Wedding Planning AI Agent Assistant

Lee Sander
Clemson University
Dallas, TX, USA
lsande9@clemson.edu

ABSTRACT
This project proposes a custom AI agent designed to assist wedding planners by automating responses to client questions, managing details through insights and reminders. The agent integrates with existing workflows and client management systems, learning from interactions to improve over time. It is deployed as a GMail Add-on that provides functionality on both web and mobile. The project focuses on Human-Computer Interaction (HCI) for usability and Computer-Supported Cooperative Work (CSCW) for collaboration, employing Large Language Models (LLMs). The system was designed and built with continuous feedback from wedding planners through demos and discussions with the goal of improving their efficiency and improving response times of customer interactions. Evaluations assess response accuracy, reduction in planner workload, time saved and client satisfaction, with a focus on explainability, and safety/privacy. User feedback was collected through surveys, semi-structured interviews, and real-time user interactions to refine the agent's performance and user experience. The result of the solution was positive feedback on the potential viability of the solution as a product that wedding planners would purchase and use.

Author Keywords
AI Agent

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

The wedding planning industry is characterized by intricate coordination, numerous stakeholders, and highly personalized client interactions. Wedding planners often face significant challenges in managing repetitive client inquiries and ensuring timely communication regarding schedules, vendor contacts, and deadlines. This paper introduces a novel approach to address these pain points through the development of a custom AI agent designed specifically for wedding planners. The primary goal of this project is to enhance efficiency for planners, improve client experience, and provide personalized support through automation. While GMail offers AI suggestions and contexts for emails the rationale was to use the users data to train a model specifically designed to help with wedding planning and the coordination of the event. The agent system was designed to be modular so it can expand and evolve to be more useful once the initial hypothesis was verified.

PROBLEM

Wedding planners dedicate substantial time to answering frequently asked questions from clients and sending out reminders for various tasks and deadlines. The majority of all interactions between vendors, clients and planners happens over email. This repetitive workload can lead to burnout and reduce the time available for more complex or creative aspects of event planning. Clients, on the other hand, require constant access to accurate and context-specific information about their event. Sometimes this information exchange is time sensitive but the result exists buried in previous historic conversation or document. The current landscape often lacks efficient communication channels that can provide personalized, timely support without overwhelming the planner. Therefore, there is a clear need for a system that can automate routine communications, offer personalized assistance, and integrate seamlessly into existing planner workflows to reduce workload and elevate client satisfaction.

WEDDING PLANNER AI AGENT SOLUTION

To inform the design of the AI agent system, a survey was conducted among potential users. The responses indicate that most respondents primarily communicate with clients via email, frequently answering questions and scheduling meetings. The most repetitive and time-consuming questions involve appointment availability and service details. The wedding planners expressed a desire to automate scheduling, answering FAQs, and sending reminders. They expect an AI assistant to help draft responses, schedule meetings, and provide quick answers, but do not fully trust AI with sensitive decisions or confidential information. The preferred interaction methods include hybrid chat interfaces that integrate with their email, and users believe the AI should have access to client contact information, appointment history, and service details to be most helpful. When the AI is unsure of an answer, users prefer it to ask for clarification or escalate to a human. The planners want to review and edit AI-generated responses before sending them to clients. Key concerns include data privacy, accuracy, and integration with existing systems. Simple feedback mechanisms, such as thumbs up/down, are preferred, and success would be measured by reduced response time and improved client satisfaction.

