iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.77921

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 14 Issue III Mar 2026- Available at www.ijraset.com

AI-Based Email Sorting and Summarization

Ashna Rachel Mathew[1], Gayathri Suri Varrier[2], Neeraj S[3], Tia Bobby[4]
Computer Scxience, Mar Baselios Christian College of Engineering and Technology, Kuttikanam, Kerala, India

Abstract: ***The rapid growth of digital communication has resulted in a significant increase in the volume of emails received by individuals and organizations. Managing large numbers of emails often leads to reduced productivity and information overload. This paper proposes an Artificial Intelligence (AI) based email management system that automatically categorizes, summarizes, prioritizes, and schedules emails using Natural Language Processing (NLP) techniques. The system integrates machine learning models with Gmail API and Google Calendar API to provide automated inbox organization and productivity support. Emails are categorized into predefined groups such as work, personal, spam, and promotions. Long email messages are summarized using the Bidirectional and Auto-Regressive Transformer (BART) model to extract the most relevant information and present it in concise form. The system also identifies meeting schedules in emails and automatically creates reminders in the user's calendar to ensure important events are not missed. The proposed architecture is implemented using Python and Flask as the backend framework. Experimental observations indicate that the proposed system significantly improves email readability and reduces the time required to process large volumes of messages. The system demonstrates the potential of AI-powered solutions in improving digital communication efficiency.***

Keywords: Artificial Intelligence, Email Classification, Email Summarization, Natural Language Processing, BART Model, Gmail API, Intelligent Email Management

I. INTRODUCTION

Email is one of the most widely used digital communication platforms in both professional and personal environments. With the rapid expansion of online services and corporate communication systems, users receive a large number of emails every day. These emails may include work-related messages, promotional advertisements, newsletters, spam content, and personal communications. Managing such a large volume of emails manually becomes time-consuming and inefficient.

Traditional email systems provide simple filtering options such as folders and rule-based sorting. However, these approaches require manual configuration and do not understand the semantic meaning of the email content. As a result, users still spend significant time reading, sorting, and prioritizing messages.

Recent developments in Artificial Intelligence and Natural Language Processing have enabled the creation of intelligent systems that can understand textual data and perform automated decision-making. Machine learning models can analyze email content, classify messages into categories, summarize lengthy text, and extract useful information.

This research proposes an AI-based Email Sorting and Summarization System that automates inbox management. The system categorizes emails into work, spam, promotion, and personal categories. It also summarizes lengthy emails to provide quick insights. Additionally, the system detects meeting schedules in emails and integrates them with Google Calendar to generate reminders.

The primary objective of this research is to reduce email overload, improve productivity, and provide users with an intelligent assistant for managing digital communication efficiently.

II. RELATED WORK

Several researchers have explored machine learning techniques for email classification and spam detection. Early systems relied on rule-based filtering and statistical models such as Naïve Bayes classifiers and Support Vector Machines.

Research by Ayodele et al. explored the use of machine learning techniques for automated email classification and summarization. Their work demonstrated that intelligent classification models could significantly reduce manual email sorting.

Alsmadi and Alhami proposed clustering and classification approaches for organizing email messages. Their system analyzed textual patterns in emails and grouped them into clusters based on content similarity.

Recent research has also focused on email summarization. Oya and Carenini introduced a model for extractive email summarization that identifies important segments of email conversations. However, extractive summarization methods often produce fragmented results.

©IJRASET: All Rights are Reserved

With the emergence of deep learning and transformer architectures, advanced language models such as BERT, GPT, and BART have significantly improved text summarization performance. Transformer-based models can understand contextual relationships between words and generate more coherent summaries.

Despite these developments, many existing systems focus on only one aspect of email management, such as spam detection or summarization. The proposed system integrates multiple functionalities including classification, summarization, prioritization, and meeting detection into a unified intelligent framework.

III. PROPOSED SYSTEM

The proposed system is designed as an intelligent email management assistant capable of automating several inbox-related tasks. The system integrates machine learning models, Gmail services, and scheduling tools to improve email organization.

