Neutralizing of Malware Sustainably using the evolution of Python's Artificial Intelligence Functionality

MSc Research Project
PGDCYB_Jan-Sept 2024

Rory Mc Crystal
Student ID: x20163371
x20163371@student.ncirl.ie

School of Computing
National College of Ireland

Supervisor: Raza UI Mustafa
raza.ulmustafa@ncirl.ie

National College of Ireland

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Rory Mc Crystal

Student ID: x20763377

Programme: MSc Cyber Security **Year:** Aug 2024

Module: Research Project + Academic Internship

Supervisor: Raza Ul Mustafa

Submission Due Date: 12/8/24

Project Title: Neutralizing of malware Sustainably using the evolution of python's artificial Intelligence functionality

Word Count: 8173 **Page Count** 33

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Rory Mc Crystal

Date: 11/8/24

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☑
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☑
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☑

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Neutralizing of Malware Sustainably using the evolution of Python's Artificial Intelligence Functionality

Rory Mc Crystal
Student ID x20163371

Abstract

The recent uptake of Artificial Intelligence (AI) systems, Generative AI and Large Language Models, as showcased in ChatGPT, gives an indication that AI is, among other things, shaping business processes and enabling bad actors from a Cyber Security point of view. Cyber Security leaders along with Chief Information Officers need to counteract AI driven Cyber Security threats (Malware delivery, phishing etc.) with their own AI driven solutions. This paper seeks to address barriers to AI Cyber Security entry while examining key business considerations which would lead to the successful implementation of an AI driven Cyber Security solution. To that end Tensor Flow and Pytorch are assessed along with fundamental infrastructure decisions that aide in prospective utilisation. Both Tensor Flow and PyTorch are examined. The fundamental educational and experience requirements that prospective staff should possess in both the AI and Cyber Security industry is addressed. This paper asks if the evolution of Pythons AI Functionality lends itself to a long term sustainable development. This with a view to realising an AI driven Cyber Security Anti Malware projects success over time.

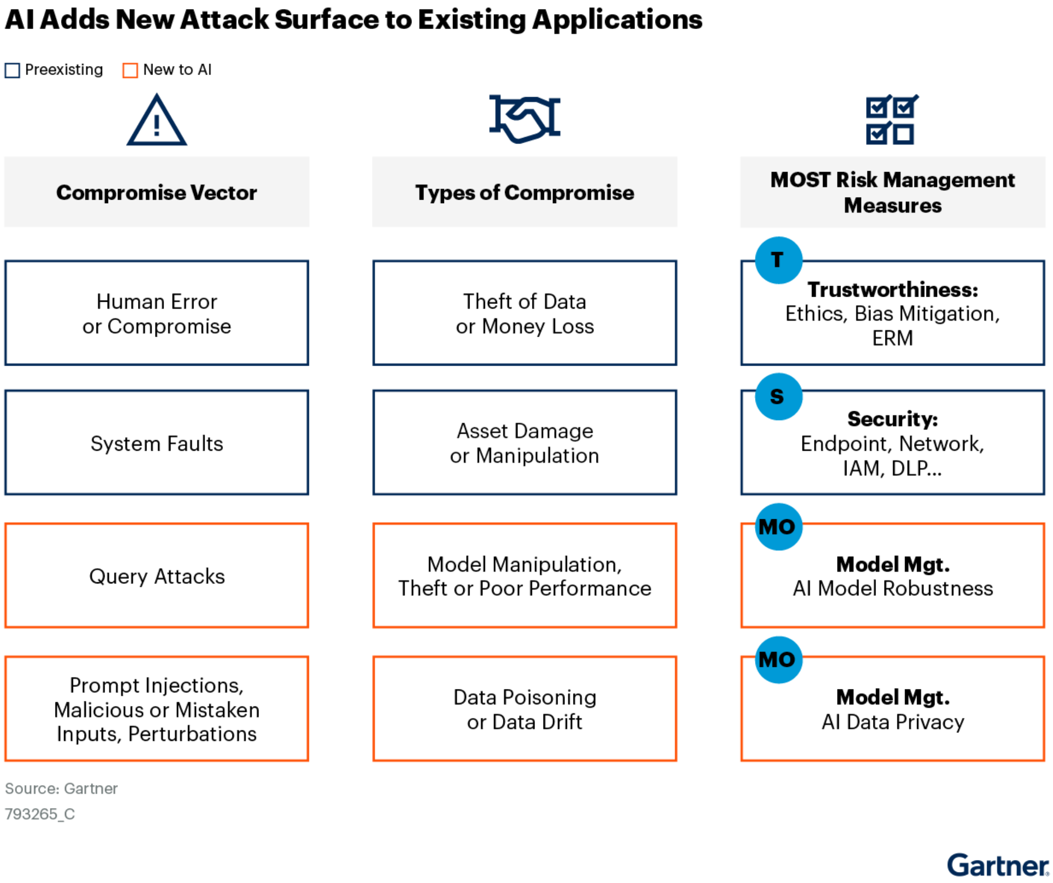
YouTube link for the video presentation:
https://youtu.be/Fkv0bfeUjYE

1 Introduction

In recent years AI has seen enormous adoption due to the immense capabilities that it offers to businesses. Along with the benefits businesses can reap, it also brings a new level of risk to Cyber Security posture. For example, D'Hoinne, Litan, Firstbrook identified the following;

"Through 2025, attacks leveraging generative AI will force security-conscious organizations to lower thresholds for detecting suspicious activity, generating more false alerts, and thus requiring more — not less — human response" (D'Hoinne, Litan, Firstbrook, 2023).

In addition, research by Gartner has identified further AI driven threat areas as shown in Figure 1 below.



AI Adds New Attack Surface to Existing Applications

□ Preexisting □ New to AI

Compromise Vector	Types of Compromise	MOST Risk Management Measures
Human Error or Compromise	Theft of Data or Money Loss	T **Trustworthiness:** Ethics, Bias Mitigation, ERM
System Faults	Asset Damage or Manipulation	S **Security:** Endpoint, Network, IAM, DLP...
Query Attacks	Model Manipulation, Theft or Poor Performance	MO **Model Mgt.** AI Model Robustness
Prompt Injections, Malicious or Mistaken Inputs, Perturbations	Data Poisoning or Data Drift	MO **Model Mgt.** AI Data Privacy

Source: Gartner
793265_C

Figure 1. : Gartner: AI Adds New Attack Surfaces to Existing Applications.

The renowned National Institute of Standards and Technology has addressed the seriousness of the risk that AI presents and as a result "NIST has published a framework called AI Risk Management Framework (AI RMF) to help organizations and individuals manage the risks associated with artificial intelligence (AI) efficiently and productively" (Jawhar, Miller, Bitar 2024).

In order to address AI’s emerging threats we need to look to AI’s own capabilities to counteract them. Cyber Security leaders and Chief Information Security Officers (CISOs) need to consider how emerging AI threats impact their current Cyber Security infrastructure. They will increasingly need to consider AI driven Cyber Security methods to enhance resilience and ensure business immunity, operational integrity and avoid large fines due to data breaches.

“It is envisaged that by 2027, Generative AI will contribute to a 30% reduction in false positive rates for application security testing and threat detection by refining results from other techniques to categorize benign from malicious events” (D’Hoinne, Litan, Firstbrook, 2023).

In order for any business to have a successful AI driven Cyber Security infrastructure a key component must be its sustainability. For a business, ensuring their security posture is up to date and capable of intercepting emerging threats is essential. In addition to this, minimizing the need for ongoing maintenance along with its associated costs is essential. To this end we look to Python, one of the most popular programing languages, and its AI functionality to assess how it can be used to create sustainable and intelligent malware neutralization.

In order to realise the given research topic, the “Neutralizing of Malware Sustainably using the evolution of Python’s Artificial Intelligence Functionality” this paper will investigate the foundations that lead to successful AI and Cyber Security projects. Fundamental business decisions relating to staff education, business management, hardware and programming languages will be investigated.

It is expected that the research will show that with the correct thought, forward planning and processes in place that a Python driven AI Cyber Security Anti Malware solution is possible and sustainable.

This report will provide an insight on AI driven Cyber Security management for companies operating in the fast paced and ever changing AI environment.

The Structure of the report is below:

2 Related Work

Scientific papers have been sourced from amongst others the NCIRL online Library and the IEEE library. Where applicable, certain notable research has been obtained from other sources including the following:

- Herbert Simon's paper on "Bounded Rationality".
- Drucker's "The Effective Executive".
- Mintzberg's "The Structuring of Organizations".
- Henderson and Venkatraman's "Strategic Alignment: A model for organizational transformation via information technology", as published in November 1990 (see Figure 2.1 below) is particularly interesting to a project such as an AI driven Cyber Security one.

