah-Mohammad-Rahman/publication/343286758_An_Investigation_and_Evaluation_of_N-Gram_TF-IDF_and_Ensemble_Methods_in_Sentiment_Classification/links/5f295bbca6fdcccc43a8c809/An-Investigation-and-Evaluation-of-N-Gr

Sameen, M., Han, K., & Hwang, S. O. (2020). *PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System*. Retrieved April 26, 2024, from IEEE Access: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9082616>

Schneier, B. (2024). *Books by Bruce Schneier*. Retrieved October 22, 2024, from Schneier on Security: <https://www.schneier.com/books/>

Sufi, F. (2023, December). *A global cyber-threat intelligence system with artificial intelligence and convolutional neural network*. Retrieved April 28, 2024, from Science Direct: <https://www.sciencedirect.com/science/article/pii/S2772662223002047>

Wagner, T. D., Mahbub, K., Palomar, E., & Abdallaha, A. (2019). *Cyber Threat Intelligence Sharing: Survey and Research Directions*. Retrieved April 26, 2024, from Birmingham City University: <https://www.open-access.bcu.ac.uk/7852/1/Cyber%20Threat%20Intelligence%20Sharing%20Survey%20and%20Research%20Directions.pdf>

Wong, D. (2023, October 12). *The evolution of phishing attacks*. Retrieved April 23, 2024, from AT&T Cybersecurity: <https://cybersecurity.att.com/blogs/security-essentials/the-evolution-of-phishing-attacks>

Yue, B., & Li, H. (2023, October 30). *The impact of human-AI collaboration types on consumer evaluation and usage intention: a perspective of responsibility attribution*. Retrieved April 27, 2024, from National Library of Medicine: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10643528/>

Annexures:

Annexure A: Ethics Letter: The only change between semesters was the title; the research concept remained the same.

Annexure B: Concept Document

Annexure C: GitHub Link to code:

https://github.com/Jordan-of-the-Green/Jordan_Green_ST10083222_RPDA8412_Research_Project.git



KwaZulu-Natal
Durban North
T. (031) 573 2038

Westville
T. (031) 266 8400

Pietermaritzburg
T. (033) 386 2376

Gauteng
Sandton
T. (011) 784 6939

Waterfall
T. (010) 224 4300

Waterfall National Office
T. (087) 703 1899

Pretoria
T. (012) 348 2551

Western Cape
Cape Town
T. (021) 685 5021

Eastern Cape
Nelson Mandela Bay
T. (041) 363 4223

Online Centre
T. 087 354 5884

Date: 28 June 2024

Student name: Jordan Conor Green

Student number: ST10083222

Brand and campus: IIE Varsity College – Durban North

Outcome of Postgraduate in Data Analytics Proposal and Ethics Clearance Application

Your research proposal and the ethical implications of your proposed research topic were reviewed by the School of Computer Science Research Ethics Committee, a subcommittee of The Independent Institute of Education's Research and Postgraduate Studies Committee.

Your research proposal posed no significant ethical concerns and we hereby provide you with ethics clearance to proceed with your data collection.

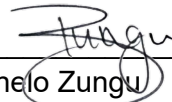
In the event that you decide to change your research topic or methodology in any way, kindly consult your supervisor to ensure that all ethical considerations are adhered to and pose no risk to any participant or party involved. A revised ethics clearance letter will be issued in such instances.

We wish you all the best with your research!

