# O-1A Visa Assessment Tool: Design Choices and Evaluation Guide

**Yash Agrawal** yash4agr@gmail.com **6th march 2025**

---

## A. System Design Overview:

The O-1A Visa Assessment Tool is designed to automate the preliminary evaluation of candidates for the O-1A visa category. This document outlines the key design choices made during development and provides guidelines for evaluating the application's output.

### Architecture

The application follows a modern microservices-inspired architecture with clear separation of concerns:

1. FastAPI Backend: Handles file uploads and assessment requests
2. RAG (Retrieval Augmented Generation) System: Core intelligence for document analysis
3. Streamlit Frontend: User-friendly interface for file uploads and displaying results

## B. Key Design Choices:

### 1. Retrieval Augmented Generation (RAG) Approach

Why RAG?

- Immigration visa assessment requires specific domain knowledge about USCIS regulations
- RAG allows the application to access specialized knowledge about O-1A criteria
- Provides transparent references to specific regulatory guidelines

Implementation Details:

- Knowledge documents are stored in the knowledge_base directory
- Documents are chunked and embedded using TogetherAI embeddings
- Retrieved context is provided to the LLM for accurate assessment

### 2. Together AI Integration

Why Together AI?

- Provides high-quality models through a simple API
- Cost-effective compared to alternatives
- Offers both embedding and LLM capabilities with consistent interfaces

Implementation Details:

- TogetherEmbeddings class for document and query embedding
- OpenAI-compatible API for LLM interactions
- Error handling for token limits and other API constraints

### 3. Document Processing Pipeline

Why Multiple Document Processors?

- Real-world CVs come in various formats (PDF, DOCX, TXT)
- Each format requires specialized extraction techniques
- Text preprocessing ensures consistency across formats

Implementation Details:

- Factory pattern for document processor selection based on MIME type
- Format-specific text extraction and normalization
- Unified preprocessing to handle common extraction artifacts

### 4. Assessment Logic

Why Criteria-Based Assessment?

- O-1A visa explicitly requires meeting at least 3 of 8 criteria
- Different evidence varies in quality and strength
- Complex regulatory framework requires nuanced evaluation

Implementation Details:

- Evidence matched to specific criteria with confidence scores
- Rating calculation based on number of criteria met with good confidence (≥0.6)
- Identification of unmet criteria for feedback

### 5. Structured Output

Why JSON-Based Response Format?

- Structured data enables consistent UI presentation
- Facilitates programmatic handling of assessment results
- Ensures all criteria are explicitly addressed

Implementation Details:

- Pydantic models for response validation
- Criteria organized by confidence for better user experience
- Evidence linked to regulatory sources for transparency

# C. Evaluation Guide:

## Evaluating Output Quality

When evaluating the application's assessment results, consider these factors:

### 1. Evidence Matching Accuracy

Excellent:

- Evidence is correctly extracted from the CV
- Evidence is matched to appropriate criteria
- Confidence scores accurately reflect match strength

Good:

- Minor misalignments between evidence and criteria
- Mostly appropriate confidence scores

- Most relevant CV sections identified

Needs Improvement:

- Evidence incorrectly matched to criteria
- Significant CV information missed
- Confidence scores don't reflect match quality

### 2. Consistency Across File Formats

Test Method: Upload the same CV in different formats (PDF, DOCX, TXT)
Excellent:

- Same rating across all formats
- Same criteria matches with similar confidence scores
- Consistent evidence extraction

Needs Improvement:

- Different ratings for the same content
- Major discrepancies in identified evidence
- Format-dependent matches

### 3. Rating Accuracy

Excellent:

- "High" rating for CVs with 5+ strong criteria matches
- "Medium" rating for CVs with 3-4 strong criteria matches
- "Low" rating for CVs with <3 strong criteria matches

Needs Improvement:

- Ratings inconsistent with criteria matches
- Failure to distinguish between strong and weak evidence
- Overemphasis on quantity over quality of evidence

## System Performance

- Processing Time: Typical assessment should complete within 15-30 seconds
- Error Handling: Application should gracefully handle malformed inputs
- Scalability: System designed for individual assessments, not bulk processing

# D. Implementation Challenges and Solutions:

## Challenge 1: Inconsistent Text Extraction

PDF and DOCX files presented challenges for reliable text extraction, especially with complex formatting.
Solution:

- Implemented specialized processors for each file format
- Added preprocessing to normalize text and remove common extraction artifacts
- Included error handling for malformed documents

### Challenge 2: Token Limits with Together API

Initial implementation encountered issues with token limits when embedding long documents.
Solution:

- Reduced chunk size to stay within token limits
- Added error handling for API limits
- Implemented fallback strategies for oversized inputs

### Challenge 3: Consistent Criteria Matching

Early tests showed inconsistent criteria matching and naming.
Solution:

- Enhanced prompt engineering to enforce consistent criteria naming
- Added post-processing to remove empty or irrelevant matches
- Used standardized confidence scoring for consistent evaluation

## E. Future Improvements:

1. Document type detection: Automatically detect file types rather than relying on MIME types
2. Multiple CV support: Allow comparison of multiple candidates
3. Enhanced explanations: Provide more detailed guidance on improving eligibility
4. Custom knowledge base: Allow users to upload reference materials
5. Batch processing: Enable processing multiple CVs in bulk

## F. Conclusion:

The O-1A Visa Assessment Tool provides a structured, consistent approach to preliminary visa eligibility assessment. By leveraging modern AI techniques like RAG and LLMs, it delivers insights that help applicants understand their eligibility profile and areas for improvement.
While not a replacement for professional immigration advice, this tool offers a valuable first step in the assessment process, helping applicants prepare more effectively for formal evaluation.