In order to realise the research goal of this paper we look to Python's AI functionality (Géron / Parisi see list of printed works below), the theory behind AI (Turing / Simon) and its implementation. We must also combine this with modern management and organisational alignment theory (Henderson and Venkatraman), organisational structure (Mintzberg) and organisational management (Drucker) in the context of implementation. The above papers have been hugely influential and have revolutionised the way technical organisations operate. In the context of implementing an AI driven Cyber Security infrastructure the theories put forth by these papers are invaluable and have influenced much of this paper. It is felt that the concepts on management theory enable management to realise the technical infrastructure and obtain the specified strategic aim.

An Enterprise Architecture (EA) model could have been selected while completing this paper and while developing an AI driven Cyber Security solution. However, it is felt that an integrated detailed EA investigation is beyond the scope of this paper.

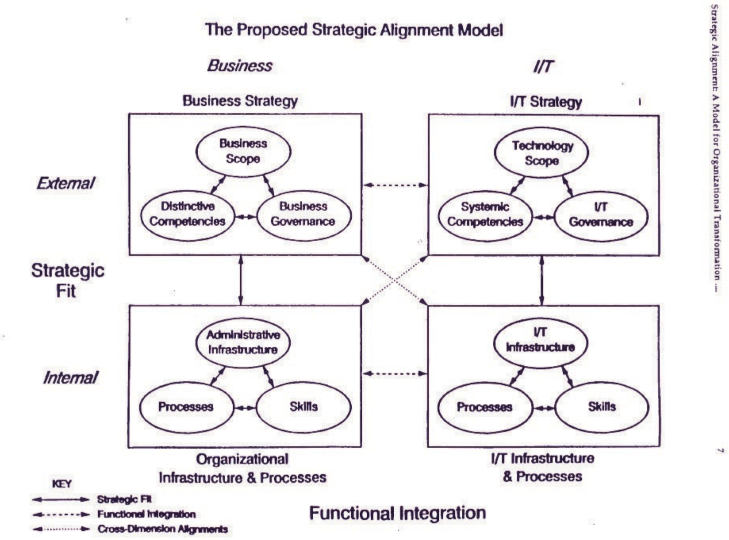


Figure 2.1: Henderson and Venkatraman, Strategic Alignment Model Nov 1990.

Herbert Simon could envisage an intelligent system as could Alan Turing. It is suggested here that we are well beyond Turing's 'Imitation game' and the idea of asking "Can machines think?".

Our ideas on AI "should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous" (Turing 1950). Machines may not be able to think in the complete human sense but intelligent systems, AI systems are capable of traversing massive amounts of data and returning an answer in real time, in a way thinking!

2.1 Scientific Research Document Sources

While searching through the IEEE and NCIRL Library search terms included Artificial Intelligence, Cyber Security and Artificial Intelligence, Dynamic Obfuscation, Dimensionality reduction, Intrusion detection, AI Based Cyber Security etc. Search returns where whittled down to 36 topic related papers.

2.2 Printed Research Document Sources

Associated published works (actual printed books) on related topics extensively used in this paper include:

- **Artificial Intelligence with Python** (Prateek Joshi, Packt Publishing 2017).
  The second edition is more relevant today, this is a great starting book for delving into Artificial Intelligence and Python.
- **Hacking The Art of Exploitation 2nd Edition** (Jon Erickson, No Starch Press 2008)
  Heavily influenced by C programming, tied in well with PyTorch and TensorFlows C++ compilers. Gave some interesting insights to Windows OS and C++ in general.
- **Cybersecurity Threats, Malware Trends, and Strategies** (Tim Rains, Packt Publishing 2023). An extremely interesting book from a general Cyber Security point of view. Referenced in this paper for comments on AI.
- **Deep Learning Algorithms 2nd Edition** (Ricardo Calix, self-published 2024). A massive influence on this paper, covering PyTorch, Deep Learning, Vector Spaces, Convolutional Neural Networks, Generative Adversarial Networks, Transformers and Artificial Intelligence as a whole. It also led to contact with the author.
- **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, Techniques to Build Intelligent Systems 3rd Edition** (Aurélien Géron, O'Reilly Media, 2022). A massive influence on the direction of this paper. Quite a complex book with a wealth of information including in-depth coverage of Scikit-Learn, decision, trees, random forests and ensemble methods.
- **Generative AI with Python and TensorFlow 2** (Joseph Babcock and Raghav Bali, Packt Publishing 2021). Extensive information on TensorFlow added to the perception of

the extensive learning curve of Tensor Flow. Fitted well with the aforementioned "Hands-on machine learning with Scikit-Learn…)

- **Hands-On Artificial Intelligence for Cybersecurity** (Alessandro Parisi, Packt Publishing 2019). Showcases identifying and predicting security threats using artificial intelligence. Also covers the development of intelligent systems and identifying unusual and suspicious patterns. This paper favored Lockheed Martin's Cyber Kill Chain over methodologies expressed in this book.

From a technical point of view, the above printed works all contain brilliant, detailed and relevant content.

2.3 Further Sources

During this research extensive web searches were completed. This with a view to placing a finger on the pulse of the $1trn Artificial Intelligence industry. Quite a cross section of AI and Cyber Security developer web sites were reviewed, with a view to finding out where researchers were concentrating their efforts, in addition to AI and Cyber Security companies, consultancies and news agencies. These organisations included but where not limited to Avanade, Avast, Gartner, Google, Microsoft, Crowdstrike, Tennable, Greenbone, Linux, Python, Mandiant, Nvidia, Paloaltonetworks, Qualys, Sentinel1, RTE, The London Times, The Sunday Business Post and The Wall Street Journal. TensorFlow Blog and PyTorch Blog.

2.4 Conclusion

There is a large volume of very influential managerial research available. Although not written for AI specifically these tried and tested alignment and management approaches are equally relevant to modern day AI and Cyber Security technical implementations as any other technical implementation.

There is an ever growing body of technical works relating to AI and Cyber Security. While a lot of these works are more recent and don't have the reputation or longevity of the aforementioned management approaches they still provide excellent content. From the technical point of view many directions are put forward and the aforementioned documents were invaluable in gathering the knowledge required to make an informed choice between PyTorch and TensorFlow. In comparing and contrasting TensorFlow and PyTorch this paper provides interesting insights to deciding AI selection.

As this paper is concerned with the evolution of Pythons AI functionality it is worth noting that it is not enough to expect that the program itself will do everything for its deployment to be successful. Rather when deploying an AI Cyber Security instance which utilises Pythons AI functionality one needs to think about the business in question, the people, changes to the operational environment or business landscape. Along with these considerations available resources must be taken into account only then can the project be a success. The deployment must work in harmony with the program, or in the terms of Enterprise Architecture, it must be in alignment. Both the technical and managerial / alignment texts are equally important however, none of these papers cover both the technical and managerial aspects as examined in this paper.

3 Research Methodology

In order to investigate the Neutralizing of Malware Sustainably using the evolution of Python's Artificial Intelligence Functionality' both a scientific review of relevant literature and a practical coding aspect are needed.

Below we examine both the hardware and software needed for project success. Python is considered along with other pervasive programming languages. TensorFlow, PyTorch (Google vs Meta) and other AI competitors are assessed along with their market share.

3.1 The Development Environment (Hardware)

It is taken into consideration that today's computer programmers are cross platform based. For the purposes of this paper Linux has been utilised. It is noted that Windows and Mac operating systems do allow for AI and Cyber Security development as do the TensorFlow and PyTorch compilers.

For the coding development aspect of this paper a physical machine was used. Its configuration is below:

- Dell Power Edge T110 II
- Processor Intel Xeon E3-1230 V2 x 8
- Memory 8gb
- 64bit
- HD SSD 2TB
- OS: Ubuntu 24.04 LTS

For further testing purposes and as a portable aide to flexible research Virtual box was used with a Kali instance via a Windows 11, Dell Latitude laptop.

3.2 Python

For the duration of this paper Python has been utilised. Various programming languages can be used for Cyber Security. Python's simplicity, the sheer weight that its development community brings and its wide range of libraries make it a well-founded choice for Cyber Security professionals, AI development professionals, students and enthusiasts alike.

It should be noted if one is interested in developing Cyber Security tools or services, C++ is also an option to consider. "C compilers exist for just about every operating system and processor architecture out there" (Erickson, 2008).

Pythons popularity can be seen below in Figure 3.1. The two primary books used during this project do not mention Java or C++. However, Java and C++ are mentioned here because of the provision in both Tensor Flow and PyTorch for them (compilers).

