

# **Maximising Efficiency: GenAI Integration in Bank Internal Audit (P000120DS)**

**Industrial Supervisor**  
Madhujith Venkatakrishna

**Academic Supervisor**  
Dr. Ke Deng

## **Project Members**

Aaron Tam COSC2667 | S3961290

Shifali Shetty COSC2667 | S3956999

Yung Qi Chin COSC2667 | S3819792

Maryam Alhowaiti COSC2777 | S3856144

## **Acknowledgement**

We extend our heartfelt appreciation to the individuals whose unwavering guidance significantly contributed to the success of our graduation project. We are deeply indebted to Dr. Ke Deng, our Academic Supervisor, whose invaluable oversight and mentorship steered us through every phase of this journey. Our gratitude also goes to Madhujith Venkatakrishna, our Industry Supervisor, and the entire team at the financial institution whose pivotal contributions shaped our project's direction and led us to achieve remarkable results. Special recognition is also due to Xiaofei Xu and Man Hou, two PHD students whose expertise and support proved invaluable as they guided us through the intricacies of our project with their wealth of knowledge and experience in the field.

Lastly, we express our gratitude to all team members, whether their names were mentioned or not, whose collective expertise and unwavering dedication have not only enriched our project but also propelled Gemini's innovative journey in the banking industry forward.

## **Table of Content**

<b>Acknowledgement</b>	<b>1</b>
<b>Table of Content</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>3</b>
<b>Partner Organisation</b>	<b>4</b>
<b>Project Aim</b>	<b>4</b>
<b>Project Deliverable</b>	<b>4</b>
<b>2 Background</b>	<b>5</b>
<b>2.1 Audit Business Process of Partner Client</b>	<b>5</b>
<b>Audit Issue</b>	<b>6</b>
<b>Business Submission and Closure</b>	<b>6</b>
<b>Internal Audit Review</b>	<b>6</b>
<b>3 Methodology</b>	<b>8</b>
<b>4 Evaluation</b>	<b>12</b>
<b>Advantages and Disadvantages of the Solution</b>	<b>14</b>
<b>5 Discussion on the Limitations</b>	<b>14</b>
<b>6 Key Recommendations to the Partner</b>	<b>15</b>
<b>7 Conclusion</b>	<b>16</b>
<b>8 Appendix</b>	<b>17</b>
<b>8.1 References</b>	<b>17</b>
<b>8.1 Code and Project Management</b>	<b>18</b>
<b>8.1 Self Reflection</b>	<b>18</b>
<b>8.2 Images</b>	<b>22</b>

## **Abstract**

In the banking and finance sector, audit and compliance are crucial for maintaining regulatory standards and mitigating risks. Traditional manual auditing processes are often time-consuming and susceptible to human error. Our project aims to transform these processes by leveraging Generative AI (GenAI) to enhance efficiency and accuracy. This project, in collaboration with a leading multinational financial services company, focuses on developing a GenAI tool to streamline audit reviews. The tool evaluates audit action items against supporting evidence by analysing textual similarities using advanced machine learning techniques.

Our solution employs a pre-trained Sentence-BERT embedding model to compute cosine similarity scores between audit actions and corresponding evidence. By summarising extensive evidence data and pinpointing the most relevant sentences, our tool assists auditors in quickly identifying whether audit issues have been addressed effectively. The system also integrates Google Gemini for providing interpretative insights, aiding auditors in making informed decisions based on the AI-generated results.

The project's primary deliverables include a GenAI system that produces a general similarity score and highlights the top five evidence sentences with the highest contextual relevance to the action items. This approach not only accelerates the review process but also reduces the likelihood of oversight, ultimately contributing to improved audit efficiency and compliance in the banking sector.

## **1 Introduction**

In the banking and finance sector, ensuring audit and compliance is paramount. With stringent regulations in place, banks are obligated to conduct regular inspections of documents and accounts while adhering to various risk-mitigating standards (Ahmad, 2020). However, these processes often demand intensive activity, involving continuous monitoring, stringent controls, and meticulous inspections by auditors. Unfortunately, manual checks are not only time-consuming but also prone to human errors (Hassan, 2018). Our project aims to leverage the power of GenAI to revolutionise audit and compliance processes within the banking and finance sector. By harnessing GenAI-enabled automation, we seek to enhance the efficiency and accuracy of these critical procedures (Jain et al., 2019).

In the context of our partner organisation, this project aims to enhance auditors' capabilities in the audit review process with artificial intelligence providing insights and guidance based on the GenAI generated outcomes. We take as input the audit action item along with supporting evidence to verify whether the issue has been adequately addressed. We process these text samples, embed them using a pre-trained embedding model and try to derive insights on the contextual similarity of the two. The higher the similarity between the Action item and Evidence the more it means that the Evidence addresses the concerns/tasks raised in the Action. The insights from our project will empower auditors to proactively identify potential risks and formulate effective strategies for risk mitigation.

## Partner Organisation

Our partner organisation is a multinational financial institution headquartered in Australia. They offer a range of services from personal and consumer banking to business and corporate investment banking services. What makes them stand out amongst their peers is their understanding of the importance of technology in the financial space and the steps that they have taken to harness the powers of artificial intelligence, block chain and big data to digitally transform the way they do business.

Our partner leverages technology to help deliver the best customer experience that is safe, secure and easy to use. Our project with them is another example of how they want to use the latest technological developments such as GenAI to help improve their business and throughput.

## Project Aim

With this project we aim to achieve the following:

- Provide our partner organisation with a solution that can help streamline their audit review process and reduce the volume of time auditors spend manually reviewing evidence or policy documents.
- Provide quick, easy to understand results that can be comprehended by people regardless of their technical background.
- Use the embedded representations of the input to calculate the cosine similarity between the action and evidence.
- Use the embedded representations of the input to get the top sentences in evidence that has the highest cosine similarity with the Action item.
- Utilise a large language model to interpret the score, providing auditors with clear explanations of the results to aid in their decision-making process.

## Project Deliverable

The deliverable of our project is:

- First, we will be providing a single cosine similarity score of the Audit Action Item and Evidence. Here we get a general idea of how similar the Action item is in meaning to the entirety of the Evidence. This will help auditors get an overall picture when reviewing the issue.
- Second, we will provide the top 5 Evidence sentences that have a high cosine similarity score with the Action. These sentences have the strongest contextual relation with the Action. This narrows down an Auditors search as now they can look at just a few sentences that provides the most insight on whether the Action was completed or not. We will also provide supporting comments from Gemini about how the scores can be interpreted. It is expected that the human auditor uses these details to arrive at an informed decision or seek additional information from the Auditee.

## **2 Background**

Artificial Intelligence (AI) has emerged as a transformative force across industries and has the potential to reshape how businesses operate and make decisions. Recent research from Deloitte reveals that in the UK alone, 4 million people who have used generative AI have done so for work-related purposes (Deloitte, 2023). This widespread adoption and interest underscores growing significance of AI technologies in organisational context which in turn presents both opportunities and challenges for internal audit functions. For internal audit teams the rise of AI introduces two key considerations: firstly, the need to provide assurance over AI implementations to ensure reliability of the generated output, integrity, compliance with regulatory standards and ethical concerns. Secondly, there is a pressing need to leverage the potential of AI to innovate and evolve internal audit methodologies and practices. However, along with the potential benefits that AI can provide for auditing, it also comes with a 'frontier of new risks' for boards and audit committees. If they would like to leverage the benefits of AI, organisations must also actively invest in risk management frameworks and controls to mitigate potential threats.

