# AAI-510 – Team Project Status Update Form

Fill out this form and submit it by the end of Module 4 in Canvas.

Team Number: **AAI-510 Group 9**

Team Leader/Representative: Gangadhar Singh Shiva

Full Names of Team Members:

1. Gangadhar Singh Shiva

2. Chandraker, Ananya

3. Sharik, Mohd

Title of Your Project: **MultiAgent Orchestration for Unlocking Semantic Processing**

**Short Description of Your Project and Objectives:**

This research addresses the challenge of performing accurate, interpretable, and domain-sensitive sentiment analysis across heterogeneous input domains using a modular, agent-based system. We propose and implement an architecture that leverages both Large Language Models (LLMs) and classical machine learning (Random Forest) in conjunction with Retrieval-Augmented Generation (RAG) to improve sentiment prediction using real-world textual datasets.

At the core of the system lies an A2A (Agent-to-Agent) Coordinator Agent that dynamically routes incoming **sentiment queries to specialized tool agents** based on **zero-shot classification**. Each tool agent is backed by either a transformer-based sentiment model (DistilBERT) or a Random Forest model trained on TF-IDF vectors, depending on the experimental configuration. The agents enrich prediction by performing semantic similarity retrieval using RAG from a local corpus (e.g., **iphone.csv, twitter\_validation.csv**) to provide contextual justification alongside sentiment classification.

**Problem Statement**

The central problem addressed in this work is: How can we build a multi-domain sentiment analysis system that balances interpretability, modularity, and accuracy by combining agent-based coordination, RAG-based knowledge retrieval, and both neural and classical machine learning techniques?

This research compares three approaches:

* **LLM-based A2A with RAG**: Routing and classification done by transformer models.
* **Random Forest + RAG**: Using classical ML models and corpus similarity search.
* **MCP-style LLM coordination**: Prompt-based tool invocation using semantic dispatch.

**Name of Your Selected Dataset: twitter Sentiment Analysis and Iphone Users Sentiment Analysis Dataset**

### Purpose: twitter Sentiment Analysis

Dataset contains the text of a tweet and a sentiment label. Training set is provided with a word or phrase drawn from the tweet (selected\_text) that encapsulates the provided sentiment.

| **Column Name** | **Description** |
| --- | --- |
| textID | unique ID for each piece of text |
| text | the text of the tweet |
| sentiment | the general sentiment of the tweet |
| selected\_text | [train only] the text that supports the tweet's sentiment |

### 

### Purpose: iphone sentiment analysis

This dataset likely contains **customer reviews or feedback** about iPhone products. It is commonly used for sentiment analysis, text classification, or feature extraction tasks.

### 

| Column Name | Description |
| --- | --- |
| reviewDescription | The raw text of a user's review or comment about the iPhone. |
| rating | A numeric score (e.g., 1–5 stars) associated with the review. |
| date | The date the review was posted. |
| reviewTitle | A short summary or title of the review. |
| verifiedPurchase | Boolean or categorical flag indicating if the review came from a verified buyer. |
| sentiment *(optional)* | Pre-labeled sentiment tag (positive, negative, neutral) for supervised learning. |

### 

Please provide GitHub the link here: **https://github.com/gshiva1975/AAI-510**

How many times have your members met in the last two weeks? 3

List the specific contributions that each team member is providing for the Final Team Project in the table below.

* **NOTE:** ALL students on the team should contribute equally to the Final Team Project.

| Gangadhar Singh Shiva | Chandraker, Ananya | Sharik, Mohd |
| --- | --- | --- |
| Create Github Repo  Data Collection & Cleaning | Data Collection & Cleaning | Data Collection & Cleaning |
| Exploratory Data Analysis | Exploratory Data Analysis | Exploratory Data Analysis |
| Feature Engineering | Feature Engineering | Feature Engineering |
| NLP & Sentiment Analysis | NLP & Sentiment Analysis | NLP & Sentiment Analysis |
| Model Optimization | Model Optimization | Model Optimization |
| A2A, RAG Implementation | MCP, RAG Implementation | Random Forest, RAG Implementation |
| Model Training & Testing,Documentation | DeploymentModel Training & Testing,Documentation | Model Training & Testing,Documentation |