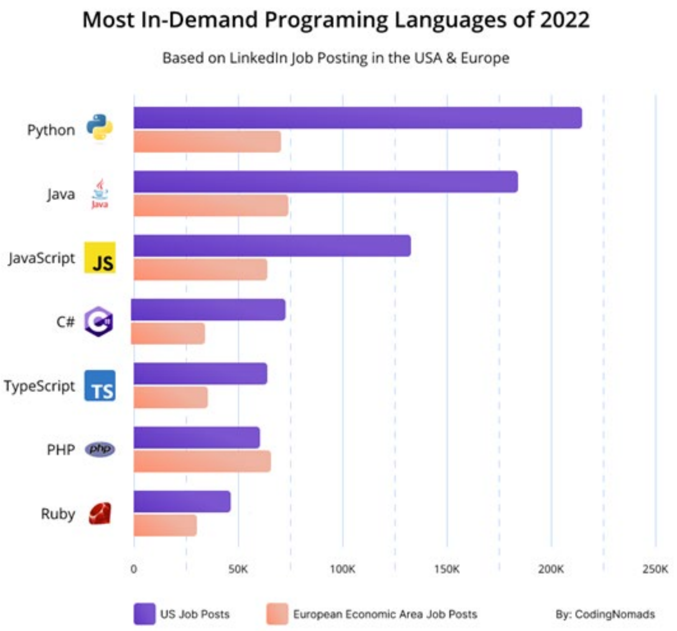


Figure 3.1: Pythons Popularity. [1]

Some of the literature acquired for this paper included Jupyter Notebook code. Jupyter allows the running of python code and shows the result of its execution in an easy view. The literature reviewed highly rates this tool and its seems to be widely used.

3.3 The Python Libraries

Below are the Python Libraries considered for the Neutralizing of Malware Sustainably using the evolution of Python's Artificial Intelligence. Most aide in AI while Volatility and Pefile are used in Cyber Security.

- Scikit-Learn (https://scikit-learn.org), easy to use implements many ML algorithms.
- Tensorflow (https://tensorflow.org). Created by Google, a complex library for mathematical computation. Makes it possible to train and run large neural networks.
- Keras (https://keras.io) a high level deep learning API for running and training a neural network. Comes bundled with Tensor flow.

[1] https://dzone.com/articles/top-10-programming-languages-to-use-in-cyber-secur

- PyTorch (https://pytorch.org/) Meta's ML framework. Used for creating Deep Neural Networks. Supports over 200 mathematical operations its popularity continues to rise.
- NumPy: For building algorithms and tools for AI, designed to aid in Linear Algebra calculations, multi-dimensional array use and matrix operations.
- Pandas: Used for data cleaning (similar to the DataFrame package in R).
- Mathplotlib: Used for data plotting
- Seaborn: For data visualization and plotting, like Matplotlib and used with Matplotlib.
- Pefile: Used for analysing windows executable files.
- Volatility: Allows for analysis of runtime memory
- Anaconda: Python environment for quick access to the most used tools and libraries aides in development activities

3.4 Python's Limitations (Interpreted v Compiled)

It should be noted that python is an interpreted language. In comparison to compiled languages/programs (Java and C++) Python's performance is slower. This is more apparent when seeking intense high performance efficiency as is seen in Cyber Security applications.

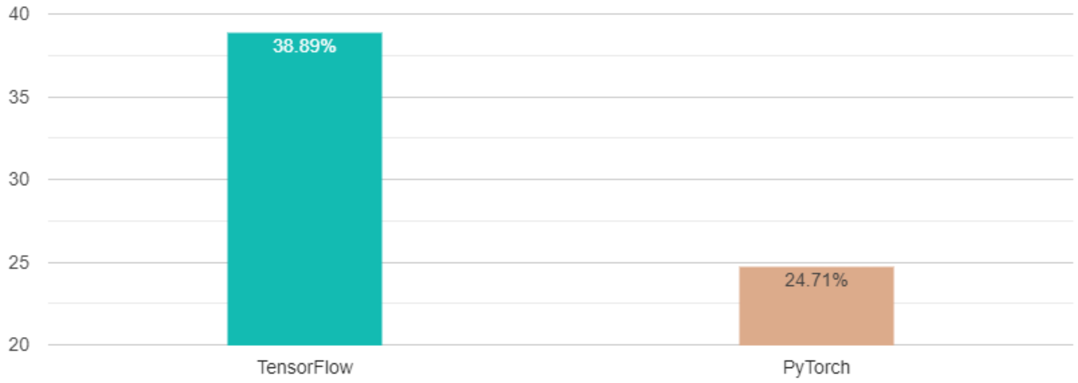
To counteract these inefficiencies Python has many libraries that aide in navigating this issue. It is strongly suggested here that all 3rd party libraries are thoroughly vetted before use. Python's flexibility, ease of use and published documentation of deployed application instances (via community etc.) are quite useful when seeking to achieve computational goals.

3.5 PyTorch (Meta) v TensorFlow (Google)

Looking to the companies who develop AI we examine two organisations which are leading the AI race. These are Google with its "Tensor Flow" and Meta AI's (Facebook) "PyTorch". These two organisations seek to use AI to harvest insights from massive data sets for corporate means. Both organisations have the cognitive brain power (people), infrastructure and the financial capability to develop AI.

3.6 Market Share – Meta v Google Plus Other Players

In this section we look at the market share of Meta, Google, Keras and OpenCV. Primarily this paper is concerned with PyTorch and TensorFlow.



“TensorFlow has a 38.89% market share in the Data Science And Machine Learning category, while PyTorch has a 24.71% market share in the same space”[2].

Tech	Domains (name)	Market Share	Versus
PyTorch	11,649	24.76%	TensorFlow vs PyTorch
OpenCV	8,746	18.59%	TensorFlow vs OpenCV
Keras	8,321	17.69%	TensorFlow vs Keras
TensorFlow	18,327	38.96%	

Top three of TensorFlow’s competitors in the Data Science And Machine Learning category are PyTorch with 24.76%, OpenCV with 18.59%, Keras with 17.69% market share [3].

3.7 Dataset selection

In order to ‘Neutralize Malware Sustainably using the Python’s Artificial Intelligence Functionality’ data must be selected which will then be used with Unsupervised (Supervised learning was also viewed) learning algorithms. This is because we are seeking to detect previously undetected malware attacks. As the system comes across anomalies they will be categorised. Data can be assessed and following this an algorithm can be fine-tuned.

“Undoubtedly, the most of real world datasets have been imperfect, noisy and very difficult to determine the behaviour of this data. Pre-processing play vital role for analysis patterns from network data for achieving accurate results. Therefore, the pre-processing steps are

[2] https://6sense.com/tech/data-science-machine-learning/tensorflow-vs-pytorch
[3] https://6sense.com/tech/data-science-machine-learning/tensorflow-market-share

essential part in IDS *(Intrusion Detection Systems)* to improve the data mining algorithms for classification of intrusions from network datasets" (Alsaadi, Almuttairi, Bayat, Ucani 2020).

An organisations IT data is spread across numerous aspects of operation. From switches to servers, to individual machines, firewalls and antivirus data returns. Some security relevant sources of data in the organisational live environment include security logs returning information from:

- **Network protection systems** – information from the likes of Data Dog and other network monitors to track spikes in cpu performance or switch performance.
- **Endpoint Protection software** – Sophos flagging individual user operations.
- **Application Security monitors** – Server Side Forgery, SQL Injection, Cross Site Scripting, Denial of service
- **Suspect User behaviour** – attempts at fraud for example multiple logins

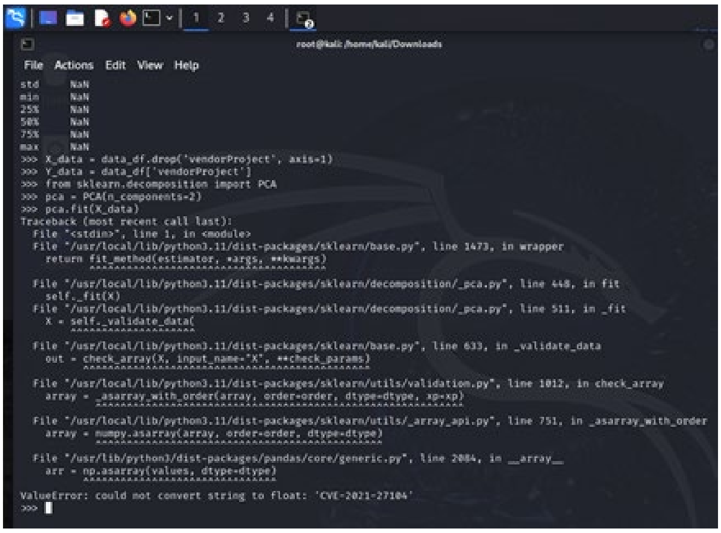
Live operational business data will not be used for this paper.

Particular attention must be given to where and what type of data is being analysed. Effective anomaly detection data assumes training data does not contain the anomaly but references a normal situation. "Handling redundant and irrelevant features in high-dimension datasets has caused a long-term challenge for network anomaly detection. Eliminating such features with spectral information not only speeds up the classification process but also helps classifiers make accurate decisions during attack recognition time, especially when coping with large-scale and heterogeneous data" (Salo, Nassif, Essex 2018)

For this project the well-known CISA CVE Dataset (KEV) has been downloaded. CISA is for 'For the benefit of the cybersecurity community and network defenders—and to help every organization better manage vulnerabilities and keep pace with threat activity—CISA maintains the authoritative source of vulnerabilities that have been exploited in the wild. Organizations should use the KEV catalog as an input to their vulnerability management prioritization framework'[4] . It should be noted that Yarra rules via Didder Stevens web site were also considered.