In navigating the complexities of AI governance and strategy, Internal Audit functions must prioritise the establishment of robust policies, standards and guidelines. This will involve maintaining an AI inventory which is a catalogue or database of AI assets within the company, ensuring compliance with regulations as well as implementing effective risk management protocols. In the current landscape of AI adoption, Davenport & Ronanki (2018) emphasise the importance of focusing on AI in terms of business capabilities rather than solely on technological skills. This highlights the key benefits that AI can help with businesses in general which are automating business processes, gaining insights through data analysis and enhancing connections with consumers and workers. Other areas of application include deep learning and machine learning where ML can "assist in transaction classification with the scope of control function" (Zhang et al, 2020). This is useful to auditors because current ML algorithms can analyse historical data to identify patterns and trends associated with specific issues or risks. By learning from past cases, ML models can analyse audit findings and recommend actions based on the outcomes of similar cases. Natural Language Processing (NLP) techniques can be employed to extract relevant information from textual documents and also has the ability to process unstructured text information, systematic and automatic retrieval and review of documents and identification of high-risk cases that deviate from the target terms (Zhang et al, 2020). Measures such as Cosine Similarity, Euclidean distance and Manhattan distance can help analyse the outputs from embeddings and help us understand the similarities between the content of documents. Embracing these AI technologies can enhance the efficiency and effectiveness of audit and compliance processes however, businesses must be mindful of the complexities of AI governance and strategy.

### **2.1 Audit Business Process of Partner Client**

The process by which the Internal Audit team identifies, addresses and resolves issues can be broken down into the following stages.

## **Audit Issue**

The bank conducts multiple internal audits throughout the different sectors or business areas of the bank in a given year. Each year, a multitude of audits (~150+) are carried out. Each audit identifies different problems or areas in the organisation that could potentially violate compliance standards. It is of utmost importance that these gaps discovered are addressed and closed by the affected business area with priority. When an audit is conducted, and an issue is identified the Internal Audit team of the bank raises an Audit Issue Register Form. This form will contain a detailed description of the issue discovered, risk involved, priority level, team that addresses the issue, list of action items and a timeline as to when the action items need to be implemented. The issue register form will also contain supporting context such as policy/procedure and information about criteria and standards. These give us more insight into why the issue was raised and what is the expected standard. This form is then submitted to the responsible business team, and they will be tasked with addressing each action item within the given timeline.

## **Business Submission and Closure**

Once the form is submitted to the business team, they will go ahead with implementing the action items. Once they have addressed each item they file an Issue Closure Form. This form details the steps taken by the team to mitigate the identified risks and will also contain supporting evidence documents that prove that these changes were in fact implemented. The evidence documents could be of any form such as image screenshots, pdfs, spreadsheets or emails. If certain items are not addressed then the appropriate reasoning for that is also provided. Once the closure form is ready it will be submitted to the Internal Audit team for review.

## **Internal Audit Review**

In this stage of the process the auditors in the Internal audit team receive the issue closure form and now need to review its contents in order to determine if the identified issues have been resolved. They will go through each section of the form along with each support evidence document manually and determine whether the gaps have been closed. They have a fixed procedure that they follow when conducting each review, they have well defined criterias and guidelines that tells them what to look for when reviewing a closure form. Once the review is completed and the expected outcome has been reached then the audit issue is closed.

Currently the third stage of the process is extremely time consuming as each auditor needs to go through the closure form manually. This could take a while as each issue may have multiple action items along with multiple evidence documents that need to be read and mapped with the expected policy or standards. The information in the closure form may also differ on a case by case basis as it is filled by multiple different people who will have different styles of writing. Some of these could be straightforward in a simple language, while others may be complex with lots of jargon. Also since the review is being conducted by a person the whole process is at risk of potential human judgement. All these factors slow down the audit review process and this is a concerning problem especially in the case of high priority issues.

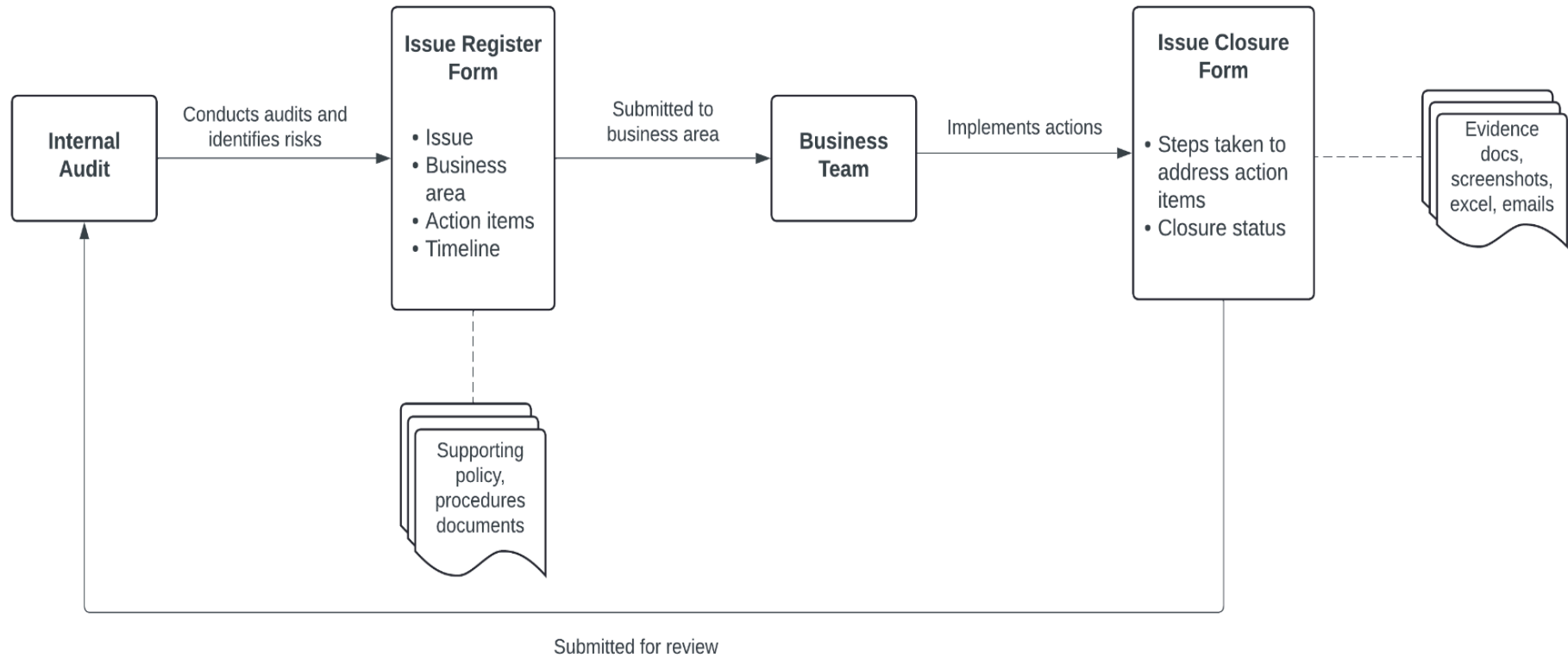


Figure 1. Flow diagram of a typical Financial Institution Audit process created by the authors in 2024



This is where the need for a software tool that can help the Internal Audit team with the review process come into play. Our partner organisation requires from us a GenAI tool that can essentially provide insight on whether an audit issue has been completed based on the evidence provided. This model can be used as a first line of action in the Internal audit review process, it can help segregate the issues and help auditors to direct their attention accordingly. For example, if the tool says,

- *Issue A* is not addressed then the auditor can immediately inform the concerned team instead of spending time reviewing it.
- *Issue A* needs more evidence then the auditor does not have to review it immediately and instead ask for more evidence from the business team.

