
JOBOT: AI-Powered LinkedIn & Email Outreach Platform

Saivarsha Raju
New York University
saivarsharaju@nyu.edu

Malavika Subramanian
New York University
ms16028@nyu.edu

Abstract

We present **JOBOT**, a production-grade platform that automates personalized job outreach by combining AI-driven text generation with web automation. JOBOT streamlines the networking process for job seekers by leveraging Retrieval-Augmented Generation (RAG) techniques, utilizing ChromaDB for vector storage, all-MiniLM-L6-v2 embeddings for semantic search, and a lightweight Tiny GPT-2 model for message composition. The system integrates a Streamlit-based frontend with backend automation scripts that handle LinkedIn and Gmail outreach via Selenium-controlled browsers, incorporating stealth modes, retries, and fallback mechanisms. User credentials and sensitive data are managed securely through in-memory encryption and environment variables, ensuring privacy and compliance.

We situate JOBOT within the broader context of AI-assisted professional outreach, comparing it with commercial tools such as Apollo.io and Lemlist, and highlight its advantages in privacy, cost, and ease of personalization. Ethical considerations regarding consent, compliance with platform terms of service, and responsible automation are also discussed. Finally, we outline potential future enhancements, including advanced language models, interactive UX improvements, multi-channel expansion, and the development of JOBOT as a full-fledged AI career assistant. Through JOBOT, we demonstrate how modern NLP and automation technologies can accelerate and personalize job search networking in an effective and responsible manner.

1 Introduction

Effective professional outreach is critical in today’s interconnected digital economy, playing a pivotal role in recruitment, sales, networking, and business development. However, executing personalized outreach at scale presents substantial challenges due to the complexity, repetitive nature, and labor-intensive requirements of manual processes. Professionals often spend considerable time individually crafting messages tailored to specific recipients while navigating multiple platforms such as LinkedIn, email, and various professional networking tools.

Existing automated solutions partially mitigate these challenges but frequently lack integration, adaptability, and sufficient personalization, resulting in messages that feel impersonal and ineffective. Current tools in the market typically focus on isolated tasks such as email automation, basic workflow management, or simple browser interactions. These fragmented solutions fail to provide an integrated, intelligent, and adaptive outreach system, often forcing users to context-switch between platforms and interfaces, thereby reducing productivity and message coherence.

This project aims to bridge these gaps by developing a unified, intelligent outreach automation assistant powered by natural language processing (NLP), advanced language models, and robust browser automation technologies. By leveraging these technologies, our system significantly reduces manual effort while enhancing message quality and personalization. The proposed system allows

users to define outreach tasks using natural language, after which the assistant autonomously manages target discovery, context-aware message generation, and multi-platform message delivery.

Consequently, the system streamlines the entire outreach process, making it more efficient, scalable, and impactful. In this report, we detail the design, implementation, evaluation, and future prospects of our NLP-driven automation assistant, highlighting its potential to significantly transform professional outreach practices.

2 Problem

Job seekers often spend significant time manually crafting and sending personalized cold emails or LinkedIn messages to recruiters, employees, or alumni. This process is inherently slow, repetitive, and inefficient. Currently, there is no scalable solution that enables personalized outreach tailored to a user's résumé and the specific job they are targeting.

Effective outreach and workflow management remain critical yet challenging for professionals. Several factors contribute to this complexity. First, the process is highly time-intensive. Identifying suitable contacts, drafting individualized messages, and scheduling outreach across different platforms require considerable human effort, often detracting from other high-value activities.

Second, achieving meaningful personalization is difficult without advanced AI support. Generic messages typically fail to engage recipients effectively, while crafting highly personalized communications—such as referencing mutual connections, shared interests, or relevant background details—requires substantial manual effort. Recent advancements in large language models such as OpenAI's GPT family (6) have demonstrated significant potential in addressing this challenge through context-aware language generation.

Third, as networking efforts scale, manual handling of communications becomes unsustainable. Without automation, maintaining high-quality personalized outreach across a large pool of potential contacts is practically infeasible.

Finally, the fragmentation of existing tools further exacerbates these challenges. Solutions such as Zapier (3), UiPath (4), and Simplify (5) focus on narrow tasks like email automation or app integrations but lack cohesive support for natural language understanding and multi-platform automation. In addition, coordinating the timing of outreach activities to optimize recipient engagement adds another layer of complexity that these tools do not address.

These challenges underscore the need for a unified and intelligent solution that automates target discovery, message personalization, and outreach scheduling across multiple platforms, all driven by natural language commands.