The main functionalities of the proposed system include:

•Automatic email classification
• Email summarization for long messages
• Email prioritization based on importance
• Meeting detection and calendar scheduling
• Notification reminders for scheduled meetings

The system architecture consists of three major components: data acquisition, data processing, and user interface modules. The Gmail API is used to retrieve email data securely. The extracted email content is then processed using Natural Language Processing techniques.

The BART transformer model is used to generate concise summaries of long email messages. The system also analyzes the content of emails to identify meeting-related keywords and extracts scheduling information such as date and time. This information is then integrated with Google Calendar using the Google Calendar API.

The final results are displayed through a web-based interface developed using the Flask framework.

IV. SYSTEM ARCHITECTURE

The architecture of the system consists of multiple layers that interact with each other to process and analyze email data.

1) Data Collection Layer: This layer is responsible for retrieving email data from the user's Gmail account using Gmail API. Secure authentication is handled using OAuth 2.0 to protect user credentials.
2) Processing Layer: The processing layer performs Natural Language Processing operations including text cleaning, tokenization, and keyword extraction. The classification module categorizes emails into predefined groups.
3) Summarization Module: The summarization module uses the BART model to generate concise summaries from long email messages.
4) Scheduling Module: The scheduling module detects meeting-related information and integrates it with Google Calendar to automatically generate reminders.
5) User Interface Layer: The final processed results are presented through a web-based interface built using Flask, allowing users to view categorized emails and summarized content.

V. IMPLEMENTATION

The system is implemented using Python as the primary programming language. Flask is used to develop the web-based backend interface that allows interaction with Gmail services. The Gmail API is used to retrieve email messages from the user's inbox. Email content is preprocessed using Natural Language Processing techniques to remove unnecessary characters and stop words.

The BART transformer model is used to generate summaries of email content. This model is capable of producing coherent and contextually accurate summaries from long text passages. The Google Calendar API is used to add meeting reminders automatically when scheduling information is detected in emails. This integration ensures that important meetings are not missed.

The system prioritizes emails by analyzing factors such as sender importance, urgency keywords, and email content relevance.

VI. RESULTS AND DISCUSSION

The proposed system was tested using a dataset of emails obtained through Gmail API. The classification module successfully categorized emails into work, personal, spam, and promotional categories.

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

The summarization module effectively reduced long email content into short summaries that preserved the key information. This significantly reduced the time required for users to read and understand messages.

Meeting detection functionality was also tested, and the system successfully identified scheduling information and added reminders to Google Calendar.

Overall, the system improved email management efficiency by automating tasks that would otherwise require manual effort.

VII. CONCLUSION AND FUTURE WORK

This research presented an AI-based email sorting and summarization system designed to reduce email overload and improve productivity. The system integrates machine learning, natural language processing, and Gmail services to provide intelligent inbox management.

