Functional Requirements

- Document Ingestion
  - Accept PDF upload (web/form/HTTP API)
  - Accept plain text inputs and email bodies
  - Validate file formats, fail with descriptive errors

- Text Extraction
  - Extract text from PDF (OCR for scanned images + native text layer)
  - Normalize text (whitespace, newlines, encoding)
  - Extract metadata (title, author, upload timestamp)

- Risk Assessment
  Evaluate each document against rules in stages:
  1. C1: data classification
  2. C2: policy matching
  3. C3: risk scoring
  4. C4: exception handling
  5. C5: summary normalizing

- AI Processing
  - RAG/LLM pipeline for context augmentation
  - Use semantic/embeddings retrieval when query requires

- API & Integrations
  - Expose REST endpoints:
  - Authenticate calls (API key or OAuth)
  - Log every request + response code for audits