**Enter your Thesis Title here**

**By**

**Enter your name here**

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**For the Degree of Bachelor of Science in The**

**Department of Mathematics and Computer Science**

**Claflin University**

**Orangeburg, South Carolina**

**Month Year**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Enter Faculty Name Here** |  | **Enter Faculty Name Here** |
| **Thesis Advisor** |  | **Committee Member** |
|  |  |  |
| **Enter Faculty Name Here** |  | **Dr. Jessica Myles** |
| **Committee Member** |  | **Faculty Liaison** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Received by:** |  | **Date:** |  |
| **Dr. Karina Liles, Chairperson, Department of Mathematics and Computer Science** |
| **Received by:** |  | **Date:** |  |
| **Dr. Verlie Tisdale, Acting Dean, School of Natural Sciences and Mathematics** |
| **Received by:** |  | **Date:** |  |
| **Dr. Verlie Tisdale, Vice Provost for Academic Affairs** |

**TABLE OF CONTENTS**

[**THESIS STATEMENT** ii](#_Toc148626126)

[**ABSTRACT** iii](#_Toc148626127)

[**KEYWORDS AND ABBREVIATIONS** iv](#_Toc148626128)

[**LIST OF TABLES** v](#_Toc148626129)

[**LIST OF FIGURES** vi](#_Toc148626130)

[**TABLE OF CONTENTS** vii](#_Toc148626131)

[**INTRODUCTION** 1](#_Toc148626132)

[Background of the Study 1](#_Toc148626133)

[Purpose of the Study 1](#_Toc148626134)

[Research Question(s) 2](#_Toc148626135)

[**LITERATURE REVIEW** 3](#_Toc148626136)

[Summary and Conclusions 3](#_Toc148626137)

[**METHODOLOGY** 4](#_Toc148626138)

[Population and Sampling Procedure 4](#_Toc148626139)

[Research Design and Development Procedures 4](#_Toc148626140)

[Instrumentation (if applicable) 5](#_Toc148626141)

[Data Analysis Plan (if applicable) 6](#_Toc148626142)

[Limitations 6](#_Toc148626143)

[Ethical Procedures & Considerations 7](#_Toc148626144)

[Summary 7](#_Toc148626145)

[**RESULTS** 8](#_Toc148626146)

[Study Results 8](#_Toc148626147)

[**DISCUSSION AND CONCLUSIONS** 9](#_Toc148626148)

[Summary of the Findings 9](#_Toc148626149)

[Global Impact of Computing Solution on Individuals, Organizations & Society 9](#_Toc148626150)

[Recommendations 10](#_Toc148626151)

[Conclusions 10](#_Toc148626152)

[**REFERENCES** 11](#_Toc148626153)

[**Appendix A: Title of Appendix** 12](#_Toc148626154)

**THESIS STATEMENT**

The problem statement is not being able to detect malware on a website posing significant risks and challenges for website owners, users, and cybersecurity professionals to know if a website is safe to browse. Malware on a website can cause the security and integrity of the website to lead to various cybersecurity threats such as data breaches, identity theft, financial fraud, and unauthorized access to sensitive information. A website infected with malware can erode user trust and confidence in the website's security. Visitors may perceive the website as untrustworthy, leading to a decline in user engagement, traffic, and conversions. Detecting malware on a website can be challenging due to the evolving nature of cyber threats and the sophistication of malware attacks. Traditional security measures such as antivirus software, firewalls, and intrusion detection systems may not effectively identify and mitigate all types of malwares. This problem requires a comprehensive understanding to website security, figuring out the

vulnerability assessments, security monitoring, and safeguarding the website to protect

against cyber threats and untrustworthiness.