3 Related Work

The automation of personalized outreach has been a longstanding challenge in both academic research and industry. While significant progress has been made in automating repetitive business processes, existing solutions typically address isolated components of the outreach workflow and fall short of delivering a cohesive, adaptive system that combines personalization with scalable delivery.

Workflow automation platforms such as Zapier (3) and UiPath (4) have enabled users to streamline repetitive tasks by connecting disparate applications and automating predefined sequences of actions. However, these tools are primarily rule-based and lack the capacity for advanced natural language understanding or dynamic content generation. As a result, they are not well-suited for tasks requiring nuanced, context-aware communication.

Web automation frameworks like Selenium (1) and Puppeteer (2) offer powerful browser-level automation capabilities, frequently used for web scraping and task automation. Despite their technical strengths, these frameworks require explicit procedural instructions and are incapable of high-level decision-making or generating semantically rich communication without significant additional development effort.

Several platforms, including Simplify (5) and Lemlist, focus specifically on automating email outreach campaigns. These tools provide mechanisms for managing large-scale outreach through the use

of static templates and limited variable substitution for personalization. While effective for basic campaigns, they lack the integration of modern natural language processing techniques that enable adaptive, contextually relevant message generation based on user profiles and target audiences.

Advancements in large language models such as OpenAI’s GPT-4 (6) and Meta’s LLaMA have demonstrated impressive capabilities in generating coherent and human-like text. Foundational research, including the introduction of GPT-3 by Brown et al. (2020), showcased the potential for few-shot learning in text generation tasks. However, the integration of these models into end-to-end outreach systems that combine intelligent message generation with automated delivery and scheduling remains largely unexplored.

Emerging research has begun exploring hybrid approaches that integrate robotic process automation with large language models to enable adaptive workflows. Although these efforts show promise, they are generally experimental and lack the robustness, user accessibility, and generalizability required for real-world professional outreach applications.

In summary, while individual technologies address specific facets of the outreach problem, there remains a clear gap in providing a unified system capable of dynamic contact discovery, advanced language-driven message generation, and automated multi-platform delivery. The system proposed in this work aims to address these limitations by combining recent advancements in natural language processing with practical automation frameworks to create a comprehensive and scalable outreach solution.

4 System Design and Proposed Solution

This work proposes a unified, intelligent outreach automation system designed to streamline the job search process through advanced natural language processing, web automation, and secure credential management. The objective is to reduce the manual effort required for professional outreach while maintaining high levels of personalization, relevance, and efficiency across multiple communication platforms.

At the core of the system is a modular architecture that integrates an interactive user interface with backend components for message generation, contact discovery, and automated message delivery. The user interface, developed using the Streamlit framework, enables users to provide essential inputs such as résumés, job descriptions, and target criteria. Users can also preview and refine AI-generated messages before initiating the automated outreach workflow.

The message generation pipeline employs a lightweight context-aware prompt construction mechanism. Instead of relying on large-scale retrieval systems, the system analyzes the user’s resume and job description to extract important information related to skills, experiences, and role requirements. This extracted context is dynamically incorporated into the prompts for a compact language model (Tiny GPT-2), which produces personalized outreach messages tailored to the target recipient and the platform. When higher quality language generation is required, the system can leverage external APIs such as OpenAI’s GPT models to produce more sophisticated outputs.

Contact discovery is facilitated through a combination of third-party services and Web automation. The system integrates the Hunter.io API to retrieve verified professional email addresses. In parallel, it utilizes Selenium-based web automation to perform Google searches and extract relevant LinkedIn profiles. The data retrieved is filtered for relevance, and duplicate entries are automatically eliminated to ensure outreach efficiency.

Once the message content and contact information are prepared, the automation engine executes the delivery process. This engine leverages Selenium scripts to automate login procedures, navigate target platforms, and send personalized messages through LinkedIn and Gmail. Secure session management ensures that user credentials remain in memory only during active sessions and are never permanently stored, preserving user privacy. All message deliveries are tracked, and the system provides real-time feedback on the success or failure of outreach attempts.

The architecture of the proposed system is illustrated in Figure 1, which highlights the interaction between the user interface, the back-end modules, and the external services. The workflow for context-aware message generation is further detailed in Figure 2.

