

AI-POWERED AUTONOMOUS ORKFORCE ASSISTANTS

Presented by:

Aishanee Sinha

Debika Choudhury

Ketki Maddiwar

Leela prasad Dammalapati

Utkarsh Tripathi

Project Advisors:

Dr. Simon Shim

Dr. Lee C. Chang

Introduction

Modern workplaces rely on a complex ecosystem of tools including Slack, email, Jira, and calendars. Coordination across these systems involves manually transforming text into tasks, scheduling meetings from emails, and extracting action items from recorded conversations—processes that are time-consuming and error-prone.

This project presents a Multi-Agent Autonomous Workforce Assistant that automates three core enterprise workflows:

- **Slack-to-Jira:** Extracting actionable tasks from Slack conversations and generating structured Jira tickets with human approval.
- **Email-to-Calendar:** Detecting meeting intent from emails, extracting scheduling details, and creating Google Calendar events.
- **Meeting Summarization:** Generating structured summaries with decisions, action items, and key points from meeting transcripts.

The system uses a unified Qwen 2.5-14B-Instruct model with task-specific QLoRA adapters, deployed on AWS with Lambda orchestration and vLLM inference on a GPU-enabled EC2 instance.

Data Engineering Pipeline

Raw Source → Samples → QLoRA Fine-Tune

- Slack Archives (S3) → 8,302 → Slack-to-Jira Adapter
- Email Emails (S3) → 14,372 → Email-to-Cal Adapter
- AWS + Meeting App (OpenAI, QLoRA) → 32,343 → Meeting Adapter

All datasets were cleaned, deduplicated, and transformed into structured JSONL format. Labels were generated using Gemini Flash 2.5 with prompt-based extraction, converting raw inputs into JSON schemas for training.

Methodology

System Architecture Overview

System Architecture Overview

Slack Events API → Gmail Watch (Push/Sub) → Google Drive → AWS API Gateway - SQS → AWS Lambda - LangChain StateGraph Orchestrator → Router Agent (Qwen 2.5-14B) → Slack Subgraph (Slack + Jira) → Calendar CoP (Email + Calendar) → Meeting Subgraph (Drive + Summary) → Reids (Reason) → ChromaDB (Feedback + RAG) → Amazon S3 (Artifacts) → Jira / Gmail / Gmail API

- **Unified Model:** Qwen 2.5-14B-Instruct with task-specific QLoRA adapters (~40 MB each) instead of separate models per workflow
- **Orchestration:** LangChain StateGraph routes events via a Qwen-based router agent to Slack-Jira, Email-Calendar, or Meeting Summarizer subgraphs.
- **Chain-of-Debate (CoD) Reasoning:** CoD generates and ranks multiple candidates for ambiguous scenarios; human-in-the-loop validates critical actions
- **Continuous Learning Pipeline:** User approvals/rejections are stored as structured telemetry in Reids and ChromaDB. Previously approved examples are retrieved via RAG to improve consistency. Feedback pairs enable future Direct Preference Optimization (DPO) retraining.
- **Serving:** vLLM 0.17 on AWS EC2 g5.xlarge (A10G), adapter swap <50ms, base model ~14 GB VRAM (4-bit AWQ)
- **Infrastructure:** Reids (sessions), ChromaDB (feedback + RAG), S3 (artifacts), integrated with Slack/Gmail/Calendar — all at ~\$25–32/month

Analysis and Results

Slack-to-Jira Evaluation

Metric	Qwen Base	Qwen Fine-Tuned
ROUGE-1	0.87	0.94
ROUGE-2	0.36	0.91
ROUGE-L	0.87	0.93
BERTScore F1	0.92	0.97
G-Eval	0.43	0.67
No-Action F1	0.75	0.93

The workforce assistant proposal for jira ticket in slack.

Thread

: X

U Utarsh Tripathi Just now

@auto-vf-assistant Leela, create a Jira ticket for adding retry logic to the Redis health check. The check should retry up to three times with a short delay before marking Redis as unavailable. Assign it to @Leela Prasad Dammalapati, set priority to Medium, and set the due date to Tuesday, May 19, 2026.