 **Figure 1.** Graph malware detected

**ABSTRACT**

The text of the Abstract is double-spaced with no paragraph breaks or citations. Briefly describe the overall research study problem you are addressing in the first couple of sentences followed by the purpose of the study. Do include a general introduction of the need for the study (identify the problem) in the first sentence, then move to a clear statement of the research problem (your specific research problem, central research question) being addressed. Identify the purpose and significance of the study. When applicable, summarize the key research question(s), and briefly describe the overall research design, data collection and data analysis procedures. [This part of the abstract should be completed at the proposal stage—before you conduct your research]

Once data collection is complete, add the following:

Identify the key results, one or two conclusions, and recommendations that capture the heart of the research. Conclude with a statement on the implications.

Should be in the neighborhood of 250 words in length

**KEYWORDS AND ABBREVIATIONS**

Malware

Artificial Intelligence

Data

Web browser detection

Malware Detached

**LIST OF TABLES**

[Table 1. Estimated Malware Detected 5](#_Toc239582816)

 Follow this format listing the caption and the page number for which the table can be found in your document. This page is optional and dependent upon if you have tables or not.

**LIST OF FIGURES**

**Figure 1.** Graph malware detected iv

Follow this format listing the caption and the page number for which the figure/image can be found in your document. This page is optional and dependent upon if you have figures/images or not.

**INTRODUCTION**

Malware is a wide ranged issue online users experience that can cause a software system to have problems and do unwanted tasks without a user’s permission. The study of detecting malware on website browsers is a critical area of cybersecurity that needs a lot of attention. Malware can manifest in various forms including viruses, worms, trojans, spyware, and ransomware. Each kind of malware can cause challenges and require special detection skills to effectively identify the threat. At its core, malware is a tool used by cybercriminals, hackers, and malicious actors to achieve various illegal objectives, ranging from stealing sensitive information and financial data to sabotaging critical foundations. Criminals may use malware to steal your login information take your online accounts to hijack your computer and use it send spam emails out. The evolution of malware continues to pose significant challenges to cybersecurity as malicious hackers constantly innovate and adapt their tactics to evade detection and exploit vulnerabilities in software and systems. Understanding the threats and issues malware causes is crucial for people to know in case something happens with there computer, and they think someone is trying to hack their software. The best and most important way to prevent malware is making sure your computer software is staying up to date. Machine learning and AI-based tools are a big topic and strong reliable resources that I believe can mitigate the issues detecting malware. The power of Artificial Intelligence (AI) on detecting malware is a great tool to identify and analyze cyber critical assets and be able defend against malicious software.

**Figure 2. Forms Of Malware Viruses**



Background of the Study

 I found my research by looking at peer-reviewed articles I looked at on malware and artificial intelligence. The authors of these articles helped allow me to gain an understanding of malware viruses that happen with users not understanding what the problem is with their software being attacked. Through this research I found the recourses that’s already available to prevent malware on software, as well as the advantages of being able to use artificial intelligence to help enhance detecting this problem. The articles gave me an understanding of the challenges that are faced with detection of malware and the strength artificial intelligence can have with cyber-attacks. The research in the articles tells me that malware plays a big part with criminals being able to commit illegal acts once hacking certain information in the software. The background of the study for malware detection highlights the nature of cyber threats, the limitations of different detection methods, and the innovative approach on malware infections. Looking at the research artless I have looked at they give a clear understanding on what malware is and that you can’t see it. AI can help identify subtle and deviations of malware activity, being able to detect it early and alleviate cyber threats. Machine learning excels at recognizing complex patterns and correlations with large datasets, enhancing detection accuracy and efficiency. The future of malware detection lies in the continued advancement of AI techniques, the development of collaborative defense mechanisms. Research efforts focus on enhancing the efficiency of AI-driven detection systems to mitigate emerging threats in real-time.

Purpose of the Study

I gathered my research for this topic through studying cyber security and taking different security courses which pursued me to research computer system issues that can be resolved. The purpose of studying malware detection on website browsers is to mitigate cyber threats and protect user’s online assets. Helping enhance trust in digital technologies for user’s that may not be confident if a website is trustworthy or not, was an interesting topic that I wanted to learn more about and do more research on to help prevent malicious software attacks. By understanding the importance of malware detection, cybersecurity professionals can develop effective strategies that can help defend against malicious activities and safeguard security of network environment.