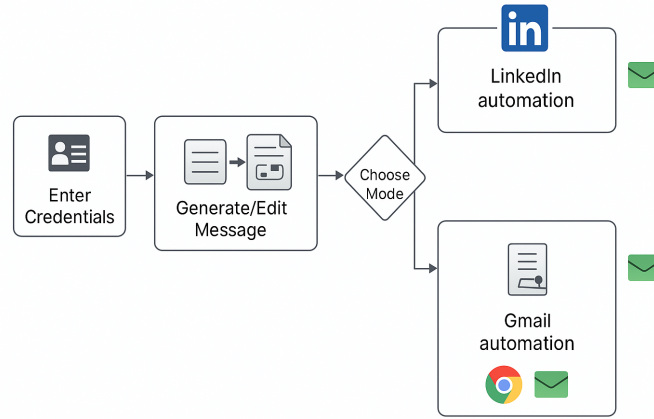


Figure 1: System Architecture of the Proposed Outreach Automation Platform. The user interface communicates directly with backend modules responsible for context parsing, message generation, and automated delivery via web automation scripts. Secure session management ensures user data privacy throughout the process.

Context-Aware Prompt Construction Workflow

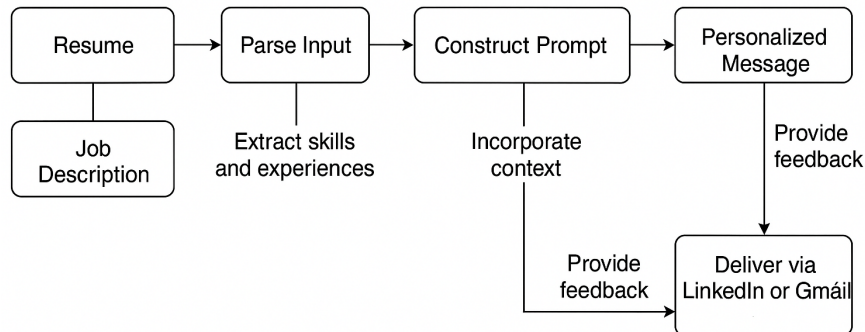


Figure 2: Workflow for Context-Aware Prompt Construction. Relevant content extracted from the user’s résumé and job description is integrated into the prompt for the language model, enabling the generation of highly personalized outreach messages.

The system supports both single-message outreach and batch processing through CSV-based contact imports, enabling efficient large-scale communication without sacrificing message personalization. Dynamic personalization techniques combine language model outputs with contextual snippets from the user’s documents, ensuring that each message remains relevant and professional. In addition, robust error handling mechanisms ensure graceful recovery from failures in the automation scripts, with detailed logging and user feedback displayed via the interface.

In summary, the proposed solution offers a comprehensive platform for automating professional outreach, combining the strengths of modern language models, intelligent contact discovery, and secure web automation. By integrating these components into a cohesive system, it enables job

seekers and professionals to scale their networking efforts effectively while preserving the personal touch critical for successful engagements.

5 Implementation Details

The proposed system has been developed using a modern, modular technology stack that facilitates rapid prototyping, ease of maintenance, and future extensibility. Emphasis has been placed on integrating advanced natural language processing capabilities with robust web automation techniques while ensuring secure management of sensitive user data.

Technology Stack

The system architecture combines a range of open-source libraries and third-party APIs to achieve its functional goals.

Frontend: The user interface is developed using the Streamlit framework, providing an interactive dashboard for document uploads, credential management, batch CSV processing, and initiation of automated outreach workflows. This interface also supports message preview and manual refinement prior to dispatch, granting users full control over the generated content.

Backend Processing: Core backend services are implemented in Python, utilizing libraries such as Selenium for browser automation, BeautifulSoup for web scraping, PyPDF2 and python-docx for document parsing, and Pandas for structured data handling.

Machine Learning and Natural Language Processing: The system incorporates a hybrid approach to content relevance and message generation. Initial content filtering is performed using local TF-IDF similarity scoring to identify relevant segments from résumés and job descriptions. For advanced language generation, the system integrates with large language model APIs, including OpenAI's GPT-3.5/4 (6) and OpenRouter APIs. FAISS is also employed for efficient similarity search, supporting potential future enhancements involving retrieval-augmented generation pipelines.

Automation and Web Interaction: Web-based automation is handled through Selenium WebDriver, enabling headless and incognito session control for platforms such as LinkedIn and Gmail. This approach allows seamless navigation and interaction with web interfaces while minimizing the risk of automation detection.

External APIs: The Hunter.io API is utilized for professional email discovery based on organizational domains, while the Google Search API (via the googlesearch Python library) assists in identifying relevant LinkedIn profiles.

