

A Multi-Agent System for Extracting and Querying Financial Documents

Sagar Gupta

Computer Science

Abstract – Financial documents—10-K/10-Q filings, earnings releases, prospectuses, invoices, and bank statements—combine dense prose, complex tables, and heterogeneous layouts. We propose a multi-agent architecture that orchestrates document ingestion, layout-aware extraction, knowledge graph construction, retrieval-augmented reasoning, and auditable question answering. Our design integrates advances in agentic LLM collaboration (AutoGen, CAMEL), reasoning-and-acting prompts (ReAct), and document AI (LayoutLMv3, Donut) with domain standards (FIBO ontology, ISO 20022 messaging) and public data interfaces (SEC EDGAR, XBRL). We specify components, routing policies, evaluation protocols (FinQA, TAT-QA), and governance for factuality, lineage, and compliance. This paper contributes: (1) a reference architecture and coordination patterns for financial-document intelligence, (2) a schema-aligned extraction pipeline that yields structured, explainable facts, and (3) a reproducible evaluation plan emphasizing numeric reasoning, table/text fusion, and citation fidelity.

Keywords – Financial Document Intelligence, Multi-Agent Systems, Large Language Models (LLMs), Document AI, Layout-aware Extraction, Knowledge Graphs.

I. INTRODUCTION

Figure 1: Multi-Agent Financial Document Intelligence Architecture

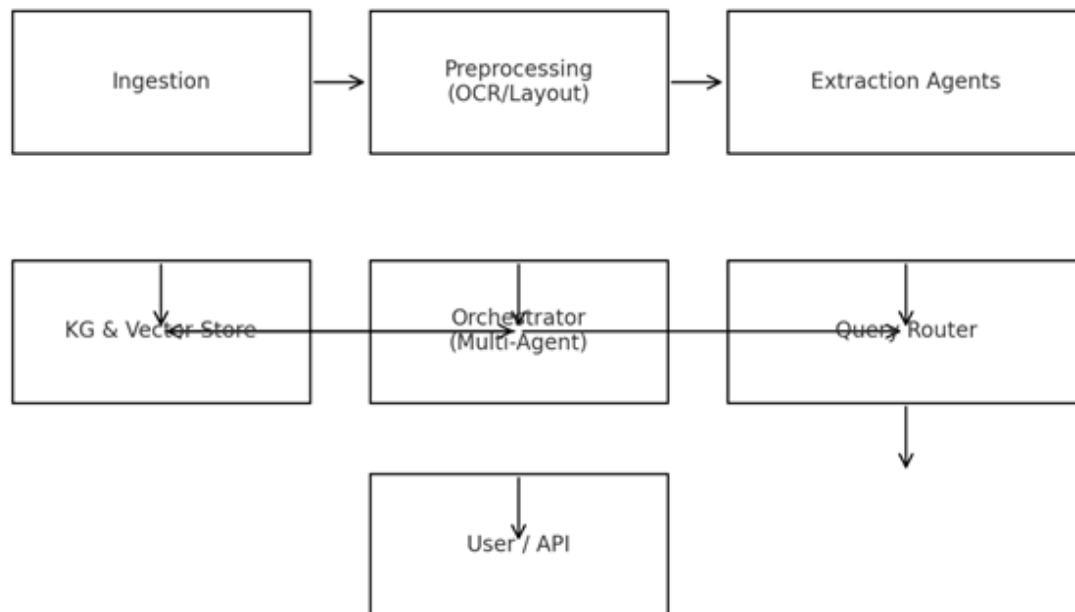
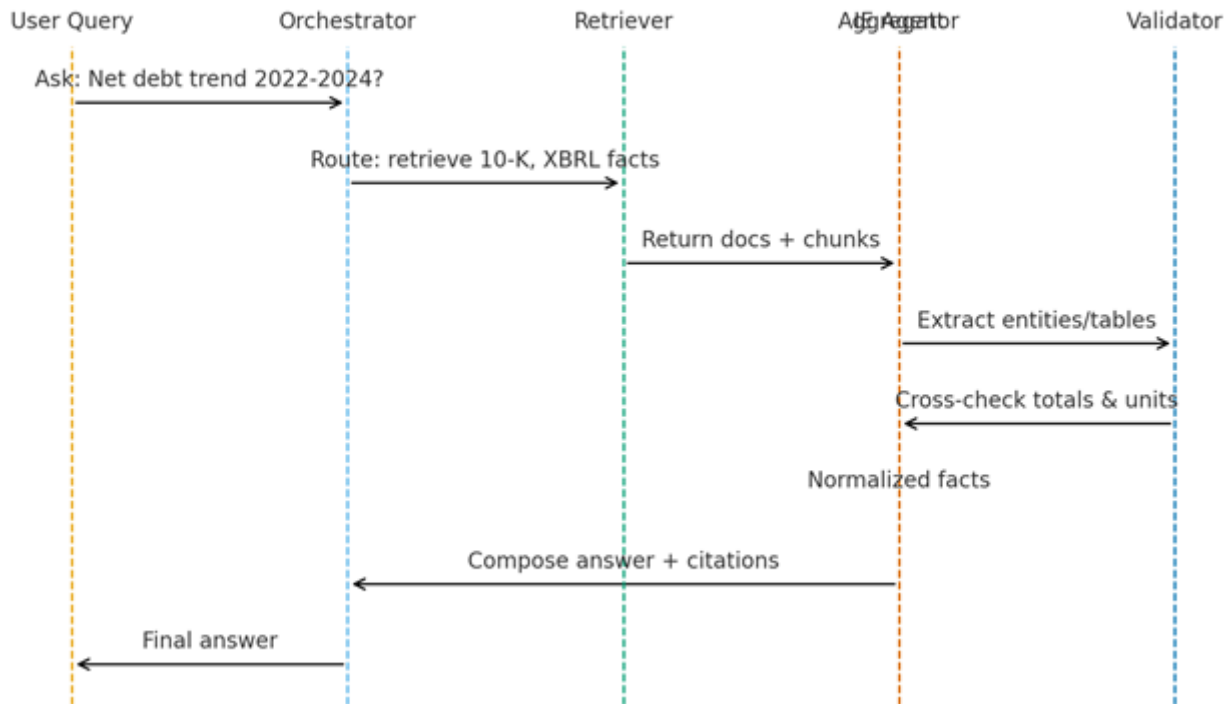


