IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CONTEXT-AWARE EMAIL AUTOMATION USING AGENTIC AI AND LARGE LANGUAGE MODELS

[1]Ramya Atchuta, [2] Chandrika Rajulapati, [3]Kurma rao Surni, [4]Yaswanth Reddy Puritipati, [5]S V Manikanta Tagore Yerrapothu

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1]Department of Computer Science and Engineering,

[1]Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: Managing large volumes of emails is time-consuming and often reduces productivity due to repetitive reading, drafting, and replying tasks. To address this challenge, this project proposes an Intelligent Email Management and Automation System using Agentic AI. The system employs Large Language Models (Gemini LLM) to produce context-specific professional email content from limited user prompts. Authentication to Gmail via OAuth 2.0 enables safe email sync and traffic control. The backend, developed in FastAPI, manages routing, intent identification, prompt assembly, and email dispatch, all presented to the user through a React frontend for drafting, rechecking, and amending content. Moving away from classic rule or robotic process automation (RPA) solutions, this API-based method guarantees quicker, more adjustable, and consistent operation. The framework facilitates tone modulation, individualization and complete email lifecycle automation, greatly lowers the time spent and increases the communication effectiveness. Empirical results show that the integrated system boosts productivity, sustains the stable email quality and exhibits the success of Agentic AI in practical email management scenarios.

Index Terms - Agentic AI, Intelligent Email Management, Email Automation, Large Language Models, Gemini LLM, FastAPI, OAuth 2.0, Context-Aware Email Generation, SMTP Automation, React Frontend.

I. INTRODUCTION

Email is a popular way to communicate professionally and academically, even with the rise of more recent digital communication methods. Managing the desktop inbox has become more difficult due to the sharp rise in the quantity of emails delivered every day. The management of massive quantity of inbound and outbound emails incurs high manual effort leading to decreased productivity and responsiveness. Industry statistics reveal a continuous increase in email volumes worldwide. Billions of emails are sent daily for different communication purposes [2]. These factors highlight the growing necessity for intelligent automation mechanisms to streamline and optimize email communication processes [1]. Basic folder-based classification, keyword detection, and filtering are offered by traditional email management systems. The increasing volume of communications is only slightly reduced by current email management solutions, which are unable to understand email context, generate appropriate responses, or automate repetitive communication tasks. Users still spend a lot of time manually handling routine contact, including creating follow-up reminders and preparing responses.

Natural language creation and understanding have significantly improved with latest improvements in Large Language Models (LLMs). Foundational research such as Transformer architecture [8] and BERT have provided powerful mechanisms for contextual language processing. Subsequent models like GPT-3 [3] InstructGPT [5] and LLaMA [6] have demonstrated impressive capabilities in generating

coherent context-aware text More recently multimodal models such as Gemini [4] and dialogue models like ChatGPT [7] have proved that LLMs can actively support users in complex language tasks including professional email composition. Parallel to LLM exploration the development of Agentic AI architectures has made it possible for autonomous task execution by sensing reasoning and acting intelligent agents with little human direction [14]. The Model Context Protocol (MCP) has unified context sharing among AI models and external tools facilitating smooth multi-step reasoning and robust task automation [13].

These developments provide a strong foundation for building intelligent email management systems that go beyond simple filtering to offer automated drafting, context-aware responses, and end-to-end workflow automation. Motivated by these advancements, this project proposes an Agentic AI-based Intelligent Email Management System named InboXpert. The system speeds up and simplifies communication. It does this by using email generation powered by LLMs, a safe connection to Gmail's API through OAuth 2.0 [22], and automation based on agents.

II. LITERATURE REVIEW

The rapid growth of email communication has introduced significant challenges in managing large volumes of messages, leading to reduced productivity and increased manual effort in handling inbox activities. Studies highlight the limitations of traditional email management approaches, which rely on basic filtering and fail to address contextual understanding and automation needs [1]. Statistical reports further emphasize the continuous rise in global email usage, projecting increasing communication burdens on users and reinforcing the necessity for intelligent automation solutions [2].

