

## MCP OCR Server Test Document

This document is created for the purpose of testing the OCR (Optical Character Recognition) capabilities of the MCP Server using:

- Tesseract Model
- Nougat Model

The test involves:

- Extracting text from PDF using direct extraction (PyMuPDF)
- Running OCR on PDF images (if direct fails)
- Summarization using AI (Groq or OpenAI)

Tasks:

1. Upload this PDF to the OCR server.
2. Test direct text extraction.
3. If no text is extracted, test image-based OCR.
4. If required, use summarization after text extraction.

Real estate fraud is a growing concern in today's digital landscape.

Many scammers use fake listings, impersonation, and identity theft to target innocent buyers.

This test ensures that your OCR pipeline accurately extracts such content and optionally summarizes it for downstream processing.

Training Phase Example (for model):

Prompt: "Detect fraud intent: 'We guarantee profit if you deposit money today.'"

Completion: "Intent detected: Fraud - urgency and money demand keywords present."

