Title: AI Document Analyzer and Keyword Extractor

**College Name: Nagarjuna College of Engineering and Technology**

**Team Members:**

1. Mohammed Musaib CAN\_32727772

**Objective**

The testing phase aimed to validate the AI Document Analyzer’s functionality, performance, and reliability, addressing issues identified during development (e.g., neutral sentiment scores, PDF download errors). The goal was to ensure the system meets requirements and is ready for deployment.

**Activities**

1. **Unit Testing**:
   * Tested text extraction for each file type (PDF, DOCX, TXT, images) using sample files, including Resume\_Musaib.pdf.
   * Verified NLP components: sentiment analysis (IBM Watson, vaderSentiment), keyword/entity extraction, summarization, and question answering.
   * Checked web interface elements: file upload, keyword input, result display, and download functionality.
2. **Integration Testing**:
   * Validated end-to-end workflows: uploading a document, processing it, displaying results, and downloading outputs.
   * Tested API integration with IBM Watson NLU, ensuring proper handling of rate limits and errors.
   * Confirmed session management for storing and retrieving analysis results across requests.
3. **Issue Resolution**:
   * Investigated neutral sentiment scores (Score: 0) by calibrating vaderSentiment thresholds and ensuring IBM Watson input text length compliance.
   * Fixed PDF download errors by sanitizing text inputs and verifying DejaVuSans.ttf font availability.
   * Addressed summary generation errors by limiting input text length to 1000 characters for transformers.
4. **Performance Testing**:
   * Measured processing time for documents of varying sizes (e.g., 1MB PDF, 500KB image), targeting <5 seconds for analysis.
   * Tested system under load with multiple simultaneous uploads (up to 5 files), ensuring no crashes.
   * Verified file size limit enforcement (10MB) and cleanup of temporary files.
5. **User Acceptance Testing**:
   * Conducted tests with sample users to validate usability of the web interface, focusing on file upload, result interpretation, and question answering.
   * Confirmed accurate answers to questions (e.g., phone number: +91-7795888591 from Resume\_Musaib.pdf).
   * Gathered feedback on UI responsiveness and error message clarity.

**Deliverables**

* **Test Plan**: Detailed scenarios covering unit, integration, performance, and user acceptance tests.
* **Test Results Report**: Summary of test outcomes, including pass/fail rates and issue resolutions.
* **Bug Fixes Log**: Documentation of issues (e.g., sentiment score, PDF errors) and their resolutions.
* **User Feedback Summary**: Insights from user testing, highlighting usability improvements.

**Outcomes**

* Achieved >95% accuracy in text extraction across file types.
* Resolved critical issues: neutral sentiment scores improved by adjusting thresholds, PDF downloads fixed by text sanitization.
* Confirmed system reliability under normal and stress conditions.
* Identified minor UI enhancements (e.g., clearer error messages) for deployment.

**Next Steps**

* Proceed to Phase 5: Deployment, to operationalize the system and implement monitoring.