The tool can also serve other /similar purposes.

- Help auditors select audit issues for sample testing based on risk.
- Help auditors select audit issues for sample testing based on risk.
- The tool can also be adopted by first line risk team / business teams themselves to verify /self-verify each action item's implementation.
- This technical feasibility test can also apply to audit fieldwork for testing a control attribute to evidence submitted.

### 3 Methodology

The primary objective of this project is to deduce whether the Audit Action item has been completed based on the Evidence provided by the business closure team. In order to do this our solution focuses strongly on deriving a quantifiable contextual similarity metric that would tell us how similar the meaning of the Action item is with the Evidence. The higher the similarity the more the content of the Action matches with the Evidence and the more reason we have to believe that the Action item was completed. Figure 2 gives an overview of our solution.

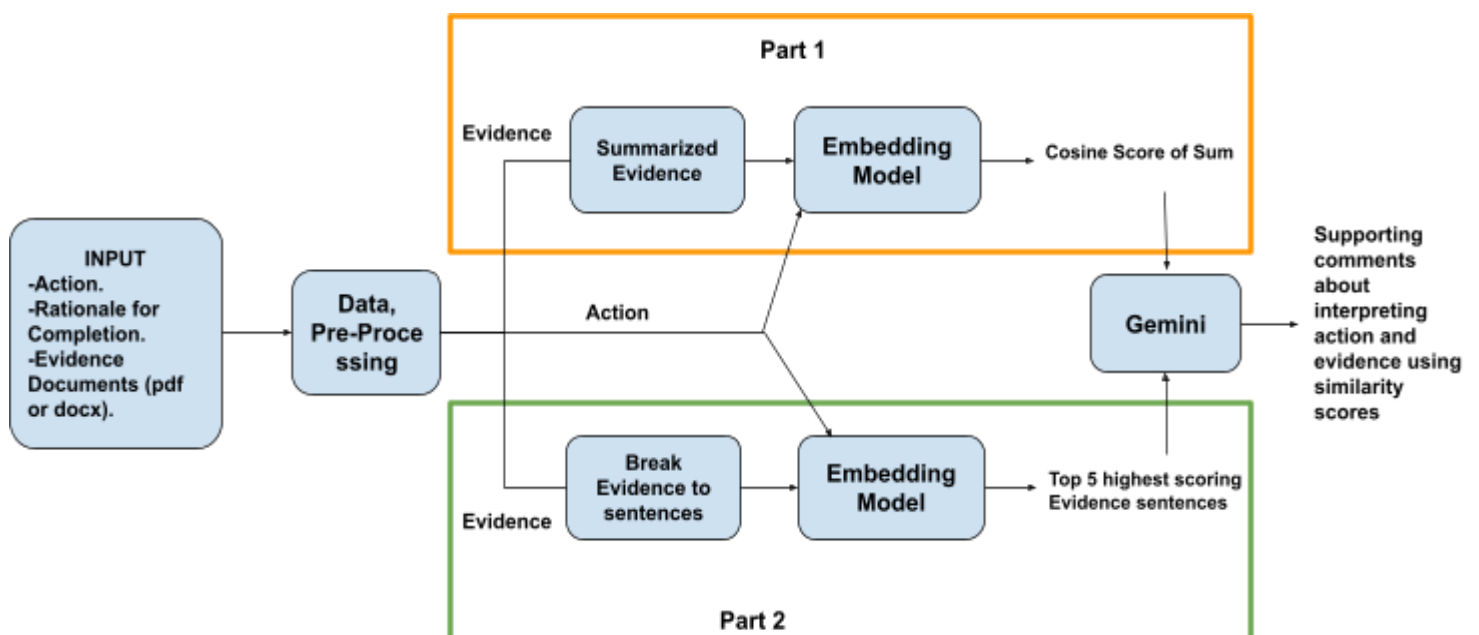


Figure 2. Project solution flow diagram created by authors in 2024

We will have three items that are taken as input through a webpage :

**Action** - Audit issue that needs to be addressed

**Rationale for Completion** - Statement of completion from Business team

**Evidence Documents** - Supporting evidence documents to support the Rationale for Completion.

Each of these items will be pre-processed to remove any junk or special characters. The Rationale for completion and text extracted from the evidence documents will be concatenated into a single evidence item. Once our input is ready we will proceed with generating the output. There are two parts to our output.

First, we will generate a cosine similarity score between the Action and Evidence item as a whole. This gives us an overall view of how the action relates to the evidence. To generate this we will first summarise the Evidence data to about a fraction of its whole content. This helps us to distil large volumes of text into concise and informative summaries, making it easier to quickly grasp the key points, insights, or trends contained within the data. The Action and Summarised Evidence will then be embedded using the SentenceBert embedding model. These embedded representations of the two will be used to calculate the cosine score.

Second, we will find the top 5 sentences in the Evidence that have a strong similarity with the Action. Human reviewers usually need to sift through large volumes of data in the Evidence to understand whether the Action has been met. By providing these top 5 sentences we greatly narrow down their search and give them recommendations of sentences to focus on when making their final decision. For this component of the output we will split the Evidence into sentences and embed each Action and Evidence sentence using SentenceBert. We will then calculate the cosine similarity score of Action against each evidence sentence and find sentences with the 5 highest scores.

The cosine scores along with the Action and Evidence will be fed to Gemini to help provide some comments on how the results can be interpreted. This is an additional resource to aid auditors in their decision-making process. In the next section we will go through each stage of our solution and the different decisions we took to arrive at our final solution which is a model that can provide valuable recommendations to support internal audit reviews

## **Step 1 : Data Extraction**

Our input consists of 3 items - Action, Rationale for Closure and Evidence Documents. The supported evidence document types are pdf and docx only. Based on the type of document submitted we use the appropriate method to extract the text from the files. The text from the evidence documents and the Rationale for Closure will be concatenated to form the Evidence data. The input data will be collected using a webpage form and used for further analysis.

## Step 2 : Data Pre-Processing

In this step we focused on the measures we would need to take to clean our data and ensure that unnecessary characters are not fed into the model. Our initial approach involved removing stop words (using nltk package), punctuations, numbers, Stemming and Lemmatization. Upon further discussion and feedback we concluded that we would not need to manually handle these cases as the pre-trained transformers model used is sophisticated enough to have its own method in place to clean and process data. We did include a separate function to be able to remove garbage/binary characters from evidence documents as these were prone to have special characters such as bullet points, page icons etc that are not handled well in text processing tasks.