Research Question(s)

State the research question(s) and/or objectives. Say:

The research question(s) guiding this study is/are:

1. What is AI contribution to help malware detection?
2. How do you validate a website using Artificial Intelligence?
3. How does a website validate safe or not safe?
4. How does a malware transfer back to a user’s computer?

**LITERATURE REVIEW**

Conducting a comprehensive literature review, helped me gain valuable insights into the current state of malware detection on websites using AI and identify areas for further investigation and innovation within cybersecurity. Being able to understand what the issue is with users having problems with detecting malware helps me with figuring out what research I need to develop to be able to come up with a solution such as me using Artificial Intelligence. The literature I have researched gives an inside on what malware is defined as, how it can spread throughout user’s software, and detecting malware downloads in network traffic.

In the article, “Malware Contaminated Website Detection by Scanning Page Links.” This article discusses the definition and understanding of what Malware is and detecting malware-contaminated websites by scanning the links within web pages. The authors propose a methodology that involves scanning the links within web pages for indicators of potential malware detection. They analyze various attributes of the links, including URL patterns, domain reputation, presence of known malware signatures, and behavior analysis. The methodology aims to identify suspicious links that may lead to malware-infected websites or malicious content. The article presents a link scanning algorithm that systematically analyzes the links embedded within web pages. The data evaluates the characteristics of each link such as the URL structure, obtained from threat intelligence sources, and historical data on malware infections associated with specific domains. The authors suggest further research to refine the link scanning algorithm and enhance its accuracy and effectiveness in identifying malware threats on the web.

The second article I did looked at is “A Study on Malware and Malware Detection Techniques.” This article explores the landscape of malware, its types, and various techniques used for malware detection. This research gives you a great understanding of malware defining it as malicious software designed to disrupt, damage, or gain unauthorized access to computer systems data. The article highlights the challenges and limitations associated with malware detection, such as the evolution of malware variants, techniques used by malware researchers, and the emergence of having an attack. The author talks about the factors such as detection rate, false positive rate, scalability, and resources that are being consumed with each detection method. Research discusses the emerging trends and future directions in malware detection, such as the use of machine learning, big data analytics, and threat intelligence integration to enhance detection capabilities.

The third article I gathered research is “Measuring and Detecting Malware Downloads in Live Network Traffic.” This article introduces the concept of malware downloads, which involve the transfer of malicious software from external sources to user devices or network infrastructure. Highlights the generality of malware downloads as common attacks used by hackers/hijackers to compromise systems and steal sensitive information. The primary objective of the study is to develop methods for measuring and detecting malware downloads in real-time network traffic. aims to enhance network security by identifying and mitigating malware threats before they can infiltrate organizational networks and compromise data integrity. The article outlines the methodology used to analyze live network traffic and identify patterns associated with malware. Research provides recommendations for enhancing network defense strategies, improving malware detection capabilities, and mitigating the risks associated with malware downloads.

The fourth article I explored information from was “Design a Mobile Game for Home Computer Users to Prevent Phishing Attacks." The literature review I gathered from this article begins with an introduction learning about phishing attacks, highlighting their impact and common techniques used by cybercriminals to deceive users and abduct their information. The challenges with the evolving nature of phishing tactics and the difficulty of educating users about potential risks. The Author explains the limitations of traditional awareness campaigns, security training programs, and static educational materials in effectively engaging and informing users about phishing threats. The role of the existing mobile games interactive applications is designed to educate users about phishing attacks and cybersecurity best practices. Research discusses key design considerations and principles for developing a mobile game specifically tailored to prevent phishing attacks among home computer users. The design used a model from Technology Threat Avoidance Theory to figure out how the game was going to be designed. This proposal fosters a great understanding of Malware training for home computer users that may not understand what phishing attacks can do to keep a safe secure environment on your software. The research identifies the existing areas for future development in the field of cybersecurity education ad gaming. The intent for this research is very efficient for home computer users dealing with phishes threat and hacks emails that may experience malicious software.