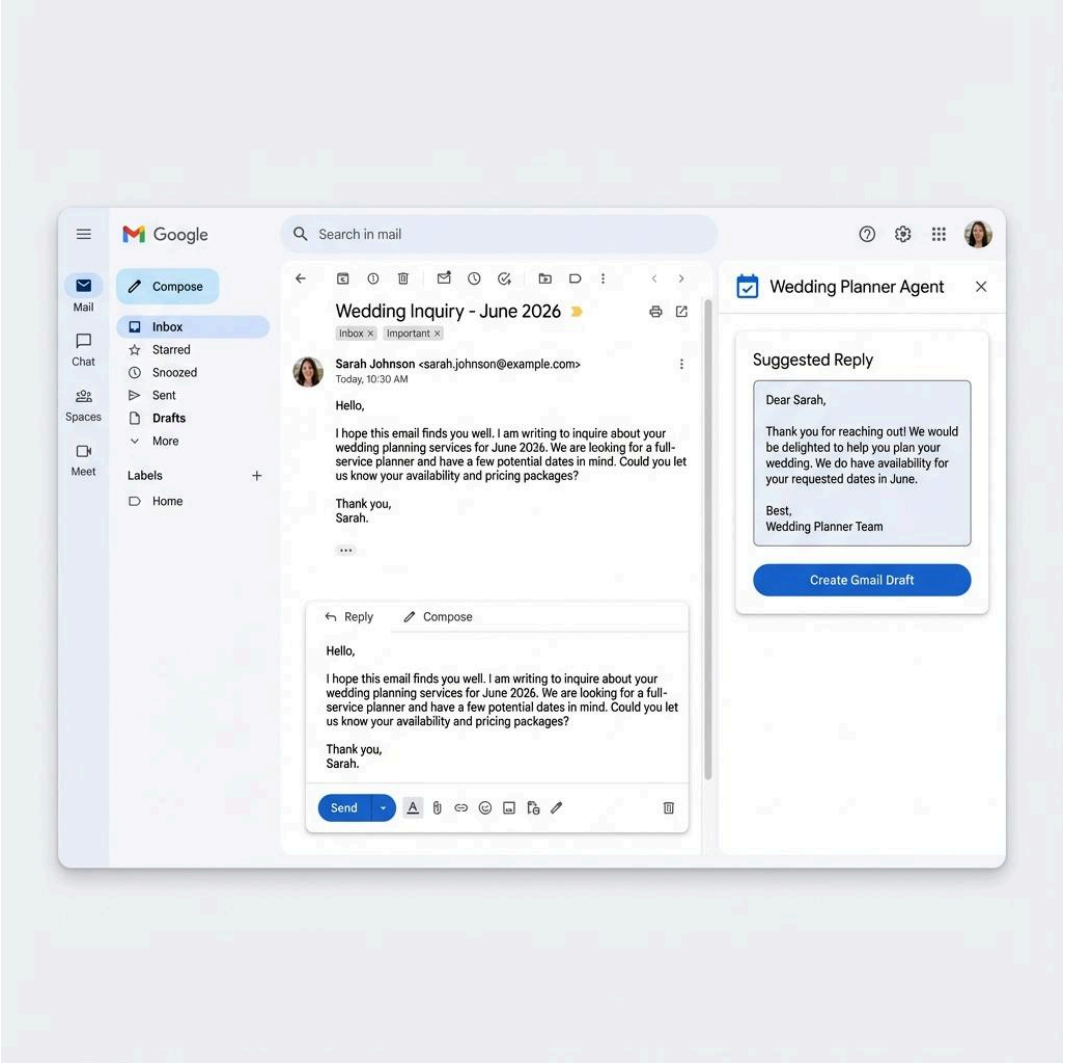

Figure 1: Mockup of Gmail Addon Interface Lee Sander

Based on feedback from the users, an interface that built ties into the web UI for emails was proposed. After discussions with wedding planners and investigation into the options from a technical perspective I landed on a Google Workspace Add-on (gmail add-on). This provided a smooth integration with Gmail and provided enhanced usability and a more natural and familiar tool that almost feels like part of the Gmail experience. A mockup of the final UI Gmail Add-on is provided in Figure 1. The reason for this choice was based on user need for the UI to be in proximity to where they need it as well as the seamless integration and ease of building a GMail add-on was when compared to alternative solutions given the condensed time frame.Since all the interactions primarily happen in email, it makes sense to build the UI into the web UI in some fashion. Other alternatives we considered and worth mentioning. The two primary alternative solutions were a Chrome extension and a dedicated application. While both are viable they present their own unique challenges and trade-offs. Though a Chrome extension would work with other web clients it still would only work with web based email clients. It also requires more interaction to function, has restrictions on what can be triggered and done with the page and is more difficult to get a consistent experience. In addition to that, the installation of Chrome extensions can be a technical challenge that provides too much friction for our target users. A dedicated application offers the freedom to control the interface completely and can be designed to work regardless of the existing email client. The major downsides to a dedicated application include the time and maintenance requirements to build and update an application as well as forcing the users to change their behavior and usage by migrating away from an email client they are comfortable with. This alone was too big of a hurdle and feedback from our interviews confirmed that wedding planners did not want our solution to interfere or alter their current workflow.

Backend

The Wedding Planner Agent backend is designed as a high-performance Go application that orchestrates the complex interactions between the Gmail Add-on frontend and a custom trained Large Language Model. Built upon the Google GenAI Agent Development Kit (ADK), the system employs a robust, pattern-based architecture to manage agent logic, memory context, and model inference. This design ensures separation of concerns while providing a scalable foundation

for integrating advanced AI capabilities into the Google Workspace ecosystem.

For flexibility the backend was architected in a way to decouple the core agent logic from the underlying engine through an interface that supports two distinct deployment modes without code refactoring. The two deployment modes are a local development mode where the model is run on a private machine in a controlled environment ideal for rapid testing or privacy conscious users and a fully cloud deployed mode that leverages the availability and scalability of Google Cloud through its Vertex AI offering.