## Step 3 : Evidence Summarisation

Since the evidence consists of data from multiple documents the volume of information may be very high. When calculating the score between the action and the whole of evidence it is important to ensure that you condense and focus on the most relevant and important content of the text. Since documents aren't often to the point we decided to summarise the Evidence before calculating cosine score.

We tried different summarisation models - summa, Gensim, BERT and T5. All of them gave us decent results but we decided to go with summa mainly because of the lightweight build and the flexibility it offered in customisations. We could easily define what percentage of the original content we wanted the summary to be. We believed this feature would come in handy if the auditors wanted to adjust this parameter based on the use case in the future. In the current solution evidence is summarised to 30% of its content. We arrived at this number after multiple tests where we summarised evidence to 20%, 30%, 40%, 50% and compared the cosine scores obtained. The scores were around the same range with maybe 0.01 decimal difference. If you see in Table 1 from 50% to 30% you see an increase in the score as the evidence is summarised to contain important points, but you see a dip at 20% which could be attributed to loss of information. Therefore we settled on 30% as this is sufficient to capture important information and not affect performance/latency of the solution.

Percentage Summarization	Cosine Similarity
50%	0.9108
40%	0.9153
30%	0.9259
20%	0.9014

Table 1. Similarity score of Action and Evidence used in section4 at different % summarization

## Step 4: Embedding Model

To compare how relevant evidence is to an action, we need to turn both action and evidence into vectors that represent meaningful semantic information. This is where the sentence embedding models comes into play. Sentence Embeddings that represent sentences as

numbers allow us to calculate cosine scores to measure the semantic similarity between the action and evidence.

We experimented with different sentence embedding models, including T5 and Sentence-BERT. In our specific case, Sentence-BERT performed the best and delivered the most effective results. Sentence-BERT is a model that is specifically designed for sentence embedding tasks. It transforms into numerical representations within a 768-dimensional space, and similar sentences will be close together in the vector space (Efimov, 2023). Other than that, it requires less computational resources than all other sentence embedding models such as T5 and consistently performs well on many sentence embedding tasks.

In this step, we used the Sentence-BERT model to generate embeddings for the given action and summarised evidence pair. In addition, we also divided the evidence text into individual sentences to generate embeddings for each sentence.

### **Step 5: Cosine Similarity**

We chose cosine similarity as the similarity metric to measure how closely the action relates to the evidence based on the sentence embeddings. Cosine similarity scores range between 0 and 1, a high cosine similarity score (closer to 1) between the action and evidence suggests that they are likely very similar in meaning, and this implies the action likely resolves the audit issue. Conversely, a low cosine similarity score (closer to 0) suggests that it is unclear whether the action resolves the audit issue.

### **Step 6: Google Gemini**

Cosine similarity scores are helpful, but numbers alone might not be clear enough for auditors to determine if the action resolves the audit issue. We need to consider other factors than the scores for auditors to make better decisions. Since large language models (LLMs) are becoming more common, we can use them to make sense of the cosine similarity scores. By feeding the action, evidence, and the cosine similarity score to the Gemini, we can get a clearer explanation of what the score means. The Gemini can even provide rationale on whether the action resolves the audit issue based on the evidence.

We have tried different LLMs including Open AI ChatGPT, Meta Llama, and Google Gemini. All the models can easily satisfy our requirements, including the ability to interpret cosine similarity scores and provide rationale. Based on different benchmarks, Google Gemini performs better than Chatgpt in terms of reasoning and understanding (Steward, 2024). Other than that, Google also offers a free API for its Gemini Pro version, unlike ChatGPT which requires a paid subscription. We provide Gemini with the action, evidence, and cosine score along with the prompt: Based on the action and evidence statements along with cosine similarity score, can you provide rationale on whether the action is completed or not?" to generate our desired output.

### **Application Web Framework: Python Flask**

The application was developed using the Python Flask framework, leveraging its simplicity and flexibility. It is currently hosted locally which allows for us to develop, test and customise the project in a controlled secure environment.

## 4 Evaluation

To evaluate the model's performance, we created two sample cases: one demonstrating a successfully closed action and another showcasing an unclosed action. Users can supply the input via the Input page shown in Figure 3 in the Appendix.

### **Case 1 – Closed Action (See Figure 4 in the Appendix for the Audit Review):**

**Action:** “Ensure that all customer transactions exceeding \$10,000 are accompanied by proper documentation and flagged for review, as per anti-money laundering regulations. Review current procedures and implement necessary controls to mitigate the risk of non-compliance and financial penalties.”

**Completion Rationale:** "Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance with anti-money laundering regulations. All necessary documentation and controls are now in place to mitigate potential risks."

**Evidence Document:** [Case1-Evidence](#)

#### **Part 1**

When an action item is closed, we expect the cosine similarity score between the action item and evidence to be high because they are semantically similar. As can be seen from Figure 4 in the Appendix, we provided the model with a sample case where the action item is closed, so we expect to get a high cosine similarity score.

As expected, the Audit Review Results page shows a high similarity score of 0.92 between the action and the summarised evidence. A high cosine similarity score indicates the more the content of the Action matches with the Evidence.

Gemini generated audit review analyses the cosine similarity score for the summarised evidence and determines if the action is likely closed based on the provided evidence. The rationale generated by Gemini aligns well with the evidence, mentioning the implementation of enhanced monitoring systems to ensure anti-money laundering compliance, which is a key criterion to complete the action. The Gemini marked this action as closed based on the summarised evidence, this aligns with our expectation.

#### **Part 2**

As mentioned in the Methodology section, we find the top 5 sentences to give auditors recommendations of sentences to focus on when making their decision. The cosine similarity scores successfully capture key completion steps as can be seen from Evidence 1 to Evidence 4. However, Evidence 5 doesn't provide enough information to confirm that the action is closed. While cosine similarity identifies highly relevant evidence sentences, some evidence sentences outside the top 5 sentences are still important to determine if the action is closed. For example, the sentence “Implemented robust controls and oversight mechanisms, including regular audits, internal reviews, and a dedicated compliance team” in the evidence should also be considered when determining whether the action is closed.

Same as before, the Gemini determined this action as closed based on the top 5 evidence sentences. In summary, the audit review results for this sample case fully aligned with our expectations.

## **Case 2 – Unclosed Action (See Figure 5 in the Appendix for the Audit Review):**

**Action:** “Ensure that all customer transactions exceeding \$10,000 are accompanied by proper documentation and flagged for review, as per anti-money laundering regulations. Review current procedures and implement necessary controls to mitigate the risk of non-compliance and financial penalties.”

**Completion Rationale:** "Despite concerted efforts, the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed. While progress has been made in reviewing procedures and exploring technological solutions, challenges remain in the implementation phase. Further coordination and resource allocation are necessary to ensure the effective deployment of enhanced monitoring capabilities and the establishment of robust controls. The issue will continue to be actively monitored and addressed until full compliance is achieved."

**Evidence Document:** [Case2-Evidence](#)

In case 2, we provided the model with a sample case where the action item is not closed. This time, we expect to get a low cosine similarity score.

