iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62292

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Document Management System Generative AI and AWS Comprehend

Niharika Tiwari[1], Prashansa Srivastava[2], Dr. Sadhana Rana[3]
[1, 2]*Department of Computer Science and Engineering*
[3]*Assistant Professor, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India*

Abstract: ***Document Management Systems that have Generative AI and cloud features integrated in them can help users to save time. The generative ai features can boost creativity, quality and cloud resources can greatly help expand the scope of the system. A document management system built entirely on the cloud offers mobility and freedom of expression and analysis. With analysis features, users can derive important insights from texts in place and cloud storage offers the mobility to access these features from anywhere. From this, comes the possibility of an application that allows users to choose their features depending upon their needs. It is a deeper stride into the pay-as-you-go-model where the users pay only for what they use. The application provides a model to what a truly dynamic feature rich application might look like.***

Keywords: Document Management System, Generative AI, AWS Comprehend, Natural Language Processing, Cloud Storage, Machine Learning.

I. INTRODUCTION

AI Cloud Analyzer is a platform that provides an all-in-one solution to create, store, build and analyze documents using cloud computing and generative ai.

Generative AI is a very new addition in the tech industry. It has added new dimensions to how we create content. This has direct implications on the content creator economy [1]. The impact of the technology is yet to be seen in its integration with existing technologies like simple document management systems. In this system we have integrated generative ai with AWS Comprehend to build one such document management system. Both generative ai and AWS Comprehend inherently use natural language processing that provides improved user experiences [2].

Smart document analysis using AI/ML has for long been in use. This can be seen in its impact in syntax detection, sematic detection and PII detection systems [3]. This is an added layer on top of NLP Systems.

Also, we see that in the recent years, cloud computing technology has witnessed a huge boom in terms of its adoption. Everywhere around us we see organizations migrating to cloud platforms. Cloud makes traditional systems like storage and processing limitless when they are delivered as services via the internet [4].

This technology can be added to simple applications to give an impression of limitless storage to the user of the application. Addition of cloud-based storage systems allows for easy access of resources on a remote basis and manage large amounts of data [5].

Following the introduction of study in this section, Section 2 describes the literature review, Section 3 explains the methodology. Section 4 presents module description and their working. Section 5 discusses the results of the work followed by the conclusion and the future scope of this work.

II. LITERATURE SURVEY

AIGC is a field of computer science that focuses on the development of systems that can generate content, such as text, images, and music, autonomously. AIGC systems are trained on large datasets of existing content, and they use this training data to learn the patterns and rules that govern the generation of new content.

AIGC systems have become increasingly sophisticated in recent years, and they are now able to generate content that is indistinguishable from human-generated content in many cases. AIGC is being used in a variety of industries, including art, advertising, and education.

Here are some of the recent advances in the field of AIGC:

1) The development of new generative AI models, such as ChatGPT, which are able to generate more realistic and high-quality text than previous models.

©IJRASET: All Rights are Reserved

2) The development of new applications for AIGC, such as the use of AIGC to generate personalized educational content for students.

Smart Document Classification

Smart document classification is a process of using AI and machine learning to automatically classify documents into different categories. This can be useful for a variety of tasks, such as organizing documents in a digital library or filtering out spam emails.

Here are some of the machine learning algorithms that are commonly used for smart document classification:

- Support vector machines (SVMs)
- Naive Bayes classifiers
- Decision trees
- Random forests
- Gradient boosting machines

AIGC and smart document classification are two rapidly developing fields with a wide range of applications. AIGC can be used to generate personalized content for users, while smart document classification can be used to organize and filter large volumes of data. As these technologies continue to develop, we can expect to see them used in even more innovative and impactful ways in the future.

Cloud storage is a model of data storage in which the digital data is stored in logical pools, the physical storage of which is spread across multiple servers (potentially in different locations) and is typically managed by the cloud storage service provider.

III. PROPOSED METHODOLOGY

Our methodology to develop this document management system consists of a layered architecture. This pattern segregates the responsibilities of the different layers of code, ensuring security, scalability, and easy feature additions.