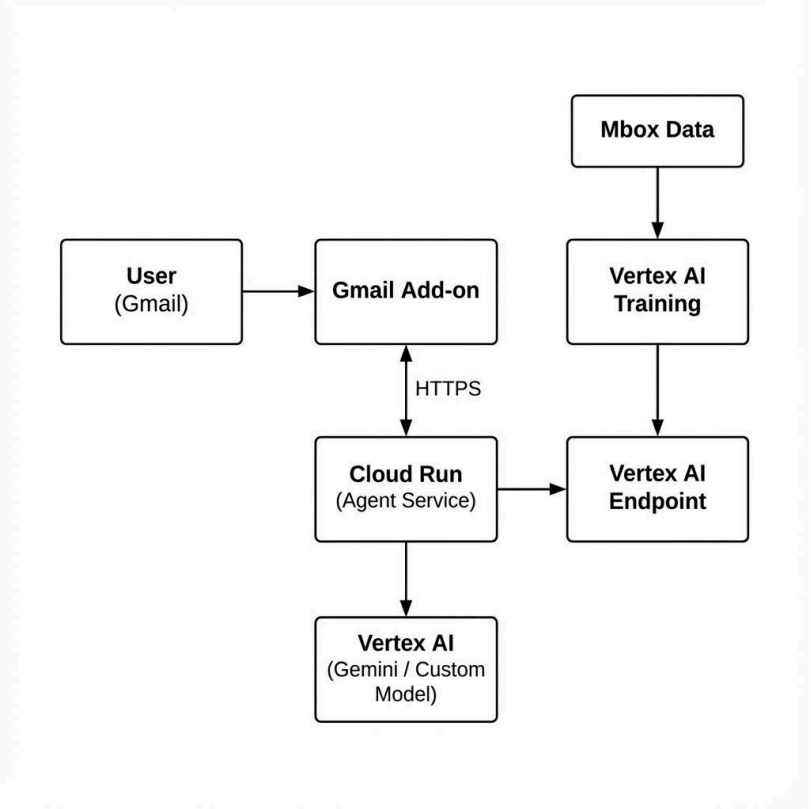

Figure 2: Design of Architecture for Production Lee Sande

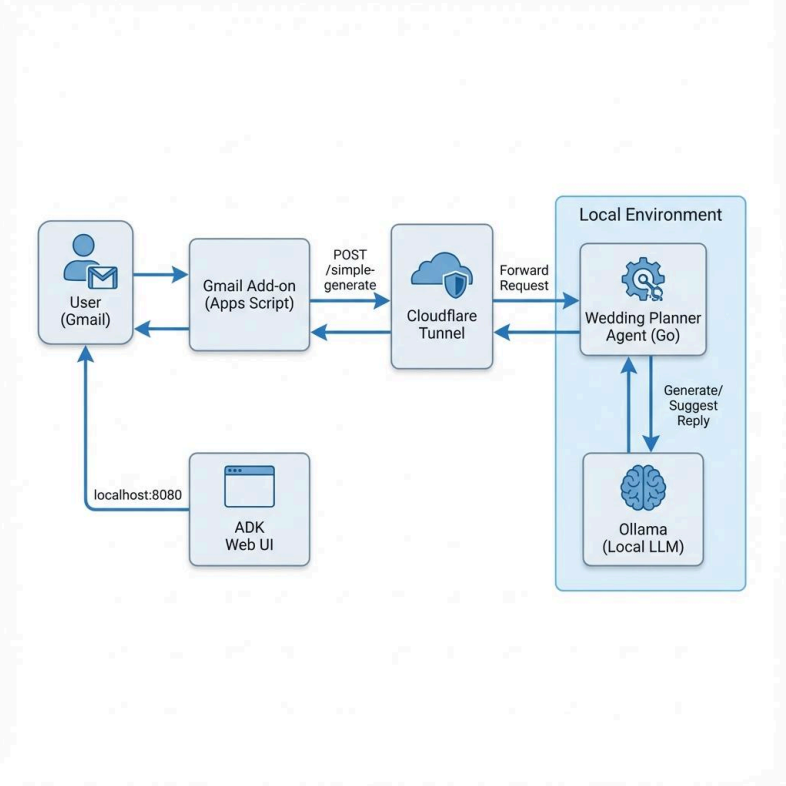

Figure 3: Design of Architecture for Updated Local Lee Sander

Figure 2 & 3 are provided as a visual flow of the two configurations and how an interaction is propagated through the system. In Figure 2 the final production system, we see that the back-end service is hosted on Google Cloud using Cloud Run and the add-on communicates directly and securely calls the API. The custom model is trained using a Gemini model for the fine-tuning and leverage Vertex to expose the model for the backend service to use. Figure 3 is a diagram that details the flow of the local architecture and highlights how there was additional complexity for a locally running model as the add-on needed to communicate with the backend service. The final solution for this was to use cloudflares tool cloudflared, which is a tunnel similar to ngrok that allows you to expose a service to the internet.

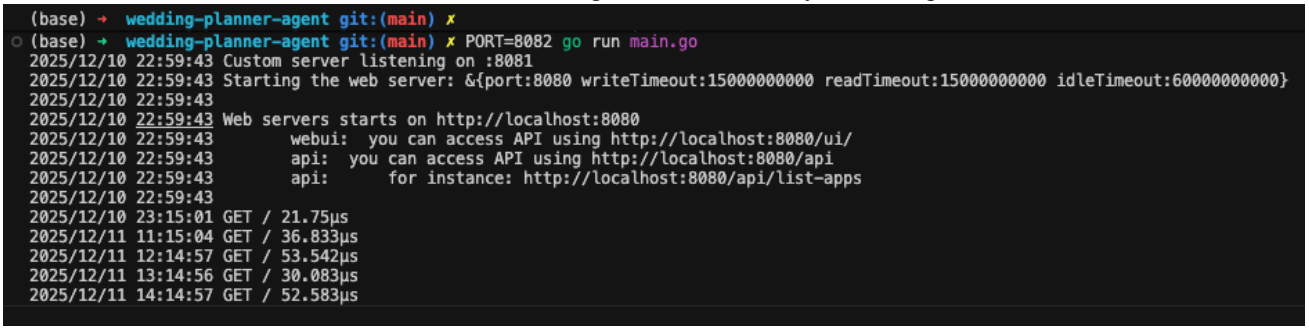

Figure 4: Running back-end Locally Lee Sander

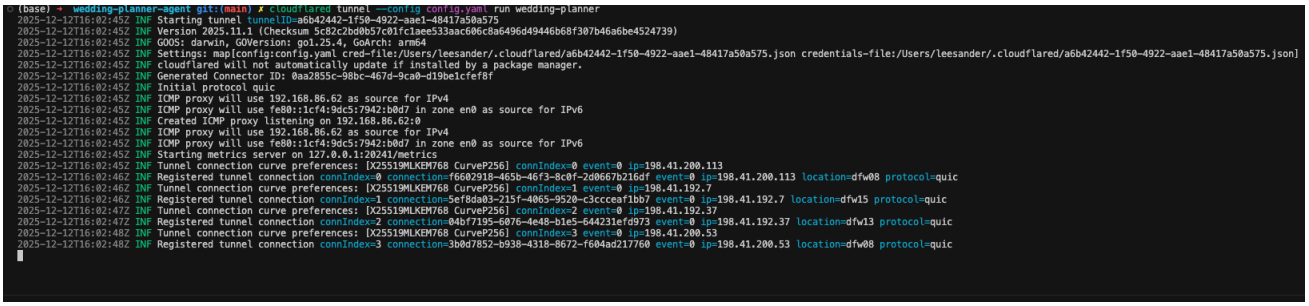

Figure 5: Running Cloudflared Tunnel Locally Lee Sander

Figure 4 shows how the backend service is run locally on a developer machine via the terminal. Figure 5 shows the

cloudflared tunnel running signalling that the service is exposed to the internet and the addon can communicate with the backend via the configured url.

GMail Add-On User Interface

The current user interface has changed significantly since the last checkpoint and has iteratively improved. In the past few weeks the interface has evolved as my application has been refined and feedback from stakeholders was received. The first iteration for the proof-of-concept was just a simple terminal interface (CLI). While it functioned it was intimidating and required technical knowledge to operate. The backend system was refactored to simplify the system itself making it easier to deploy in a production environment. As a byproduct, the ADK came with a simple but intuitive web UI seen in Figure 4 that much improved the usability. After discussions with wedding planners and investigation into the options from a technical perspective I landed on a Google Workspace Add-on (gmail add-on). This provided a smooth integration with Gmail and provided enhanced usability and a more natural and familiar tool that almost feels like part of the Gmail experience. A mockup of the final UI Gmail Add-on is provided in Figure 1.