### **Part 1**

As can be seen from Figure 5 in the Appendix, we still get a high cosine similarity score of 0.71 (closer to 1), but the cosine similarity score is not as high as what we observed in Case 1. In this case, the cosine similarity score is high even if the evidence indicates that the action is not closed. This means the context of evidence might still be relevant to the action, even though the evidence doesn't show that the action is closed. Therefore, we might need to consider setting an optimal cosine similarity threshold as a cut-off point in the future to determine if the cosine similarity score is high or low.

### **Part 2**

Overall, the top 5 sentences identified by the Sentence Bert model are sufficient to determine the action is not closed. While Evidence 3 and 4 mention the difficulties encountered, they lack the details needed to determine whether the action is closed or not. All the top 5 evidence sentences have a moderate degree of similarity with the action (around 0.5). Similar to Case 1, some key sentences were missed because they weren't very similar according to the cosine similarity scores.

The audit review generated by Gemini aligns well with the evidence, mentioning that the action is not completed because the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed. The Gemini model provided clear explanations for its decision.

## Advantages and Disadvantages of the Solution

Advantages	Disadvantages
<b>Approach Structure</b> <ul style="list-style-type: none"> <li>- With both top 5 sentence embedding and summarised evidence embedding, we can provide a comprehensive view that facilitates a more thorough understanding of the context for the human reviewer</li> <li>- Access to multiple layers of information that can increase confidence in the decision-making process</li> <li>- Improved efficiency from streamlined review process</li> <li>- Summarised evidence helps in understanding the main points as a whole while top-5 sentences helps narrow down the focus</li> <li>- Uses contextualised embeddings that can capture nuances and contextual meanings leading to more accurate similarity comparisons</li> <li>- sentenceBert is pre trained on large-scale text data</li> <li>- The Gemini model can help auditors understand the context of the evidence and draw more meaningful conclusions.</li> </ul> <b>Cosine Similarity</b> <ul style="list-style-type: none"> <li>- Effective for high-dimensional data making it suitable for embeddings generated by SentenceBert</li> <li>- Can capture semantic similarities between vectors in high-dimensional space</li> <li>- Calculation is computationally efficient especially compared to other similarity measures such as Euclidean distance</li> <li>- Suitable for processing large datasets</li> </ul>	<b>Approach Structure</b> <ul style="list-style-type: none"> <li>- Takes time if affect the efficiency if we showed all the steps</li> <li>- Processing both summarised evidence and sentence pairs requires more computational resources</li> <li>- Risk of redundant information or overwhelming the reviewer with too much information</li> <li>- Final decision still relies on human judgement which may be prone to subjectivity</li> <li>- Pre Trained sentenceBert may not be optimised for the specific task. Fine-tuning may be helpful to adapt the model to the auditing context</li> <li>- Fine-tuning sentenceBert model will require significant computational resources for training and inference</li> <li>- The performance depends on the quality and representativeness of the pretraining data</li> <li>- Using the summarisation model might lose important information</li> <li>- Top 5 sentences might not be able to capture all the important sentences in the evidence.</li> <li>- The outputs of the Gemini model might not always be consistent.</li> </ul> <b>Cosine Similarity</b> <ul style="list-style-type: none"> <li>- Limited to semantic similarity, may not capture the other aspects of similarity such as syntactic or structural similarity</li> <li>- Can be sensitive to noise or irrelevant features in the data</li> <li>- Dependent on embedding quality. If embeddings do not adequately represent the semantics of the text, cosine similarity may not accurately reflect the similarity</li> </ul>

## 5 Discussion on the Limitations

Throughout this project, we faced several challenges that shaped our approach and provided valuable learning experiences.

- The primary issue was the lack of real-world data from the company, necessitating the generation of synthetic test data to train and validate our model. This required careful simulation of realistic scenarios and data structures.
- Handling diverse and unstructured data formats like PDFs and Word documents involved meticulous preprocessing to extract meaningful text. Selecting and



optimising models, particularly Sentence-BERT for embedding and cosine similarity for text relevance, was crucial for balancing computational efficiency and accuracy.

- Summarising extensive evidence data without losing critical information, interpreting cosine similarity scores, and integrating Google Gemini for insightful comments were additional hurdles.
- Managing significant computational demands and designing a user-friendly interface further added to the complexity. Despite these challenges, the project demonstrated the feasibility and benefits of leveraging Generative AI for audit and compliance in the banking sector, paving the way for future advancements in AI-driven auditing solutions.

## 6 Key Recommendations to the Partner

Our current method of gathering data is too reliant on manual entry. Instead of relying on auditors to input action and evidence manually from the Audit Closure Form, the bank could explore other approaches that can break these issue actions into individual lines and directly collect data from the source using a new or existing database with an issue creation to closure workflow. An ideal scenario would be for auditors to simply submit the Audit Closure Form along with the supporting evidence documents for the specific issue action items and, the GenAI tool can directly gather evidence from the database. This would streamline the process and minimise the need for manual data entry, making the process faster and structured from a data creation, use and consumption perspective across Internal Audit and Business Teams.

We used a basic version of a pre-trained model called SentenceBERT. While it performed well with our test cases, the bank's real-world data might contain unfamiliar banking terms. The bank should consider fine-tuning the SentenceBERT model on their specific dataset to improve its understanding of relevant banking terms. In addition, the bank can experiment with more powerful versions of SentenceBERT or other models like RoBERTA that might handle the bank's data more effectively.

The bank can consider setting an optimal threshold for cosine similarity scores as a cut-off point to determine if the action is closed or unclosed. The threshold can be adjusted based on historical data analysis and feedback from human reviewers, allowing it to adapt to real-world use cases. The bank can leverage statistical analysis tools such as ROC Curve to determine the optimal cosine score threshold, as well as the number of top sentences that are considered as relevant to the action.

Other than that, the bank can explore various summarization techniques and their parameters to keep the summaries brief without losing important information. Human reviewers can be involved in the process to provide valuable feedback, so the system can learn to ensure the summaries are clear and informative.

In our specific scenario, the Gemini model is already capable of generating meaningful rationale for audit closures. However, if the bank's focus is on prioritising ethical and safety considerations, alternative solutions such as Claude AI might be better suited. While cost can be a consideration for our project, the bank has greater flexibility to explore various options to find the best fit for their needs. Moreover, the bank can also consider building its



own large language model (LLM) specifically tailored to the banking industry. This custom model could potentially generate more relevant and nuanced rationales for audit closures.

**Recommended Next Action** - Examine this technical feasibility study and develop a business case that considers the whole audit review process, integrating with a Gen-AI feature to improve audit review abilities with a human reviewer for final approval until fully tested .

## 7 Conclusion

In conclusion, this project successfully demonstrated the potential of leveraging Generative AI to revolutionise audit and compliance processes within the banking and finance sector. By developing a sophisticated AI tool that evaluates audit action items against supporting evidence, we significantly enhanced the efficiency and accuracy of audit reviews. The use of Sentence-BERT for embedding and cosine similarity for evaluating textual relevance, combined with the interpretative capabilities of Google Gemini, provided auditors with precise and actionable insights. This project not only streamlined the audit process but also reduced the risk of human error and time consumption. The integration of advanced AI techniques with robust data handling and preprocessing frameworks underscores the transformative impact of technology in improving regulatory compliance and risk management. This project sets a foundation for further innovations in AI-driven audit processes, ultimately contributing to the integrity and stability of the banking and finance industry.