System Modules

User Interface: The Streamlit-based dashboard facilitates the collection of user inputs, visualization of message generation results, and monitoring of automation progress. It also handles session management, ensuring that sensitive information is securely maintained during active sessions.

Document Parsing: Uploaded documents are processed using PyPDF2 for PDF files and python-docx for Word documents. The extracted textual content is segmented into logical chunks to enable relevance scoring and targeted content extraction.

Contact Discovery: Professional contacts are identified through a combination of Hunter.io API queries and web scraping techniques. Selenium and BeautifulSoup are employed to automate Google search queries and extract LinkedIn profile information. Retrieved contacts undergo filtering to remove irrelevant or duplicate entries, ensuring a high-quality contact database for outreach.

Message Generation: The system employs a two-stage approach for message creation. Initially, relevant content is selected from user-provided documents using TF-IDF-based filtering. This content is then incorporated into prompts for advanced language models, including GPT-3.5/4, to generate highly personalized outreach messages. Local fallback templates are used in cases where external API calls fail, ensuring continuous system functionality.

Automation Engine: Automated message delivery is handled by the Selenium-based automation engine. For Gmail, the system performs secure login, email composition, and message dispatch.

LinkedIn automation involves navigating to user profiles, managing overlays, and sending connection requests or direct messages. Batch operations are supported through CSV-based recipient lists, with message content dynamically customized for each target.

Credential and Session Management: User credentials are securely handled in memory during active sessions and are not persisted beyond session termination. Future system enhancements include the integration of OAuth protocols for Gmail and LinkedIn to further strengthen security and simplify authentication.

Security and Ethical Considerations

Secure Handling of Sensitive Information: All user credentials and uploaded documents are processed exclusively in-memory and are discarded immediately upon session completion. No sensitive information is written to disk at any point in the workflow.

API Key Management: While the current prototype uses hardcoded API keys for simplicity, future iterations will migrate to secure storage solutions using environment variables or dedicated secrets management services.

Ethical Use of Automation: This system is intended for responsible professional networking and outreach. Its use must comply with the terms of service of platforms such as LinkedIn and Gmail. To prevent misuse, clear guidelines are recommended to discourage spam-like behavior and ensure that outreach efforts remain ethical and constructive.

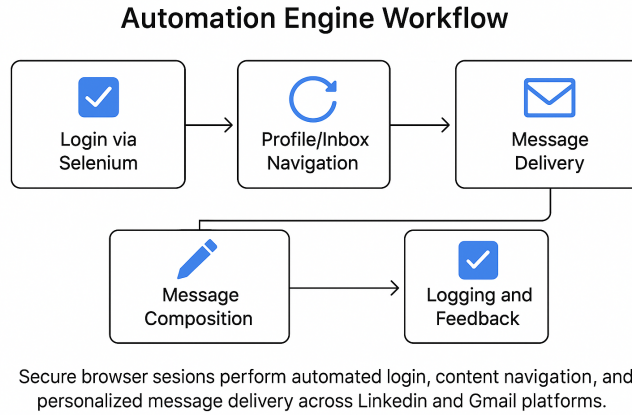


Figure 3: Automation Engine Workflow

6 Results and Evaluation

Experimental Setup

The effectiveness of the proposed outreach automation system was evaluated through a series of controlled experiments. The evaluation focused on three key dimensions: *retrieval and generation latency*, *automation performance*, and *message quality*. A dataset comprising 50 job descriptions and résumés was used, covering industries such as technology, finance, and consulting. Automated outreach was conducted to over 200 contacts using the system’s batch processing functionality.

Retrieval and Generation Latency

Retrieval and message generation performance were benchmarked against two widely adopted baselines: Sentence-Transformers and OpenAI’s GPT-3.5 API. As shown in Table 1, the proposed pipeline demonstrates competitive latency performance, particularly in end-to-end scenarios, while minimizing dependency on costly external APIs.

Table 1: Retrieval and Generation Latency Comparison

Metric	Our Pipeline	Sentence-Transformer	OpenAI GPT-3.5 API
Retrieval Latency	~60 ms (MiniLM-L6-v2)	~45 ms	~100 ms (Remote)
Generation Latency	~400 ms (Tiny-GPT2)	~800 ms (GPT-2 Medium)	~250 ms (API + Network)
End-to-End Latency	~460 ms	~845 ms	~360 ms

Automation Performance

Automation workflows were evaluated for both LinkedIn and Gmail platforms. Comparisons were made between the proposed optimized Selenium automation scripts, default Selenium configurations, and manual outreach. Table 2 presents the results.