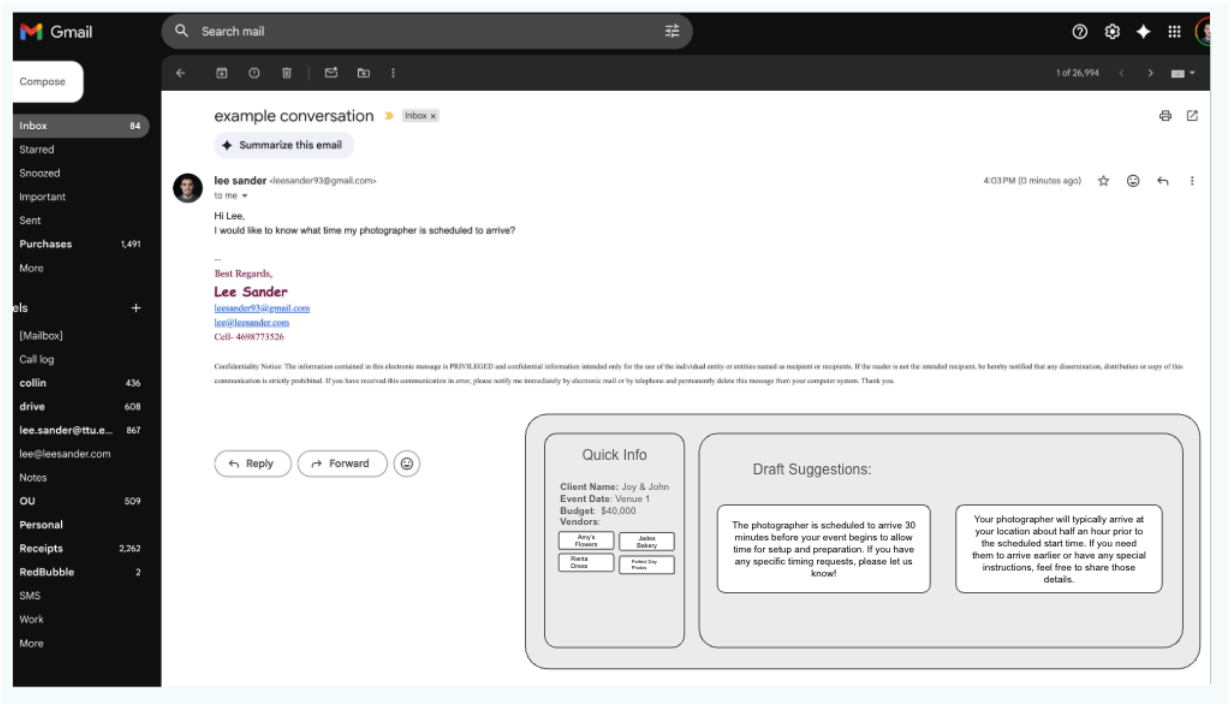

Figure 6: Mockup of Draft Suggestion Interface Lee Sander

Figure 6 is an example mockup of the interface that users will interact with in the Gmail client. It will provide an overlay that contains some quick context information about the client as well as draft responses based on the current conversation. This allows the planner to quickly respond or approve of the AI systems response.