Advancements in natural language processing have been largely driven by the development of Transformer-based architectures, which enable efficient modeling of contextual relationships in large text sequences [8]. Models such as BERT introduced bidirectional contextual understanding, significantly improving performance in various language tasks including classification and sentiment analysis [9]. Generative models like GPT-3 further enhanced text generation capabilities through large-scale pretraining, while InstructGPT improved alignment with human intent using reinforcement learning techniques [3], [5]. Additionally, LLaMA demonstrated efficient and scalable model performance, making deployment more practical for real-world applications [6]. More recently, multimodal and conversational models such as Gemini and ChatGPT have expanded the capabilities of AI systems by enabling context-aware reasoning and human-like interaction, making them highly suitable for tasks such as professional email generation [4], [7]. Supporting research also shows that advanced deep learning and optimization techniques improve contextual modeling and prediction accuracy across diverse domains, reinforcing the applicability of intelligent systems in automation tasks [28]–[31].

Semantic understanding plays a critical role in intelligent email systems, where approaches like Sentence-BERT enable efficient generation of semantically meaningful embeddings, allowing accurate similarity comparison and improved intent detection [10]. Furthermore, recent developments in tool-augmented language models have significantly enhanced automation capabilities. Frameworks such as ReAct integrate reasoning with action execution, enabling models to dynamically interact with external tools [11]. Toolformer extends this concept by allowing models to autonomously learn tool usage through self-supervision, facilitating seamless API interaction [12]. The Model Context Protocol (MCP) further strengthens these systems by ensuring consistent context management across multi-step workflows, which is essential for maintaining coherence in automated communication tasks [13].

The evolution of Agentic AI has introduced autonomous agents capable of reasoning, planning, and executing tasks with minimal human intervention. Generative agent frameworks demonstrate how agents can maintain memory, adapt behavior, and perform complex decision-making processes [14]. Practical implementations of AI agents in enterprise environments highlight their effectiveness in automating application-level workflows, including email communication systems [15]. Techniques such as Chain-of-Thought prompting enhance reasoning capabilities in language models, improving their ability to generate contextually accurate and structured responses [16]. Foundational theories of intelligent agents define key characteristics such as autonomy and proactivity, which are essential for designing effective automation systems [17]. Advanced frameworks like HuggingGPT further demonstrate how language models can coordinate multiple specialized tools to accomplish complex tasks, while studies on agent interaction models provide insights into scalable and adaptive system design [18], [19]. Comprehensive surveys on LLM-based agents also outline current challenges and future directions in building reliable autonomous systems [20].

In addition to AI advancements, supporting technologies play a crucial role in enabling intelligent email automation. The Gmail API provides controlled access to email data, while OAuth 2.0 ensures secure authentication without exposing user credentials [21], [22]. The SMTP protocol defines standardized mechanisms for reliable email transmission across networks [23]. Modern backend frameworks such as FastAPI enable efficient API development with asynchronous processing capabilities, and frontend technologies like React support the creation of responsive and user-friendly interfaces for email management systems [24], [25]. Furthermore, AI development frameworks such as LangChain facilitate modular integration of LLMs, agents, and tools, while experimental systems like AutoGPT demonstrate the potential of fully autonomous AI agents capable of task decomposition and iterative execution [26], [27].

III. METHODOLOGY

The proposed methodology follows a modular, API-driven workflow to achieve intelligent email automation using Agentic AI principles. Starting at the React frontend, the user enters the new email, or chooses an existing email to reply to, and the request is routed to the FastAPI backend that takes care of all routing and processing. Safe authentication and email sync follow the OAuth 2.0 via the Auth and Sync routers to enable safe access to Gmail services. Retrieved emails, user information and generated drafts are stored and handled by a SQLAlchemy database. For email generation, the backend constructs a context-aware prompt and forwards it to the Gemini LLM API via the Generate Router, which produces a professional and relevant email draft. The created material is returned to the user for inspection and editing. After approval, the Send Router automatically sends the email via SMTP automated delivery, closing the circle of the email process. This formal approach guarantees safe integration, smart answer creation, scalability, and less human work in e-mail correspondence.