To showcase that raw data needs to be optimised the CISA dataset was brought straight into the test system as can be seen below in Figure 3.6.1 Raw Data Input Result.

[4] https://www.cisa.gov/known-exploited-vulnerabilities-catalog



```
std      NaN
min      NaN
25%      NaN
50%      NaN
75%      NaN
max      NaN
>>> X_data = data_df.drop('vendorProject', axis=1)
>>> Y_data = data_df['vendorProject']
>>> from sklearn.decomposition import PCA
>>> pca = PCA(n_components=2)
>>> pca.fit(X_data)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/usr/local/lib/python3.11/dist-packages/sklearn/base.py", line 1473, in wrapper
    return fit_method(estimator, *args, **kwargs)
           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

  File "/usr/local/lib/python3.11/dist-packages/sklearn/decomposition/_pca.py", line 448, in fit
    self._fit(X)
  File "/usr/local/lib/python3.11/dist-packages/sklearn/decomposition/_pca.py", line 511, in _fit
    X = self._validate_data(
        ^^^^^^^^^^^^^^^^^^^^

  File "/usr/local/lib/python3.11/dist-packages/sklearn/base.py", line 633, in _validate_data
    out = check_array(X, input_name="X", **check_params)
          ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

  File "/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py", line 1012, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
            ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

  File "/usr/local/lib/python3.11/dist-packages/sklearn/utils/_array_api.py", line 751, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
            ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

  File "/usr/lib/python3/dist-packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
          ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

ValueError: could not convert string to float: 'CVE-2021-27104'
>>> 
```

Figure 3.6.1: Raw Data Input Result

Our process for data cleaning will include editing the data in Excel. Some columns and data will be formatted from floats to text and some text will be turned into columns.

On viewing the data and in order to work with this data over (hypothetically speaking) our network a code matrix has been attached for each manufacturer and product. An example of a manufacturer code would be Accellion who are numbered 100001. The product code has been broken into year and number. For example, 20240001 is FTA and 202400012 is iOS, iPadOS, and macOS.

Interesting to note that Coldfusion is on the CVE list along with Flash Player. Also that various Apple IOS are broken into numerous different categories: "iPadOS, and macOS", "iOS and macOS", "iOS, iPadOS, and watchOS" and "iOS". The breaking of apple IOS into numerous categories is interesting and highlights any decision on grouping other Operating Systems (OS).

If this was a live return of data one would have to consider the security posture of a given organisation and its use of both Coldfusion and Flashplayer. Flashplayer was popular 10 years ago and Adobe Flex led to an argument with Apple over severe security concerns which were not addressed. "Symantec recently highlighted Flash for having one of the worst security records in 2009. We also know firsthand that Flash is the number one reason Macs crash," [5]

To further iterate over the data, regression visualisations are run and further Python code driven analysis has been used. As can be seen below in Figure 3.6.2 Python Coding and Visualization. In Figure 3.6.3 colours are introduced for better visualisation.

[5] https://www.pcworld.com/article/512482/apple_v_adobe_something_just_doesnt_add_up.html

The full code matrix “DataMatrix.xlsx” is attached with this submission along with the exploits CVE csv database and code samples.

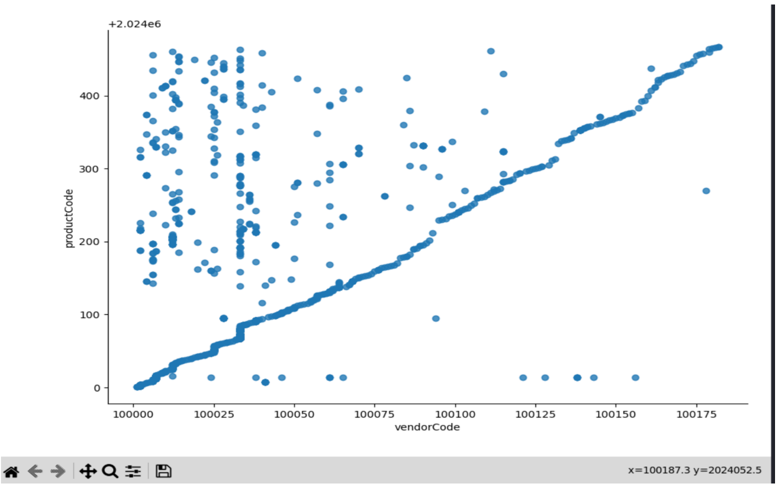


Figure 3.6.1: Python Coding and Visualization.

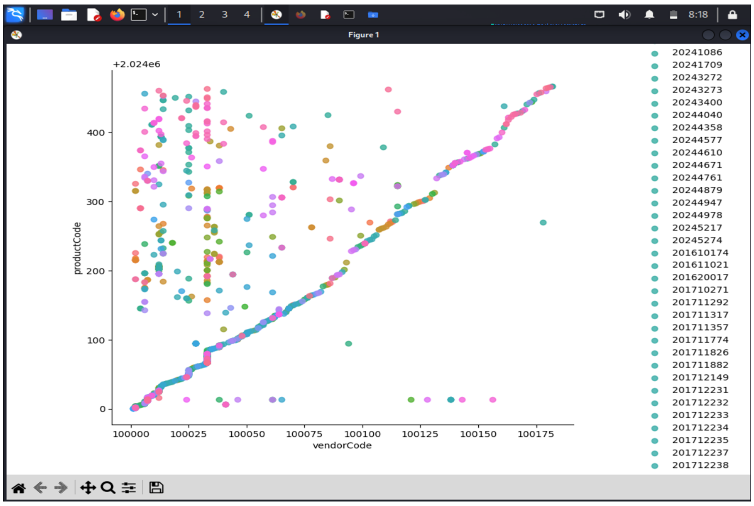


Figure 3.6.3: Python Coding, Visualization and colour.

Both figures 4.18.3 and 4.18.4 show Mathpotlib being used for visualization of data and results. Pandas has been used for data pre-processing. Seaborn has also been used. In section ‘5 Implementation’ further data analysis will be showcased.

4 Design Specification

This section charts the considerations for implementation of an AI driven Cyber Security anti malware program across a number of topics with a view to realising Neutralization of Malware Sustainably using Python's AI functionality.

4.1 Review of a successful AI implementation: The Spam Filter

The first step in researching and planning a successful AI cyber security implementation is to look at examples of previous successful implementations. One of these examples that stands out is the email spam filter. "The first ML application that really became mainstream, improving the lives of hundreds of millions of people, took over the world back in the 1990s: the *spam filter*", "it has actually learned so well that you seldom need to flag an email as spam anymore" (Géron, 2019).

In seeking to neutralize malware lessons can be taken from the spam filter and include:

- **How it makes use of training data:** In order for the spam filter to be trained to know the difference between spam and an actual email the filter uses user flagged (now marked as spam) emails and regular emails to differentiate a difference. These examples are training emails. Each training example is known as a training instance.

  When it comes to the neutralisation of malware the training data could be normalised user traffic and abnormal network behaviour. This would be a very similar methodology as above.

- **The selection of an effective formula:** Task or *T* flag an email, experience *E* is the training data and then the measure needs to be defined. This is defined as *P* or performance. Using a ratio for corrected classified emails results in a measure known as accuracy.

  Again in the neutralisation of malware users traversing a network could be categorised and benchmarked against training and network performance data in a similar approach.

- **Its algorithm usage:** "A spam filter based on machine learning techniques automatically learns which words and phrases are good predictors of spam by detecting unusually frequent patterns of words in spam examples compared the ham (non-spam) examples" (Géron, 2019). Models can be assessed to see what they have learned.

  There is a huge range of algorithms, decision trees and neural networks which can be utilised for malware neutralisation in a similar manner to the spam filter.

- **The Perceptron algorithm:** One of the most frequent AI Neural Network (NN) models used in spam filtering is the **Perceptron**. NNs operate via classifying samples that are linearly separable. The aim of the algorithm is to identify the precise optimal weight vector to be applied On the given estimated values to acquire reliable future predictions on unknown future data.

Taking from the above lessons the widely used Perceptron Algorithm is used in the implementation section of this paper to process information in the context of an unsupervised learning task.

Developing on the work started with the spam filter on the subject of algorithms in the modern context it is interesting to note that "MITRE (US National Cyber Security Firm) and others have developed a taxonomy for describing cyber threats based upon heuristics, signatures, techniques and practices. Researchers have worked on tasks associated with creating vector descriptions that can be utilized by current learning algorithms" (Haass, 2022).

4.2 Management Decisions and Cost

Management need to make a decision on an AI driven Cyber Security infrastructure. It needs to navigate risk. We look to research to aide in business decisions. If we look to Herbert Simon's theory of 'Bounded Rationality' when trying to understand the decision making process of individuals involved with the strategic apex of a given businesses management, Simon "emphasized the limitations of the cognitive system, the change of processes due to expertise, and the direct empirical study of cognitive processes involved in decision making" (Campitelli, Gobet, 2010).

"We will need to reflect upon whether we are investing enough in our ability to defend against and be resilient in the face of AI used in a malign manner, given what we are investing in AI itself." (Creese 2020).