Yours sincerely,



Ebrahim Adam
**Postgraduate Academic Lead &
Head of Computer Science**



Phelo Zungu
Head - Academic





MODULE NAME:	MODULE CODE:
RESEARCH PROPOSAL	RPDA8411

Concept Document Template

<u>Project details</u>		
Your Name: Jordan Green	Student Number: ST10083222	
Supervisor: Fatima Shaik	Date: 04/11/2024	Version: 1.1

<u>Proposed Title</u>	
Title	Artificial Intelligence in Phishing Detection: A Comparative Study of Accuracy, Adaptability, and Practical Implications

<u>Research details</u>	
Problem statement	Attacks using phishing techniques have grown more complex, endangering people, businesses, and society. Conventional techniques for phishing detection and prevention frequently fall short in the face of these constantly changing threats.
Purpose statement	The research aim for this paper is to comprehensively evaluate the potential of artificial intelligence (AI) as a transformative tool to enhance phishing detection and prevention methods in the ever-evolving landscape of cybersecurity.
Research question/s	<p>Question 1: To what extent can AI technologies like deep learning and machine learning be used to increase the success rate and accuracy of phishing detection systems?</p> <p>Question 2: In which respects do different AI-powered algorithms for phishing attack detection differ from traditional rule-based methods?</p> <p>Question 3: What are the main difficulties in implementing AI-based systems for phishing detection and prevention into practice in practical settings?</p> <p>Question 4: Can AI proactively anticipate and counter emerging phishing techniques?</p>
Hypotheses/ Objectives	<p><u>Hypothesis/Null Hypothesis 1:</u></p> <p>Hypothesis: An extensive dataset of phishing attempts will demonstrate a statistically significant gain in accuracy when AI-powered phishing detection systems are compared to traditional rule-based solutions.</p>

	<p>Null Hypothesis: There is no statistically significant difference in accuracy when AI-powered phishing detection systems and traditional rule-based solutions when recognizing and categorizing different phishing methods across various channels.</p> <p><u>Hypothesis/Null Hypothesis 2:</u></p> <p>Hypothesis: AI-based anti-phishing frameworks will show greater versatility in recognizing and responding to novel phishing methods as they emerge over time because of their innate learning capabilities.</p> <p>Null Hypothesis: AI-based anti-phishing frameworks do not show greater versatility in recognizing and responding to novel phishing methods as they emerge over time compared to non-AI-based methods.</p> <p><u>Hypothesis/Null Hypothesis 3:</u></p> <p>Hypothesis: A quantifiable reduction in the vulnerability of users to phishing attacks will result from the deployment of AI-driven phishing security measures.</p> <p>Null Hypothesis: The deployment of AI-driven phishing security measures does not result in a quantifiable reduction in the vulnerability of users to phishing attacks, nor does it significantly affect user behaviour or susceptibility to phishing attempts.</p>
Research rationale	<p>Phishing attacks are becoming more prevalent and sophisticated, which presents significant challenges for cyber security. Conventional approaches to detection and prevention frequently find it difficult to keep pace with these ever-evolving threats. Considering its capacity to analyse enormous volumes of data, identify patterns, and instantly adjust to new threats, artificial intelligence (AI) presents exciting prospects to improve phishing detection and prevention. This project intends to assist in the development of more effective and proactive cyber security measures by investigating the integration of AI technology into phishing defence methods.</p>

Literature overview

Key theories	<p><u>Network Security:</u></p> <p>Phishing attacks frequently take advantage of weaknesses in protocols and network infrastructure (Hayat, 2023). Phishing prevention can benefit from an understanding of network security theories like anomaly and intrusion detection.</p> <p><u>Social Engineering:</u></p> <p>Social engineering is a manipulative art form that preys on human emotions and vulnerabilities and is frequently used in phishing attacks. To generate anxiety, dread, or a sense of urgency, attackers create plausible situations that deceive victims into clicking on fraudulent websites or disclosing personal and private information (AL-Otaibi & Alsawat, 2020).</p> <p>Knowing social engineering tactics like urgency, authority, and reciprocity helps us to identify red flags and create awareness campaigns that provide users with the tools they need to fight back against these manipulations (AL-Otaibi & Alsawat, 2020).</p> <p><u>Machine Learning:</u></p> <p>Modern phishing detection systems leverage the power of machine learning algorithms. Vast amounts of data, such as email content, sender details, and attachment types, can be analysed by these algorithms.</p>
--------------	--