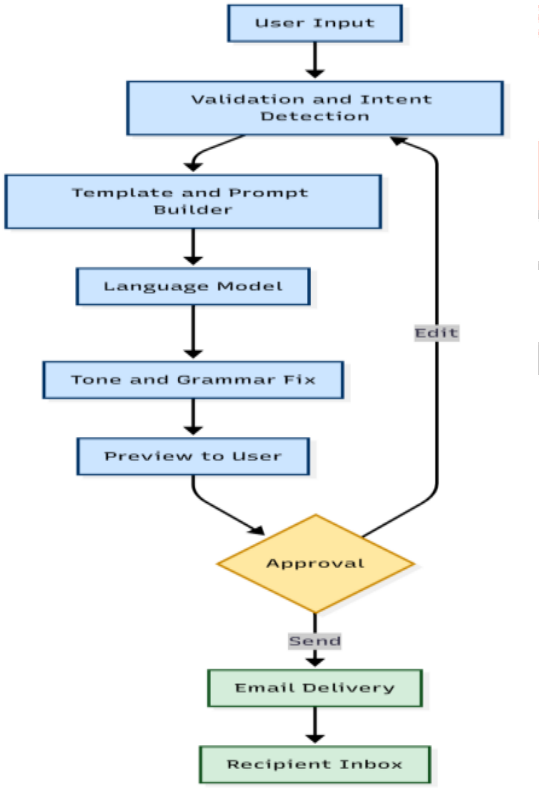


Fig. 1. Proposed system architecture showing the flow from user input through intent detection, prompt construction, LLM-based generation, refinement, user review, and automated email delivery.

A. Proposed Undertaking

The proposed system aims to develop an Intelligent Email Management and Automation solution using Agentic AI to minimize manual effort in daily email communication. Unlike traditional systems that rely on rigid templates or rule-based automation, the proposed approach leverages a Large Language Model (Gemini LLM) to generate contextually relevant, personalized, and professional email responses from minimal user input. Secure integration with Gmail through OAuth 2.0 enables efficient email synchronization while maintaining data privacy. The FastAPI-based backend manages authentication, email retrieval, intent detection, and response generation, while the React frontend provides an intuitive interface for composing, reviewing, and managing emails.

The system follows an end-to-end agentic workflow in which intelligent agents collaboratively perform context extraction, prompt generation, email composition, and automated delivery. Model Context Protocol (MCP) principles ensure consistent context handling across multi-step processes, enabling reliable and coherent outputs. Generated emails are presented to users for validation, ensuring human-in-the-loop control before final submission, after which emails are automatically delivered using SMTP-based automation. The architecture is designed to be scalable and flexible, supporting features such as tone customization, personalization, multilingual capabilities, and intelligent email prioritization. Overall, the system enhances productivity, maintains consistent communication quality, and demonstrates the practical effectiveness of Agentic AI in real-world email automation.

B. System Architecture

The proposed architecture enables intelligent, agent-based email automation with a focus on precision, personalization, and user control. The process begins with user input in the form of a short prompt or an existing email, which is analyzed to identify intent such as inquiry, request, invitation, or response. Based on this intent, a dynamic prompt construction module generates an optimized input by incorporating relevant context and user preferences. This structured prompt is then processed by the Gemini LLM, which utilizes a Transformer-based architecture to generate a coherent, context-aware, and professionally structured email draft.