Traditional security solutions can't anticipate new AI-driven attacks, possibly even in a computational capacity. The grim fact that malign actors are adopting AI means that Cybersecurity providers must develop "defensive AI" to counter the threat. Below Figure 4.2.1. shows that when MIT surveyed respondents from the World's 20 largest economies in relation to Cybersecurity initiatives and technologies 59% of respondents tagged Artificial Intelligence.

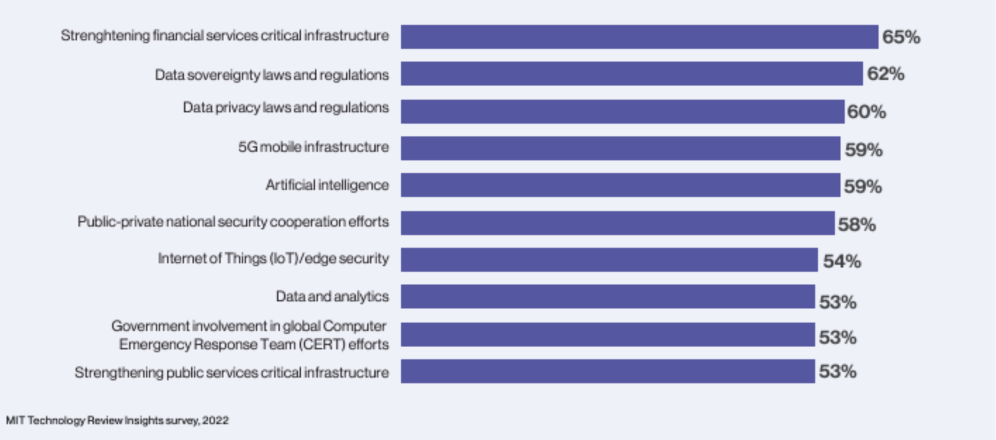


Figure 4.2.2: Above Respondents rank the above initiatives and technologies as the most important to bolster cybersecurity. MIT Technology Review Insights Survey 2022. Surveyed the world's 20 largest and most digitally forward economies.

There is “the potential of I/T strategy to influence key dimensions of business strategy. Within the 'competitive role' for I/T, this perspective is concerned with the exploitation of emerging I/T capabilities to impact new products and services (i.e., business scope), influence the key attributes of strategy (distinctive competencies) as well as develop new forms of relationships (i.e., business governance). (Henderson, Venkatraman 1990)”

4.3 Realisation of Resources

The estimated costs to embrace AI varies. Take for instance, Latitude, a company which provided an AI Generative game before Chat GPT rose to fame. Its CEO Nick Walton recalled that in 2021: “Latitude was spending nearly $200,000 a month on OpenAI’s so-called generative AI software and Amazon Web Services in order to keep up with the millions of user queries it needed to process each day.” And “By the end of 2021, Latitude switched from using OpenAI’s GPT software to a cheaper but still capable language software offered by startup AI21 Labs” [6].

An uncomfortable truth is that the cost to develop and maintain AI software can be extraordinary high. This brings down the margin and puts up a large barrier to entry. For larger organizations who can afford the cost, they can create a bespoke, high performing and sustainable solution over which they have greater control. Smaller organizations will most likely not train their own Large Language Models (LLM). It will be cheaper to have Microsoft or some such company do this in the future for a fee. This would result in a solution which is less tailored to their specific business but at a lower cost and with a greater level of sustainability than a small company could afford to achieve independently.

We can say with certainty that Nvidia has the leading market share for GPUs used in the AI industry, and its primary data center workhorse chip costs $10,000. Interestingly to note that in in May 2024 “Nvidia quarterly profits soar on demand for AI power” [7]. Followed by Reuters “After the buzz investors are doing their own home work on AI” [8]. Lastly Meta had a situation where it announced a $3 billion dollar spend in AI, Meta generates $50 billion profit a year, after the announcement $132 billion drop off its share price: “Meta, Alphabet Investors View Spending Through Different Lenses” [9]. It is not that markets are volatile, it is that harnessing AI and making a return off it is challenging in the short term. The $1trn AI boom continues and “The risk of under-investing is dramatically greater than the risk of over-investing, (Sundar Pichai, Alphabet CEO)” [10].

[6] https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html

[7] https://www.rte.ie/news/business/2024/0523/1450750-nvidia-quarterly-results/

[8] https://www.reuters.com/technology/after-buzz-investors-are-doing-their-own-homework-ai-2023-06-26/

[9] https://www.bloomberg.com/news/newsletters/2024-04-29/meta-s-ai-spending-plan-spurs-pushback-from-investors

[10] https://www.economist.com/business/2024/07/28/what-are-the-threats-to-the-1trn-artificial-intelligence-boom

The high cost of training and "inference" that is actually running AI's large language models is a cost that differs from previous computing booms. Even when the AI infrastructure is built, trained and deployed, it will still require a huge amount of computing power to run. This is because AI's large language models do billions of calculations every time they return a response.

4.4 Operations and the Ability to Make Informed Decisions

Aside of the strategic apex management decision makers, IT needs to be involved with technical decisions. To this end company structure must be defined. In this we look to Mintzberg's paper "The Structuring of Organizations". Figure 4.4.1 below shows Mintzberg's organizational diagram.

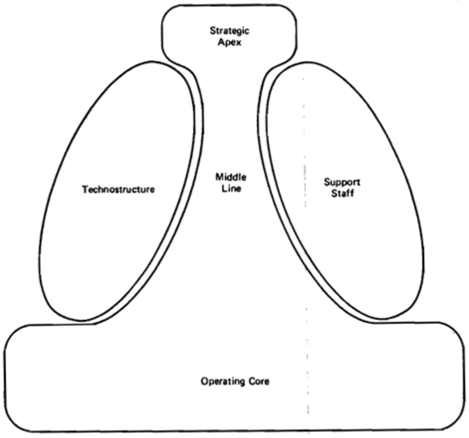


Figure 4.4.1: Mintzberg's 5 parts of Organizational Structure.

In particular, we are interested in the relationship with between the Strategic Apex and the Technostructure. The Technostructure's autonomy and the effect (guidance / input / experience) of the strategic apex on its operation. The Technostructure in today's tech driven companies although at times small in size (given the entity), numbers wise can have more of an impact on business than any other area. "In the Technostructure we find the analysts (and their supporting clerical staff) who serve the organization by affecting the work of others. These analysts are removed from the operating work flow-they may design it, plan it, change it, or train the people (author note* or machines) who do it, but they do not do it themselves. Thus, the Technostructure is effective only when it can use its analytical techniques to make the work of others more effective." (Mintzberg, 1979)

"A decision is a judgment. It is a choice between alternatives. It is rarely a choice between right and wrong. It is at best a choice between "almost right" and "probably wrong"—but much more often a choice between two courses of action neither of which is probably more nearly right than the other. Most books on decision-making tell the reader: "First find the facts." But executives who make effective decisions know that one does not start with facts. One starts with opinions." (Drucker, 2006).

In the live environment a Security Policy may well aide in decision making. "The policy and procedure highlight rules and assign roles for implementation and monitoring. A list of responsibilities should also be set and a risk register must be provided to encourage responsible parties to implement the policy and understand the possible harm of different security breaches" (Jawhar, Miller, Bitar 2024).

4.5 Education and Staff Requirements

In looking at key skills and attributes needed for members of an implementation team a formal education in Computer Science with experience in Python would be required. Added to this maths concepts such as Liner Regression and Linear Algebra.

It is suggested here a visual imagination is becoming more important. This is extremely useful when editing meshes with linear algebra and understanding the essential matrices and vectors for algorithm creation. Knowledge of Statistics would be beneficial.

Aside of AI knowledge some Cyber Security particulars will be needed including some knowledge of ethical hacking/malware analysis and its procedures. Knowledge of Linux with possibly another formal qualification ideally at Postgraduate or Masters level would be optimal.

One of the most important aspects of being effective is to be a good communicator. Communication between staff and a clearly defined hierarchy can go a long way towards project and business success.

Needless to say difficult positons to fill.

4.6 Algorithms

AI driven infrastructure with its expansive resource heavy and more complicated algorithms provide an added incline when it comes to the barriers for entry. Cyber Security staff utilising AI will need to understand the logic of algorithms. This will aide in the 'fine tuning' phases based on acquired results. Classification of data, clustering of data along with Predictive analysis (identifying threats as they emerge) requires a dynamic approach allowing algorithms to optimise learning capabilities.

Algorithms are quite valuable "Even the smallest patterns can bring in millions to the first investor who unearths them. And they'll keep churning out the profits until one of two things happens: either the phenomenon comes to an end or the rest of the market catches on to it, and the opportunity vanishes. By that point, a good quant will be hot on the trail of dozens of other tiny wrinkles" (O'Neill, 2016).

Algorithms will be show cased in the next section of this paper 'Implementation'.