2 replies

auto-vf-assistant 4PM Just now

Proposed Jira Action

Action: Create

Summary: Add retry logic to Redis health check, retry up to 3 times

with short delay.

Issue Key: New Ticket

Assignee: USAL4PR3ADH

Confirm Create

Cancel

Meeting Summarizer UI

Action Item Triage (6)

- High – [Ketki Kulkarni] Add a Jira proposal duplication test case to duplicate Jira proposals can lead to system inefficiency and clutter. | O
- High – [Aishwnee Sinha] Write a short test checklist for the mee Meeting summarizer agent incorrectly identifies vague phrases as action
- Medium – [Leela Prasad] Implement retry logic for Redis health Triage unavailable. | On track
- Medium – [Utarsh Tripathi] Schedule a bug triage follow-up m Backend Redis health check failures and meeting summarizer agent acc
- Medium – [Debika Choudhury] Prepare a small evaluation note

7/5

Proposed Jira Tickets (5)

- Task – [Ketki Kulkarni] Add a Jira proposal duplication
- Task – [Leela Prasad] Implement retry logic for Redis
- Task – [Utarsh Tripathi] Schedule bug triage follo
- Task – [Aishwnee Sinha] Write a short test checkli
- Task – [Debika Choudhury] Prepare a small evalua

After confirming, each ticket will appear for individual approv

Ready to send to: utkarsh.tripathi@sjvsu.edu, leela.prasad.c.ketkikarni@gmail.com, utkarsh.tripathi@sjvsu.edu, leela.prasad.c.ketkikarni@gmail.com

Click Confirm to email summary + triage + ICS + CSV + re Click Cancel to dismiss.

Confirm - Send Email

Cancel

Meeting Summarizer Bot 4:13 PM

Confirmation: I'll generate agent updates for scheduling. Reviewing 5 Jira tickets (1/1).

Consolidated summary email will be sent to participants after Jira review. (1/1)

I'll generate agent updates for scheduling. Reviewing 5 Jira tickets (1/1).

Summary: Add a Jira proposal duplication test case to duplicate Jira proposals can lead to system inefficiency and clutter. | O

Assignee: Ketki Kulkarni | Type: Task

Implement a duplication test case for Jira proposals to prevent duplicates based on title and content

Confirm Create

Cancel

Email-to-Calendar Evaluation

Metric	Qwen Base	Qwen Fine-Tuned
Full Exact Match	0.63	0.75
G-Eval	0.84	0.87
Accuracy	0.90	0.93
F1 Score	0.84	0.89
BERTScore F1	0.74	0.91
ROUGE-L	0.85	0.94

COD proposals for meetings based on email intent

auto-vf-assistant 4PM 4:13 PM

Meeting detected in incoming email

Title: Integration Review Meeting - Hypothesis Project

Location: N/A

Attendees: utkarsh.tripathi@sjvsu.edu, madd.madd@sjvsu.edu, leela.prasad.dammalapati@sjvsu.edu

CoD proposed the following slots – pick one:

#1 – Mon May 11, 8:00 AM UTC-7 This slot is the earliest available and respects the morning preference.

✔ Slot 1

#2 – Tue May 12, 8:00 AM UTC-7 This slot is the second earliest available and respects the morning preference.

✔ Slot 2

#3 – Mon May 11, 4:00 PM UTC-7 This slot is later in the day but is fully available.

✔ Slot 3

All times shown in UTC-7. Calendar invites will be sent to each participant's local timezone.

Cancel

Meeting Summarizer Evaluation:

DeepEval LLM-as-a-Judge (GPT-4o)

Category	Overall	Summary	Details
Summary	0.85	0.85	0.85
Context	0.85	0.85	0.85
Details	0.85	0.85	0.85
Action	0.85	0.85	0.85
Tone	0.85	0.85	0.85

Consolidated Email

Notes meeting details consolidated. Tue, Tuesday, May 19, 2026. Priority: Medium.

