Evaluate the use of Artificial Intelligence (AI) and Natural Language Processing (NLP) to bridge the gap between security policies and employees in large enterprises.

MSc Research Project

MSc Cyber Security

Lisa Brockman

Student ID: x20185758

School of Computing

National College of Ireland

Supervisor: Michael Pantridge

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Lisa Brockman……………………………………………………………..…………

Student ID: X20185758..……………………….…………………..…………………………………

Programme: MSc Cyber Security ..…………………………… **Year:** 2021...……..

Module: Final Project……………………………………………………………………………

Supervisor: Michael Pantridge..…………………...…………….…………………………………

Submission Due Date: 24th April 2024………………………………..………………………………….

Project Title: Evaluate the use of Artificial Intelligence (AI) and Natural Language Processing (NLP) to bridge the gap between security policies and employees in large enterprises.………

Word Count: 7286…………………………..…. **Page Count** 18……………………….……

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: ……………………………………………………………………………………………………

Date: 24/04/2024……………………………………………………………………………………………

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):	

Contents

1 Abstract

Security policies are crucial for establishing the security posture of an enterprise, with significant number of research papers attributing a reduced level of compliance to those enterprises where there is not sufficient dissemination, communication, understanding and clarity of ask.

The factors contributing to a reduced level of compliance are by and large thought to be people problems, rather than technical ones. Despite significant research into these factors, the key challenge which persists is bridging the gap between the security policies of the enterprise and the employees who need to comply with them. The importance of having a securely educated workforce cannot be underestimated. A 2020 research paper from Stanford University found that approximately 88% of all data breaches are caused by employee mistakes, in many cases attributed to a lack of security policy knowledge, with 45% of employees attributing distraction as the top reason for falling for security threats such as phishing. The study went on to conclude that employees report they are primarily focused on the job they have been hired to do, rather than having the time to find, read, understand, and comply with security policies. [1]

An opportunity to bridge this gap requires the simplification of the process for employees to find security policy information and avoid having to read through pages and pages of "security speak" to attempt to decipher an answer to their problem.

This research paper evaluated the potential for the use of Artificial Intelligence (AI) and Natural Language Processing (NLP) to bridge the gap by carrying out a full literature review on both the factors affecting successful security policy adoption, and the state of the art in Chatbots. A technical model for an enterprise specific security policy Chatbot was designed, implemented, and trained on a limited set of ISO27001 policies. The Chatbot: PolicyPal, was also fine tuned during an iterative set of phases to continuously improve the quality and accuracy of the answers provided by the Chatbot. At the end of the fine-tuning phase, the PolicyPal was able to answer 72% of test cases effectively, with the remaining 28% being partially effective.

2 Introduction

Security policies are the foundation of a strong Security Management System (SMS) in a large enterprise (more than 200 employees). They outline the way in which the business must behave (covering both the things that the enterprise (it's employees and suppliers) must and must not do) to align with the objectives of the Security function, and the overall strategy of the enterprise to protect its data, systems, and people.

However, simply having a library of Security Policies is not enough to deliver the security objectives for the enterprise. Frequently an enterprise will have the library of documents available; having been created at a point in time, sometimes for a compliance exercise or as the result of an audit finding, but they are not fully communicated, maintained, implemented, or understood.

People are the cornerstone of security in an enterprise – the first line of defence. To be able to take appropriate action when a security concern arises, people (the employees of the enterprise) need to understand the enterprises requirements for security as documented in their security policy.

In a large enterprise, having a successful approach for the adoption of Security policies by employees, can be challenging. There are many stakeholders to manage, multiple levels of review and approval, which can take considerable time to complete, meaning that sometimes the relevance and applicability of the policy is not as timely or as appropriate as it should be.

This is supported by the most frequently used phrases to describe security policies from employees, as shown in the word cloud on the right.

Due to the negative perceptions and concerns of employees across large enterprises, one of two scenarios typically occur when the employee has a security question.

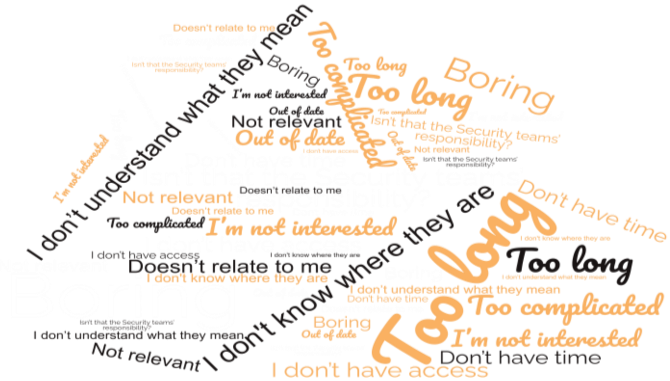

Figure 1: Employee perceptions of security policies

In the first Scenario, the employee takes no action (usually due to one of the barriers shown above) and does nothing, they don't get their security question answered and potentially may cause a policy or data breach, start using shadow IT, and potentially increase the likelihood of a cyber event to the enterprise. In the second scenario, the employee reaches out to a member of the security team, poses their question, and gets the correct answer to enable them to continue with their activity securely.

For the enterprise, scenario 2 is without doubt preferable; however, security teams have finite resources and are frequently not able to give timely responses to every question they receive. Answers are also subject to human errors and biases. An additional concern is that employees report being reluctant to reach out to security professionals due to a fear of having to ask what may be perceived as a basic question that they should know and are also concerned there may be repercussions if they may have done something wrong, and so stay quiet.

2.1 Research Question

Given the challenges outlined in the introduction, and with recent developments in Artificial Intelligence (AI), Natural Language Processing (NLP) and Chatbot tools and techniques, this research will focus on the potential for an NLP solution to help reduce the barriers and bridge the gap between security policy and the employees within an enterprise.

In addition, since there is not a one size fits all model for Security in an enterprise, and the policies of one enterprise differ from those of another; the main aim of this research is to identify if there is potential for the use of an AI and NLP system that would enable an enterprise to customise the system to the extent that it would align with their own policies and enterprise requirements.