5 Implementation

In order to complete a detailed assessment of code sustainability and to truly investigate the idea of 'Neutralizing of Malware Sustainably using the evolution of Python's Artificial Intelligence Functionality' we need to get as close to a real world deployment as possible. In this section we look at a Code management and the selection of PyTorch. Data sources that are in the live environment are discussed along with the algorithms that are used. The Perceptron, linear, logistical and decisions trees are discussed. Management and its communication with development is considered. Lastly the Cyber Kill Chain is discussed.

5.1 Code Management infrastructure

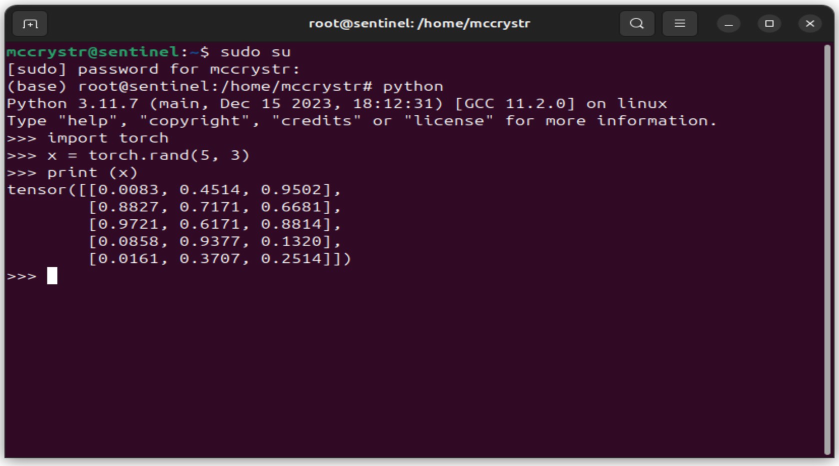
In order to manage code collaboratively and consistently from the testing environment through the live environment Anaconda has been used. Anaconda allows for version control making version management easy along with ensuring that packages and libraries are reproduced. External git repos can also be added.

5.2 PyTorch

The primary coding for this project utilised PyTorch. During the initial stages of this paper, while reviewing TensorFlow, it was felt that TensorFlow debugging was more complex than PyTorch. This led to PyTorch and Scikit Learn being the chosen option.

Further to the above, and with sustainability in mind, it was noted that "Google did a complete rewrite of TensorFlow to create, at least, two ecosystems which include: low level TensorFlow 2 and high Level Keras" (Calix, 2023). This lead to a significantly increased learning curve when utilising TensorFlow.

Below in Figure 5.2.1 is a quick check on the PyTorch install.



```
mccrystr@sentinel:~$ sudo su
[sudo] password for mccrystr:
(base) root@sentinel:/home/mccrystr# python
Python 3.11.7 (main, Dec 15 2023, 18:12:31) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import torch
>>> x = torch.rand(5, 3)
>>> print (x)
tensor([[0.0083, 0.4514, 0.9502],
        [0.8827, 0.7171, 0.6681],
        [0.9721, 0.6171, 0.8814],
        [0.0858, 0.9377, 0.1320],
        [0.0161, 0.3707, 0.2514]])
>>> 
```

Figure 5.2.1 A quick check of the PyTorch install.

5.3 Data Sources

The CISA CVE database is used in this deployment. “it is crucial for deep learning models to become able to make connections and associations in raw data, and to process them in a way suitable for further use.” (Onchis, Istin, Eduard-Florin, 2022)

Firewall alerts via email can be seen below. It is envisaged that a Python script would take these from Outlook automatically and add the retuned information to our main data:

```
Message meets Alert condition
The following intrusion was observed: Nmap.Script.Scanner.
date=2024-07-31 time=20:28:35 devname=UCDCSL-FW-Pri devid=FG200ETK18904043 eventtime=1722454115195803551
tz="+0100" logid="0419016384" type="utm" subtype="ips" eventtype="signature" level="alert" vd="root" severity="low"
srcip=10.90.18.68 srccountry="Reserved" dstip=192.168.1.241 dstcountry="Reserved" srcintf="VLAN212" srcintfrole="lan"
dstintf="wan1" dstintfrole="wan" sessionid=146463912 action="dropped" proto=6 service="SSL" policyid=12
poluuid="b697d790-aabe-51e8-31cf-12315fe4c962" policytype="policy" attack="Nmap.Script.Scanner" srcport=59207
dstport=21 url="/" direction="outgoing" attackid=45360 profile="all_default"
ref="http://www.fortinet.com/ids/VID45360" incidentserialno=76136226 msg="tools: Nmap.Script.Scanner" crscore=5
craction=32768 crlevel="low"

Message meets Alert condition
The following intrusion was observed: Nmap.Script.Scanner.
date=2024-07-31 time=20:28:35 devname=UCDCSL-FW-Pri devid=FG200ETK18904043 eventtime=1722454115195693729
tz="+0100" logid="0419016384" type="utm" subtype="ips" eventtype="signature" level="alert" vd="root" severity="low"
srcip=10.90.18.68 srccountry="Reserved" dstip=192.168.1.216 dstcountry="Reserved" srcintf="VLAN212" srcintfrole="lan"
dstintf="wan1" dstintfrole="wan" sessionid=146463916 action="dropped" proto=6 service="SSL" policyid=12
poluuid="b697d790-aabe-51e8-31cf-12315fe4c962" policytype="policy" attack="Nmap.Script.Scanner" srcport=59211
dstport=21 url="/" direction="outgoing" attackid=45360 profile="all_default"
ref="http://www.fortinet.com/ids/VID45360" incidentserialno=76136224 msg="tools: Nmap.Script.Scanner" crscore=5
craction=32768 crlevel="low"
```

Firewall alerts via email can be seen below, note common ports 80 (HTTPS) and 211 (TCD/UDP):

From IP	To IP	Source VLAN	Source Port	Destination Port	Direction	Score
192.168.1.241	10.90.18.68	212	59209	21	In	9
52.31.67.171	10.90.107.4		42633	80	Out	11
139.59.72.49	10.90.107.4		35226	80	Out	5
193.177.182.119	10.80.107.4		58892	80	Out	12

Each alert instance would be assigned a score as can be seen in the table above. Subsequently frequently occurring IP addresses showing excessive traffic could be classified as suspicions and checked. Armed with this information in the matrix we turn to the implementation offered by the numpy library. When we go to assess this information the linear algebra will look like this:

$$y = w_0 x_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = \sum w_i x_i = wTx$$

Malware Dataset

It can be seen from the above that data across a network can come from many sources. When looking at IP addresses we could use data from Active Directory (if it is being used for network management). This would assist in tracking down a given machine which is exhibiting signs of malicious behaviour. Wireshark could also be used. Depending on what way assets are distributed to users, most organisations track Mac addresses and a ledger is kept on who registered the machine, who users it, or at least what room it is in.

Static Malware

There are two types of malware analysis static and dynamic. When assessing static malware analysis one might download a malware sample from Malware-Traffic-Analysis.net. The sample may possibly be run through Volatility. This author prefers the Python 2 version of volatility. When assessing malware both Python 2 and Python 3 can be installed on the assessing machine. Some older legacy tools only run on Python 2 as that is what they were built on. The malware sample for example is saved as a disc image and run through the volatility.

Dynamic Malware

As we are monitoring a network we are concerned with Dynamic analysis. It is becoming increasingly important that malware is correctly identified and associated with classes or a family type even if the signatures are not compatible with each other. A clustering algorithm can be used for example K-Means which is showcased in Code_9_Clustering.py.

Its return is:

```
Predicted      0        1
Observed
0         83419 13107
1         7995    32923
```

Silhouette coefficient: 0.975

Decision trees can also be used here.

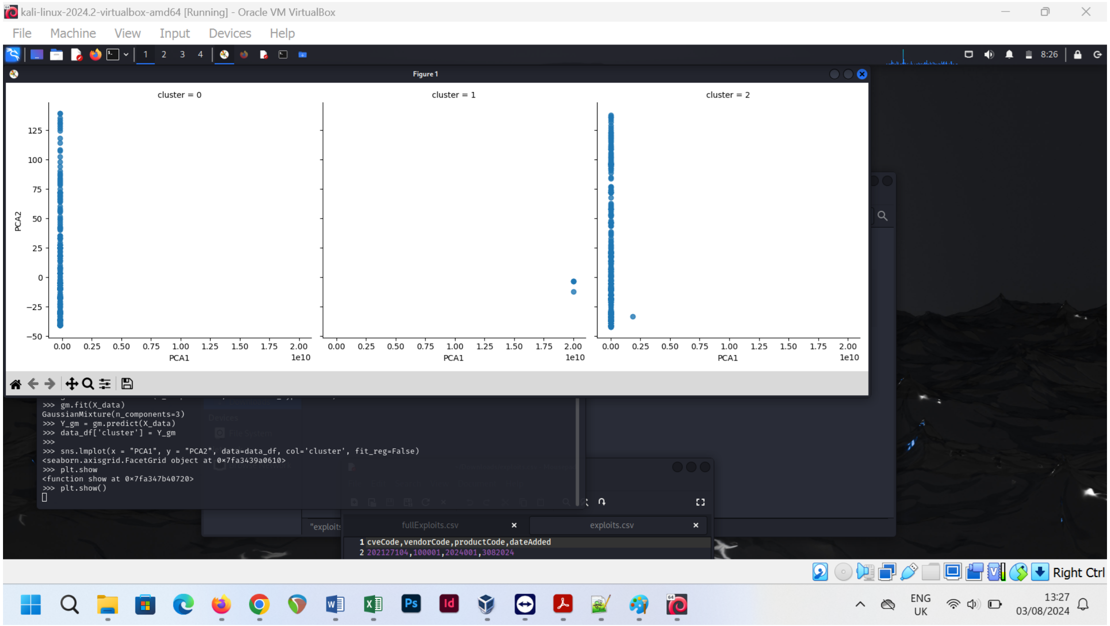
Finally, it should be mentioned that with the data inputs it is network anomalies that are sought. In simple terms an example would be - that strange IP that seems to be returning on a strange port.