The layered architecture will consist of the following layers:

1) *Presentation layer:* This layer will be implemented using Vaadin, providing a user-friendly interface for users to manage their work on the cloud.

2) *Application layer:* This layer will contain the business logic of the application, such as processing user requests and generating responses. It will also use the AWS SDK for Java to establish a connection to the cloud and access the necessary services.

3) *Data layer:* This layer will be responsible for accessing and managing data, such as storing and retrieving data from a database. It will also use the AWS SDK for Java to interact with AWS services.

Each layer will be decoupled from the other layers, making the code more modular and reusable. This will also make it easier to scale the application and add new features in the future.

IV. MODULE DESCRIPTION

A. Prompt Module

The prompt module is responsible for handling all generative AI related content. It provides a number of features, including:

1) Prompt generation: The prompt module can generate prompts for a variety of generative AI tasks, such as text generation, image generation, and code generation.

2) Prompt editing: The prompt module allows users to edit prompts to improve their accuracy and specificity.

3) Prompt evaluation: The prompt module can evaluate prompts to identify language, tone and sentiments.

B. Comprehend Module

The comprehend module is a powerful tool for analysing and understanding large amounts of text data using AWS Comprehend. It makes it easy to extract entities, identify sentiment, model topics, and extract key phrases.

The comprehend module can be used for a variety of tasks, such as:

1) Customer insights: Analysing customer feedback to identify trends and patterns.

2) Market research: Analysing social media data to understand public opinion about a product or service.

3) Risk analysis: Analysing financial data to identify potential risks.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Storage Module

The storage module is responsible for providing cloud storage to users using Amazon S3. It provides a number of features, including:

1) Document storage: The storage module can store documents in Amazon S3, which is an object-based database that is well-suited for storing and managing large amounts of data.

2) Document retrieval: The storage module can retrieve documents from Amazon S3 quickly and efficiently.

V. IMPLEMENTATION

AI Cloud Analyzer is a project that has been implemented using the programming languages, Java. It has a frontend, backend, and cloud connectivity.

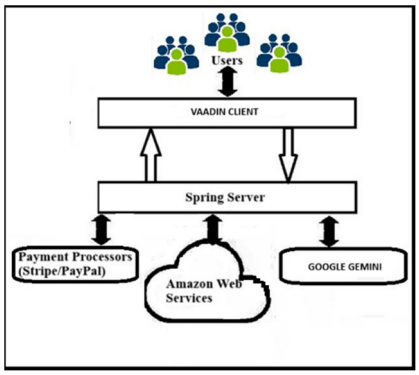


Figure 1: Client-Server Architecture

The frontend of the project has been built using Vaadin that is a very user-friendly Java framework. The backend of this project is implemented using the Spring Framework in java. The cloud used is Amazon Web Services. The storage feature is the courtesy of the Amazon Simple Storage Service called as S3 and the APIs used for analysis is the AWS Comprehend API.



Figure 2: Use Case Diagram of The System

The main aim of the project is to provide the facility to generate document templates and give suggestions for improvement using the Google Gemini API. It provides analysis and storage. The payment gateway is implemented using Stripe. The robustness and security of the application is ensured by the use of the Spring Framework which is an ideal choice for such an application.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. RESULT

The AI Cloud Analyzer website provides a platform for users to interact with their documents in an all-in-one dashboard.

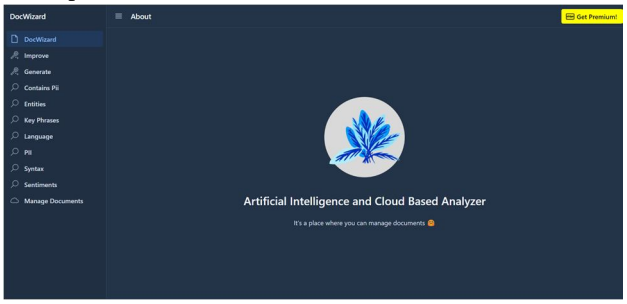


Figure 3: The Dashboard

When the user decides what document, he wishes to generate, he can simply enter it in the textbox and click on the 'Generate' button. This carries the user input from the UI to the backend and makes an API call to the Gemini API by Google. Then the result is then provided in the given space.

