Project Proposal – Team 18

Ruoqian Huang (ruoqian5@illinois.edu)
Lin Yang (linyang6@illinois.edu) *(Coordinator)*
Ye Yu (yeyu4@illinois.edu)
Longsheng Yan (lyan11@illinois.edu)

March 29, 2025

1 Functions and Users

Our project aims to develop an **Intelligent Email Assistant**, designed as an enhancement to the existing Gmail system. Seamlessly integrated with Gmail via Google Workspace APIs and OAuth 2.0 authentication, the assistant will significantly augment user email management.

Major Functions:

- **Enhanced Spam Filtering:** Leveraging machine learning algorithms and/or custom-trained large language models (LLMs) to improve spam detection accuracy.
- **Automatic Email Categorization:** Using text preprocessing and NLP techniques to automatically classify and organize incoming emails into relevant categories.
- **Intelligent Conversational Email Retrieval:** Employing a conversational interface powered by our own LLM (potentially enhanced with Retrieval-Augmented Generation, RAG) to intuitively retrieve, summarize, and manage emails.

Target Users:

- **Heavy Email Users:** Individuals dealing with high daily email volume needing improved organization and retrieval efficiency.
- **Professionals and Enterprises:** Users requiring productivity enhancements through automated email sorting and critical message prioritization.
- **Tech-Savvy Consumers:** Individuals interested in advanced AI and machine learning applications to optimize digital communication.

2 Significance

The Intelligent Email Assistant addresses the prevalent issue of email overload, significantly streamlining inbox management.

Addressing Existing Pain Points:

- **Overwhelming Email Volume:** Enhances efficiency by managing large volumes of emails.

- **Ineffective Spam Filtering:** Improves detection accuracy, reducing inbox clutter and the risk of misclassification.
- **Inefficient Email Retrieval:** Simplifies and accelerates the process of finding specific emails via conversational queries.

Impact and Societal Need:

- **Enhanced Productivity:** Reduces cognitive load, prioritizing urgent messages.
- **Improved Email Security:** Decreases phishing risks through advanced filtering.
- **Accessible AI-Powered Assistance:** Democratizes AI-powered email management, making it intuitive and accessible.

This tool contributes significantly to a broader societal shift towards AI-enhanced productivity and efficient digital communication.

3 Approach and Implementation Plan

Project Scope and Goal: Develop an advanced Email Assistant integrated seamlessly into Gmail with enhanced spam filtering, automatic categorization, and intelligent conversational retrieval.

Technologies and Tools:

- **Programming Languages:** Python (backend, NLP, ML), JavaScript/TypeScript (frontend, Gmail plugin).
- **Libraries and Frameworks:** TensorFlow/PyTorch, Hugging Face Transformers (BERT models), SpaCy, React, Google Workspace APIs, OAuth 2.0.

Existing Resources to Leverage:

- **The Enron Email Dataset [1]:** Realistic email dataset for robust model training.
- **Pre-trained Models:** Leveraging BERT for accelerated training and improved accuracy.
- **Google APIs:** Efficient Gmail integration support.

Development Procedure:

- **Data Preparation:** Cleaning, preprocessing, and categorizing the Enron Email Dataset.
- **Model Training and Evaluation:** Training supervised ML models; evaluating via accuracy, precision, recall, and F1-score.
- **Email Retrieval via Chat Interface:** Developing a conversational interface using our custom-trained LLM.
- **Integration with Gmail:** Developing a Gmail plugin, securely integrating models via REST APIs.

Risks and Mitigation Strategies:

- **Data Privacy and Security:** Mitigated by encryption, data anonymization, Google OAuth adherence, and GDPR compliance.
- **Model Accuracy and Generalization:** Addressed through transfer learning, comprehensive testing, and incremental model retraining.

4 Evaluation

We will validate our email assistant through comprehensive user testing, simulating realistic usage scenarios to ensure practical effectiveness. Specifically, the evaluation will consist of the following aspects:

- **Functional Testing:** Verify key functionalities including email summarization, conversational retrieval, and smart replies across multiple devices and platforms.
- **User Experience Assessment:** Collect feedback via user surveys and interviews focusing on ease of use, interface design, responsiveness, and overall satisfaction.
- **Performance Metrics:** Measure response time, output accuracy, and reliability, comparing results with initial baseline performance metrics.
- **Security and Compliance Verification:** Ensure adherence to privacy standards, data protection regulations, and perform comprehensive security checks.
- **Continuous Improvement:** Establish a continuous feedback mechanism for iterative improvements post-launch, facilitating ongoing enhancements based on user input.

5 Timeline

- **March 29:** Project proposal submission
- **April 7:** Dataset preparation and background research
- **April 14 & 21:** Implementation phase
- **April 28:** Evaluation and testing
- **May 10:** Final project submission

6 Task Division (tentative)

- **Ye Yu:** Backend server setup, email interface development, data processing.
- **Ruoqian Huang:** LLM integration, email categorization, retrieval UI interaction design.
- **Lin Yang:** Frontend development, user-friendly design, UI/UX management.
- **Longsheng Yan:** System testing oversight, security implementation, documentation.

References

[1] "The Enron Email Dataset." Available: https://www.kaggle.com/datasets/wcukierski/enron-email-dataset, Accessed: Mar. 28, 2025.