The research question that this paper will seek to address is to "Evaluate the use of Artificial Intelligence (AI) and Natural Language Processing (NLP) to bridge the gap between security policies and employees in large enterprises."

2.2 Research Objectives

To answer the overall research question, the following research objectives were derived:

Objective one: Identify the needs case. To identify a potential solution to the barriers of effective security policy adoption in large enterprises, we must first understand what the challenges are. Only when we have a clear understanding of the challenges, can we design a solution to mitigate them.

Objective two: Investigate the start of the art of Artificial Intelligence, Natural Language Processing (NLP) and chat bots and determine a target architecture - To arrive at a configuration plan for the system, a full review of research and current opportunities for AI, NLP and Chatbots was carried out. The aim of this investigation was to determine the most appropriate architecture and components for the system to support the speciality focus on security policies.

Objective three: Design and implement the system in alignment with the architecture and components identified in objective one - Once an overall architecture was identified as the output of Objective one, the system was designed in alignment with the target architecture, ensuring compatibility between each of the modules and versions required to ensure a fully functioning system. The system was built and accessible for training.

Objective four: Train the system on a specific library of security policies - Following successful build of the system it needed to be loaded with the library of security policies and trained to be able to respond effectively to a suite of test cases designed to test the efficacy of the system.

Objective five: Evaluate the accuracy of answers provided by the Chatbot to determine if the system is feasible and beneficial for large enterprises - With the system built and trained on security policies, a suite of test cases is documented – with an input (question) and the expected output (answer). Executing the test cases will involve entering a test question and determining if the system returns the expected answer partially, fully, or not at all.

2.3 Criteria for success

Criteria one: Documentation of a needs case - a set of problem statements which are identified as the barriers to policy adoption in large enterprises which will be used to shape the system designed in Criteria two – (Objective one)

Criteria two: Determine an appropriate architecture for the system, based on the state of the art of NLP and Chatbots – as assessed by Objective two. This will be evidenced by an architecture schematic of the system.

Criteria three: Build and implement the system, in alignment with the architecture schematic proposed. This will be evidenced by the availability of the system, to use as detailed in the configuration manual. (Objective three)

Criteria four: Execution of documented test cases on the system to determine if the responses provided by the system are in alignment with the expected answers which would be provided by a security professional. This will be evidenced by the percentage accuracy and reliability of the system. (Objective four and five)

2.4 Structure of this report

This report is structured into eight further sections. A review of related research and work on the topic is discussed and evaluated in section three, along with detail of the research methodology which has been followed in section four and the design specification in section five. Once the system has been built, we will move on to discussing the approaches taken to implementation in section six, before evaluating the implementation in section seven, and finally, section eight contains conclusions and recommendations for future work. A configuration manual is also available and submitted to accompany this final report.

3 Related Work

In 2021, Uchendu et al, in their systematic review of the previous 10 years of research on security culture, identified that the top two success factors in building and maintaining a cyber security culture are having top management support, leadership or involvement, and security policies. [1]

The SANS Institute defines a security policy as a formal document that *"establishes what must be done to protect information stored on computers. A well-written policy contains sufficient definition of "what" to do so that the "how" can be identified and measured or evaluated…."*. Whereas the National Institute of Standards and Technology (NIST) defines a security policy as *"Aggregate of directives, regulations, rules, and practices that prescribes how an organization manages, protects, and distributes information"*.

What is clear from both definitions, is that security policies represent the top tier of security direction within an enterprise, setting the "tone from the top" as to how the enterprise must act to discharge its security responsibilities and comply with relevant legal and regulatory requirements. When employees don't receive this direction, the security culture and therefore security posture suffers, with an estimated 10 to 20% lower security culture (and therefore lower overall security posture) when employees have not read security policies. [2], [3], [4].

Much research has been carried out to identify the challenges associated with poor levels of compliance with security policies. To identify the major findings, and how they can be resolved, an academic literature review was completed covering over 30 research papers and articles.

Following the review, identified factors for security policy success were categorised into four key areas: culture, awareness and training, clarity of ask and applicability and timelines. The overview of the related work is summarised by factor in the following table:

Factor	Indicators of Success	Source (s)
Culture	The enterprise has a positive security culture which is embedded in working practices from the highest level of leadership to entry level employees. Negligent or malicious behaviour is monitored, identified, and punished, and good behaviour should be rewarded. The use of positive motivation, such as job promotion, financial, recognition for team or individuals for identifying near misses or implementing improvements.	[4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [1]
Awareness and training	Policies are clearly communicated and disseminated throughout the enterprise. All new employees are briefed on and educated on the security policies of the enterprise and their roles and responsibilities related to security. Employees are regularly trained as threat landscapes change and as the operating environment of the enterprise changes to maintain currency and applicability.	[5], [7], [9], [19], [20], [21], [22], [13], [16], [1], [23], [24]
Clarity of ask	Policies are well designed and clearly articulate the requirements which employees must and must not do. Policies require or describe security mechanisms which are available, appropriate, useable, and easy to understand and translate into actions by a non-security professional.	[9], [19], [25], [26], [13], [27]
Applicability and timeliness	Policies are reviewed and updated regularly, and tailored to suit the enterprise, its threat profile and legal or regulatory requirements: providing a single source of truth for required behaviours. Responses to questions are received in a timely manner to avoid employees resorting to shadow IT solutions.	[9], [21], [27], [26]

Table 1: Indicators of success for information security policies

Culture and Awareness and Training are well documented factors that contribute to a successful implementation of security policies.

However, research into clarity of ask, and applicability and timeliness of security policies is less mature. Of the 29 research papers reviewed, just 7 of them considered these as significant factors in the success of security policies. Could this be a contributing factor as to why security policies are poorly understood and adopted – these factors are overlooked and not well implemented in enterprises.

Many researchers have pointed out that a clear clarity of ask leads to a better understanding of actions that a person needs to take, this is not limited to security policies, but is a general indicator of the likelihood the end user (in this case the employee) can clearly understand what is expected of them.