Summary and Conclusions

The possibility of using AI-driven techniques to improve malware detection on online browsers while considering the difficulties and complications that may arise during implementation. These issues and trends can be used to improve cybersecurity defenses and protect against ever-evolving dangers in digital software by utilizing AI technology. To create malware detection systems that are effective, researchers stress the significance of feature extraction and selection. Inputs for AI models that detect malicious activity and identify potential risks come from features that are derived from the content, structure, behavior, and traffic patterns of websites. Several machine learning algorithms are investigated for the purpose of detecting malware on websites. These algorithms include supervised, and unsupervised learning techniques. To increase detection performance and accuracy, researchers test out algorithms such as support vector machines (SVM), decision trees, and neural networks. Training and assessing malware detection models heavily depends on the creation and labeling of datasets. Researchers stress that to properly train AI models and guarantee strong detection capabilities across various malware there must be diversified datasets containing both benign and harmful samples are essential.

**METHODOLOGY**

Detecting malware on websites using Artificial Intelligence (AI) integrates various techniques and algorithms to detect any suspicious web content or user behavior. The purpose of studying malware detection on website browsers is to mitigate cyber threats and protect user’s online assets. Helping enhance trust in digital technologies for user’s that may not be confident if a website is trustworthy or not, was an interesting topic that I wanted to learn more about and do more research on to help prevent malicious software attacks. The questions I am attempting to answer are the following: What is AI contribution to help malware detection? Secondly, how do you validate a website using Artificial Intelligence? Thirdly, how does a website validate safe or not safe? Fourthly, how does a malware transfer back to a user’s computer?

**SOME OF THESE SECTIONS MAY NOT APPLY TO YOUR RESEARCH PROJECT. IF THEY DO NOT, THEN THEY CAN BE DELETED.**

Population and Sampling Procedure

Define the target population. State target population size (if known) or approximate/estimated size. Identify and justify the type of sampling strategy. Explain specific procedures for how the sample will be drawn. Describe the sampling frame (inclusion and exclusion criteria). *Note*: This section may be about less than 1 page in length.

Research Design and Development Procedures

The solution I am proposing is a database software that allows user to get an AI-generated response tool that tests if a specific website detects malware. Until figuration the user will want to know if the website can trance malware and AI comes in to play being able to scan if any viruses can be traced back onto suspicious website. The AI’s computing foundation would be constructed using Python as the programming language to calculate the requested results. The program would have the ability to scan the user’s website and generate if the website is trustworthy or not upon request. I chose to do Python because of python’s reputation as an easy-to-learn foundational programming language. Different developers using the application would be able to learn the language quickly if they have knowledge of other languages, and if they have no programming experience, python is a language that won’t require extensive training in learning. This program will be built with the understanding of the collected data from a website. Start by inputting the URL of the website to be tested for malware detection. Then extract features from the website data such as Frequency of certain keywords associated with malware, noticing suspicious scripts or external links, and website's hosting server and domain. I will then conduct an analysis based on the initial data picked up by checking for known malware signatures in the website’s files. After that I will apply machine learning algorithms to identify datasets of known malware and benign websites. If the data predicts a high probability of malware presence, then it will tell you flag the website as potentially malicious software. If the data predicts a low probability of malware presence, consider the website safe, but continue monitoring for changes.

Thoroughly describe your research plan. It should be laid out like a recipe that anyone can pick up, follow and recreate. Things to think about: programming platform used for software development. Decisions made in software development (and rationale). Describe any recruiting procedures and particular demographic information that will be collected. Describe how participants will be provided informed consent. Describe how data will be collected. Explain how participants exit the study (for example, debriefing procedures, etc.). Describe any follow-up procedures (such as requirements to return for follow-up interviews, treatments, etc.).