## 8 Appendix

### 8.1 References

- Ahmad, S. (2020). Regulatory Compliance in Banking Industry: A Review of Literature. *International Journal of Business and Management*, 15(5), 49-59.
- Davenport, T.H., Ronanki, R. (2018). Harvard Business Review: Artificial intelligence for the real world. [Harvard Business Review Article](#)
- Dey, A., Sharma, A., & Shukla, A. (2021). Leveraging Big Data Analytics for Efficient Audit and Compliance in Banking Sector. *International Journal of Advanced Computer Science and Applications*, 12(5), 341-348.
- Dey, S., Ghosh, S., & Sharma, R. (2021). Extracting Insights from Multiple Datasets for Improved Audit and Compliance in Banking: A Case Study. *International Journal of Banking and Finance*, 7(1), 23-35.
- Deloitte (2023). Navigating the artificial intelligence frontier: An introduction for internal audit. [Deloitte Article](#)
- Efimov, V. (2023). Large Language Models: SBERT — Sentence-BERT. [Towards Data Science Article](#)
- Hasan, A. R. (2022). Artificial Intelligence (AI) in Accounting & Auditing: A Literature Review. *Open Journal of Business and Management*, 10(01), 440–465. <https://doi.org/10.4236/ojbm.2022.101026>
- Hassan, M. S. (2018). Audit Automation and Its Impact on Audit Efficiency: Evidence from the Banking Sector. *Journal of Accounting and Finance*, 21(3), 78-93.
- Jain, R., Singh, A., & Gupta, R. (2019). Artificial Intelligence in Banking Sector: A Comprehensive Overview. *International Journal of Recent Technology and Engineering*, 8(2S3), 643-648.
- Jain, A., Raj, A., & Shekhar, S. (2019). Leveraging AI for Audit and Compliance Processes in the Banking Sector: A Review. *Journal of Finance and Accounting*, 3(2), 45-58.
- Steward, E. (2024). Is Google's Gemini Better than ChatGPT? A Comparison. [EM360 Article](#)
- Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. *IEEE Access*, 8(8), 110461–110477. <https://doi.org/10.1109/access.2020.3000505>

## 8.2 Code and Project Management:

We used github to manage and collaborate on the code. The repository can be found here <https://github.com/Yqisme490/GenAI-for-Bank-Internal-Audit/tree/main>. The repository contains our code and all necessary information needed to run the application.

We primarily used Microsoft Teams to manage our tasks and schedules. We created a Teams group which allowed us to discuss the project and set up meetings. We had weekly standup meetings internal to our team and bi-weekly meetings with our partner organisation and course coordinators where we presented the project progress update. You can find the Minutes of the Meetings here - [Minutes Of Meetings for PGP](#)

We adopted a hybrid approach for the project, combining virtual meetings and in-person collaboration. Virtual meetings on Teams facilitated spontaneous discussions, while RMIT's collaborative space allowed for effective face-to-face work. Most of our project time was spent working together as a team of four, though we also dedicated individual hours to our assigned tasks. Each team member contributed equally to all aspects of the project, including documentation, presentation, application development, and design. We shared our responsibilities and assisted one another in completing each task. The roles and responsibilities of each team member are detailed below.

Name	Responsibilities
Aaron Tam	<ul style="list-style-type: none"><li>• Lead Data analysis and extraction</li><li>• Application development &amp; design</li><li>• Report and presentation preparation</li></ul>
Maryam Alhowaiti	<ul style="list-style-type: none"><li>• Lead Quality and Test</li><li>• Application development</li><li>• Report and presentation preparation</li></ul>
Shifali Shetty	<ul style="list-style-type: none"><li>• Lead Project Management</li><li>• Application development &amp; design</li><li>• Report and presentation preparation</li></ul>
Yung Qi Chin	<ul style="list-style-type: none"><li>• Lead Generative AI integration</li><li>• Application development &amp; design</li><li>• Report and presentation preparation</li></ul>

## 8.3 Self Reflection

### Shifali Shetty (S3956999) :

This project was a great real-world example of how one builds a solution from the ground up. When I read the description of the project I was really interested as it was a domain I had previously worked in but not in the capacity of a data scientist. Our initial meeting with representatives of the financial institute was very positive and slightly overwhelming. The requirements provided and the expectations seemed hard to accomplish in the limited time we had. However as we slowly started the work I realised that the goal was accomplishable and that it was important to clearly communicate what can and cannot be achieved with the

client. Once the deliverables were clearly defined and we were all on the same page, I felt a lot more confident.

I enjoyed trying out different approaches and being able to experiment with different models. When I initially started working with the data I felt that it is important to follow the data preparation steps exactly as what we were taught in our courses. But I learnt later on that there is no hard and fast rule. The approach needs to be modified and adapted based on the task. Our project didn't require traditional data cleaning steps like Stemming, Lemmatization or Stop word removal and it was ok to not have to do this.

There were many hurdles that we faced with the project, one of which was that we did not get a lot of sample audit issues from our client due to data security reasons. This was disappointing as we felt the true performance and usability of the solution would be best known if we tested on real audit issues that were faced by the bank. However we were able to make do with AI generated samples and we communicated all limitations with the client so they knew what to expect.

Another hurdle we had was with finding the best embedding model for our data. We started with T5 as this was a widely popular model but this did not give us good results. Our analysis helped us understand that we need a model that was more fine-tuned for understanding text similarity such as SentenceBERT rather than T5 which was more designed for text transformation tasks. This particular instance helped me understand that the true intention with which a model was built needs to be considered before applying it to your task.

Overall I feel that this project was a great example of how we can apply what we learnt in our program in the real world. I learnt that I liked interacting with clients and building a good rapport with them. I also discovered that I enjoyed leadership and project management opportunities and collaborating with a team.

### **Maryam Alhowaiti (S3856144)**

This project offered a unique platform to refine and advance both my technical acumen and interpersonal skills. On the technical front, I deepened my understanding of Artificial Intelligence and machine learning, specifically gaining hands-on experience with advanced NLP techniques using Sentence-BERT for precise text similarity analysis and Google Gemini for sophisticated decision-making processes. This practical exposure was complemented by a focus on data preprocessing and summarization tasks, where I mastered the integration of complex algorithms to enhance both efficiency and accuracy of our models.

From a soft skills perspective, the project was a crucible for developing key professional competencies. I honed my abilities in effective communication and collaboration, particularly with diverse stakeholder groups, ensuring that all project deliverables were aligned with our client's expectations. The environment demanded rigorous problem-solving and adaptability, pushing me to navigate through challenges with strategic foresight and resilience.

Furthermore, the project management aspect of this role illuminated the significance of leadership within a tech-driven landscape. I took pride in steering our project through various phases, from conceptualization to execution, ensuring meticulous attention to detail at every

step. This was vital not only in achieving our goals but also in upholding the integrity of our results, especially given the reliance on AI-generated data.

The experience underscored the importance of a thorough and methodical approach in the field of data science. It reinforced my belief in the necessity of maintaining a high standard of quality and reliability in AI outputs, which are crucial for crafting solutions that stakeholders can trust and rely upon.

