

Project Proposal Final

by Cai Perry

Submission date: 19-Oct-2025 08:48PM (UTC+0100)

Submission ID: 265695527

File name: 50391_Cai_Perry_Project_Proposal_Final_695044_738921014.docx (118.38K)

Word count: 1453

Character count: 8196



APPLYING MACHINE LEARNING TO PHISHING DETECTION

Project proposal

Cai Perry (2304939)
University of Wales Trinity Saint David

Contents

1. Introduction 2

2. Background to the Proposed Research..... 2

 2.1 – Phishing as a Threat 2

 2.2 – Traditional Prevention Methods 2

 2.3 – The Rise of AI in Phishing 2

3. Research Problem statement 3

4. Research Aim & Objectives 3

 4.1 Aim 3

 4.2 Objectives..... 3

5. Methodology..... 3

 5.1 Methods..... 3

 5.2 Materials 4

6. Project Plan 4

7. References 4

1. Introduction

Phishing is one of the most common yet damaging cyber scams today. According to a study published by IEEE [1], phishing was said that over 90% of cyber-attacks start with phishing, this could be done through an unverified link, or a replica website mean to trick the victim into entering sensitive data, under the impression that they are posing as a reputable name. Phishing is one of the most successful cyber-attacks, because counter measures such as firewalls, AV and SSL certificates cannot prevent these scams as it plays on the human element in security, often the most exploitable, and if the victim doesn't know exactly what to look for in these scams, the results can often be devastating.

The use of AI in this field of research is currently developing rapidly [2], and opposed to other more traditional methods of ways to detect phishing, such as the use of rule based systems, or blacklist databases, AI can adapt and remove the human error element of detection, which ultimately ends up being the reason most victims end up falling for phishing scams.

This project aims to create an AI model that can detect Phishing URLs and differentiate them from normal links, with the expectation that it will reduce the overall percentage of modern phishing victims.

2. Background to the Proposed Research

2.1 – Phishing as a Threat

Phishing involves cyber attackers using deceptive techniques such as fake email links or URLs, to trick victims into giving up personal or private information [3]. It is the most common online scam, effecting over 21,000 people in the United States in 2024 [4]. Despite extensive methods to combat phishing, such as antivirus, firewalls and SSLs [5] being some of the most common, it still manages to be successful as it targets the human element, which remains the most vulnerable part of cybersecurity. The evolution of phishing makes the scam more effective, where cyber-criminals will use AI to create more convincing messages [6].

2.2 – Traditional Prevention Methods

There are two main types of detection methods in phishing prevention. Historically, phishing prevention had relied on 'rule-based system' [7], this is where specific features of URLs or emails are evaluated by preset rules while blacklists will store a catalogue of known phishing sites. Whilst these methods are effective against threats that are known, they lack the nuance again faster acting and evolving threats, with many modern phishing methods being able to bypass these systems [8].

2.3 – The Rise of AI in Phishing

The application of machine learning in phishing prevention had been explored in more recent research. The addition of both deep learning (DL) and machine learning (ML) have been tested to try and address the limitations of current traditional methods of phishing prevention. ML models such as 'Random forests, SVMs and Logistic regression' [8] can utilise large datasets to learn and adapt to illegitimate links and URLs and differentiate them to real links. This allows them to identify emerging patterns that rule based systems may miss.

Whilst generative AI can be used to aid phishing prevention, the rise of 'Phishing as a service' allows phishing attacks to be more dynamic and they have become harder to detect [9]. Generative AI can generate more convincing messages that can mimic real organisations in real time. Because of this, modern detection systems require the use of adaptive models, which are capable of real time

learning. Recent research also points to a growing emphasis on Explainable AI (XAI) [10], this ensures that the detection decisions that the AI makes, can be understood and trusted by endpoint users.

3. Research Problem statement

This project aims to investigate how machine learning and artificial intelligence can improve the accuracy of detecting phishing attempts. This research focuses on the development of artificial intelligence for a specific task, making sure learning models are able to distinguish between legitimate and false links and URLs. The study aims to assess not only the improvements that machine learning has the potential to bring in this problem area, but how it compares to traditional phishing methods, and if it can be added onto them as an improvement.

4. Research Aim & Objectives

4.1 Aim

The overall aim of this project is to develop an AI that is capable of determining legitimate links and URL'S from fake phishing scams by using training data sourced from email, URLs and other metadata. This project aims to compare how effective the AI is, and demonstrate how it can improve on the accuracy, adaptability and overall efficiency of traditional phishing methods.

4.2 Objectives

Objective 1: Conduct background research on both AI and traditional phishing methods to gather information and identify gaps in existing solutions

Objective 2: Collect and process training and testing data from open sources.

Objective 3: Implement and train AI model using Python and random training URLs

Objective 4: Test trained AI on more random based URLs and record accuracy and precision rates. (if results aren't as expected revise design)

Objective 5: Compare results alongside other phishing prevention methods. Analyse strengths, weaknesses and note future improvements for the project.

5. Methodology

5.1 Methods

This study contains a mix of both qualitative and quantitative research; the main body will consist of quantitative, as it focusses mainly on examining measurable data through previous AI tests, examining numerical performances when testing it against phishing and legitimate URLs. More quantitative data includes analysing the testing and training data for the AI.