Hi everyone, please find the below information consolidated for the Jira ticket open against attendees and action items.

Integration Review Meeting - Hypothesis Project | Mon, Monday, May 11, 2026. Priority: Medium.

Attendees: utkarsh.tripathi@sjvsu.edu, madd.madd@sjvsu.edu, leela.prasad.dammalapati@sjvsu.edu

CoD proposed the following slots – pick one: #1 – Mon May 11, 8:00 AM UTC-7 This slot is the earliest available and respects the morning preference.

#2 – Tue May 12, 8:00 AM UTC-7 This slot is the second earliest available and respects the morning preference.

#3 – Mon May 11, 4:00 PM UTC-7 This slot is later in the day but is fully available.

All times shown in UTC-7. Calendar invites will be sent to each participant's local timezone. Tue, Tuesday, May 19, 2026. Priority: Medium.

Cancel

Jira Tickets

5 of 6 Jira tickets are currently created.

Task –

Abstract

AI-based automation is becoming increasingly important in modern workplaces where significant time is spent on repetitive coordination tasks. This project develops a Multi-Agent Autonomous Workforce Assistant that automates tasks across application boundaries through three workflows: task extraction from Slack to Jira, meeting scheduling from emails to Google Calendar, and meeting transcript summarization.

The system uses a fine-tuned Open 2.5-14B model with QLoRA for efficient training and structured outputs. A Multi-agent architecture with LangGraph orchestration coordinates specialized agents. Human-in-the-loop validation ensures reliability. The system is deployed on AWS with Lambda orchestration and vLLM inference, achieving production-grade performance at ~\$25-32/month.

Summary/Conclusions

The system validates that production-grade multi-agent assistants can be built using a single base model with modular LoRA adapters. Key achievements include:

- Unified architecture: One Open 2.5-14B model serves all three workflows via lightweight task-specific adapters
- Strong performance: ROUGE-1 of 0.94 (Slack-Jira), 88.38% meeting detection F1 (Email-Calendar), and 85% action quality (Meeting Summarizer)
- Production deployment: Serverless AWS architecture with ~\$25-32/month operational cost
- Continuous learning: Feedback-driven improvement through RAG retrieval and DPO-ready telemetry

Key References

Future work includes reducing inference latency via speculative decoding, automating DPO retraining, and expanding to domain-specific applications.

- [1] Jiang et al., Mistral 7B (2023) – Baseline LLM
- [2] Owen Team, Owen2.5 Technical Report (2024) – Production model
- [3] Hu et al., LoRA: Low-Rank Adaptation (ICLR 2022) – Adapter design
- [4] Dettemer et al., QLoRA (NeurIPS 2023) – Training method
- [5] Rafailov et al., DPO (NeurIPS 2023) – Feedback learning
- [6] Lewis et al., RAG (NeurIPS 2020) – Context memory
- [7] Lin, ROUGE (ACL 2004) – Evaluation metric
- [8] Liu et al., G-Eval (EMNLP 2023) – LLM-based evaluation
- [9] Kwon et al., vLLM (SOSP 2023) – Inference backend
- [10] Du et al., Multi-Agent Debate (ICML 2023) – CoD inspiration

Acknowledgements

We would like to thank our project advisors Dr. Simon Shih and Dr. Lee C. Huang for their guidance and support throughout this project. We also acknowledge the Department of Applied Data Science at San Jose State University for providing the resources and infrastructure needed for this work.

This project was built using open-source tools and public datasets including the Enron Email Corpus, AMI Meeting Corpus, MeetingBank, QMSum, and Python Developer Stack archives. We thank the maintainers of these datasets and the open-source community.

Interested in partnering with us on a capstone project or exploring other ways to collaborate with the College of IDS? Contact CIDS_CAP@sjsu.edu to discuss potential projects and partnership opportunities.