Following generation, the system applies tone and grammar refinement to enhance clarity, formality, and linguistic quality. The generated draft is presented to the user through a preview interface, allowing further editing and validation. An integrated feedback loop enables iterative refinement without repeating the entire workflow, ensuring efficiency and flexibility. Once approved, the finalized email is automatically delivered using SMTP-based automation, ensuring reliable transmission. Overall, the architecture combines intelligent automation with human oversight to deliver scalable, context-aware, and high-quality email communication using Agentic AI.

IV. IMPLEMENTATION

The proposed system consists of multiple integrated modules that collectively enable intelligent email automation. A React-based user interface provides an interactive environment where users can compose new emails, reply to existing ones, and review generated drafts while selecting preferences such as tone, length, and language. Secure authentication and email synchronization are handled through OAuth 2.0, allowing safe access to Gmail services without exposing user credentials. The system further incorporates intent detection and validation mechanisms that analyze user input and email context to identify the purpose of communication, such as requests, responses, or invitations, ensuring accurate and meaningful processing.

Based on the identified intent, a prompt and template construction module dynamically generates optimized inputs by combining user data, context, and preferences. These structured prompts are processed by the Gemini Large Language Model, which leverages Transformer-based architecture and self-attention mechanisms to produce coherent, context-aware, and professionally structured email drafts. To enhance output quality, a tone and grammar refinement component ensures clarity, correctness, and stylistic consistency. All generated drafts, user data, and contextual information are efficiently managed using a SQLAlchemy-based database, enabling persistent storage and seamless retrieval across sessions.

The system's functionality is further strengthened by intelligent algorithms that drive each stage of the workflow. The Gemini LLM serves as the core engine for natural language generation, while intent detection relies on semantic understanding to interpret user inputs accurately. A prompt construction algorithm ensures structured and personalized inputs, and a refinement algorithm enhances linguistic

quality and tone consistency. These components work together to maintain high accuracy, relevance, and personalization in generated email content.

Finally, the automated email dispatch mechanism ensures reliable delivery through SMTP-based execution. Once the user approves the generated draft, the system formats the email according to standard protocols and securely sends it to the recipient while handling potential transmission errors. Overall, the integration of modular components and intelligent algorithms enables a scalable, efficient, and context-aware email automation system powered by Agentic AI.

V. RESULTS AND DISCUSSION

The proposed Agentic AI-based Intelligent Email Management System was evaluated through multiple real-world email interaction scenarios, including formal requests, appointment scheduling, and casual communication. Users interacted with the system via a React-based interface by providing prompts of varying granularity, and the system consistently generated contextually relevant, grammatically correct, and professionally structured email drafts. The evaluation highlights the effectiveness of the Gemini LLM in accurately capturing user intent and contextual information, producing high-quality outputs even from minimal input.

To evaluate the quality of text generation, standard Natural Language Generation (NLG) metrics were employed instead of traditional classification metrics. These metrics focus on assessing both system efficiency and the quality of generated email content. Response latency measures the time taken from prompt submission to draft generation, indicating system responsiveness and API efficiency. It is calculated as:

$$\text{Response Latency} = T_{draft\_received} - T_{prompt\_submitted}$$

Human evaluation scores were obtained using a 5-point Likert scale across four dimensions: relevance (R), fluency (F), professionalism (P), and completeness (C). The overall quality score is computed as:

$$\text{Quality Score} = \frac{R+F+P+C}{4}$$

The edit rate represents the percentage of generated drafts that required user modification before sending, reflecting generation accuracy and alignment with user expectations. It is calculated as:

$$\text{Edit Rate} = \frac{\text{Number of Edited Drafts}}{\text{Total Generated Drafts}} \times 100\%$$

The task completion rate measures the proportion of emails successfully sent without regeneration, indicating the robustness and efficiency of the system workflow. It is calculated as:

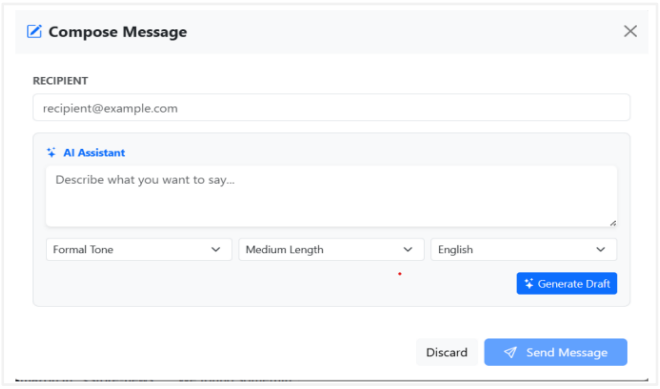
$$\text{Task Completion Rate} = \frac{\text{Successfully Sent Emails}}{\text{Total Attempted Emails}} \times 100\%$$

The system performance was further analyzed in terms of end-to-end automation efficiency, including email synchronization, draft generation, user review, and SMTP-based delivery. The FastAPI backend ensured low response latency and reliable API communication, while OAuth-based authentication enabled secure email synchronization without exposing user credentials. These components collectively contribute to a stable and efficient email automation pipeline.

Table 1: Performance Evaluation Metrics

Metric	Value	Interpretation
Average Response Time	< 3 seconds	Fast generation suitable for real-time use
Relevance Score	4.2 / 5.0	High alignment with user intent
Fluency Score	4.5 / 5.0	Grammatically correct and natural
Professionalism Score	4.3 / 5.0	Appropriate tone and formatting
Completeness Score	4.1 / 5.0	Comprehensive draft coverage
Edit Rate	< 15%	Minimal corrections required
Task Completion Rate	> 90%	Reliable end-to-end workflow

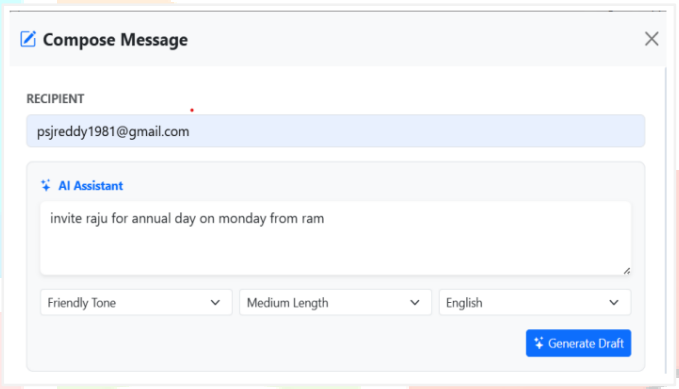
This GUI is the React frontend for the email composition. Users fill in the recipient email address, and a concise text instruction describing the message to be sent. Additional options for tone, email length, and language can be set, to generate a customized email. "Generate Draft" button pushes current user input



to the backend, to process it with the AI email generator.

Fig. 2. Email Composition Interface

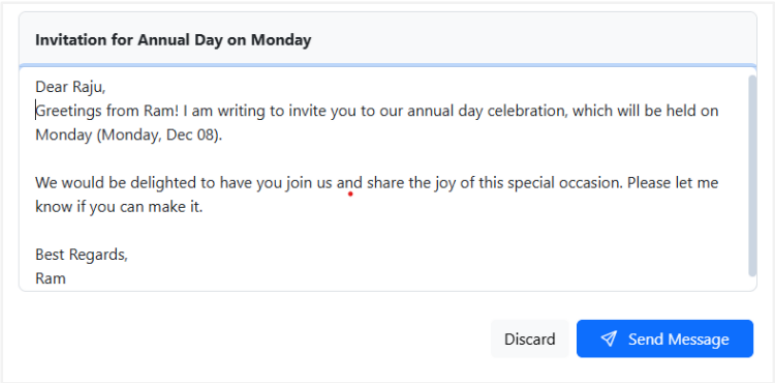
This illustration shows the AI assistant input step, in which the user provides a short prompt (e.g. an invitation request) and chosen preferences (e.g. friendly tone, medium-length answer). Based on this input, the system infers the user's intention and formulates an improved prompt for the Gemini LLM. This



step exemplifies that the user intervention can be kept to a minimum for producing the entire, relevant email body, thus confirming the benefits of the upcoming Agentic AI concept.