Overall, the project was not just a testament to what we have learned academically; it was a profound journey that expanded my professional horizons and solidified my passion for data science and AI. It was truly rewarding to see theoretical knowledge come to life and to witness firsthand the impact of our work on real-world business processes.

### **Aaron Tam (S3961290)**

Working on this project has allowed me to develop more of an understanding of the complex nature of banking audit processes as well as the critical role of evidence in supporting findings. Initially, one of the most difficult aspects of the project for me was fully understanding the intricate audit process which was essential before we could start working.

Collaboration with audit professionals and supervisors was invaluable, offering insights and strategic direction that helped with the approach. This teamwork also fostered a collaborative environment which allowed me to strengthen communication skills and ensured our solutions were both practical and technically sound.

The project allowed me to understand the current developments in the field of large language models and AI generators and showed me what could be achieved with these in real life scenarios. The project also provided the opportunity to improve my technical skills applying the models as well as performing analysis.

This project has been an enriching experience, blending technical innovation with practical application, and the knowledge and skills gained will serve as a strong foundation for future endeavours in AI and financial auditing.

### **Yung Qi Chin (S3819792)**

Before starting this project, I had no background knowledge about auditing. It provided a valuable opportunity to understand the banking audit process and how we can leverage the capabilities of AI to make the complex audit process more efficient. In addition to auditing, I also explored NLP techniques like Sentence BERT, T5, and the Gemini model, along with their practical use cases. I also learned how to effectively communicate with clients and understand their needs and expectations.

Initially, we tried using the rationale for closure along with the evidence documents to calculate the cosine similarity scores with the action item. However, the cosine similarity scores were low even when the evidence supported the action being closed. Therefore, we decided to break down the evidence into individual sentences and calculate the cosine similarity between each sentence and the action item.

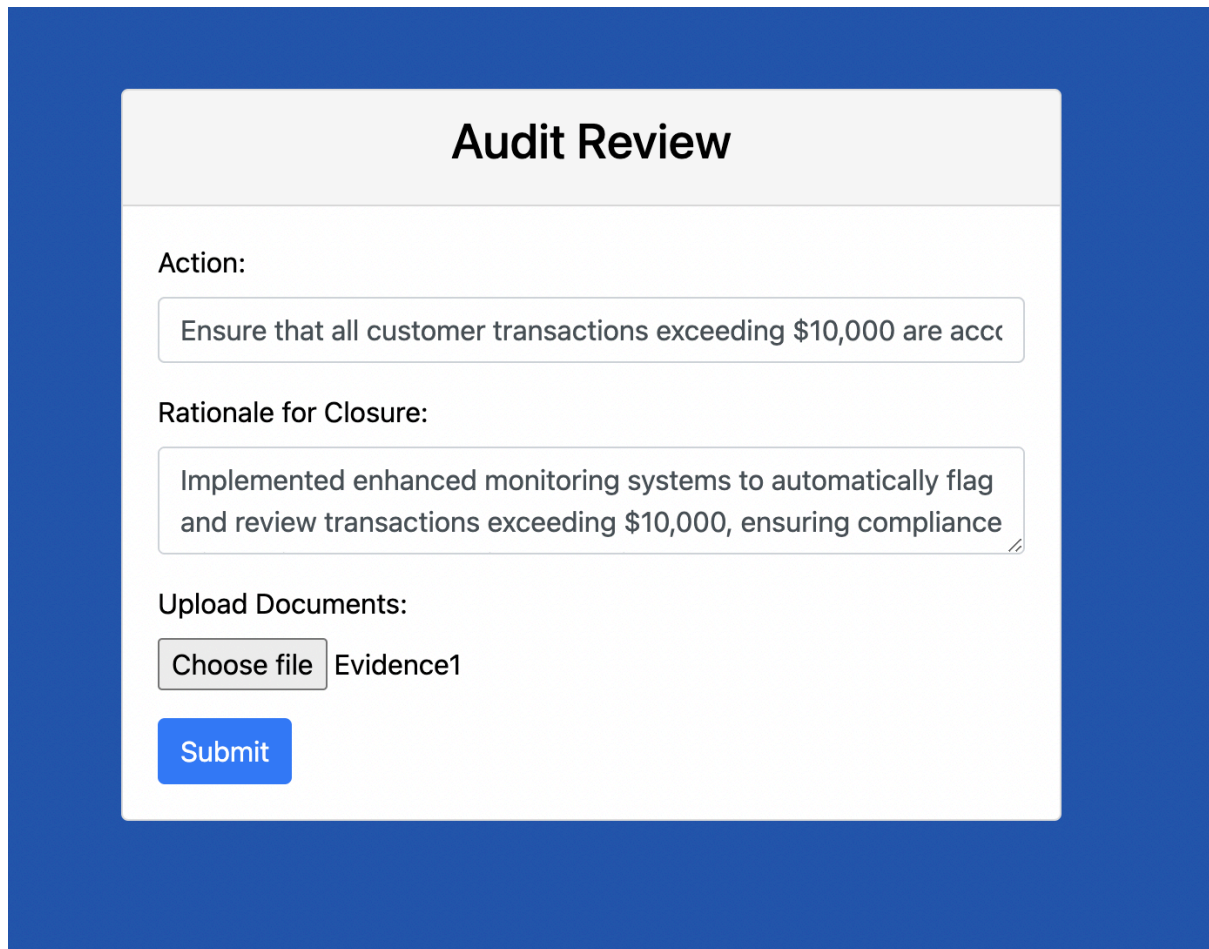
During the system development process, we experimented with various approaches, some of which did not work. We attempted to implement the T5 model to get sentence embeddings and calculate the cosine similarity score between the action and evidence. However, the results from the T5 model were not as effective as those from the Sentence BERT model.

I conducted research on using Large Language Models (LLMs) to generate rationales for audit action closure and interpret cosine similarity scores. I attempted to implement the Llama3 model using Hugging Face libraries, but this failed due to limited computational resources. Additionally, we considered using the Open AI GPT model during development, but we did not use it because it requires a paid subscription. As a result, we opted for the Gemini model, which offers a free API.

We didn't have sufficient samples to test our system due to safety considerations on the client's end. To overcome this, we generated sample data using Chat GPT to test our system. Consequently, we did not have the opportunity to develop our system using real-life cases.

Overall, this project provided me with valuable experience in implementing generative AI in real-life applications. Through this project, I have gained essential soft and hard skills, as well as knowledge specific to the banking industry.

## 8.4 Images



The screenshot shows a web form titled "Audit Review" set against a blue background. The form is a white rectangle with a light gray header. It contains three main sections: "Action:", "Rationale for Closure:", and "Upload Documents:". Each section has a text input field. The "Action" field contains the text "Ensure that all customer transactions exceeding \$10,000 are acco". The "Rationale for Closure" field contains the text "Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance". The "Upload Documents" section includes a "Choose file" button, the text "Evidence1", and a blue "Submit" button at the bottom left.

**Audit Review**

Action:

Ensure that all customer transactions exceeding \$10,000 are acco

Rationale for Closure:

Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance

Upload Documents:

Choose file Evidence1

Submit

Figure 3. Audit Review Input page



## Audit Review Results

### Input Provided:

**Action:** Ensure that all customer transactions exceeding \$10,000 are accompanied by proper documentation and flagged for review, as per anti-money laundering regulations. Review current procedures and implement necessary controls to mitigate the risk of non-compliance and financial penalties.