Describe the procedure for gaining access to the data set. Describe necessary permissions to gain access to the data (with permission letters located in an appendix.) If historical or legal documents are used as sources of data, demonstrate the reputability of the sources and justify why they represent the best sources of data. *Note*: This section may be 1-2 pages in length.

Data Analysis Plan (if applicable)

How will you collect the information gathered from your project and assess it regarding your questions. Identify software used for analyses. Provide an explanation of data cleaning and screening procedures as appropriate to the study. Restate the research questions and hypotheses.

Describe in detail the analysis plan including the elements below.

* Statistical tests that will be used to test the research question(s)/hypothesis (es)
* Procedures used to account for multiple statistical tests, as appropriate
* How results will be interpreted (e.g., key parameter estimates, confidence intervals and / or probability values, odds ratios).
* Assumptions related to the selected hypotheses

*Note*: This section may be about 1 page in length.

Limitations

The significant challenges in malware detection using AI are dealing with imbalanced datasets. Integrating AI-based malware detection systems into existing security browsers can be complex and challenging. It may require compatibility with other security tools such as logging systems and updates to keep the system effective against evolving threats. Malicious software can be detected but AI can sometimes give out mistakes without understanding some information. Imbalanced data can mess up the overall dataset making it difficult for AI models to learn effectively and generalize to new malware variants. AI models may suffer from overfitting, where they memorize patterns in the training data rather than learning to generalize to new examples. This can lead to poor performance on unseen data and increased false positives or false negatives in malware detection. Some AI techniques used for malware detection, such as deep learning models, can be intense and require significant resources for training and inference. This may limit their scalability and real-time detection on large data websites or networks.

Ethical Procedures & Considerations

To establish good ethical procedures for the security of Websites there should be some kind of transparency about the use of AI for malware detection. Users should be informed about the technology being used and how it may affect their browsing experience or data privacy. obtaining website users' informed consent before analyzing their data to look for malware can help provide a clear explanation of the data being gathered, its intended use, and any associated threats. By minimizing the collection and storage of personal data it can protect users’ privacy. Implementing security measures to protect the data collected for malware detection from breaches is important in case the data is at rest and needs to regularly be audit and updated for security purposes. These ethical concerns require clear data collection, robust security measures, unbiased AI algorithms, meaningful user consent mechanisms, user's oversight, and compliance with relevant laws and regulations. The security issues that came up were storing and processing large amounts of data for AI detecting risks of data breaches and unauthorized access. Navigating the legal issues effectively, for deploying AI with malware detection must conduct thorough assessments such as, ensure compliance with specific regulations and addressing any illegal concerns that deal with illegal use from cybercriminals. The third part software’s I will be using is database systems to store and manage data related to malware detection, Python libraries for the data processing AI algorithms, and web servers host for interacting with malware detection systems. In conclusion, detecting malware on websites using AI presents a range of ethical concerns that must be addressed to ensure responsible and transparent deployment of these technologies.

**RESULTS**

Review briefly the purpose, research questions, and/or objectives. *Note*: This introduction to the chapter should be less than 1 page in length.

Study Results

Report findings, organized by research questions and/or objectives. Include tables and figures to illustrate results, as appropriate, and per the current edition of the *Publication Manual of the American Psychological Association*.

For statistical analyses include:

* Exact statistics and associated probability values.
* Confidence intervals around the statistics, as appropriate
* Effect sizes, as appropriate.

Report results of post-hoc analyses of statistical tests, if applicable. Report any additional statistical tests of hypotheses that emerged from the analysis of main hypotheses, as appropriate for the study.

*Note*: This section should be about 5-7 pages in length, or as many needed to present the results.