Fig. 3. AI-Assisted Prompt Input

This screen shows the generated email draft produced by the Gemini LLM after processing the user prompt and context. The draft includes a relevant subject line, structured body content, and a professional closing. The user is provided with options to review, edit, discard, or send the email. This stage ensures human-in-the-loop validation, allowing users to maintain control over the final message



before it is delivered through SMTP automation to the recipient's inbox.

Fig. 4. Generated Email Draft

Overall, the experimental results confirm that the proposed system enhances productivity, ensures consistent communication quality, and validates the practical applicability of Agentic AI for intelligent email automation.

VI. CONCLUSION

This project presented an Intelligent Email Management and Automation System using Agentic AI to address the challenges of handling large volumes of email communication. By leveraging a Deep Learning–based Large Language Model (Gemini LLM) built on Transformer architecture, the system enables context-aware, professional email generation from minimal user input. The system establishes a complete email system that combines multiple technologies including FastAPI, React, OAuth 2.0, SQLAlchemy, and Gmail API with SMTP automation to provide secure and scalable and efficient email synchronization and draft generation and user review and delivery processes. The Model Context Protocol (MCP) principles provide a framework that maintains context consistency throughout multi-step processes, while the agentic architecture enables automatic task management. The experimental results together with system evaluation confirm that the proposed method decreases manual work requirements, enhances response accuracy, and boosts total work efficiency when compared with conventional template-based and rule-driven email systems. The human-in-the-loop design maintains system reliability while granting users the ability to control system operations, and the modular architecture supports future development of multilingual capabilities together with intelligent email prioritization and advanced summarization features. The combination of semantic embeddings and contextual understanding enables the system to accurately detect user intent through short user prompts.

The project demonstrates how Agentic AI functions effectively in real-world email automation systems which provide businesses with modern communication management solutions. Future work may explore integration with additional email providers, enhanced personalization using user behavior patterns, and deployment of lightweight local LLM alternatives for privacy-sensitive environments.

REFERENCES

[1] J. Burkel and D. Anderson, "The evolving landscape of email communication and management strategies," J. Inf. Syst. Technol. Manage., vol. 16, no. 2, pp. 135–148, 2019.

[2] Statista, "Number of emails sent and received per day worldwide from 2017 to 2026," 2023. [Online]. Available: https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/

[3] T. Brown et al., "Language models are few-shot learners," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 1877–1901. [Online]. Available: https://arxiv.org/abs/2005.14165

[4] R. Anil et al., "Gemini: A family of highly capable multimodal models," Google DeepMind, Dec. 2023. [Online]. Available: https://arxiv.org/abs/2312.11805

[5] L. Ouyang et al., "Training language models to follow instructions with human feedback," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), 2022, pp. 27730–27744. [Online]. Available: https://arxiv.org/abs/2203.02155

[6] H. Touvron et al., "LLaMA: Open and efficient foundation language models," Meta AI, Feb. 2023. [Online]. Available: https://arxiv.org/abs/2302.13971

[7] OpenAI, "GPT-4 Technical Report," arXiv preprint, Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[8] A. Vaswani et al., "Attention is all you need," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008. [Online]. Available: https://arxiv.org/abs/1706.03762

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics (NAACL), 2019, pp. 4171–4186. [Online]. Available: https://arxiv.org/abs/1810.04805

[10] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2019, pp. 3982–3992. [Online]. Available: https://doi.org/10.18653/v1/D19-1410

[11] S. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," in Proc. Int. Conf. Learning Representations (ICLR), 2023. [Online]. Available: https://arxiv.org/abs/2210.03629

[12] T. Schick et al., "Toolformer: Language Models Can Teach Themselves to Use Tools," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), 2023. [Online]. Available: https://arxiv.org/abs/2302.04761