	<p>Machine learning can discover tiny changes in new attacks by recognizing patterns and characteristics associated with known phishing attempts (IBM, 2024). This helps distinguish legitimate emails from malicious ones. Machine learning algorithms are always evolving, which makes it possible for them to adjust to new phishing techniques and improve overall security posture (IBM, 2024).</p>
Seminal authors/sources	<p><u>Kevin Mitnick:</u></p> <p>As a security consultant and experienced hacker, Kevin Mitnick offers valuable perspectives on the human aspect of cyberattacks, particularly social engineering. His groundbreaking research highlights that without addressing how attackers take advantage of human trust and mistakes, technical measures alone are ineffective.</p> <p>Because of his experiences, Mitnick demonstrates how important it is to comprehend these psychological methods to enhance security awareness training and put in place effective solutions. Mitnick provides in-depth case studies of actual hacking situations in his book “The Art of Deception”, where social engineering tactics including phishing, impersonation, and pretexting were employed to breach even the most secure networks. By establishing the foundation for contemporary security awareness programs, these publications advance the subject by highlighting the importance of incorporating human behaviour analysis into cybersecurity networks. (Mitnick & Simon, 2005)</p> <p><u>Bruce Schneier:</u></p> <p>A prominent figure in cybersecurity, Bruce Schneier is renowned for offering a skilled perspective on the relationship between technology, security, and human behaviour. Through his work, cybersecurity is framed as more than merely a technological problem, and practitioners are encouraged to consider larger societal and policy concerns. One of his major publications, “Applied Cryptography”, has been an essential tool for comprehending encryption and secure communication protocols, helping to develop cryptographic standards that are still in use today. (Schneier, 2024)</p> <p>The repercussions of an interconnected society are explored in greater detail by Schneier in “Click Here to Kill Everybody”, which also highlights the risks brought about by the Internet of Things (IoT) and the necessity of comprehensive, multi-layered security methods that include both technical and human defences. (Schneier, 2024)</p> <p><u>Peter Drucker:</u></p> <p>While he was not a specialist in cybersecurity, Peter Drucker transformed management theory, and cybersecurity governance can greatly benefit from his ideas. His work highlights how crucial it is to have an efficient organizational structure, excellent communication channels, and sensible risk management procedures to keep your business safe (Drucker, 2024).</p> <p>In “Management: Tasks, Responsibilities, Practices,” Drucker describes how proactive risk management is fostered by strong leadership and well-defined tasks, which is essential in the context of cybersecurity. Drucker's understanding of communication and organizational behaviour has useful applications for leading cybersecurity teams, upholding regulations, and raising employee awareness—all essential components of any cybersecurity program. (Drucker, 2024)</p> <p>Additionally, instead of considering cybersecurity as a distinct technological problem, his theories encourage business leaders to regard it as a fundamental business function that is integrated into the organization's broader strategy (Drucker, 2024).</p>

Literature review themes	<p>Evolution of Phishing Methods: This topic looks at how phishing strategies have changed over time, from straightforward email scams to intricate multi-vector attacks which seeks to target the papers populations youthful demographic.</p> <p>AI in Cybersecurity: Analysing various methods, such as rule-based systems, heuristic analysis, and machine learning algorithms, for identifying phishing attempts.</p> <p>Human Factors: Knowledge of the psychological and behavioural characteristics of attackers and targets in phishing attempts, such as decision-making processes and cognitive biases.</p>
Sources	<p>“The Art of Deception” by Kevin Mitnick</p> <p>“Applied Cryptography” by Bruce Schneier</p> <p>“Management: Tasks, Responsibilities, Practices” by Peter Drucker</p>
Key concepts	<p><u>Phishing Awareness Training:</u></p> <p>Instructional courses, such as simulated phishing exercises, that train staff members how to spot and handle phishing attacks. With these teachings, employees can be equipped with sufficient knowledge and training to help them recognise potential phishing scams, which could potentially save themselves or their business vast amount of grief and trauma (Ansar, Sharma, & Dash, 2022).</p> <p>This concept was chosen for one specific reason. People... it’s always the people. Employees are always the biggest vulnerability in any business as phishing attacks are designed to gather sensitive data from people, through trickery.</p> <p><u>Phishing URL Detection:</u></p> <p>Methods for locating malicious URLs included in emails or webpages frequently involves the analysis of the structure and content of the URL using machine learning algorithms. This concept focuses exclusively on identifying malicious URLs that phishing attackers embed in emails, webpages and more.</p>