Accessibility has also been proven to foster engagement. That is to say that when policies are written in simple English language, they are easier to understand, and therefore people are more likely to read them. The trade off between having a thoroughly secure policy which meets security requirements, but the average employee cannot understand is discussed by Faily and Flechais in their 2010 study on Security through usability. This study also argued that involving the employees early in the policy development process also yielded a more usable security policy. [28]

Reduced ambiguity is also a contributing factor in increasing the clarity of ask – ensuring the policies are clear and concise in their contents, with no ambiguous language – reduces the potential for likelihood of different interpretations, and ultimately non-compliance. This is not an issue restricted to security policies, and the same has been found to be the case for other disciplines where policies and complex documentation is used such as the scientific community, accountancy and legal. Hotaling in 2020 outlined 10 rules for scientific writing. This approach interestingly includes (amongst others) suggestions aligned to those required for fine tuning of data for ingestion into AI, such as removal of unnecessary words and simplify your language. [29]

The impact of poor adoption of timeliness (that is to say, the time to get an answer to a security policy question) is discussed by Kirlappos et al (2015), where they posture that where employees are unable to gain quick, simple solutions to security dilemmas, they resort to the use of shadow IT solutions to fulfil their need. These shadow IT solutions are by their nature, outside of the reach of the enterprise and are therefore highly likely to be non-compliant with the enterprises' security policies. [26]

Furnell et al (2018) propose that there are two ways to support users to enhance security behaviours – passive and active. Passive support usually takes the form of provision of upfront guidance or instruction (e.g a policy) which is expected to steer the user toward the desired behaviours, whilst active support utilises psychology, human-computer interaction and behaviour economics to actively communicate with and guide users during the course of their action. [13]

Other research has identified that having real-time applicable guidance to a user at the point of action, can contribute significantly to the security of the decision-making process by the user. For example, the incorporation of real-time information on suspicious domain names was identified as being an effective deterrent in preventing users from visiting phishing sites. [30]

In summary, research has identified that to be successfully implemented, policies need to be accessible and clearly worded in simple English that is capable of being understood by employees of all levels of technical knowledge and is best delivered at the point of decision for maximum benefit.

This research project will now determine if it is possible to create a Chatbot to be at the point of decision, armed with the correct security policy information, and able to respond accurately to users' security questions.

4 Research Methodology

At the start of the project, I identified the high-level components which needed to be completed and defined a seven-phase approach. Namely, the needs case, system design and build, data collection, data pre-processing, model training, fine tuning and testing and evaluation as shown below:

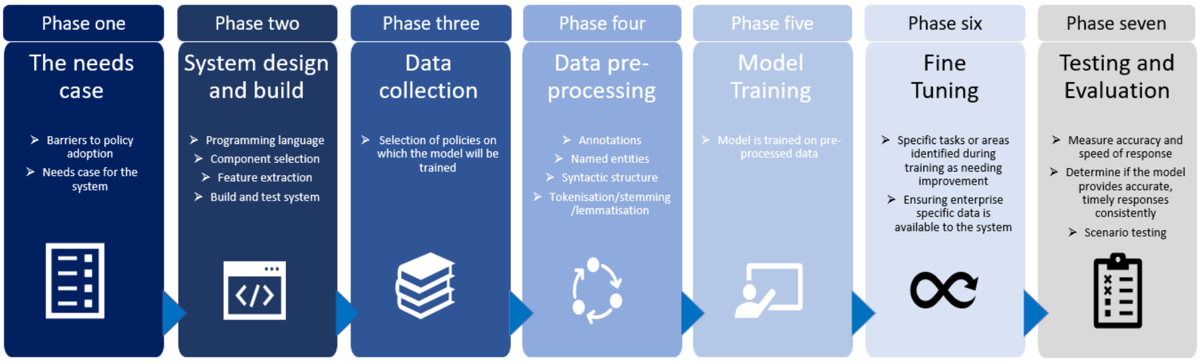
Phase one
The needs case
Barriers to policy adoption
Needs case for the system
Phase two
System design and build
Programming language
Component selection
Feature extraction
Build and test system
Phase three
Data collection
Selection of policies on which the model will be trained
Phase four
Data pre-processing
Annotations
Named entities
Syntactic structure
Tokenisation/stemming /lemmatisation
Phase five
Model Training
Model is trained on pre-processed data
Phase six
Fine Tuning
Specific tasks or areas identified during training as needing improvement
Ensuring enterprise specific data is available to the system
Phase seven
Testing and Evaluation
Measure accuracy and speed of response
Determine if the model provides accurate, timely responses consistently
Scenario testing

Figure 2: Research methodology

4.1 Phase one - the needs case

Following on from the review of related work, as outlined in section 3, it was determined that there has already been significant research and recommendations for best practices into the factors of culture and awareness and training, that I would focus on the less well researched factors – the clarity of ask, and applicability and timeliness.

Researching these factors in further detail it became apparent that there may be an opportunity to take advantage of recent advances in AI and NLP to identify potential ways to deliver indicators of good practice to enterprises.

A questionnaire was developed for professionals to identify the appetite for a potential use of AI and NLP in the future to bridge the gap between security policies and employees.

The results of the questionnaire (n=30) identified that there was appetite for an AI, NLP Chatbot, with 93% of respondents stating they would be at least somewhat comfortable in using the PolicyPal Chatbot. Only 2% of respondents signalled they would be somewhat uncomfortable. The perceived benefits and potential negatives of using a Chatbot are summarised in the figure below:

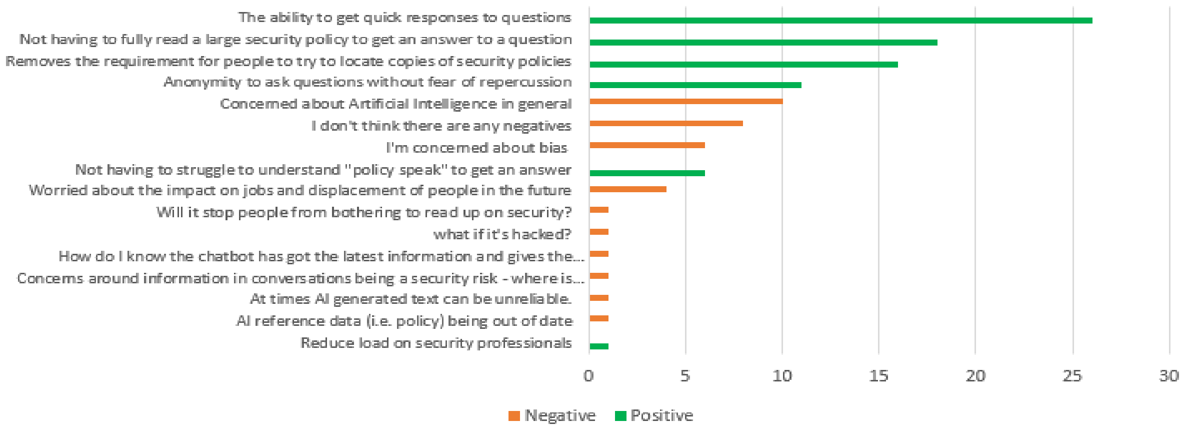

Figure 3: positive and negative aspects of a security policy focused Chatbot

4.2 Phase two - system design and build

During this phase, the various elements of the system were researched to identify a suitable design that could be created to help answer the research question and determine if there is a valid use case for NLP to bridge the gap between security policies and employees; ***Objective three***.

Three distinct areas were researched – what programming language is best suited to the application (bearing in mind what is available to the project, since programming skills are limited), what components are needed to support the system, and finally, the features (the type of data e.g word documents, pdf, databases, or spreadsheets) that need to be supported as part of the system.

The first step was to select the programming language that I was going to use. My choices in this space were somewhat limited due to my programming experience, however I have previously studied both Java and Python applications. The pros and cons of each language were identified, along with the suitability of the language for the use case. On reviewing both languages, I determined that Python was going to be more appropriate since it is more commonly suited to Artificial Intelligence (AI) and Machine Learning applications, is very quick to develop and has a large library of components available.

Following the selection of the programming language, the components of the Python application, and the respective version need to be selected to ensure intra compatibility with other components of the application. The first component to be selected was the Natural Language Processing model. During the review of related work, I identified that a Hybrid generative type model was going to best suit the use case for security policies. An overview of the different NLP models is shown in the figure below.

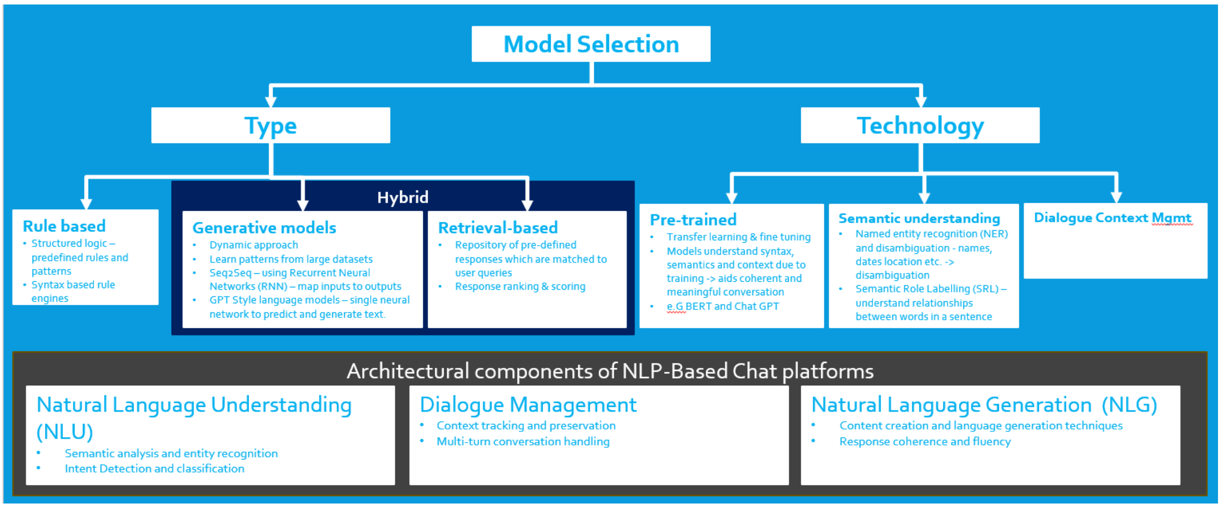
Model Selection
Type
Technology
Rule based
• Structured logic – predefined rules and patterns
• Syntax based rule engines
Hybrid
Generative models
• Dynamic approach
• Learn patterns from large datasets
• Seq2Seq – using Recurrent Neural Networks (RNN) – map inputs to outputs
• GPT Style language models – single neural network to predict and generate text.
Retrieval-based
• Repository of pre-defined responses which are matched to user queries
• Response ranking & scoring
Pre-trained
• Transfer learning & fine tuning
• Models understand syntax, semantics and context due to training -> aids coherent and meaningful conversation
• e.G BERT and Chat GPT
Semantic understanding
• Named entity recognition (NER) and disambiguation - names, dates location etc. -> disambiguation
• Semantic Role Labelling (SRL) – understand relationships between words in a sentence
Dialogue Context Mgmt
Architectural components of NLP-Based Chat platforms
Natural Language Understanding (NLU)
• Semantic analysis and entity recognition
• Intent Detection and classification
Dialogue Management
• Context tracking and preservation
• Multi-turn conversation handling
Natural Language Generation (NLG)
• Content creation and language generation techniques
• Response coherence and fluency

Figure 4: Model Selection [31], [32]

NLP models use techniques like Word Embeddings (e.g., Word2Vec, GloVe) to represent words as numerical vectors. These vectors capture semantic relationships between words, which is crucial for understanding context. Security policies are typically either in Word or Adobe PDF format and therefore I selected docx2txt and PyPDF2 to parse and extract text from word and adobe documents.

During this phase, the Python application was assembled and tested. Despite having documented a set of system requirements earlier in the phase, compatibility issues were encountered during system build. Within Python, there are excellent compatibility features, with compatibility being checked when each component is installed. This did lead to some back and forth with specific versions of components, and a number having to be rolled back to previous versions due to error messages received during installation, before arriving at a stable system which is documented in section 5.

