AI-Powered Parts Manual Search System

Project Understanding & Implementation Guide

Project Overview

The client requires a system that uses AI to search equipment parts manuals (PDFs) to help customers find correct part numbers based on descriptions. The system would:

1. Store a library of PDF manuals (~7000) on AWS
2. Allow users to search for parts by description (e.g., "brake cylinder")
3. Return accurate part numbers without giving direct access to the PDFs
4. Link these part numbers to products on the client's e-commerce website

---

Current Challenges

- **Manual Selection**: Ensuring the correct PDF is searched when multiple manuals exist for similar models
- **Content Complexity**: Handling multiple models covered within a single PDF
- **Data Quality**: Processing scanned PDFs with poor image quality
- **Format Variation**: Managing part numbers with inconsistent formatting (dashes, leading zeros)
- **AI Reliability**: Preventing hallucination of non-existent part numbers
- **Scale**: Maintaining accuracy across approximately 7000 manuals

---

Example Workflow

Customer Scenario

A maintenance technician needs a brake cylinder for a Hyster H50XM forklift with serial number C177B12345.

1. User Input

The technician enters:

- Equipment: Hyster H50XM
- Serial Number: C177B12345
- Part needed: "brake cylinder"

2. Backend Process

- System identifies the correct PDF manual based on model AND serial number prefix (C177)
- AI searches only that specific manual for "brake cylinder"
- AI extracts all relevant part numbers with their descriptions
- System verifies these part numbers against the "universe" Excel file for Hyster
- System returns only verified, legitimate part numbers

3. Result Displayed to User

```
Found 3 brake cylinder components for your Hyster H50XM (Serial C177B12345):

1. Master Brake Cylinder (Part #1234567) - Located in brake assembly section, page 45
2. Left Wheel Brake Cylinder (Part #1234568) - Located in wheel assembly section, page 62
3. Right Wheel Brake Cylinder (Part #1234569) - Located in wheel assembly section, page 62

Select a part to view purchasing options.
```

4. Purchase Flow

When the user selects a part, they're taken to that product page on the client's website, without ever accessing the PDF manual directly.

Technical Components Required

1. **PDF Processing System**
   - Extract and index text from all PDFs
   - Handle OCR for scanned documents
   - Create searchable text representations
2. **Equipment Database**
   - Map models and serial number prefixes to specific manuals
   - Handle the complexity of multiple manuals per model
3. **AI Query System**
   - Process natural language part descriptions
   - Search the correct PDF for matching parts
   - Extract part numbers, descriptions, and locations
4. **Verification System**

- Check extracted part numbers against "universe" Excel files
- Normalize part number formats (removing dashes, leading zeros)
- Prevent hallucination of invalid part numbers

5. **Web Interface**
   - Allow users to specify equipment details and parts needed
   - Display results in a clear, actionable format
6. **Product Linking System**
   - Connect identified parts to purchasable products on the website
   - Provide direct purchase pathways

Solution Approaches for Key Challenges

PDF Selection Issue

- Require users to provide both model AND serial number
- Create a lookup table mapping serial number prefixes to specific PDF documents
- Implement fuzzy matching for model numbers to account for user input variations

Multiple Models in One PDF

- Have the AI first identify which section of the manual applies to the specific model
- Only search relevant sections for part information
- Use section headers and model references to partition the document

Image Quality

- Run OCR improvement preprocessing on the PDFs
- Store pre-extracted text alongside PDFs for better search results
- Implement confidence scoring for extracted text

Part Number Formatting

- Implement normalization rules (removing dashes, leading zeros)
- Create a mapping system between formatted and normalized part numbers
- Match normalized versions against the product database

PDF Protection

- Extract and index PDF content server-side only

- Never expose PDF URLs or download options to end users
- Implement strict access controls on the AWS storage

Implementation Roadmap

1. **Phase 1: Proof of Concept**
   - Test with a small set (~10) of diverse PDF manuals
   - Validate accuracy of part number extraction
   - Evaluate model selection accuracy
2. **Phase 2: Core System Development**
   - Build the equipment-to-manual mapping database
   - Develop the AI search and extraction system
   - Create the part number verification system
3. **Phase 3: Integration**
   - Connect to e-commerce product database
   - Develop user interface for searching
   - Implement results display and purchasing flow
4. **Phase 4: Scaling**
   - Expand to process all 7000 manuals
   - Optimize for performance with large document sets
   - Implement caching for common searches
5. **Phase 5: Refinement**
   - Add learning mechanisms to improve over time
   - Implement user feedback loops
   - Fine-tune accuracy and response time

Testing & Quality Assurance

- **Manual Testing**: Compare AI results with manual lookups in PDFs for accuracy
- **Edge Cases**: Test with poorly scanned documents, unusual part descriptions
- **Volume Testing**: Ensure performance with multiple simultaneous users
- **Feedback Loop**: Implement system to flag and correct errors

---

Success Metrics

- **Accuracy**: >95% correct part number identification
- **Coverage**: Successful handling of >99% of user queries
- **Speed**: Results returned in <5 seconds
- **Conversion**: Increase in parts sales from manual references

---

This document serves as an initial understanding of requirements. Further technical specifications and system design details will need to be developed in consultation with the client.