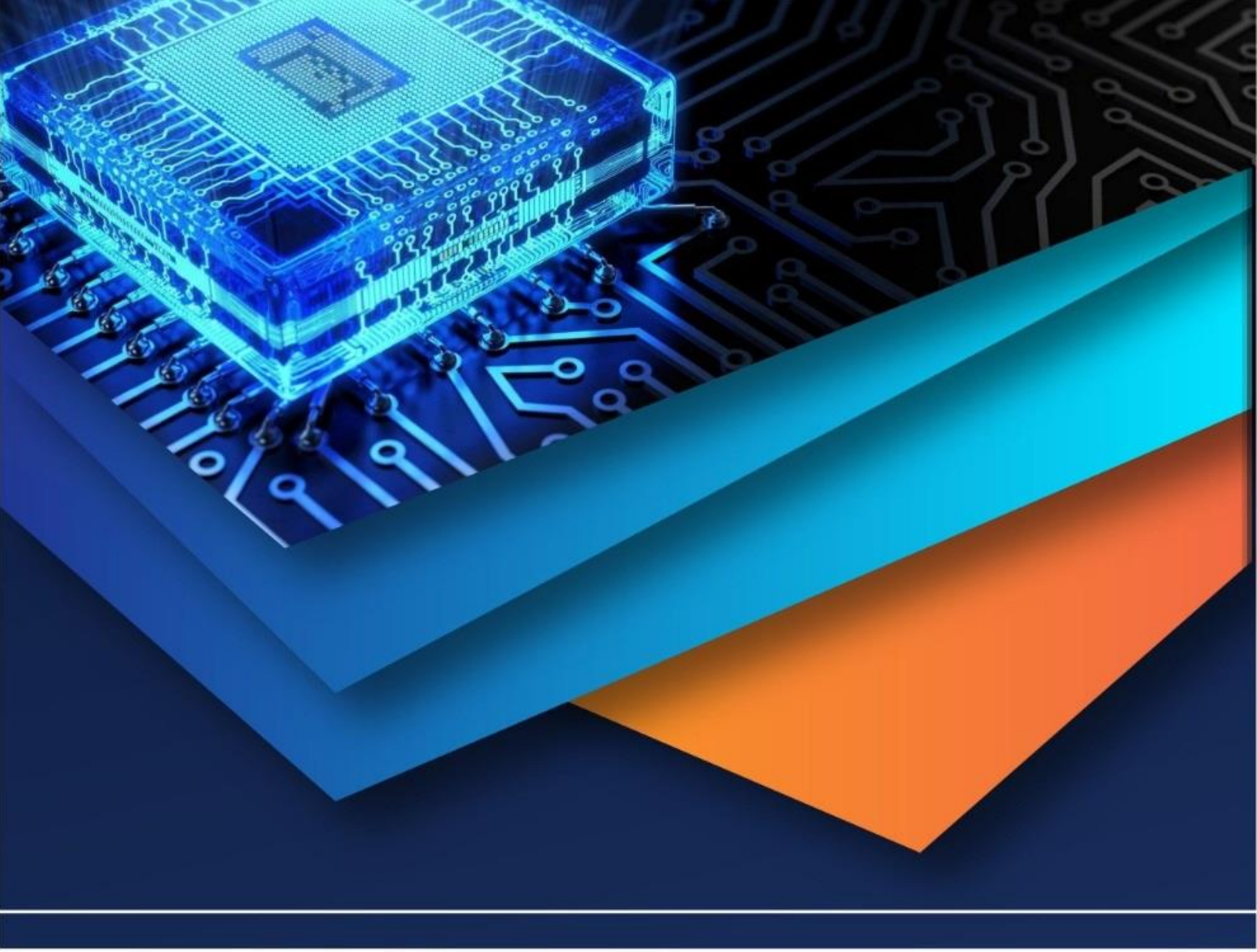
The proposed system demonstrates the potential of AI technologies in improving digital communication systems by automating repetitive tasks and organizing large volumes of information.

Future work may include integrating voice-based interaction systems similar to digital assistants, implementing smart reply generation for automated responses, and extending the system to support multilingual email processing. Additional improvements may include mobile application integration and enhanced machine learning models for improved classification accuracy.

REFERENCES

[1] T. Ayodele, R. Khusainov, and D. Ndzi, “Email Classification and Summarization: A Machine Learning Approach.”
[2] T. Ayodele, Z. Zhou, and R. Khusainov, “Email Grouping and Summarization: An Unsupervised Learning Technique.
[3] I. Alsmadi and I. Alhami, “Clustering and Classification of Email Contents.
[4] T. Oya and G. Carenini, “Summarization and Dialogue Act Modeling on Email Threads.
[5] E. G. Dadaa, J. Bassia, H. Chiroma, S. Muhammad, and A. Adetunmbi, “Machine Learning for Email Spam Filtering: Review, Approaches and Open Problems.
[6] Y. Zhang et al., “EMAILSUM: Abstractive Email Thread Summarization.
[7] K. Zhang, J. Chen, and D. Yang, “Focus on the Action: Learning to Highlight and Summarize Jointly for Email To-Do Items Summarization.
[8] F. Zhu, “Automatic Classification for Unlabeled Email Messages into Folder.
[9] N. S. Khedkar, P. Kumar, and A. K. E. Rojas, “Email Intelligence: An AI-Driven Framework for Context-Aware Office Communication Management.

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



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)