4.2 Phase three - data collection

To train the PolicyPal on security policies, a selection of ISO 27001 template policies was used, where were in a combination of Word and Adobe PDF formats.

4.3 Phase four - data pre-processing

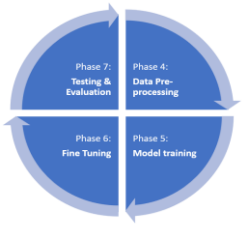

Figure 5: Continuous Improvement

The next four phases ran on an iterative cycle, to yield a continuous improvement of the responses of the PolicyPal Chatbot.

During each phase, the performance of the PolicyPal Chatbot was monitored and evaluated to determine if further improvements were needed to the data collection to improve the quality of the results to questions posed to the Chatbot.

Tabassum and Patil (2020), and Kannan and Gurusamy (2014) identified the use of text preprocessing techniques as a major factor in the quality and accuracy of responses provided by a Chatbot, with the most efficient techniques being tokenisation (breaking text into individual words of tokens), Stop Words removal (commonly used words such as "a", "the", "are"), punctuation removal and lemmatization (grouping words together so they can be analysed as one). [33], [34]

4.4 Phase five - model Training

Following data pre-processing, the model was trained on the data. With the design of the PolicyPal, data modelling is executed each time the application is launched by GPT Index and Langchain.

4.5 Phase six - fine tuning

Once the PolicyPal was trained on the source data, the next phase was to start fine tuning the data, to enable better quality answers. To determine what could be improved on, (via data pre-processing techniques discussed in 4.4) 25 test cases were created, which were designed to test the accuracy of the PolicyPal response, based on the data on which it was trained.

4.6 Phase seven - testing and evaluation

The final stage involved repeated testing and evaluation of question sets and determining the accuracy of responses from the Chatbot.

Before fine tuning, only 24% of test cases were answered satisfactorily by the Chatbot, with 48% being somewhat satisfactorily answered, and 28% not being answered satisfactorily at all. By combining iterative loops over Phase four to seven, I was able to deliver 72% of test cases to meet the expected answer, with a further 28% being partially successful. I observed a significant increase in the quality of answers (increase of 25% of answers meeting expectations) by implementing Named Entity Recognition (NER). [33]

At the end of the final cycle (phases four to seven), there were no test cases failed to deliver at least some response to the questions posed.

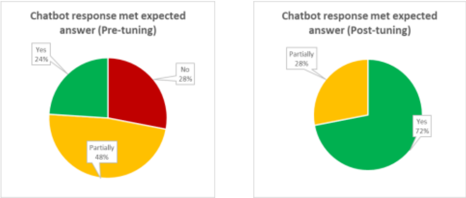

Figure 6: success of chatbot response, pre and post tuning

5 Design Specification

This section details the design of the PolicyPal which was implemented.

5.1 System architecture

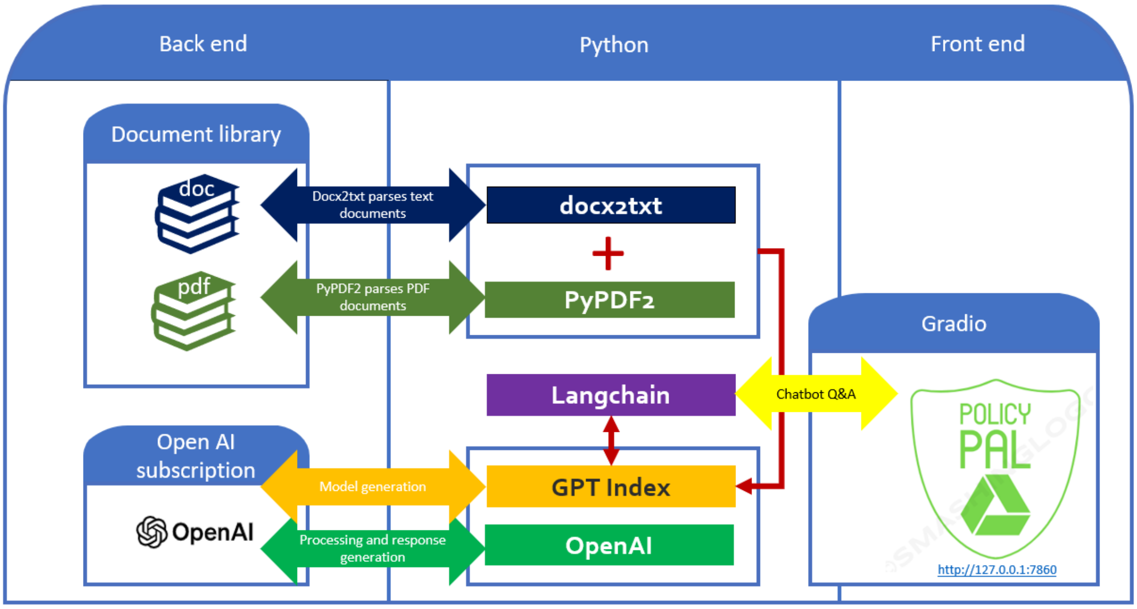

Figure 7: System architecture

5.2 Python and python components

Component	Usage	Version
Python	Programming language	3.11.2
Langchain	Large language model Integration Framework. Carries out the work of processing and responding of user queries.	0.0.148
Docx2txt	Enables the parsing of word documents to extract text and build the language model in GPT Index	0.8
PyPDF2	Enables the parsing of PDF documents to extract text and build the language model in GPT Index	3.0.1
PyCryptodome	Required to support the usage of PyPDF2	3.19.0
GPT Index	GPT Index, also known as Llama Index ingests data provided to categorise and sort text. Langchain interfaces with GPT Index to coordinate questions and answers to the end user via the Gradio front end	0.4.24
OpenAI	The provider of GPT Index. All data is funnelled via Open AI for processing and training. The OpenAI component within Python enables the utilisation of GPT Index and manages the API connections required to utilise the service.	3.8.6
Gradio	Enables the build and sharing of the Chatbot interface via a locally hosted URL. Provides the front end for the end user.	3.36.1

5.3 OpenAI Account and subscription

The application is configured to connect via API to OpenAI where the model generation takes place. To access this, a subscription is required to OpenAI. The system is linked to my personal account, and no additional accounts or subscriptions are needed for the end user.