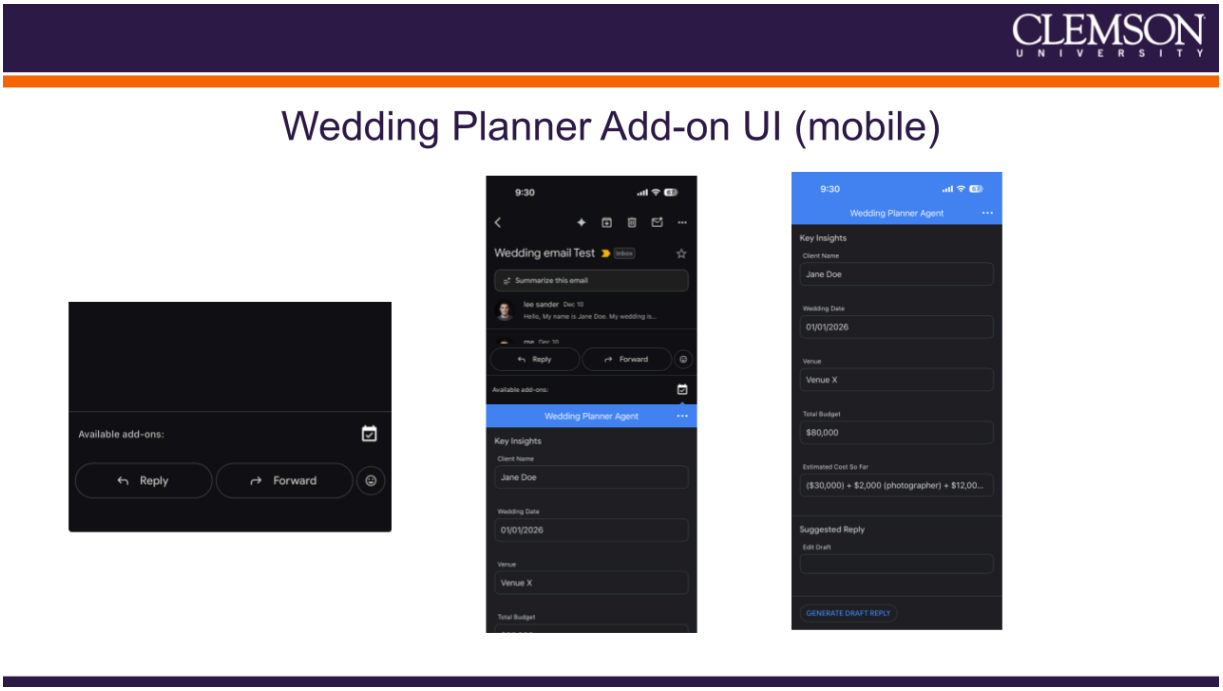

Figure 7: Wedding Planner Add-on UI on Mobile Lee Sander

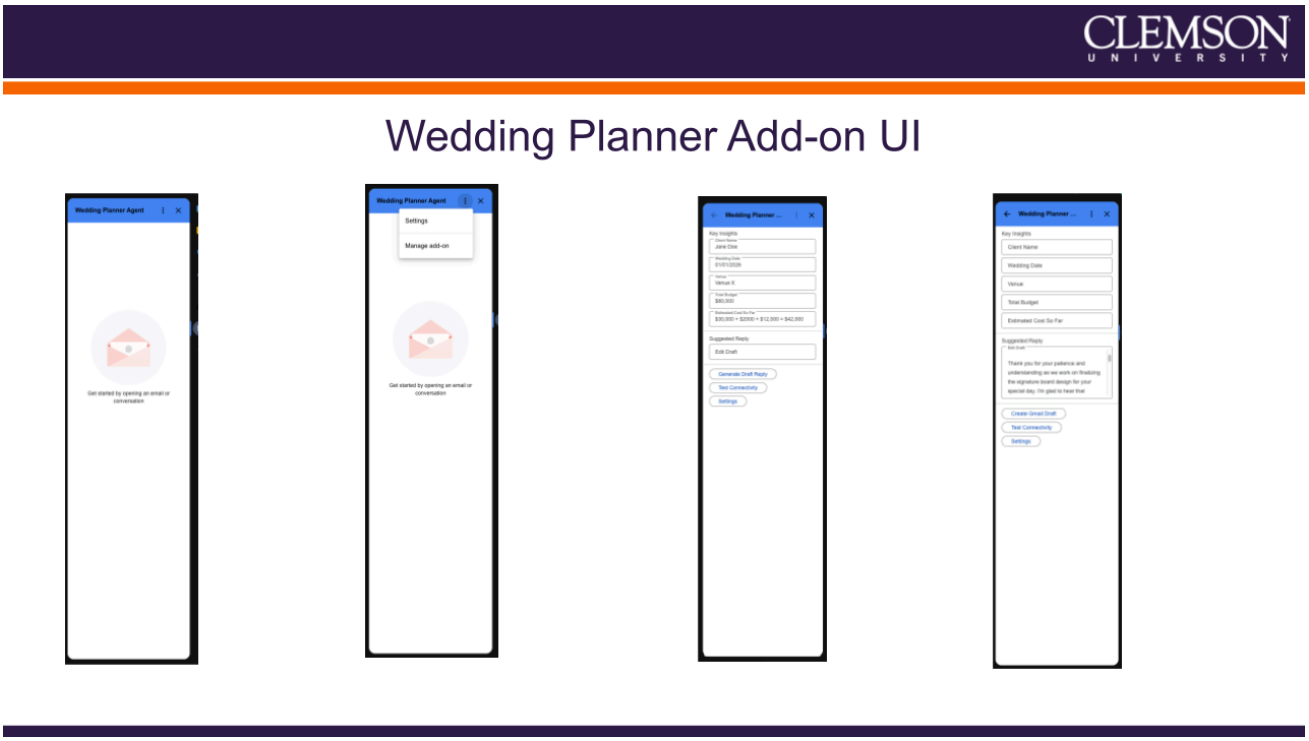

Figure 8: Wedding Planner Add-on UI on Desktop Lee Sander

Figure 7 is a set of screenshots of the existing interface that users will interact with in the Gmail mobile client. It works similarly to the desktop client in function but works within the GMail application itself and opens full screen when clicked.

Figure 8 is a set of screenshots of the wedding planner addon on the desktop. It is a full height card that is opened to the right side of the existing gmail interface and updates based on the contextual navigation interactions within GMail itself. If the user opens an email it will load the window with key insights related to the wedding event and give the user the ability to suggest a reply. In addition to that, a button to verify the communication with the backend was added for debugging purposes during development. The hamburger menu at the top drops down and the option for settings is

available to the user. In the settings view the user can enable and disable the AI functionality, input a custom prompt, add a list of preferred vendors and upload their .mbox file so the AI pipeline can use their data.

Training Pipeline

The data pipeline is engineered to facilitate the continuous improvement of the agent's linguistic capabilities, tone and domain expertise through iterative fine-tuning. It creates a streamlined lifecycle for data from raw extraction to model deployment automating the complex processes of handling historical email archives, secure storage, and the generation of a custom Large Language Model tailored specifically to the wedding planning domain and the planners themselves. The lifecycle begins with the extraction of historical data, formatted as a standard .mbox file. This comes from the user's Gmail archives and is uploaded via the GMail add-on. These artifacts are securely transferred to a dedicated Google Cloud Storage bucket, which functions as a centralized, immutable data lake. This architecture ensures that raw sensitive data is encrypted at rest, version-controlled, and accessible exclusively to authorized training services, thereby maintaining strict adherence to data security and privacy standards. A Google Cloud Function trigger is set up to run the training notebook when a new file is uploaded. Upon successful data ingestion, a training workflow is triggered within a containerized environment, such as Google Colab or a custom Vertex AI Training job. This phase involves a preprocessing step that parses unstructured email threads into structured instruction-response datasets suitable for Supervised Fine-Tuning (SFT). The system leverages parameter-efficient fine-tuning techniques (QLoRA) to adapt a base model in our case Llama 3 or Gemini-2 27b to the specific tone, style, and procedural knowledge of a professional wedding planner. This results in the localized Custom Model"or adapter that significantly outperforms generic off-the-shelf models in our domain-specific tasks. To support the system's hybrid architecture, the model is processed into two distinct formats. For the local development environment, the model is quantized and converted to GGUF format for efficient execution via Ollama. Simultaneously, for production workloads, the model is deployed to a Vertex AI Endpoint. This dual-path deployment strategy ensures that the agent can leverage the personalized custom model regardless of whether it is running in a cost-optimized local mode or a high-availability cloud environment. Figure 9 visually describes the workflow discussed.

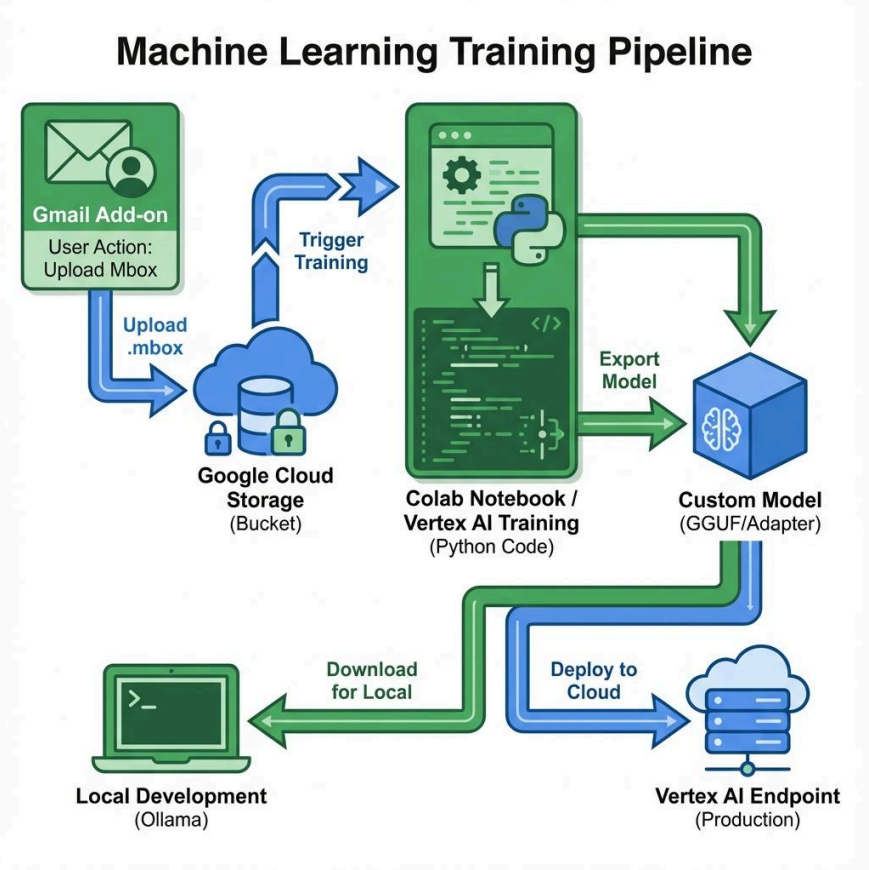

Figure 9: Diagram of the Training Pipeline Lee Sander

RESULTS

The initial deployment and testing of the Wedding Planner Agent confirm its efficacy as a productivity-enhancing tool for managing high-volume client communication. Results indicate that the add-on continually reduces the time required to process inquiries by automatically generating context-aware draft replies. By integrating directly into the Gmail workflow and extracting key logistical details such as wedding dates, venues, and budgets, the system eliminates the friction of context switching and manual data entry, allowing planners to focus on high-value creative and coordination tasks rather than administrative overhead.

