UDV-TR-001. UniDocVerse System Overview

Abstract

(Already written earlier, but here is a polished version if you want consistency.) Organizations increasingly require AI-driven document intelligence without compromising privacy or compliance. UniDocVerse introduces a fully offline, multi-agent system capable of processing any document type—PDFs, Word files, spreadsheets, images, and more. Built on LangGraph, pgvector, and adaptive parsing, the system delivers summaries, insights, semantic search, entity linking, and domain-specific analytics. This report presents the architecture, implementation, evaluation, and future roadmap of UniDocVerse, demonstrating its applicability across enterprise, legal, financial, healthcare, and government environments.

UDV-TR-002. Procurement Document Intelligence

Abstract

Procurement workflows generate diverse document types—purchase orders, quotations, vendor agreements, and RFPs—each containing critical operational and financial information. Traditional cloud-based tools struggle with inconsistent formats, privacy constraints, and limited cross-document analysis. This report presents UniDocVerse's procurement intelligence module, a fully offline, multi-agent system that extracts structured data, identifies key entities, links related documents, and generates actionable insights. Using adaptive parsing, domain-specific analyzers, and hybrid semantic-keyword search, the system delivers reliable analytics across heterogeneous procurement documents while maintaining strict privacy and compliance guarantees.

UDV-TR-003. Operational Document Intelligence (Logistics)

Abstract

Logistics operations rely on documents such as bills of lading, delivery notes, packing lists, and shipment records—each with unique structures and compliance requirements. UniDocVerse introduces a privacy-first, offline intelligence layer that standardizes extraction, validates shipment data, identifies discrepancies, and links related operational documents. Through a multi-agent pipeline and specialized logistics analyzers, the system provides structured insights, cross-document relationships, and real-time metrics. This report details the architecture, extraction strategies, and evaluation of UniDocVerse's logistics intelligence module.

UDV-TR-004. Academic Document Intelligence

Abstract

Academic institutions manage transcripts, certificates, mark sheets, and institutional records, often in inconsistent formats and across multiple systems. UniDocVerse’s academic intelligence module provides a unified, offline solution for parsing, validating, and analyzing academic documents. Using adaptive parsing, domain-aware analyzers, and entity linking, the system extracts structured academic data, computes performance metrics, and identifies cross-document relationships. This report outlines the architecture and evaluation of UniDocVerse’s academic document processing pipeline, demonstrating its applicability across global education systems.

UDV-TR-005. Insurance Document Intelligence

Abstract

Insurance workflows involve complex documents such as claims, policy summaries, coverage details, and risk assessments. These documents often contain dense, unstructured information that is difficult to analyze using traditional tools. UniDocVerse’s insurance intelligence module provides a fully offline, multi-agent system for extracting policy attributes, identifying claim details, detecting risk indicators, and linking related documents. This report describes the architecture, domain-specific extraction strategies, and evaluation of the insurance analyzer, highlighting its ability to deliver accurate insights while preserving data privacy.

UDV-TR-006. HR Document Intelligence

Abstract

Human resources departments manage resumes, offer letters, employment contracts, payroll documents, and compliance records—each requiring precise extraction and interpretation. UniDocVerse’s HR intelligence module automates this process through a privacy-first, offline pipeline that extracts structured data, identifies employment attributes, analyzes compensation patterns, and links related documents. This report presents the HR analyzer architecture, domain-specific extraction logic, and evaluation results, demonstrating its effectiveness across diverse HR document formats.

UDV-TR-007. Legal Document Intelligence

Abstract

Legal workflows depend on contracts, tax forms, regulatory filings, and vendor agreements, all of which require high-accuracy extraction and strict privacy controls. UniDocVerse’s legal intelligence module provides a deterministic, offline system for parsing legal documents, identifying obligations,

extracting clauses, detecting compliance risks, and linking related records. This report details the legal analyzer’s architecture, clause-level extraction strategies, and evaluation, demonstrating its suitability for enterprise and regulated environments.

UDV-TR-008. Financial Document Intelligence

Abstract

Financial documents—including bank statements, payroll stubs, and credit card summaries—contain structured and semi-structured data that must be extracted with high precision. UniDocVerse’s financial intelligence module uses multi-signal validation, transaction-level parsing, and domain-specific analyzers to extract balances, transactions, income details, and spending patterns. This report describes the architecture, extraction logic, and evaluation of the financial analyzer, highlighting its accuracy, privacy guarantees, and applicability across global financial formats.

UDV-TR-009. Invoice & Receipt Intelligence

Abstract

Invoices and receipts vary widely in layout, currency, tax structure, and line-item formatting. UniDocVerse’s invoice and receipt intelligence module provides deterministic line-item extraction, multi-currency parsing, tax detection, and totals reconciliation through a fully offline, multi-agent pipeline. This report outlines the architecture, table-extraction strategies, and validation mechanisms that enable consistent, high-accuracy extraction across diverse invoice and receipt formats.

Limitations

UniDocVerse has not yet been tested on 1,000+ documents, and large-scale stress behavior is unverified. Domain-specific analyzers continue to evolve, and edge-case formats may require additional tuning. Cross-document reasoning and entity linking accuracy have not been benchmarked at scale. Performance varies based on local hardware, as all processing is fully offline.