Within the Open AI platform, usage statistics are available to monitor the usage, in terms of general activity on the platform and the financial impact of the activity.

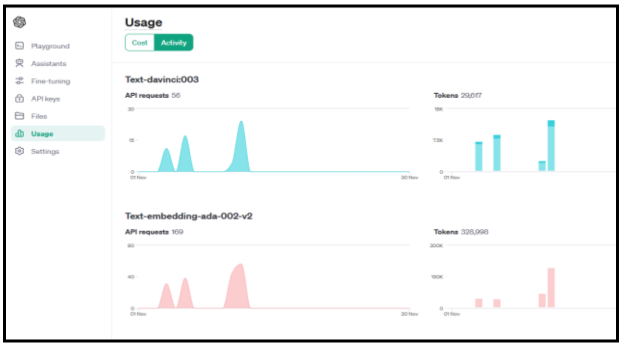

Figure 8: OpenAI activity usage data

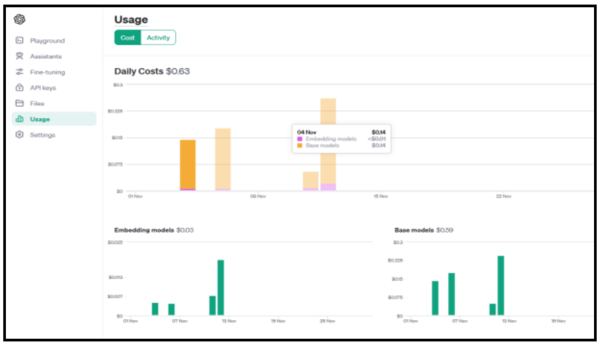

Figure 9: OpenAI financial usage data

6 Implementation

Questionnaire

Created: Output of phase one – the needs case

Description: A Microsoft forms questionnaire containing a mix of open ended and closed ended questions to identify attitudes and thoughts on security policies along with opinions regarding the use of AI powered Chatbots in the future.

Design artefacts: Microsoft forms questionnaire.

User(s): Researcher

System Architecture

Created: Output of phase two – system design and build.

Description: During Phase 2 - system design and build**,** the system architecture that was required to support the needs case for the system was identified, documented, and tested to ensure inter-operability between components. The final design was then documented, detailing how each component of the system interacts with each other. This is available in section 6.1.

Design artefacts: System architecture schematic drawing.

User(s): The system administrator for initial system build, and support post release.

Python Application

Created: Output of phase two – system design and build.

Description: The application is written in Python, with an app.py file containing the required information to run the system.

Design artefacts: (1) App.Py file which includes the code required to run the application.

(2) Executable Python application installer for end users.

User: (1) The application administrator for initial system build, and future changes.

(2) End users of the application.

Library of Policies

Created: Output of phase three – data collection, updated during phase four - data pre-processing and phase five – model training and phase six – fine tuning.

Description: These are the documents that have been used to train the Chatbot.

Design artefacts: (1) A collection of word and excel base documents which contain the policy documents on which the Chatbot has been trained.

(2) A collection of documents assembled during Phase four – data pre-processing.

User: The Python application. Each time the python application is launched, the documents are ingested by the components of the system, specifically GPT Index, OpenAI, Langchain, docx2txt, PyPDF2 and Gradio.

Test Cases

Created: Output of phase

Description: 25 documented test cases to test the responsiveness and effectiveness of the Chatbot. A spreadsheet detailing the test questions which are to be used to test and measure the efficacy of the Chatbot. Each question has a desired outcome, which is measured against the response given by the Chatbot.

Design artefacts: Test case file, with test results split into two categories – pre and post data processing and model training.

User: The system administrator for testing the application.

7 Evaluation

Both the questionnaire and research into the state-of-the-art support the fact that security policies, whilst critical to an enterprise, are largely not sufficiently understood, communicated, implemented, or maintained in a timely manner, with the main reasons being attributed to four key factors: Culture, Awareness and training, clarity of ask and applicability and timeliness.

As a result of the execution of ***objective one*** (needs case) and ***objective two*** (identifying the state of the art) two key findings have emerged. Firstly, the data gathered during research phase supports the fact that there is a use case for AI/NLP for security policies in large enterprises, with 70% of questionnaire respondents indicating positive overall attitude towards the use of AI/NLP. The implementation of a security focused Chatbot, has the potential to assist an organisation with improving the outcomes of the two identified factors which were focused on: the clarity of ask and applicability and timeliness of security policies; by delivering on the positive benefits which end users expect to yield from a system such as a Chatbot.

Objective three (design and implement) was initially very successful. A full design was established, documented, and tested and the success criteria were met as evidenced in section five and seven of this report. However, during the writing of the evaluation and conclusions, the risk of the pace of change within the AI and NLP realms, and the lack of control over deprecations and unsupported components experienced by a Chatbot developer has come to light. The PolicyPal Chatbot was designed and built in late 2023 with OpenAI 3.8.6 which utilised the text-davinci-002 model. At the time of writing (April 2024) the OpenAI text-davinici-002 model which was used in the build of the PolicyPal is now deprecated, and the PolicyPal Chatbot is unfortunately no longer functioning as it was. [35]

This does raise a concern over the maintenance and supportability of models such as the PolicyPal. Enterprises may have a considerable overhead in running and maintaining models such as this, until such time as they become ubiquitous and are available integrated within our enterprise tools such as procurement or IT Service Management. In addition, the end user has very little control over how and when models are deprecated, when used in a "model as a service" mode from developers such as OpenAI.

Objective four and five (train the system and evaluate the accuracy of results) were assessed together. The results of training and evaluation of answers evidence the importance of the language, structure, and formatting of policies for the Chatbot to be trained on. Policies and standards need to be written with AI in mind, (reference before and after fine-tuning stats) to ensure the correct information is available within the data provided to the Chatbot. It's not enough to simply collate the library, documents need to be written to reflect and enhance the methodology language models use to index and catalogue data. This was evidenced in the positive impact (a 48% increase in the numbers of questions accurately answered by the Chatbot) realised during the iterations of phases four to six during design. Continuous tailoring, rewording, restructuring and fine tuning of the data collection was needed to yield effective responses from the Chatbot.