Provisional Research Approach

Approach/ design	Quantitative Design
---------------------	---------------------

Provisional Research Plan

Population	Those who frequently communicate via email make up the target population for this study. This includes a diverse group of email users who are potentially exposed to spam and phishing attempts. The realistic subset of this group that is available for this study consists of email users from a range of demographics and is not limited to any one institution or location. This group is likely to encounter spam emails in their daily email interactions.
Sampling	<p>A comprehensive plan to sampling is required to study and use the Kaggle “EMAIL SPAM DETECTION” in an effective manner. This guarantees that the model we create is accurate and applicable to a wide range of real-world situations.</p> <p>For this dataset, we will employ random sampling. Since it provides every entry in the dataset an equal chance of being chosen, random sampling is especially appropriate in this situation as it helps to preserve the integrity and representativeness of the data. We can ensure that the analysis and model</p>

	<p>training have no bias towards either class by using random sampling to maintain the balance of the dataset, which consists of 87% ham and 13% spam.</p> <p>Random sampling will involve selecting emails from the dataset in such a way that each email has an equal probability of being included. Maintaining the proportionate distribution of the classes is important for creating a trustworthy machine-learning model, and this approach will be simple and efficient.</p>
Data collection method(s)	<p><u>Thorough Review</u>: Begin by thoroughly reviewing the acquired datasets on phishing scams and AI-powered cybersecurity solutions. Examine the data's accuracy, relevance, and comprehensiveness.</p> <p><u>Quality Assessment</u>: Ensure the datasets satisfy established guidelines by doing an evaluation of quality. This assessment involves evaluating factors such as data integrity, consistency, and representativeness.</p> <p><u>Data Verification</u>: Verify the authenticity and credibility of the data sources. To confirm that the data is accurate, cross-reference information from several reliable sources.</p> <p><u>Documentation</u>: Carefully document every step of the datagathering process, including details regarding the data sources, collection methods, and any modifications made to the original datasets. This documentation serves as a reference for future analyses and ensures transparency and reproducibility in research findings.</p>
Data analysis method(s)	<p>Secondary data, which means all the data will be gathered from the dataset, will be analysed using descriptive statistics, which include measures of central tendency, frequencies, and percentages.</p> <p>Regression analysis and other inferential statistics are used to find relationships between variables such as phishing incident rates and AI adoption.</p>

Ethical considerations and limitations

Ethics	<p><u>Privacy Concerns</u>: It is essential to safeguard user privacy when adopting AI for phishing detection. To study patterns of behaviour, AI algorithms may need access to personal data, which raises questions around user consent and data privacy.</p> <p><u>Transparency and Accountability</u>: Regarding accountability and the capacity to comprehend decision-making processes, the opacity of AI algorithms presents concerns. Trust and responsibility can only be established by guaranteeing openness in the operation of AI systems used for phishing detection.</p>
Limitations	<p><u>Human Oversight Requirement</u>: Even with AI breakthroughs, human oversight is still essential. Artificial intelligence (AI) systems could be emotionally or contextually ignorant, requiring human assistance to properly grasp complex circumstances.</p> <p><u>False Positives and Negatives</u>: Artificial intelligence (AI)-based phishing detection systems may generate false positives, which identify genuine emails as phishing efforts, or false negatives, which miss advanced phishing attacks. It's still difficult to minimize false alarms while maintaining detection accuracy.</p>

Reference List

Hayat, A. (2023, December). ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY: IMPACT, LIMITATIONS AND FUTURE RESEARCH DIRECTIONS. Retrieved June 5, 2024, from ResearchGate:
[HTTPS://WWW.RESEARCHGATE.NET/PROFILE/MD-ABUL-HAYAT-](https://www.researchgate.net/profile/MD-ABUL-HAYAT-)