[13] Anthropic, "Introducing the Model Context Protocol," Anthropic News, 2024. [Online]. Available: https://www.anthropic.com/news/model-context-protocol

[14] J. S. Park et al., "Generative Agents: Interactive Simulacra of Human Behavior," in Proc. 36th Annu. ACM Symp. User Interface Softw. Technol. (UIST), 2023, pp. 1–22. [Online]. Available: https://arxiv.org/abs/2304.03442

[15] Amazon Web Services, "Automate tasks in your application using AI agents," AWS Documentation, 2024. [Online]. Available: https://docs.aws.amazon.com/bedrock/latest/userguide/agents.html

[16] J. Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), 2022. [Online]. Available: https://arxiv.org/abs/2201.11903

[17] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice," The Knowledge Engineering Review, vol. 10, no. 2, pp. 115–152, 1995. [Online]. Available: https://doi.org/10.1017/S0269888900008122

[18] Y. Shen et al., "HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face," in Proc. Advances Neural Inf. Process. Syst. (NeurIPS), 2023. [Online]. Available: https://arxiv.org/abs/2303.17580

[19] J. Floch et al., "A survey of flexible agent interaction approaches," Multiagent and Grid Systems, vol. 9, no. 3, pp. 159–185, 2013. [Online]. Available: https://doi.org/10.3233/MGS-130208

[20] L. Wang et al., "A Survey on Large Language Model based Autonomous Agents," arXiv preprint, 2023. [Online]. Available: https://arxiv.org/abs/2308.11432

[21] Google Developers, "Gmail API documentation," Google, 2024. [Online]. Available: https://developers.google.com/gmail/api

[22] D. Hardt, "RFC 6749: The OAuth 2.0 Authorization Framework," Internet Engineering Task Force (IETF), Oct. 2012. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc6749

[23] J. Klensin, "RFC 5321: Simple Mail Transfer Protocol," Internet Engineering Task Force (IETF), Oct. 2008. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc5321

[24] S. Ramírez, "FastAPI: Modern, fast web framework for building APIs with Python," FastAPI Documentation, 2024. [Online]. Available: https://fastapi.tiangolo.com/

[25] Meta Platforms, "React: A JavaScript library for building user interfaces," React Documentation, 2024. [Online]. Available: https://react.dev/

[26] LangChain, "LangChain: Building applications with LLMs through composability," LangChain Documentation, 2024. [Online]. Available: https://python.langchain.com/

[27] Significant Gravitas, "AutoGPT: An Autonomous GPT-4 Experiment," GitHub, 2023. [Online]. Available: https://github.com/Significant-Gravitas/AutoGPT

[28] A. Koteswaramma, M. Babu Rao, and G. Jaya Suma, "An intelligent adaptive learning framework for fake video detection using spatiotemporal features,"Signal, Image and Video Processing, Springer Nature, 2024.

[29] M. Baburao, CH. Kavitha, D. N. V. S. L.S. Indira, and M. V. N. S. S. R. K. S. Somayajulu, "Content-based video recommendation system (CBVRS): a novel approach to predict videos using multilayer feed forward neural network and Monte Carlo sampling method," Multimedia Tools and Applications, vol. 82, no. 2, pp. 6965-6991, 2022.

[30] G. Keerthi, Y. Ashvitha, V. S. R. Reddy, R. M. Krishna, and P. Sandeep, "Integrating Convolutional Neural Networks (CNN) and Machine Learning for Accurate Identification of Autism Spectrum Disorder Using Facial Biomarkers," 2024 IEEE International Conference on Emerging Systems and Intelligent Computing (ESIC), 2024.

[31] G. Keerthi, G. Ramachandran, and G. V. S. N. R.V. Prasad, "Autism spectrum disorder prediction using LASSO regularised bat search optimisation, " International Journal of Services Operations and Informatics, 2024.