On reflection, if the resources had been available to support it, my research should have included a pilot phase within an enterprise, with the PolicyPal having been trained on their policies to determine if there is a realised benefit from within the enterprise. Testing a capability which relies on user behaviours and experience, is vital to be tested hands on by a representative sample of users. This was not possible within the time scale of this research but would be recommended as a next step for future work.

8 Conclusion and Future Work

The research question that this paper sought to address is to "Evaluate the use of Artificial Intelligence (AI) and Natural Language Processing (NLP) to bridge the gap between security policies and employees in large enterprises."

Five objectives were established at the start of the research, which were all successfully delivered during the research; however, it has not been possible to fully evidence Objective Three, Four and Five as part of the presentation of the research due to the recent deprecation of the OpenAI model (text-davinci-002) as referenced in section 7 of this report.

In summary, the research was able to identify, build and implement a successful model for implementation of a Security focused Chatbot, including the data collection, training and fine tuning of the content to enable the Chatbot to provide expected quality of answers which could be of use to an employee querying security policies.

By having security policies accessible from within a Chatbot, the factors that were previously analysed can be improved within an organisation. Namely, the Clarity of ask – can be enhanced by having questions returned to employees in simple English language, rather than sometimes convoluted "security speak". In addition, having to train the Chatbot, on specific areas such as entities (e.g names, contact numbers, responsibilities) focuses the trainer on ensuring the data collected is accurate and relevant for the enterprise.

Using technology and a Chatbot to provide the front end to security policies reduce the time taken to produce, share, communicate and disseminate changes (new or updated policies) with employees. Since employees will only need to visit one location to seek their answers, there is less likelihood of the wrong policy being reviewed, it not being up to date, or the employee looking at the incorrect policy. There is also less dependence on the members of the security team to provide responses, so responses are provided at the time the employee needs it, (to enable them to continue on with their normal day to day activities), rather than when the security employee is available to answer their query.

Future work

Following on from the recent deprecation of the OpenAI model, the first thing that must be done is to carry out an upgrade of the components of the PolicyPal Chatbot. The key to a successful implementation of any new product or feature, is making it easy for an end user to find and use. The next logical steps for this project are to integrate it with enterprise communication tools such as Microsoft Teams or Slack.

Further improvements could be made to facilitate continuous learning and improvement: The system should be designed for ongoing learning. Feedback loops should be established to gather user feedback and update the model for better performance over time. E.g Add a “rate this answer” feature so we can continuously monitor the helpfulness of the answers and address any gaps identified where additional training or fine tuning of the model is needed.

Integration with enterprise ticket management systems such as ServiceNow or Maximo to enable users to seamlessly raise an incident ticket where necessary, (following on from their conversation with the Chatbot) and potentially for the Chatbot to be able to identify when there is an incident in progress which may be causing unusual user experience within the enterprise.

Finally, the concept of a PolicyPal Chatbot is not limited to security policies – further work could consider expanding the use to general company policies, not just security, providing a single source for all employee questions within a company, which would contribute greatly to standardisation within companies.

9 References

[1] B. Uchendu, J. R. C. Nurse, M. Bada, and S. Furnell, ‘Developing a cyber security culture: Current practices and future needs’, *Computers & Security*, vol. 109, p. 102387, Oct. 2021, doi: 10.1016/j.cose.2021.102387.

[2] V. A. Da, ‘Comparing the information security culture of employees who had read the information security policy and those who had not: Illustrated through an empirical study’, *Information & Computer Security*, vol. 24, no. 2, pp. 139–151, Jan. 2016, doi: 10.1108/ICS-12-2015-0048.

[3] E. Sherif, S. Furnell, and N. Clarke, ‘An Identification of Variables Influencing the Establishment of Information Security Culture’, in *Human Aspects of Information Security, Privacy, and Trust*, T. Tryfonas and I. Askoxylakis, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 436–448. doi: 10.1007/978-3-319-20376-8_39.

[4] M. Siponen, M. A. Mahmood, and S. Pahnila, ‘Technical opinionAre employees putting your company at risk by not following information security policies?’, *Commun. ACM*, vol. 52, no. 12, pp. 145–147, Dec. 2009, doi: 10.1145/1610252.1610289.

[5] B. Bulgurcu, H. Cavusoglu, and I. Benbasat, ‘Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness’, *MIS Quarterly*, vol. 34, no. 3, pp. 523–548, 2010, doi: 10.2307/25750690.

[6] K. Parsons, A. McCormac, M. Butavicius, and L. Ferguson, ‘Human Factors and Information Security: Individual, Culture and Security Environment’.

[7] ‘The European Network and Information Security Agency (ENISA), "The new users ’ guide : How to raise information security awareness’. Accessed: Apr. 23, 2024. [Online]. Available: https://www.enisa.europa.eu/publications/archive/copy_of_new-users-guide

[8] ‘A framework and assessment instrument for information security culture - ScienceDirect’. Accessed: Apr. 23, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404809000923

[9] M. Alotaibi, S. Furnell, and N. Clarke, ‘Information security policies: A review of challenges and influencing factors’, in *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, Dec. 2016, pp. 352–358. doi: 10.1109/ICITST.2016.7856729.

[10] ‘(PDF) SECURITY CULTURE: PRESENT AND FUTURE STRATEGIES, POLICIES AND CHARACTERISTICS’. Accessed: Nov. 19, 2023. [Online]. Available: https://www.researchgate.net/publication/374228177_SECURITY_CULTURE_PRESENT_AND_FUTURE_STRATEGIES_POLICIES_AND_CHARACTERISTICS?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InNpZ251cCIsInBhZ2UiOiJzZWFyY2giLCJwb3NpdGlvbiI6InBhZ2VIZWFkZXIifX0

[11] K. Padayachee, ‘Taxonomy of compliant information security behavior’, *Computers & Security*, vol. 31, no. 5, pp. 673–680, Jul. 2012, doi: 10.1016/j.cose.2012.04.004.