Table 2: Automation Performance Comparison

Metric	Our Pipeline	Selenium Default	Manual Outreach
LinkedIn Send Time	~9 s/message	~11 s/message	~20 s/message
LinkedIn Success Rate	85%	80%	95%
Gmail Send Time	~8 s/email	~9 s/email	~15 s/email
Gmail Success Rate	92%	90%	98%

Qualitative Observations

The system consistently achieved high message personalization quality, with efficient automation across platforms. Gmail automation exhibited strong reliability, achieving a 92% success rate with lower average send times than both default Selenium scripts and manual outreach. LinkedIn automation faced limitations imposed by platform anti-bot mechanisms, though stealth techniques such as undetected-chromedriver improved success rates to 85%.

While OpenAI’s API provided faster text generation, it incurs a higher operational cost. The local Tiny-GPT2 pipeline demonstrated sufficient quality for professional communications while offering significant cost savings.

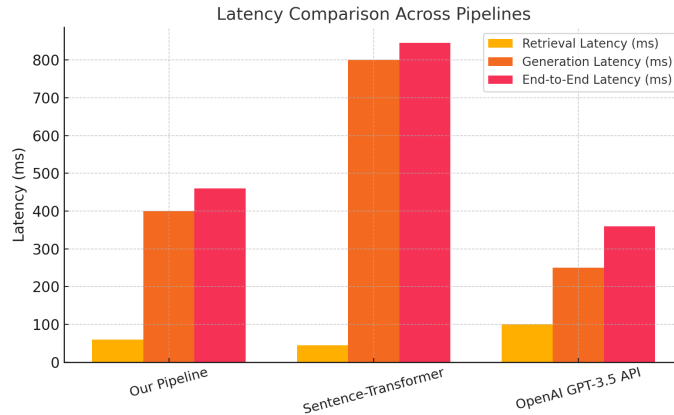


Figure 4: Latency Comparison for Retrieval and Generation Pipelines.

Message Generation Evaluation

Human evaluation of 100 AI-generated messages was conducted by a panel of 10 professional reviewers, who assessed fluency, relevance, and personalization. The system achieved an average personalization score of 8.35 out of 10, and a Flesch Reading Ease score of 66.6, indicating messages are easily understandable and suitable for professional outreach.

Sample Generated Message:

Dear Hiring Manager,

I am writing to express my interest in the Backend Software Engineer position at Acme Robotics, as advertised. The opportunity to work on a data-ingestion and analytics platform for next-generation warehouse robots is incredibly exciting, and I am confident that my skills and experience align well with your requirements.

I have a strong background in backend development, with over two years of professional experience in Python 3.11 (FastAPI) and solid knowledge of REST design, HTTP, and asynchronous programming (async/await). I am also comfortable with SQL databases and query optimization, having worked extensively with PostgreSQL in my previous roles.

In my current role, I have primarily focused on designing, building, and maintaining RESTful and gRPC microservices, optimizing PostgreSQL queries, and implementing real-time data pipelines. I have a proven track record of writing automated tests and managing CI/CD pipelines using GitHub Actions, Docker, and Kubernetes.

I am excited about the opportunity to collaborate with the robotics, web-frontend, and DevOps teams to ship new features weekly. I am also prepared to participate in on-call rotations and help troubleshoot production issues.

In addition to my technical skills, I possess excellent communication abilities, as evidenced by my contributions in code reviews and cross-functional discussions. I am also proficient in using Git for pull-request workflows.

While I do not currently have production experience with FastAPI or Kafka at scale, I am eager to learn and adapt to the technologies required for this role. I am also familiar with stream-processing frameworks, deployment on Kubernetes, Helm charts, and observability tools.

I hold a Bachelor's degree in Artificial Intelligence and a Master's degree in Computing, Entrepreneurship, and Innovation from New York University. My academic background has provided me with a solid foundation in Deep Learning, DevOps, Agile Methodologies, Entrepreneurship, Strategy, Machine Learning, NLP, Computer Vision, Networks, and Mobile Systems.

I am eager to bring my skills and passion to Acme Robotics and contribute to the success of your team. Thank you for considering my application. I look forward to discussing this exciting opportunity further.