Multiple AI

Another possible approach would be: “two Artificial Neural Networks (ANN) are used – one serving as an IDS, capable of detecting usual attacks. The other ANN is trained using the test-time neuron activations of the first ANN to perform binary classification into samples displaying the characteristics of adversarial attacks and normal, non-adversarial samples. This way, a detector is formed capable of raising an alarm whenever a sample trying to circumvent the IDS with an adversarial evasion attack is spotted” (Pawlicki, Kozik, Choras, 2022)

5.4 The Perceptron algorithm at work

Before using the Perceptron, a clustering algorithm has been used on the data. As we are concerned with unsupervised learning “detecting a previously undetected malware attack” the GaussianMixture clustering model will be used. Using this approach we want to get the data as a collection of Gaussian Blobs.



```
>>> gm.fit(X_data)
GaussianMixture(n_components=3)
>>> Y_gm = gm.predict(X_data)
>>> data_df['cluster'] = Y_gm
>>>
>>> sns.lmplot(x = "PCA1", y = "PCA2", data=data_df, col='cluster', fit_reg=False)
<seaborn.axisgrid.FacetGrid object at 0x7fa3439a0610>
>>> plt.show
<function show at 0x7fa347b40720>
>>> plt.show()
```

```
1 cveCode,vendorCode,productCode,dateAdded
2 202127104,100001,2024001,3082024
```

Figure 5.4.1: A clustering algorithm at work.

In the above figure (5.4.1: A clustering algorithm at work) the clustering algorithm succeeded and has successfully classified the data automatically, without previous information as to the labels in the data and various samples.

$$if\ wx \geq \theta \rightarrow f(y) = +1;$$
$$if\ wx < \theta \rightarrow f(y) = -1;$$

The Perceptron (Code_6_Perceptron3.py) algorithm learns by initializing weights to a predefined value (usually 0). Then the output value is calculated for each training sample. The Perceptron algorithm then updates the weights on the distance between the expected output value and the predicted value. The Perceptron is a simple algorithm. Its results can only be used if the data can be linearly separable.

5.5 Linear + Logistical Regression and Decision Trees

Continuing with the algorithms and to further explore the sustainable evolution of Python we will look at Logistical Regression and Decisions trees. Regression models are extensively used as learning algorithms. The most used model is linear regression due to its simplicity and its predicative integrity.

Linear Regression
Linear regression can be described as:

$$y = wX + \beta$$

Y is the predicted values, the result of single features X, with the weight of the estimate w, with β the predicted value. All values are 0 or missing.

(Code_6_linearReg.py) Returns a coefficient between 0 and 1, measuring the predictions returned when compared to the simple mean. It can only be used with quantitative data. Its model assumes features are unrelated, that is to say they do not influence each other. If data being analysed is complex the linear regression model can lead to distorted predictions.

Logistical Regression
Linear regression can lead to classification errors. Using the Perceptron with linear regression can lead to skewed classification accuracy. Logistical regression estimates the probability of samples belonging to individual classes.

Logistical regression can be defined as:

$$P(y = c|x) = \frac{e^z}{(1 + e^z)}$$

Logistical regression can be used for phishing detection along with spam filtering. The (test) data that we have contains 30 features that characterize phishing websites. The accuracy of the Logistical regression classifier is good and detects 90% of URL.

The Logistical regression model can be trained efficiently. It can be used effectively with a large number of features and is quite scalable. Features need to be linear independent (minus multicollinearity). It requires more training samples.

Decision Trees
Looking at data in non-numeric qualitative form (words, descriptions). When the code accompanying this paper (Code_8_decision_trees.py) is run on the exploit_5.csv we can see that it has an improved performance in comparison to the afore mentioned logistic regression.

A decision tree is a flow chart like structure used to make decisions or predictions as can be seen below in Figure 5.5.1: A decision tree example.

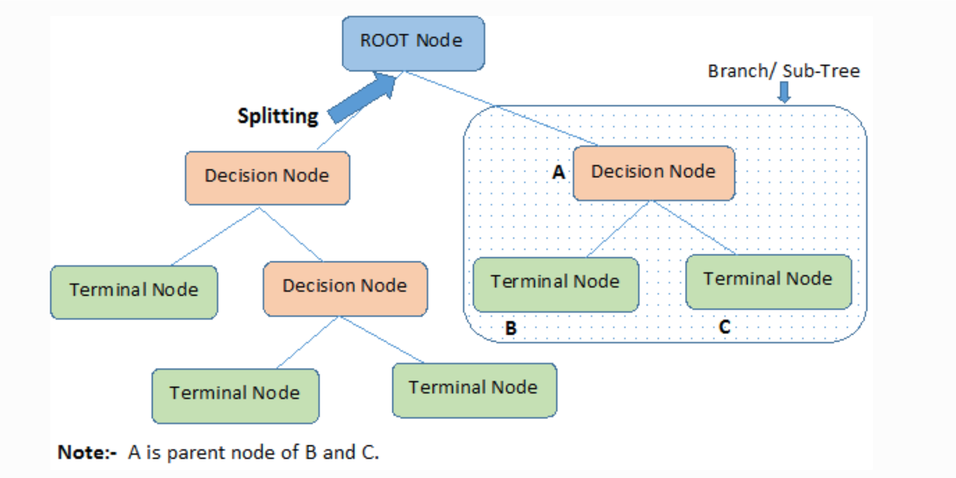


Figure 5.5.1: A decision tree example [11]

It should be noted the data in the exploit_5.py dataset does not have words in it. It is used here to showcase its use.

5.6 Strategic Management and Development communication

Communication is key. Having viewed Mintzberg's '5 parts of Organizational Structure' and considering Drucker's 'The Effective Executive' along with taking Henderson and Venketraman's 'Strategic Alignment: A model for organizational transformation via information technology' into account it is suggested here that the strategic apex of the organization should be involved with both knowledge and direction. There should be a free flow exchange of knowledge in all directions with a view to greater resource utilization and allocation. Such a cohesive communication structure would aide in return on investment and make the utilization of Python and its AI more sustainable.

5.7 The Cyber Kill Chain

The Cyber Kill Chain Framework (figure 5.7.1 below) Developed by Lockheed Martin can be used as part of an Intelligence Driven Defence model for identification and prevention of cyber intrusions activity. It is selected here as a blueprint for identifying aspects of an organisations IT footprint which might be exposed to threats and subsequently what data can be returned to maximise in malware neutralization.

[11] https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

“It might not be possible to attribute any particular threat actor’s activities in the Reconnaissance phase, prior to their attack. With so much network traffic constantly bombarding all internet-connected devices, it is typically challenging to pick out specific probes and reconnaissance activities conducted by specific attackers. But it is not impossible. This is an area where the combination of AI and ML, good threat intelligence, and granular logs is very promising” (Rains, 2023).

AI enabled Cyber Threat Intelligence (CTI) can return information across a wider threat landscape faster than a human thus lowering response times. As can be seen with the amount of planning that is involved in deploying an AI Cyber Security solution and “like many aspects of computer science and cybersecurity, the value derived here is a function of the effort that is put into it” (Rains, 2023).