[12] S. Furnell and K.-L. Thomson, ‘From culture to disobedience: Recognising the varying user acceptance of IT security’, *Computer Fraud & Security*, vol. 2009, no. 2, pp. 5–10, Feb. 2009, doi: 10.1016/S1361-3723(09)70019-3.

[13] S. Furnell, W. Khern-am-nuai, R. Esmael, W. Yang, and N. Li, ‘Enhancing security behaviour by supporting the user’, *Computers & Security*, vol. 75, pp. 1–9, Jun. 2018, doi: 10.1016/j.cose.2018.01.016.

[14] X. Chen, D. Wu, L. Chen, and J. K. L. Teng, ‘Sanction severity and employees’ information security policy compliance: Investigating mediating, moderating, and control variables’, *Information & Management*, vol. 55, no. 8, pp. 1049–1060, Dec. 2018, doi: 10.1016/j.im.2018.05.011.

[15] Y. Chen, W. Xia, and K. Cousins, ‘Voluntary and instrumental information security policy compliance: an integrated view of prosocial motivation, self-regulation and deterrence’, *Computers & Security*, vol. 113, p. 102568, Feb. 2022, doi: 10.1016/j.cose.2021.102568.

[16] J. D’Arcy and G. Greene, ‘Security culture and the employment relationship as drivers of employees’ security compliance’, *Information Management & Computer Security*, vol. 22, no. 5, pp. 474–489, Jan. 2014, doi: 10.1108/IMCS-08-2013-0057.

[17] T. Herath and R. Rao, ‘Protection motivation and deterrence: A framework for security policy compliance in organisations’, *EJIS*, vol. 18, pp. 106–125, Apr. 2009, doi: 10.1057/ejis.2009.6.

[18] Q. Hu, T. Dinev, P. Hart, and D. Cooke, ‘Managing Employee Compliance with Information Security Policies: The Critical Role of Top Management and Organizational Culture*’, *Decision Sciences*, vol. 43, no. 4, pp. 615–660, Aug. 2012, doi: 10.1111/j.1540-5915.2012.00361.x.

[19] F. L. Greitzer *et al.*, ‘Unintentional Insider Threat: Contributing Factors, Observables, and Mitigation Strategies’, in *2014 47th Hawaii International Conference on System Sciences*, Waikoloa, HI: IEEE, Jan. 2014, pp. 2025–2034. doi: 10.1109/HICSS.2014.256.

[20] H. Chan and S. Mubarak, ‘Significance of Information Security Awareness in the Higher Education Sector’, *IJCA*, vol. 60, no. 10, pp. 23–31, Dec. 2012, doi: 10.5120/9729-4202.

[21] N. S. Sulaiman, M. A. Fauzi, W. Wider, J. Rajadurai, S. Hussain, and S. A. Harun, ‘Cyber–Information Security Compliance and Violation Behaviour in Organisations: A Systematic Review’, *Social Sciences*, vol. 11, no. 9, Art. no. 9, Sep. 2022, doi: 10.3390/socsci11090386.

[22] R. Torten, C. Reaiche, and S. Boyle, ‘The impact of security Awareness on information technology professionals’ behavior’, *Computers & Security*, vol. 79, pp. 68–79, Nov. 2018, doi: 10.1016/j.cose.2018.08.007.

[23] M. Siponen, S. Pahnila, and M. A. Mahmood, ‘Compliance with Information Security Policies: An Empirical Investigation’, *Computer*, vol. 43, no. 2, pp. 64–71, Feb. 2010, doi: 10.1109/MC.2010.35.

[24] P. Ifinedo, ‘Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition’, *Information & Management*, vol. 51, no. 1, pp. 69–79, Jan. 2014, doi: 10.1016/j.im.2013.10.001.

[25] S. Pahnila, M. Siponen, and A. Mahmood, ‘Employees’ Behavior towards IS Security Policy Compliance’, in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, Waikoloa, HI: IEEE, Jan. 2007, pp. 156b–156b. doi: 10.1109/HICSS.2007.206.

[26] I. Kirlappos, S. Parkin, and M. A. Sasse, ‘“Shadow security” as a tool for the learning organization’, *SIGCAS Comput. Soc.*, vol. 45, no. 1, pp. 29–37, Feb. 2015, doi: 10.1145/2738210.2738216.

[27] H. Paananen, M. Lapke, and M. Siponen, ‘State of the art in information security policy development’, *Computers & Security*, vol. 88, p. 101608, Jan. 2020, doi: 10.1016/j.cose.2019.101608.

[28] S. Faily and I. Flechais, ‘Security through usability: a user-centered approach for balanced security policy requirements.’, Dec. 2010, Accessed: Apr. 25, 2024. [Online]. Available: https://rgu-repository.worktribe.com/output/1427760

[29] S. Hotaling, ‘Simple rules for concise scientific writing’, *Limnology and Oceanography*, Jun. 2020, doi: 10.1002/lol2.10165.

[30] A. Xiong, R. W. Proctor, W. Yang, and N. Li, ‘Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages?’, *Hum Factors*, vol. 59, no. 4, pp. 640–660, Jun. 2017, doi: 10.1177/0018720816684064.

[31] G. Caldarini, S. Jaf, and K. McGarry, ‘A Literature Survey of Recent Advances in Chatbots’, *Information*, vol. 13, no. 1, Art. no. 1, Jan. 2022, doi: 10.3390/info13010041.

[32] S. Chaurasia, S. Jain, H. O. Vishwkarma, and N. Singh, ‘Conversational AI Unleashed: A Comprehensive Review of NLP-Powered Chatbot Platforms’, vol. 7, no. 3, 2023.

[33] Tabassum, Ayisha and Patil, Dr Rajendra R, ‘A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing’, *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 702–718, Jul. 2021, doi: 10.1016/j.jksuci.2019.04.001.

[34] D. S. Kannan and V. Gurusamy, ‘Preprocessing Techniques for Text Mining’.

[35] ‘OpenAI Platform’. Accessed: Apr. 25, 2024. [Online]. Available: https://https://platform.openai.com/docs/deprecations