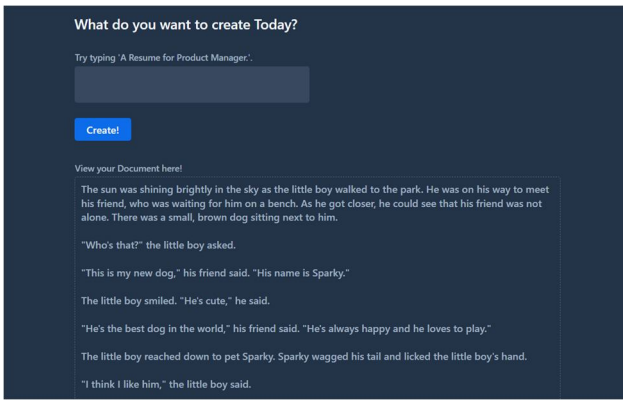


Figure 4: Document Generation - Generated a short story on pets

A similar implementation is of the Improvement feature. The two differ in the kind of prompt that has been used in the backend as input to Gemini.

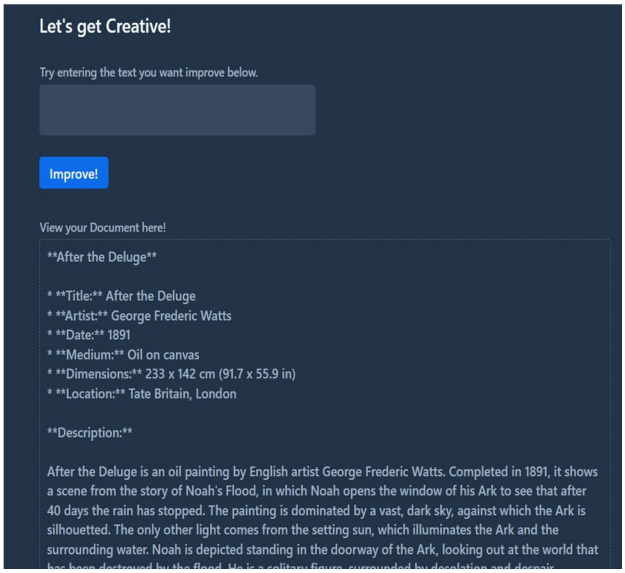


Figure 5: Text Improvement

The analysis API works by taking text input and supplying it to AWS Comprehend as input. It then fetches the output and displays

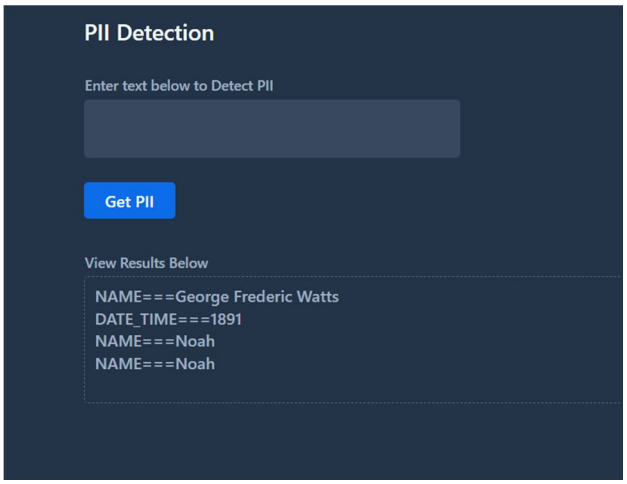


Figure 6: PII Analysis of Wikipedia Text Sample

It even provides the feature to extract all the parts of speech from a given piece of text.