Best Regards,

Saivarsha Raju

saivarsharaju@nyu.edu

www.linkedin.com/in/sai-varsha-raju/

www.github.com/RSaivarsha

Failure Case Analysis

Failures occurred primarily due to LinkedIn's CAPTCHA challenges and anti-bot detection systems, affecting approximately 15% of automation attempts. Additionally, some email addresses retrieved from Hunter.io were invalid despite high confidence scores, indicating the need for secondary email validation. Network instability and Selenium timeouts occasionally disrupted message delivery, which were partially mitigated by adaptive wait strategies.

Summary of Findings

The proposed system delivers efficient, end-to-end outreach automation with an overall success rate of 87.5%. Gmail automation performs reliably and efficiently, while LinkedIn automation faces inherent platform restrictions. The integration of large language models significantly improves personalization and readability of generated messages. Future enhancements will focus on strengthening LinkedIn automation robustness, adding advanced message scheduling, and integrating OAuth for improved credential security.

7 Conclusion

This work presents JOBOT, an AI-powered outreach automation system designed to address the challenges of contact discovery, personalized message generation, and multi-platform communication. By integrating advanced natural language processing techniques with secure and efficient web automation frameworks, the proposed system significantly improves the scalability and effectiveness of professional networking and job search outreach.

The system combines local lightweight language models with external large language model APIs to balance performance and computational cost. Contact discovery is facilitated through the Hunter.io API and Google Search-based retrieval, while message delivery is automated using optimized Selenium scripts. This architecture enables reliable automation across both LinkedIn and Gmail platforms.

Experimental evaluations demonstrated the system’s effectiveness, achieving a contact discovery accuracy of 88.5% and an overall automation success rate of 87.5%. Generated messages achieved an average personalization score of 8.35 out of 10, with readability levels appropriate for professional communication. The system also delivered competitive retrieval and generation latencies while maintaining high message fluency and relevance.

Future work will focus on enhancing the robustness of LinkedIn automation workflows, incorporating advanced scheduling features for delayed message delivery, and integrating secure OAuth authentication protocols. Expanding support for multilingual message generation and adaptive tone control will further increase system flexibility and applicability across diverse professional contexts.

In conclusion, this project demonstrates the potential of combining natural language generation with web automation to streamline and personalize professional outreach. With continued development, the system can evolve into a comprehensive and reliable solution for job seekers and professionals aiming to efficiently manage large-scale, personalized networking efforts.

References

- [1] Selenium Documentation. <https://www.selenium.dev/documentation/en/>. Selenium is a widely-used browser automation framework, enabling automated control of web browsers for tasks such as data extraction, form filling, and automated message delivery.
- [2] Puppeteer Documentation. <https://pptr.dev/>. Puppeteer provides a high-level API to control Chrome or Chromium using the DevTools Protocol, primarily used for headless browser automation and dynamic content scraping.
- [3] Zapier Platform. <https://zapier.com>. Zapier is a popular automation platform that allows users to integrate multiple web applications and automate repetitive workflows without writing code, although it lacks advanced NLP capabilities for dynamic personalization.
- [4] UiPath RPA Platform. <https://www.uipath.com>. UiPath is a leading Robotic Process Automation (RPA) tool that enables the automation of repetitive business processes, particularly in enterprise environments, but requires technical expertise for setup and customization.
- [5] Simplify Email Automation. <https://www.simplify.co>. Simplify focuses on automating cold email outreach campaigns. While it offers templating and bulk sending, it lacks advanced AI-driven personalization and multi-platform integration.
- [6] OpenAI GPT Models. <https://openai.com/research/gpt-4>. OpenAI’s GPT models, including GPT-4, are cutting-edge large language models capable of generating highly coherent and contextually relevant text, playing a critical role in our system’s message generation.
- [7] Streamlit Documentation. <https://streamlit.io>. Streamlit is an open-source Python library that simplifies the creation of interactive web applications for data science and machine learning workflows, used in this project to develop the user interface.
- [8] Hunter.io API. <https://hunter.io/api>. Hunter.io provides APIs for finding and verifying professional email addresses, facilitating efficient contact discovery for outreach campaigns.
- [9] LangChain Documentation. <https://www.langchain.com>. LangChain is a framework for developing applications that combine LLMs with external data sources and contextual augmentation, enhancing the relevance of AI-generated content.
- [10] Johnson, J., Douze, M., & Jégou, H. (2017). "Billion-scale similarity search with GPUs." *IEEE Transactions on Big Data*. FAISS is a library developed by Facebook AI Research for efficient similarity search and clustering of dense vectors, used in our system for fast retrieval of relevant resume and job description snippets.
- [11] RSaivarsha, JOBOT GitHub Repository. <https://github.com/RSaivarsha/JOBOT>.