**Completion Rationale:** Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance with anti-money laundering regulations. All necessary documentation and controls are now in place to mitigate potential risks.

**Evidence Document:** Evidence1.docx

### Part 1 : Similarity Score with Summarised Evidence

**Summarised Evidence:** Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance with anti-money laundering regulations. All necessary documentation and controls are now in place to mitigate potential risks. Addressing Compliance Concerns: Enhancing Anti-Money Laundering Measures In today's global financial landscape, ensuring compliance with anti-money laundering (AML) regulations is paramount for banking institutions. One significant aspect of AML compliance involves monitoring and reviewing customer transactions exceeding a certain threshold, typically \$10,000. To provide evidence of our commitment to compliance and the effectiveness of our measures, we have compiled comprehensive documentation outlining the steps taken to address the action item concerning high-value transactions. Through investments in technology, enhanced documentation requirements, and rigorous oversight, we have fortified our defenses against money laundering risks and demonstrated our unwavering commitment to compliance.

**Similarity Score for Action and Summarised Evidence** 0.9259813

### Gemini generated Audit Review for Summarised Evidence:

**Output:** **\*\*Closed:\*\*** The evidence statement demonstrates that the institution has implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance with anti-money laundering regulations. Necessary documentation and controls are now in place to mitigate potential risks. This indicates that the action has been effectively addressed and closed.



## Part 2 : Key sentences in the Evidence which align closely with the Action item:

**Evidence1:** Implemented enhanced monitoring systems to automatically flag and review transactions exceeding \$10,000, ensuring compliance with anti-money laundering regulations.

**Similarity Score:** 0.882421

**Evidence2:** Additionally, we revamped our documentation requirements for high-value transactions, instituting a rigorous process to ensure that all transactions exceeding \$10,000 are accompanied by proper documentation and are flagged for review.

**Similarity Score:** 0.86632717

**Evidence3:** All necessary documentation and controls are now in place to mitigate potential risks. Addressing Compliance Concerns: Enhancing Anti-Money Laundering Measures In today's global financial landscape, ensuring compliance with anti-money laundering (AML) regulations is paramount for banking institutions.

**Similarity Score:** 0.7967143

**Evidence4:** One significant aspect of AML compliance involves monitoring and reviewing customer transactions exceeding a certain threshold, typically \$10,000.

**Similarity Score:** 0.74318737

**Evidence5:** By capturing this information at the outset, we not only facilitate compliance but also strengthen our due diligence efforts in combating financial crime.

**Similarity Score:** 0.68782663

## Gemini generated Audit Review for key sentences in the Evidence:

**Output:** The action is **"closed"**. \* Evidence 1 and Evidence 2 both describe the implementation of new systems and procedures to monitor and review transactions exceeding \$10,000, and to ensure that they are accompanied by proper documentation. \* Evidence 3 states that all necessary documentation and controls are now in place to mitigate potential risks. \* Evidence 4 and Evidence 5 explain the importance of monitoring and reviewing customer transactions exceeding \$10,000 for AML compliance. Taken together, these evidence statements provide strong support for the conclusion that the action has been closed.

Figure 4. Example Audit Review Results - Closed Action

## Audit Review Results

### Input Provided:

**Action:** Ensure that all customer transactions exceeding \$10,000 are accompanied by proper documentation and flagged for review, as per anti-money laundering regulations. Review current procedures and implement necessary controls to mitigate the risk of non-compliance and financial penalties.

**Completion Rationale:** Despite concerted efforts, the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed. While progress has been made in reviewing procedures and exploring technological solutions, challenges remain in the implementation phase. Further coordination and resource allocation are necessary to ensure the effective deployment of enhanced monitoring capabilities and the establishment of robust controls. The issue will continue to be actively monitored and addressed until full compliance is achieved.

**Evidence Document:** Evidence1.1.docx

### Part 1 : Similarity Score with Summarised Evidence

**Summarised Evidence:** Despite concerted efforts, the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed. The issue will continue to be actively monitored and addressed until full compliance is achieved. Navigating Compliance Challenges: Understanding Hurdles in Addressing Anti-Money Laundering Measures The journey towards achieving full compliance with anti-money laundering (AML) regulations is often fraught with challenges and complexities. Despite our institution's earnest efforts to enhance monitoring systems for transactions exceeding \$10,000, several hurdles have impeded the completion of this crucial action item. In conclusion, the journey towards addressing AML compliance concerns is marked by numerous hurdles, ranging from technological constraints and regulatory complexities to resource limitations and internal resistance. While the action item pertaining to enhanced monitoring systems for high-value transactions may not have been completed as anticipated, the challenges encountered have provided valuable insights and opportunities for growth. By acknowledging these hurdles and remaining steadfast in our commitment to compliance, we will continue to overcome obstacles and fortify our AML framework to safeguard the integrity of the financial system.

**Similarity Score for Action and Summarised Evidence:** 0.711368

### Gemini generated Audit Review for Summarised Evidence:

**Output:** The action is not closed. The evidence statement states that "the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed". This means that the action has not been fully implemented and is still in progress. The action will remain open until it is fully completed and all necessary controls are implemented to mitigate the risk of non-compliance and financial penalties.

## Part 2 : Key sentences in the Evidence which align closely with the Action item:

**Evidence1:** Despite our institution's earnest efforts to enhance monitoring systems for transactions exceeding \$10,000, several hurdles have impeded the completion of this crucial action item.

**Similarity Score:** 0.679669

**Evidence2:** The issue will continue to be actively monitored and addressed until full compliance is achieved. Navigating Compliance Challenges: Understanding Hurdles in Addressing Anti-Money Laundering Measures The journey towards achieving full compliance with anti-money laundering (AML) regulations is often fraught with challenges and complexities.

**Similarity Score:** 0.6738731

**Evidence3:** Adhering to stringent regulatory standards while balancing operational efficiency and customer experience necessitated meticulous planning and coordination across multiple departments.

**Similarity Score:** 0.57018906

**Evidence4:** By acknowledging these hurdles and remaining steadfast in our commitment to compliance, we will continue to overcome obstacles and fortify our AML framework to safeguard the integrity of the financial system.

**Similarity Score:** 0.5465019

**Evidence5:** Despite concerted efforts, the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed.

**Similarity Score:** 0.52910316

## Gemini generated Audit Review for key sentences in the Evidence:

**Output:** Based on the provided document, the action is **\*\*not\*\*** closed. The action is to "Ensure that all customer transactions exceeding \$10,000 are accompanied by proper documentation and flagged for review, as per anti-money laundering regulations. Review current procedures and implement necessary controls to mitigate the risk of non-compliance and financial penalties." Evidence 1 and 5 indicate that the action item has not been fully completed. Evidence 1 states that "several hurdles have impeded the completion of this crucial action item." Evidence 5 states that "the action item pertaining to the enhancement of monitoring systems for transactions exceeding \$10,000 has not been fully completed." Therefore, the action is not closed.

Figure 5. Example Audit Review Results - Unclosed Action