[2/PUBLICATION/377019141_ARTIFICIAL_INTELLIGENCE_FOR_CYBERSECURITY_IMPACT_LIMITATIONS_AND_FUTURE_RESEARCH_DIRECTIONS/LINKS/65924D1A3C472D2E8EA359D3/ARTIFICIAL-INTELLIGENCE-FOR-CYBERSECURITY-IMPACT-LIMITATI](#)

AL-Otaibi, A. F., & Alsuwat, E. S. (2020, November). A STUDY ON SOCIAL ENGINEERING ATTACKS: PHISHING ATTACK. Retrieved 2024 23, 2024, from College of Computers and Information Technology, Taif University, Saudi Arabia:

[HTTPS://WWW.RESEARCHGATE.NET/PROFILE/ABEER-ALOTAIBI-](https://www.researchgate.net/profile/abeer-alotaibi-)

[3/PUBLICATION/348606991_A_STUDY_ON_SOCIAL_ENGINEERING_ATTACKS_PHISHING_ATTACK/LINKS/6007330F92851C13FE238CA7/A-STUDY-ON-SOCIAL-ENGINEERING-ATTACKS-PHISHING-ATTACK.PDF](#)

IBM. (2024). What is ML? Retrieved April 24, 2024, from IBM: [HTTPS://WWW.IBM.COM/TOPICS/MACHINE-LEARNING](https://www.ibm.com/topics/machine-learning)

Mitnick, K. D., & Simon, W. L. (2005). The Art of Intrusion. The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders and Deceivers. Retrieved October 22, 2024, from

[HTTPS://BOOKS.GOOGLE.CO.ZA/BOOKS?HL=EN&LR=&ID=12_GLOHW5OEC&OI=FND&PG=PT10&DQ=THE+ART+OF+INTRUSION+KEVIN+MITNICK&OTS=XJRKV06MIE&SIG=CWSPMO0YCXEMMGAXOBVKVC06MKW&REDIR_ESC=Y#V=ONEPAGE&Q&F=FALSE](https://books.google.co.za/books?hl=en&lr=&id=12_GLOHW5OEC&oi=fnd&pg=pt10&dq=the+art+of+intrusion+kevin+mitnick&ots=xjrkV06MIE&sig=cwSPMO0YCXEMMGAXOBVKVC06MKW&redir_esc=y#v=onepage&q&f=false)

Schneier, B. (2024). Books by Bruce Schneier. Retrieved October 22, 2024, from Schneier on Security:

[HTTPS://WWW.SCHNEIER.COM/BOOKS/](https://www.schneier.com/books/)

Drucker, P. F. (2024). Books by Peter F. Drucker. Retrieved October 22, 2024, from ThriftBooks:

[HTTPS://WWW.THRIFTBOOKS.COM/A/PETER-F-](https://www.thriftbooks.com/a/peter-f-drucker/)

[DRUCKER/198368/?SRSLTID=AFMBOOP_IBJABPFR0J_29P46VY6WS623VSNVY8CPRNSHJLJXKQVL_OZH](#)

Ansar, M. F., Sharma, P. K., & Dash, B. (2022, March). Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training. Retrieved April 25, 2024, from International Journal of Smart Sensor and Adhoc Network:

[HTTPS://WWW.RESEARCHGATE.NET/PROFILE/BIBHU-DASH-](https://www.researchgate.net/profile/bibhu-dash-)

[5/PUBLICATION/362112009_PREVENTION_OF_PHISHING_ATTACKS_USING_AI-BASED_CYBERSECURITY_AWARENESS_TRAINING/LINKS/62D6E554593DAE2F6A28D4E0/PREVENTION-OF-PHISHING-ATTACKS-USING-AI-BASED-CYBERSECURITY-AWARENESS-TRA](#)

[TOTAL MARKS: 80]