



# Microsoft OCR Pre-Deployment Audit Checklist

Use this functional checklist to map your enterprise content, determine the appropriate Microsoft OCR service, and identify potential hidden architectural costs before initiating a pilot.

## Phase 1: Content Repository Mapping

Identify where your target files currently reside to determine the most cost-effective OCR engine.

- [ ] **Microsoft 365 Native:** Over 80% of target files reside in SharePoint Online, OneDrive, or Exchange. (*Recommendation: Microsoft Purview OCR*)
- [ ] **External Cloud/On-Premises:** Files reside in AWS, Google Cloud, local file shares, or custom databases. (*Recommendation: Azure AI Document Intelligence*)
- [ ] **Hybrid Environment:** Data is split between M365 compliance borders and custom application storage. (*Recommendation: Split deployment model*)

## Phase 2: Document Shape & Volume Estimation

Calculate the structural complexity of your documents, as Microsoft bills heavily rely on page counts and item processing.

Metric	Estimated Volume	Notes / Risk Factors
Total Target Files	_____	Pure image formats (JPEG/PNG) count as single items.
Average PDF Page Count	_____	Long, multi-page PDFs multiply Azure transaction costs.
Blank Page Frequency	_____	High frequency requires page-level processing optimization.



<b>Handwritten Content</b>	Yes / No	Requires advanced zero-shot extraction capabilities.
----------------------------	----------	--

### Phase 3: Hidden Infrastructure Costs Assessment

Evaluate secondary architectural expenses that are often excluded from initial software budgets.

- **Storage Allocation:** We have provisioned Azure Blob Storage to house extracted JSON data and bounding box coordinates.
- **Data Egress Calculation:** We have mapped out bandwidth fees if processing documents in a different Azure region than our primary data center.
- **Validation Workflows (HITL):** We have factored in the operational costs and labor hours for a Human-in-the-Loop team to review low-confidence scores.
- **Geographic Compliance:** Our selected Azure region complies with local data residency laws (e.g., GDPR, CCPA).

### Phase 4: Pilot Project Definition

Structure your initial 30-day testing phase to establish a reliable cost baseline.

- **Sample Set Selected:** We have isolated a small, diverse sample set of 1,000–2,500 "ugly" files (crooked scans, faxed PDFs) rather than polished exports.
- **Success Metrics Defined:** We have established acceptable thresholds for false positives and manual rework time.
- **Billing Alerts Configured:** Budget limits and automated alerts are configured within the Azure portal/Purview dashboard to prevent overage charges during testing.