

This is a test file

SYSTEM_PROMPT_MERGE = ''''''

You merge two extracts that describe the same PDF page:

(1) vis_text – produced by Vision OCR, rich in tables/figures

(2) raw_text – produced by a PDF text extractor, reliable for plain text

Return strict JSON with keys:

merged_text : best reading order text

omissions : list of fragments present in only one source

table_json : list of tables; each table is a list of rows (each row is list of cells)

Rules:

- Prefer vis_text for tables and images.
- Prefer raw_text for headers, footers, running paragraphs when duplicates occur.
- Do NOT add or hallucinate content.

''''''

''''''

mixed_layout_extractor.py

Author: Jun Xu

Robust PDF extractor that routes pages with large figures or tables

through GPT-4o Vision and merges the result with classical text extraction.

''''''



item	categories		comments
	A	B	
1	a.1	b.1	Test result1
2	a.2	b.2	Test result 2
3	a.3	b.3	Test result 4

COMPLIANCE-RELATED USE CASES	
Automating compliance processes	AI can automate compliance processes such as documentation, audits, and risk assessment to reduce human errors and help employees use time more strategically.
Predictive analytics for proactive issue identification	You can use predictive analytics to identify issues proactively before they happen and deal with individuals and departments accordingly.
Chatbots and virtual assistants for compliance	You can also use chatbots and virtual assistants to answer compliance questions from employees.
Changing compliance requirements	AI can help keep pace with changing compliance requirements by staying up-to-date and proposing improvement areas in compliance programs.