**DISCUSSION AND CONCLUSIONS**

To identify malware on a website presents serious hazards and difficulties for website owners, visitors, and cybersecurity experts in determining if a website is safe to visit. A website's security and integrity may be compromised by malware, opening the door to several cybersecurity risks including financial fraud, identity theft, data breaches, and unauthorized access to private information. My research for this area of study came from researching cyber security and attending various security classes, which pushed me to investigate problems with computer systems that could be fixed. The goal of researching malware detection on web browsers is to reduce online dangers and protecting users' personal information. To prevent harmful software attacks, I thought it would be interesting to help users who are unsure about the trustworthiness of websites by enhancing their trust in digital technologies. The challenge remains between cybersecurity professionals and cybercriminals dealing with detection methods becoming more advanced. As the digital world evolves so does the culture of cyber threats, making malware detection on websites a challenge needed to be solved. This requires constant innovation from cybersecurity professionals. One emerging approach is the behavioral analysis examining the patterns of behavior on websites to predict and identify potential malware.

Summary of the Findings

This is where you will summarize the findings in your own words. Organize it by the research questions and/or objectives. This is where you are discussing how your results either answer or do not answer your research question (or support your research objective). Think of how you would describe your study findings to an interested colleague or family member. Keep the writing scholarly, but not overly technical or statistically heavy. Describe in what ways findings confirm, refute, or extend knowledge in the discipline by comparing them to the literature featured in your literature review. *Note:* This section can be at minimum 1-2 pages in length, but he length of this section varies depending on the study and design.

Global Impact of Computing Solution on Individuals, Organizations & Society

 Discuss the local and global impact of computing solutions on individuals, organizations or society regarding your problem and proposed solution. Remember, computing solutions are software programs, applications, and systems that leverage technology to solve problems and streamline processes. Choose at least two of the five impacts to discuss their implications (local, global, individuals, organizations, society). Be sure to discuss potential benefits and challenges posed by these impacts.   *Note:* This section can be at about 1 pages in length.

Recommendations

Describe recommendations for further research that are grounded in the strengths and limitations of the current study as well as the literature reviewed in earlier sections. This is where you can bring a bit of yourself into the research. Think of ways that, in hindsight, you may have approached the study to elicit different results.

Ensure recommendations do not exceed study or ethical boundaries. *Note*: This section should be less than 1 pages in length.

Conclusions

Close with a strong “take away” message where you capture the key essence of the study for your reader. Be sure to mention the Significance of the Study as described in terms of (a) advancing theory, (b) advances in practice, and/or (c) filling a gap in the literature. **Why did this study matter.** *Note*: This section should be less than 1 page in length.

**REFERENCES**

Dadkhah, Amin., Dadkhah Mehdi. & Deng, Jie. Malware Contaminated Website Detection

 <http://www.iteejournal.org/archive/vol3no6/v3n6_1.pdf>

Vadrevu, P., Rahbarinia, B., Perdisci, R., Li, K., Antonakakis, M. (2013). Measuring and Detecting Malware Downloads in Live Network Traffic. In: Crampton, J., Jajodia, S., Mayes, K. (eds) Computer Security – ESORICS 2013. ESORICS 2013. Lecture Notes in Computer Science, vol 8134. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-40203-6_31>

N. A. G. Arachchilage and M. Cole, "Design a mobile game for home computer users to prevent from “phishing attacks”," International Conference on Information Society (i-Society 2011), London, UK, 2011, pp. 485-489, doi: 10.1109/i-Society18435.2011.5978543.

Rabia Tahir, “A Studyon Malware Detection Techniques”, International Journal of Education and Management Engineering(IJEME), Vol8, No.2, pp.20-30, 2018.DOI: 10.5815/ijeme.2018.02.03

Djenna A, Bouridane A, Rubab S, Marou IM. Artificial Intelligence-Based Malware Detection, Analysis, and Mitigation. Symmetry. 2023; 15(3):677. https://doi.org/10.3390/sym1503067

**Appendix A: Title of Appendix**

If you are using a survey or particular documents, you can list them as an appendix.

YOUR CODE SHOULD BE INCLUDED AS AN APPENDIX