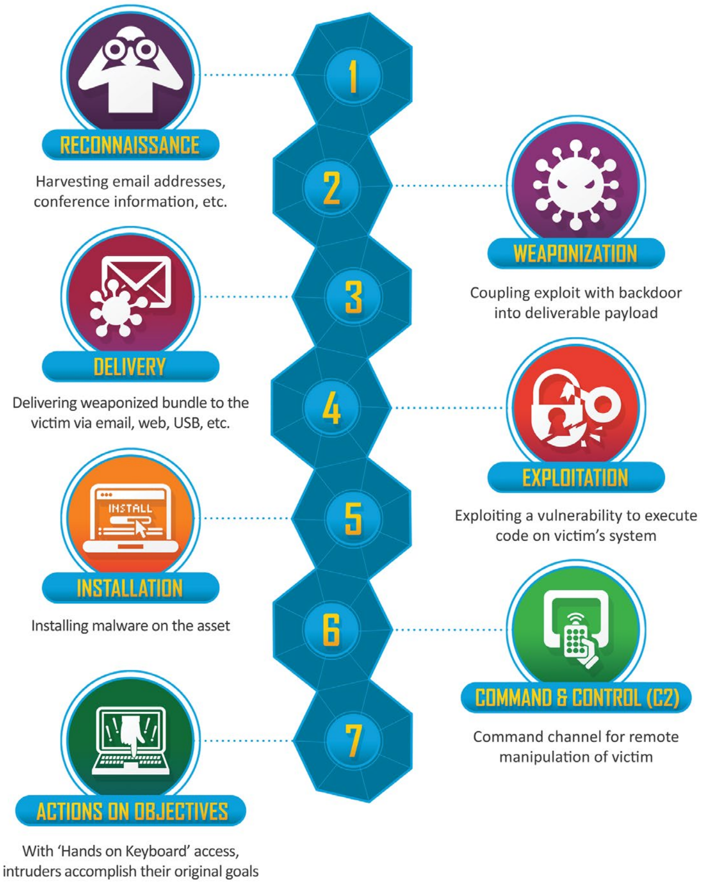

1
RECONNAISSANCE
Harvesting email addresses, conference information, etc.
2
WEAPONIZATION
Coupling exploit with backdoor into deliverable payload
3
DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.
4
EXPLOITATION
Exploiting a vulnerability to execute code on victim’s system
5
INSTALL
INSTALLATION
Installing malware on the asset
6
COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim
7
ACTIONS ON OBJECTIVES
With ‘Hands on Keyboard’ access, intruders accomplish their original goals

Figure 5.7.1: Lockheed Martin’s Cyber Kill Chain

6 Evaluation

The ‘Neutralizing of Malware Sustainably using the evolution of Python’s Artificial Intelligence Functionality’ is achievable by navigating change effectively. “As environments change, the administrative process must deal not just with which domain, but how and how fast to change the design, structure, or technology of the organization” (Henderson, Venkatraman, 1990).

AI use in Cyber Security has a great potential for enabling both Security Operations Centres (SOC) and Security Information and Event Management (SIEM) plans. “Malware cannot be easily identified and most attackers use the best algorithms to build advanced malware therefore SIEMs need to be capable of using the best and strongest process analysis system to identify, remove malware in the systems” (Perera, Rathnayaka, Madushanka, Senarathne 2021).

Python and PyTorch are flexible, easy to learn and implement. Data sources and their optimization will always be a sticking point as organisations evolve. Concepts of alignment and organisational transformation are very important to sustainability. Educated staff capable of harnessing the power of an AI driven Cyber Security implementation along with those making the decisions at the strategic apex add to the sustainability of Pythons AI and its ability to Neutralize Malware.

Cost is a barrier to entry into AI and its use. However, sustainability on a smaller scale is possible via third party suppliers. Using object orientated programming methodology AI algorithms can be better utilised and new ones constructed. Algorithm refactoring leads to improved performance. The algorithms don’t change, new ones may emerge and the code behind their processing may change but in the norm it is the methodology that defines their success over time. How staff, technology hardware, software and cyber security threats are managed in the organisational context is vital to sustainability. Python code is sustainable with the correct machinations in place for version tracking and code updating.

The algorithms and coding libraries used proved a steep learning curve. However, it is hypothesised that the investigation was complete. Time was a large constraint. With more time more in depth investigation could be carried out. Given the virtual computing business Amazon AWS and Azure it would have been interesting to run multiple machines and test such platforms further however this was not possible within the financial constraints of the research.

AI and Cyber Security are here to stay and huge amounts of investment continue to enter this market. Returns may be low initially but those not in the game will likely struggle to keep pace with the changing technology environment. AI take up has reached the general public as has it reached malicious actors. In order to counter the darker elements of AI Cyber Security needs to follow its own AI path.

The AI the Cyber Security landscape has changed and the option of continuing with non AI or incumbent technology and methods is unlikely to keep pace with the ever evolving threat risk.

7 Conclusion and Future Work

The ‘Neutralizing of Malware Sustainably using the evolution of Python’s Artificial Intelligence Functionality’ has been an interesting research topic and quite a learning endeavour. The research topic has been subject to a detailed investigation.

Cost, education, communication and a given organisations ability to manage change is vital to success. There is a learning curve from algorithms and industry jargon but these issues can be overcome.

The benefits of realising an AI Cyber Security instance that can aide in the neutralization of malware makes this a project worth considering. The information garnered from the project endeavour would produce a significant return and aide in informed decisions and organisational security posture and overall direction.

Future work would include more extensive use of Anaconda and obtaining a working knowledge of Jupyter note books.

AI driven Cyber Security along with “Deep learning is a promising machine learning-based approach that can address the challenges associated with the design of intrusion detection systems as a result of its outstanding performance in dealing with complex, large-scale data” (Musafer, Alessa, Faezipour, Abuzneid 2019).

8 References

D'Hoinne J., Litan, A. and Firstbrook, P. (2023) '4 Ways Generative AI Will Impact CISOs and Their Teams', *Gartner*, pp. 2.

Jawhar, S., Miller, J. and Bitar, Z. (2024), 'AI-Based Cybersecurity Policies and Procedures', *3rd IEEE International Conference on AI in Cybersecurity (ICAIC*), pp2

D'Hoinne J., Litan, A. and Firstbrook, P. (2023) '4 Ways Generative AI Will Impact CISOs and Their Teams', *Gartner*, pp. 2.

Turing, A. (1950) 'Computing Machinery and Intelligence', *Computing Machinery and Intelligence. Mind 49:* 433-460. Pp

Erickson, J. (2008) 'The Art of Exploitation', 2nd edition, *No Starch Press*, pp19

Alsaadi, H,I., Almuttairi, R. M, Bayat, O. and Ucani, O. M. (2020) 'Computational Intelligence Algorithms to Handle Dimensionality Reduction for Enhancing Intrusion Detection System', *Journal of Information Science and Engineering 36, 293-308 (2020)*, pp296

Salo, F., Nassif, A. B. and Essex, A. (2018) 'Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection', *Department of Electrical and Computer Engineering,* pp1.

Géron, A. (2019) 'Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow', *O'Reilly Media, Inc*. pp3.

Géron, A. (2019) 'Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow', *O'Reilly Media, Inc*. pp6.

Haass, J.C. (2022), 'Cyber Threat Intelligence and Machine Learning', *Fourth International Conference on Transdisciplinary AI (TransAI)*, pp1.

Campitelli, G. and Gobet, F. (2010) 'Herbert Simon's Decision-Making Approach: Investigation of Cognitive Processes in Experts', *Review of General Psychology*, pp 2.

Creese, S. (2020), 'Artificial Intelligence and the Law', *Routledge, also MIT Technology Review* 2022.

Henderson, J. and Venkatraman, N. (1990), 'Strategic Alignment: A model for organizational transformation via information technology', *Massachusetts Institute of Technology, Sloan School of Management* pp16.

Mintzberg, H. (1979), 'The Structuring of organisations', *Prentice Hall*, pp30

Drucker, P.F. (2006) 'The Effective Executive', *Harper Collins Publishing*, P143.

Jawhar, S., Miller, J. and Bitar, Z. (2024), 'AI-Based Cybersecurity Policies and Procedures', *3rd IEEE International Conference on AI in Cybersecurity (ICAIC*), pp1

O'Neill, C. (2016), 'Weapons of math destruction, how big data increases inequality and threatens democracy', *Penguin Publishing*, pp33

Calix, R.A. (2023), 'Deep Learning Algorithms, [GPTS to MLPs] In Pytorch', Ricado Calix, 2nd Edition, *Self-Published*, pp xv.

Onchis, D.M., Istin, C. and Eduard-Florin, H. (2022), 'Advantages of a neuro-symbolic solution for monitoring IT infrastructures alerts', *24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing* (SYNASC), pp190

Pawlicki, M., Kozik, R. and Choras, M. (2022), 'A survey on neural networks for (cyber-) security and (cyber-) security of neural networks', *Neurocomputing 500 (2022) 1075–1087*, pp 1082

Rains, T. (2023) 'Cybersecurity Threats, Malware Trends, and Strategies', 2nd Edition', *Packt Publishing*, pp480

Rains, T. (2023) 'Cybersecurity Threats, Malware Trends, and Strategies', 2nd Edition', *Packt Publishing*, pp32

Henderson, J. and Venkatraman, N. (1990), 'Strategic Alignment: A model for organizational transformation via information technology', *Massachusetts Institute of Technology, Sloan School of Management* pp27.

Perera, A., Rathnayaka, S., Perera, N. D. , Madushanka, W.W., and Senarathne, A.N. (2021) 'The Next Gen Security Operation Center', *2021 6th International Conference for Convergence in Technology (I2CT), IEEE*, pp2

Musafer, H., Alessa, A., Faezipour, M. and Abuzneid, A. (2019) 'Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection', *Electronics 2019*, 8, 322, MDPI pp2