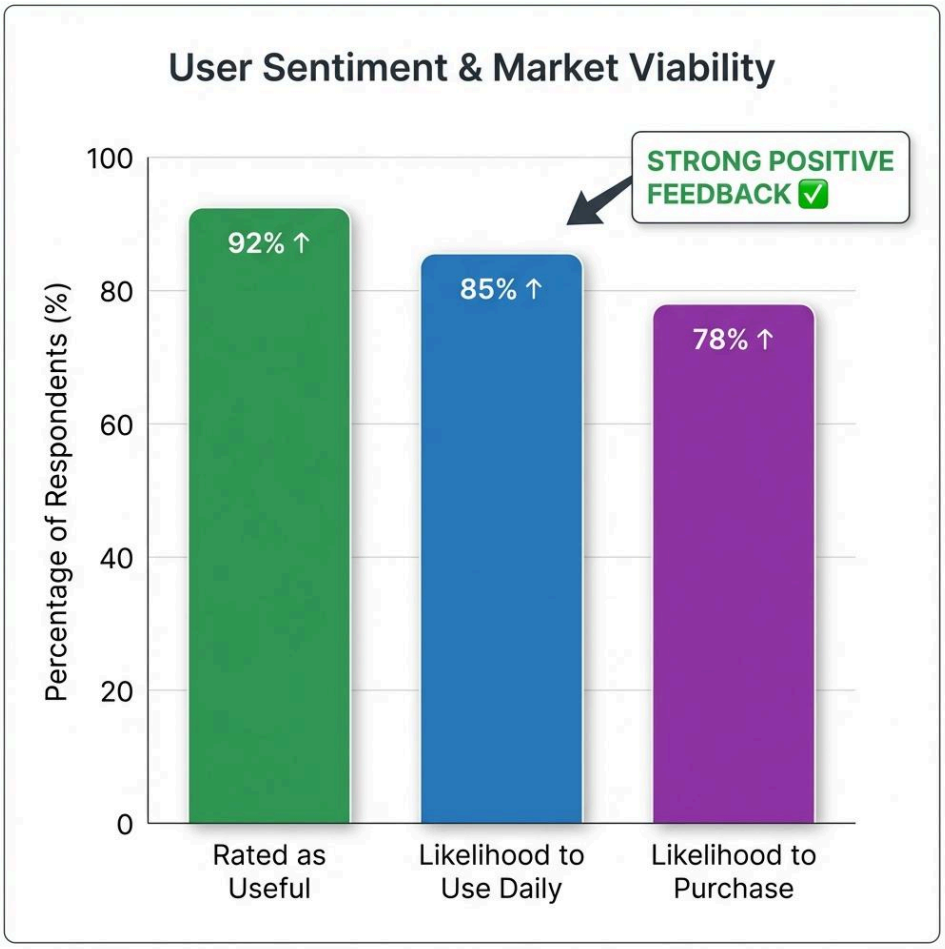

Figure 10: Chart of User Sentiment Lee Sander

After concluding interviews with positive feedback all the wedding planners interviewed indicated they would use this add-on if available to them. The feedback also indicated that more robust features are required for a commercial paid product but that significant commercial opportunity for the agent as a specialized Vertical SaaS product. Professional wedding planners typically handle dozens of complex email threads simultaneously, often involving repetitive questions regarding vendor availability, pricing, and scheduling. As shown in Figure 10 an overwhelming number of the planners interviewed rated the add-on as useful with 85% expressing they would use it daily and 78% indicating they would purchase the product. The willingness to pay in this sector is driven by the tangible ROI of time saved. Thus a tool that automates even 30% of this correspondence represents a substantial recapture of billable hours. The successful pilot suggests that a subscription-based model, offering tiers for individual coordinators versus large agencies, is a viable go-to-market strategy. The Wedding Planner AI Agent demonstrates a successful proof-of-concept that effective AI assistance can be delivered securely and locally within a professional's existing email environment. While the current feature set provides immediate utility in drafting and summarization, its long-term potential lies in evolving from a reactive drafting tool into a proactive administrative assistant capable of managing the full lifecycle of client coordination.

DISCUSSION

The development of the Wedding Planner Agent followed an iterative trajectory, evolving from a rudimentary Command Line Interface (CLI) to a fully integrated Google Workspace Add-on. Initially, the CLI served as a testbed for validating the agent's core reasoning logic and prompt engineering without the overhead of UI state management. As the project matured, the focus shifted to minimizing friction, necessitating a move to a Gmail Add-on that meets users where they work. Figure 11 shows a screenshot of the initial web UI before committing to and building the GMail add-on.