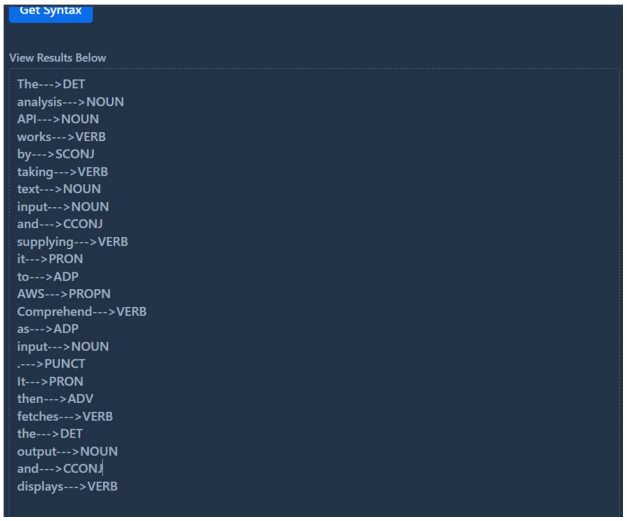


Figure 7: Syntax Detection Analysis

The payment button takes the user to Stripe checkout that is currently set to the testing mode.

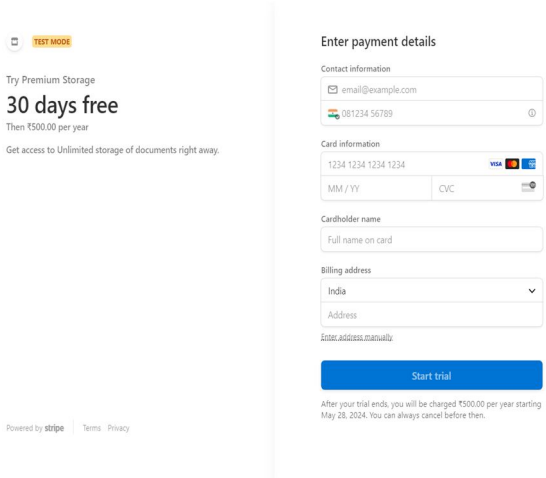


Figure 8: Stripe Checkout

Lastly the document management system lists all documents stored in the cloud storage and allows user to perform CRUD operations on them.