This project will also include qualitative data through the primary research methods, such as conducting a local survey to gather people's opinions on the idea of the project, and their stance on the ethics and morality of the usage of AI.

The chosen research strategy for this project is Software development combined with experimentation. This is an appropriate choice for this project as it involves the design, implementation and testing of an AI system. Surveys will also be used to gather context about other thoughts on the project and the topic of AI.

5.2 Materials

This project will contain the use of secondary data from publicly available repositories such as Kaggle in order to source data to train and test the AI system. These repositories contain both legitimate and false phishing links and downloaded through python-based import tools such as pandas.

Ethically, this data is all publicly available and completely anonymous, it will contain no personal or sensitive data of any kind. Python is the programming language that will be used, and it will be supported by libraries within it such as pandas and scikit learn.

6. Project Plan

October 2025

- Finalise and submit project proposal

November – December 2025

- Identify suitable training datasets
- Collect training and testing datasets
- Begin processing the data into workable sets

January 2026

- Draft and design initial AI model
- Conduct preliminary testing
- Prepare for Interim Viva (Thurs 15/ Fri 16 Jan 2026)

February – March 2026

- Revise and refine model design
- Perform in depth evaluation and unit testing
- Record results
- Compare accuracy reports against other phishing detection methods

April 2026

- Continue writing final report, add final project in
- Finalise final report with complete bibliography

May 2026

- Turn in final report (May 1st)
- Prepare for final viva (6th/7th/8th May)
- Complete poster for submission
- Prepare for degree show (End of May, exact date TBA)

7. References

[1] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: a Literature Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013, doi: <https://doi.org/10.1109/surv.2013.032213.00009>. [acc 15]

- [2] B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting against phishing attacks: state of the art and future challenges," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3629–3654, Mar. 2017, doi: <https://doi.org/10.1007/s00521-016-2275-y>.
- [3] M. Kosinski, "What is Phishing?," IBM, May 17, 2024. <https://www.ibm.com/think/topics/phishing>
- [4] "BEC victim numbers U.S. 2023 | Statista," Statista, 2023. <https://www.statista.com/statistics/1389991/bec-victim-number-us/>
- [5] "Phishing | KnowBe4," www.knowbe4.com. <https://www.knowbe4.com/resource-center/phishing>
- [6] A. Krishnan, "Generative AI is making phishing attacks more dangerous | TechTarget," *Security*, Dec. 18, 2023. <https://www.techtarget.com/searchSecurity/tip/Generative-AI-is-making-phishing-attacks-more-dangerous>
- [7] "Rule-based system explained," *Explained.today*, 2018. https://everything.explained.today/Rule-based_system/ (accessed Oct. 19, 2025).
- [8] Alexandre Nesic, "Beyond the Hook: A Technical Deep Dive into Modern Phishing Methodologies - Quarkslab's blog," *Quarkslab.com*, 2025. <https://blog.quarkslab.com/technical-dive-into-modern-phishing.html> (accessed Oct. 19, 2025).
- [9] "What is Phishing-as-a-Service (PhaaS) and How To Protect Against It," *Heimdal Security Blog*, Oct. 19, 2023. <https://heimdalsecurity.com/blog/what-is-phishing-as-a-service-phaas/>
- [10] IBM, "Explainable AI," IBM, Mar. 29, 2023. <https://www.ibm.com/think/topics/explainable-ai>

Project Proposal Final

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

75/100

PROJECT PROPOSAL – ACCA6020 INDEPENDENT PROJECT

Student Cai Perry
Mark

Intro / A&O (400 words)		
Project title; Identification of a specific problem to be investigated; Well-defined aims and SMART objectives; Explaining why this project is worth doing; Discussion of what the project needs to do in order to be a success; Description any deliverables that will be produced at the end of the project	20	16
Great introduction clearly identifying the problem and why the research is needed. With the support of the aim and objectives covering the SMART format. It would have been a good addition to briefly mention what makes this project		

unique compared to others of a similar theme.		
Background to the Project / Literature Review (400 words)		
Description of how the work to be done by in the project relates to the work identified in the literature review; Discussion of any ethical issues raised by the project Identifying what has already been done within the field; Topic familiarity; Brief literature review using appropriate sources – 6-8 journal papers and book chapters rather than just websites; Selection of relevant models & concepts Good coverage of phishing as a threat, traditional methods, and AI-based approaches. Nice selection of references including reputable sources but lacks analysis of existing work and how your research will compare.	20	14
Project Design (500 words)		
Clear statement of which development methods are to be used and why they are appropriate; Clear statement of which tools are to be used and why they are appropriate; Clear description of how data collected will be analysed Defined methodology combining software development and experimentation.	30	22

Tools are clearly stated with data sources where appropriate. Though the design lacks detail on model selection and could include how survey data will be used in your findings.		
Resources / Planning (200 words)		
List of any deliverables; Clear diagram showing both timing and resource use; Completed ethics form; Limited and relevant appendices Timeline is comprehensive and organised for the project proposed. A Gantt chart would have been a nice addition to provide visual representations of the work and stages of progress.	15	11
Document Presentation		
Correct spelling, grammar and punctuation; Flow of narrative/fluency; Coherence; Word processed ; Correct use of referencing with bibliography; Strict word limit Well-structured with clear headings and fluency keep to the word count. Some minor grammar and breaking up of the longer sentences have cost a few points.	15	12
TOTAL	100	75

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6