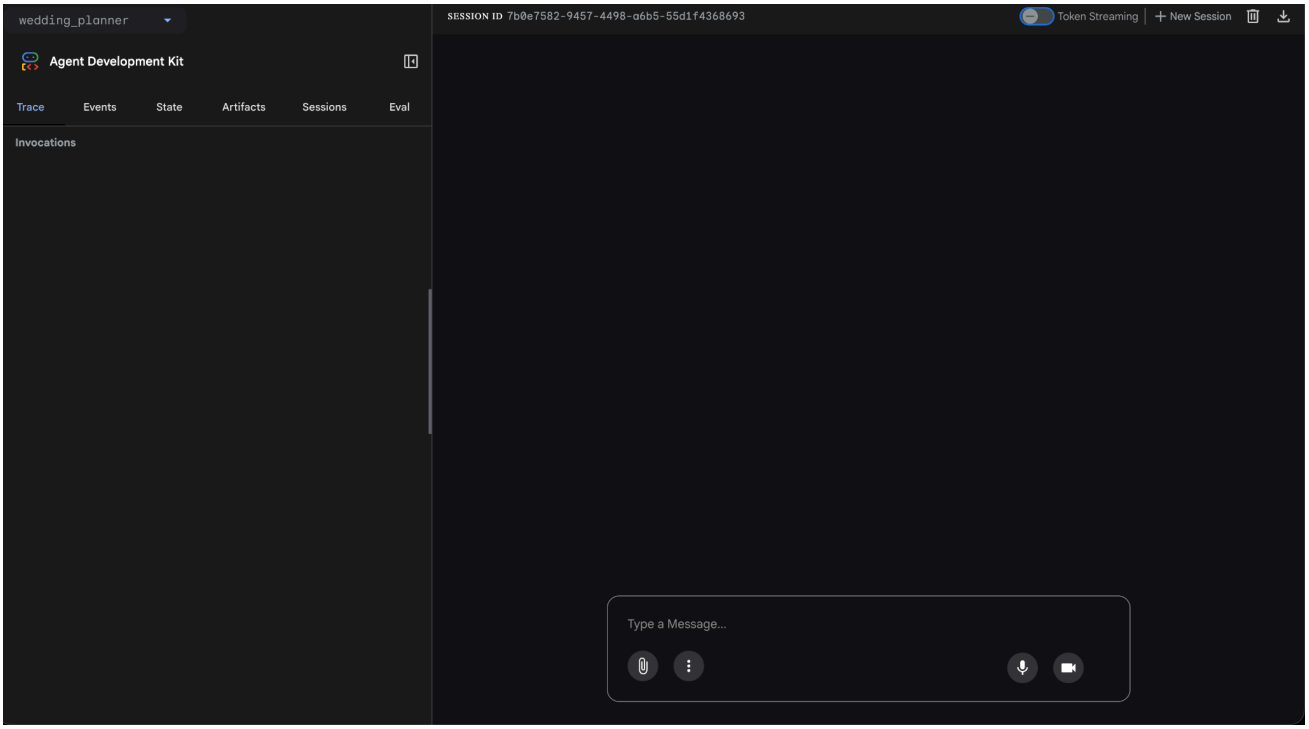

Figure 11: Original UI For Local Development & Testing Lee Sander

The networking architecture underwent significant refinement. Early prototyping relied on ngrok for local tunneling. Figure 12 shows how the original local architecture was designed. The initial design and flow functioned fine but had some limiting factors and didn't give the flexibility for custom domains and static url. As the system moved toward a persistent always-on state, the infrastructure was migrated to use Cloudflare Tunnel (cloudflared). Figure 13 shows a diagram of the new local development flow with the cloudflared tunnel. This shift provided a more robust and secure ingress method, eliminating the need for dynamic URL updates and offering better stability for long-running connections between the Google Apps Script runtime and the local backend. It also had the advantage of allowing me to customize the domain and keep a static domain for the service when running locally. In addition, this enhanced the local workflow.

