

# Builder Spec - ASCII Edition

Version 2025-09-01

## Executive Summary

Build a Dockerized REST API that ingests an in-force illustration PDF, extracts key fields (carrier-agnostic), PII scrubs content, computes a confidence score, generates proof snips, and returns decision-ready JSON plus a redacted PDF. If confidence  $\geq 0.80$  mark Final; otherwise mark Provisional - Needs Review but still return full output. No user interaction beyond upload.

## Deliverables

- Endpoints: POST /analyze (main), POST /extract\_text (optional PDF to text)
- Extractor: anchors + synonyms + table ROI + shape and math checks
- PII scrub: JSON, snips, redacted PDF; strip metadata
- Proof snips: rate header, first-year row, current-year row, surrender row, loan callout
- Auto-profiles: saved layout hints after success
- QA: 20-30 PDF golden set; PII audit view; README + Postman + Dockerfile

## Request/Response (POST /analyze)

Request: multipart 'file' = PDF. Response JSON: decision\_ready, needs\_manual\_review, confidence\_overall, fields{...}, series{...}, verifications{...}, proof\_snips (base64 PNG), redacted\_pdf\_b64, notes[].

## Schema and Scenario

Columns: year, age, premium, death\_benefit, cash\_value, surrender\_charge, net\_surrender\_value, loan\_balance, loan\_interest, withdrawals. Scenario priority: Current, Current Assumption, Non-Guaranteed, Current Interest.

## No Touch Parsing

- 1) Find candidate ledgers (Year + money columns)
- 2) Auto label columns via synonyms + shape rules (year+, age 18-100+, surrender\_charge down to 0, net\_sv  $\leq$  cash\_value)
- 3) Pick Current column; verify identity: net\_sv  $\approx$  cash\_value - surrender\_charge - loan - interest ( $\leq 1$  percent error)
- 4) Choose current-year row; stitch pages; normalize numbers

## Confidence and Gating

Blend: header 20 percent, shape 25 percent, reconciliation 35 percent, rows parsed 20 percent. Threshold  $\geq 0.80 \rightarrow$  decision\_ready=true; else needs\_manual\_review=true and still emit full output.

## PII Scrub

Names  $\rightarrow$  CLIENT/INSURED/OWNER. Policy IDs  $\rightarrow$  \*\*\*\*-\*\*\*\*-1234 (keep last 2-4). Remove phones, emails, SSN, DOB; drop street address (optionally keep city, state ZIP). Blur barcodes/QR/signatures. Strip PDF metadata.

## Proof Snips

Embed small screenshots: rate header; ledger first-year row; current-year row; surrender-charge current row; loan callout.

## Auto Profiles

Save carrier/product tokens, ROI hints, scenario index, label map. Reuse on similar uploads. No user clicks.

## Acceptance

Accuracy  $\geq$  95 percent; PII leaks = 0; runtime  $\leq$  30s incl. OCR; confidence threshold enforced; docs and Postman provided.