**Proposed Architecture**

**1. LLM-Based A2A with RAG**

* **Coordinator**: Uses a zero-shot classifier (e.g., BART, RoBERTa) to select iphone or twitter agent.
* **Agents**:
  + Use DistilBERT for sentiment classification.
  + Use sentence-transformers (MiniLM-L6-v2) for RAG.
  + Load respective CSV corpus.
* **Dispatch**: Via ToolCall/ToolResult mechanism.
* **Strengths**: Modular, generalizes well, explainable with retrieved examples.
* **Limitations**: Higher compute cost, slower inference.

**2. Random Forest + RAG**

* **Manual Routing**: Based on test case selection.
* **Classifier**: RandomForestClassifier trained on TF-IDF features from corpus.
* **RAG**: TF-IDF vectors + cosine similarity to retrieve matching examples.
* **Strengths**: Fast, interpretable, no GPU needed.
* **Limitations**: Sensitive to vocabulary, lower semantic capacity.

**3. MCP-Style LLM Coordination + RAG**

* **LLM (e.g., GPT)** reads prompt and selects tool via language or function call style.
* **Tools/Handlers**:
  + One for each domain: use DistilBERT + SBERT-based RAG.
* **Workflow**:
  + All context passed in a single LLM call.
  + Output sentiment and most similar examples.
* **Strengths**: Seamless tool orchestration, easy to scale.
* **Limitations**: Requires large prompt context, traceability of tool usage less transparent.

**Outcomes:**

* A reusable, interpretable, agent-based sentiment analysis system.
* Comparative evaluation across transformer-based, classical ML, and unified LLM approaches.
* Quantitative evaluation supported by metrics including accuracy, F1, and ROC-AUC.
* RAG integration improves explainability by justifying predictions.

Comments/ Roadblocks: Using Cloud account, Training model with RAG, Agent to Agent Protocol, Message Context Protocol and NLP

**Workflow**

1. **User Input**: The system receives a sentiment query.
2. **Routing Decision**:
   * *A2A*: Coordinator uses zero-shot classification to select the appropriate agent.
   * *MCP*: A single LLM interprets the prompt and selects a tool via natural language or function call.
3. **Agent Execution**:
   * Loads domain-specific corpus from CSV (iphone or twitter).
   * Performs sentiment classification using either DistilBERT or Random Forest.
   * Performs RAG retrieval from the same corpus.
4. **Response**: The agent returns sentiment label and top-matching retrieved example.
5. **Evaluation**: Prediction is validated against expected sentiment, agent selection, and retrieval relevance.

**Evaluation Metrics**

| **Metric** | **Description** |
| --- | --- |
| **Routing Accuracy** | Whether the coordinator sends the prompt to the correct agent |
| **Sentiment Accuracy** | Whether the predicted sentiment matches the ground truth label |
| **F1 Score** | Harmonic mean of precision and recall, useful in imbalanced datasets |
| **ROC-AUC Score** | Measures the model’s ability to distinguish between classes |
| **RAG Relevance** | Semantic similarity score between input and retrieved corpus sentence (cosine) |
| **Explainability** | Qualitative alignment of matched examples with prediction |

**Metric Implementation Details:**

* **Sentiment Accuracy**: Computed using accuracy\_score() from scikit-learn.
* **F1 Score**: Calculated via f1\_score() to handle class imbalance.
* **ROC-AUC Score**: Measured using roc\_auc\_score() for binary sentiment classification.
* **RAG Relevance**: Uses cosine similarity scores between query and top-k retrieved sentences.
* **Routing Accuracy**: Boolean comparison between expected and selected tool agent.

**Sample Evaluation Case:**

* **Prompt**: "The iPhone battery drains too fast"
* **Expected Agent**: iphone\_sentiment
* **Expected Sentiment**: NEGATIVE
* **Score**: 1 if both sentiment and agent match.