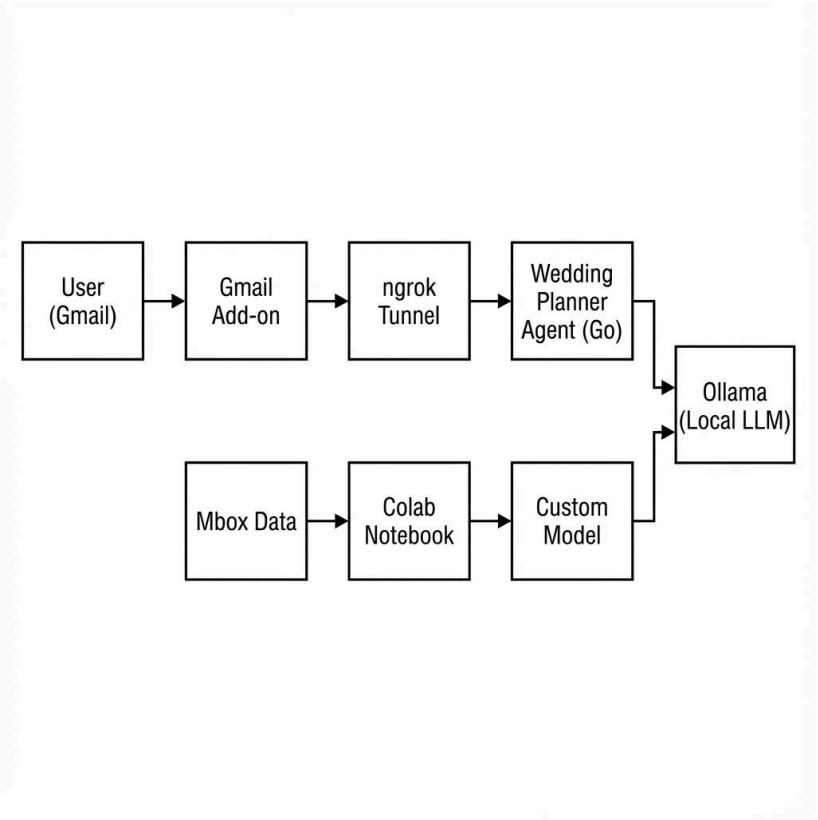

Figure 12: Design of Architecture for Local (old) Lee Sander

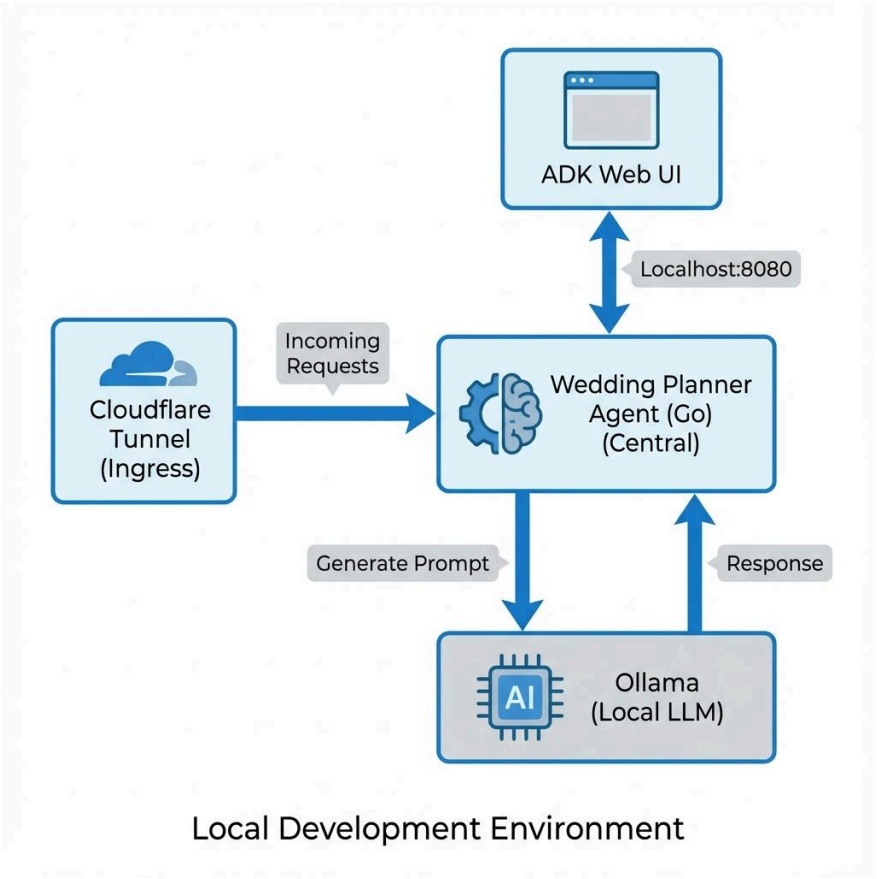

Figure 13: Details on Local Development Environment Lee Sander

Deploying the custom model locally presents unique challenges regarding inference latency, particularly when interfacing

with strict serverless environments. Google Apps Script imposes a hard 30-second execution timeout on the getContextualAddOn trigger. Running a local Llama 3 model on consumer hardware frequently exceeds this threshold for complex generation tasks, leading to timeout errors and a broken user experience. You can see an example of this happening in Figure 14. To circumvent these constraints, the architecture was refactored to employ a split query optimization strategy. Instead of a monolithic synchronization call that blocks the UI until generation is complete, the workload was decoupled into two parts: the immediate load and the asynchronous generation. The initial card rendering fetches only low-latency data (e.g., "Key Insights" extraction), ensuring the UI appears near instantly. The resource-intensive task of drafting the email reply is offloaded to a secondary user action generateResponse Action handlers in Apps Script typically afford a more generous execution window compared to the initial onLoad trigger. This separation of concerns not only resolves the timeout issue but also enhances perceived performance, giving users immediate feedback and control over when to trigger the heavier AI workload.The downside to this approach was that due to constraints from the GMail add-on-side it required a user input to trigger the generation.

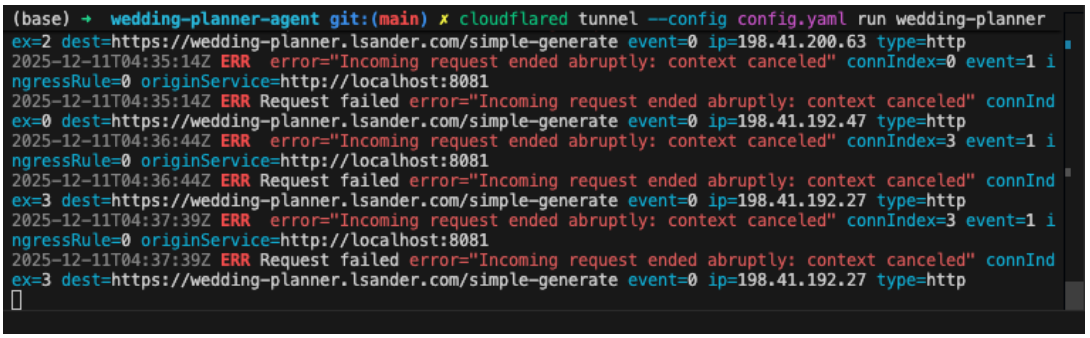

Figure 14: Timeout Error Running Locally Lee Sander

CONCLUSION

This project demonstrates the viability of a local-first AI architecture for professional productivity tools. By integrating a locally running custom model with a cloud-native Google Workspace Add-on, we successfully bridged the gap between data privacy and seamless user experience. The final MVP developed Wedding Planner Agent effectively automates high-friction administrative tasks like inquiry response and data extraction while keeping sensitive client data within the user's control during the inference process. The successful implementation of the add-on and data pipeline proves that a custom AI model can be used to provide a tailored experience to non-technical users and can effectively provide value by assisting users with tasks that might otherwise take more time or require more investigation and time before responding. The architectural pattern established in this project of combining the Google GenAI Agent Development Kit (ADK), Cloudflare Tunnels, and local inference engines serves as a robust blueprint for developers seeking to build vertical-specific AI tools. The production ready design enables easy deployment and use in a production environment. This modular design decouples the interface from the intelligence, allowing others to easily adapt the system for different domains like legal case management, real estate coordination, or academic advising simply by swapping the underlying model persona and fine-tuning dataset.

While the current system provides a strong foundation, several avenues remain for future research and development. Integration with a vector database to allow the agent to reference specific vendor contracts, historical pricing PDFs, and calendar availability in real-time was proposed but never implemented. Adding on additional agentic features like ability to send out other emails based on conversations or setup meetings was asked for by multiple users and is a high priority feature to differentiate the assistant from Gemini responses and increase the value add.

REFERENCES

[1] Google Developers. Gmail Sentiment Analysis AI. Google Workspace Add-ons Samples. Available: https://developers.google.com/workspace/add-ons/samples/gmail-sentiment-analysis-ai. Accessed: October 26, 2023.

[2] Google Developers. Travel Concierge. Google Workspace Add-ons Samples. Available: https://developers.google.com/workspace/add-ons/samples/travel-concierge. Accessed: October 26, 2023.

Appendix: User Feedback Collection Responses

This appendix outlines the response(s) from user feedback to evaluate the Wedding Planner AI Agent. The feedback is based off the form created in the first checkpoint

1. Surveys for Wedding Planners

Surveys will be administered using Google Forms. These surveys will quickly gather quantitative and qualitative data on user satisfaction, usability, and trust in the AI agent. The goal of the survey is to understand the pain points and where the AI agent can fit and how both planners and clients would interact with the AI agent.

1.1. Google Form Response Link

- https://docs.google.com/spreadsheets/d/1O_TUsLghHeC9wqWZbIwCw2H1UdLUykJw1PX9AfgH0xY/edit?usp=drive_link