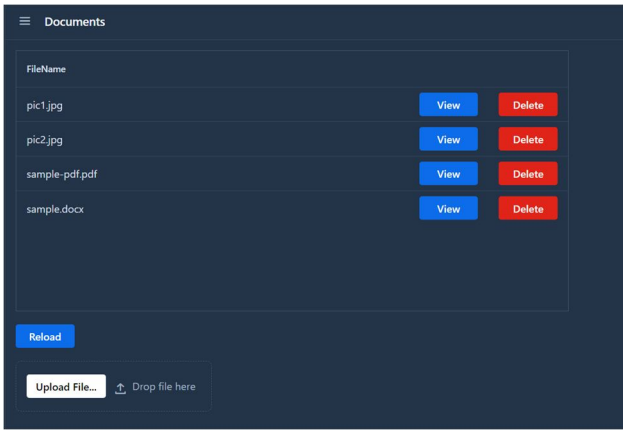


Figure 9: Document Management System

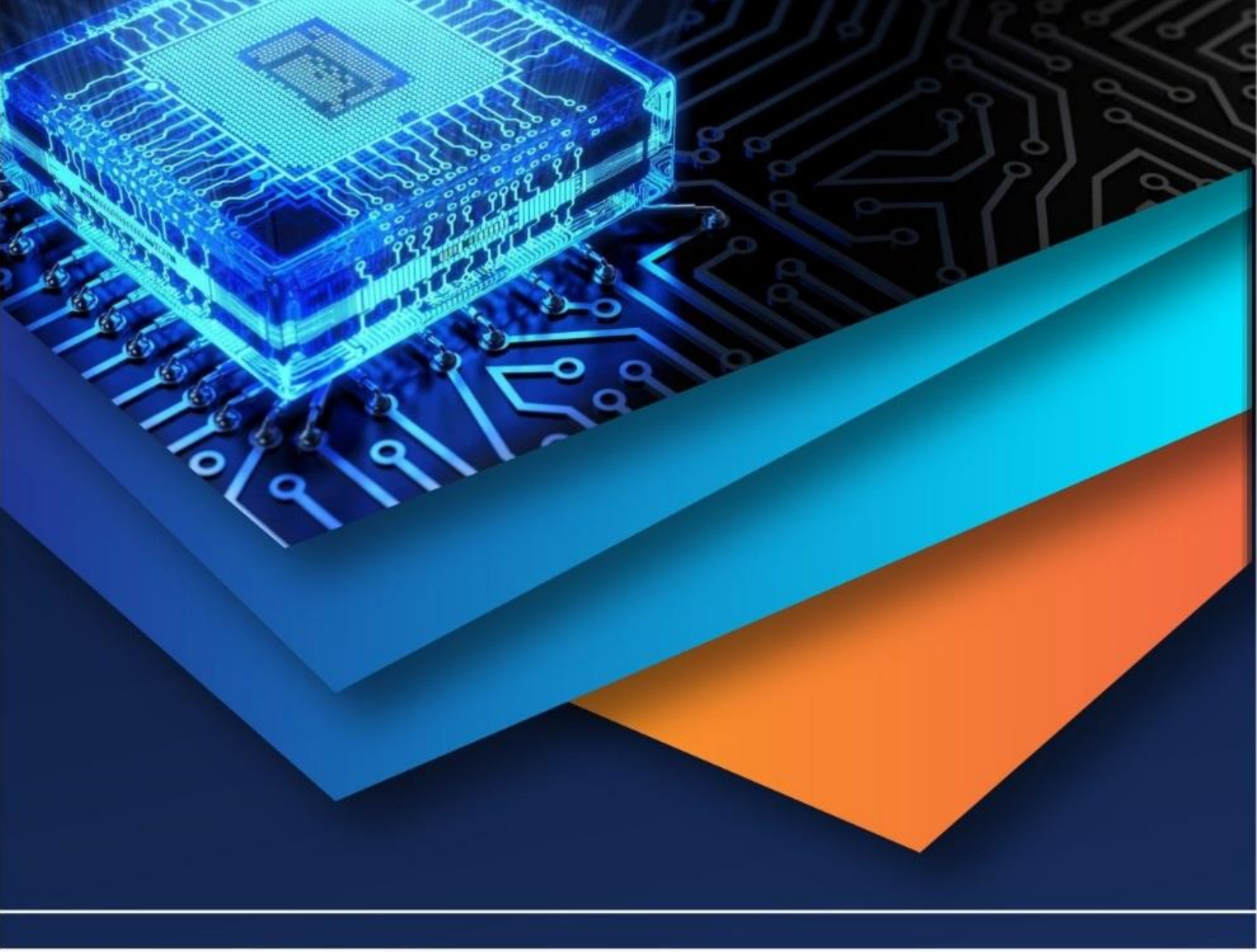
VII. CONCLUSION AND FUTURE SCOPE

To conclude, we can see that the application accomplishes a fair number tasks in an elegant way. It allows users to create, improve, analyze, upload, download and delete the documents.

The scope of expanding upon this field of work is huge. This involves adding technologies like AWS Quicksight for data visualization, DALL-E models for generative ai image processing, and other analysis models that we were limited to use by the free-tier of Amazon Web Services [7].

REFERENCES

[1] https://papers.ssrn.com/sol3/papers. cfm?abstract_id=4670714
[2] https://link.springer.com/chapter/10 .1007/978-3-031-48057-7_14
[3] https://www.ijircst.org/DOC/6_irp6 81.pdf
[4] https://www.researchgate.net/publi cation/257723864_Research_on_C loud_Data_Storage_Technology_a nd_Its_Architecture_Implementatio n
[5] https://typeset.io/papers/enterpriseweb-based-file-management-asystem-architecture-o51krvvtkz
[6] https://aws.amazon.com/comprehend/
[7] https://aws.amazon.com/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)