Figure 2: Agent Interaction Sequence for Query Execution



Financial analysis depends on information that is often unstructured (narratives), semi-structured (tables, footnotes), and multi-modal (scanned PDFs). Large language models (LLMs) are strong at synthesis, but reliable enterprise use requires tool-use, structured extraction, and traceable reasoning. Recent work shows agents can coordinate specialized roles and tools to solve complex tasks, offering a natural fit for end-to-end document understanding and question answering in finance. Guohao Li+3arXiv+3OpenReview+3

We target two hard problems:

P1 — Robust extraction. Layout, OCR errors, and cross-page references degrade quality. Layout-aware models (e.g., LayoutLMv3) and OCR-free models (Donut) address these issues. ECVA+3arXiv+3arXiv+3

P2 — Faithful QA. Financial QA often requires numeric reasoning across text+tables (e.g., trend deltas, ratio proofs) with source citations. Benchmarks like FinQA and TAT-QA stress these capabilities. nextplusplus.github.io+3arXiv+3ACL Anthology+3

II. RELATED WORK

Agentic LLMs. AutoGen formalizes multi-agent conversations with tool calls; CAMEL explores role-playing to coordinate specialized agents; ReAct couples reasoning traces with actions for tool use. These inform our

orchestrator and dialogue policies. arXiv+5arXiv+5Microsoft+5

Document AI. LayoutLMv3 unifies text+image masking for multimodal pretraining; Donut achieves OCR-free extraction; FUNSD and DocLayNet provide layout/field supervision for forms and diverse page designs. ACM Digital Library+7arXiv+7arXiv+7

Financial QA and reasoning. FinQA emphasizes program-style numeric reasoning; TAT-QA blends tables and text; newer FinTextQA expands to long-form regulatory and textbook sources. arXiv+4arXiv+4ACL Anthology+4 Open data and standards. SEC provides EDGAR and extracted XBRL via REST APIs; the XBRL US API complements this. FIBO standardizes financial concepts for knowledge graphs; ISO 20022 enriches payment messages with structured semantics—useful for mapping transaction-level artefacts. Swift+8Securities and Exchange Commission+8Securities and Exchange Commission+8

III. SYSTEM OVERVIEW

Architecture

Figure 1 illustrates the pipeline:

1. Ingestion. Pull PDFs/HTML from EDGAR, data rooms, or S3; fetch machine-readable XBRL where available. Securities and Exchange Commission+1

2. Preprocessing. Apply OCR (if needed), de-duplication, page segmentation, and layout detection. Prefer LayoutLMv3 features and optional Donut for OCR-free pages. arXiv+1



3. Extraction Agents. Specialists for: (a) entity & fact extraction (balances, revenues, dates, units), (b) table structure (row/column headers, spanning cells), (c) footnote linking (policies, adjustments), (d) XBRL normalization (concept, unit, period). arXiv

4. Knowledge Graph (KG) & Vector Store. Facts map to FIBO classes with provenance edges (document, page, bbox) and embeddings for RAG retrieval. EDM Council+1

5. Orchestrator (Multi-Agent). Implements role-based dialogues (CAMEL-style) with ReAct tool steps (search, retrieve, parse, verify). arXiv+1

6. Query Router. Selects between (i) KG/SPARQL for precise facts, (ii) programmatic calculators for metrics, (iii) RAG QA for narrative answers.

7. Answer Composer. Returns grounded responses with citations to page spans or XBRL facts.

Agent Roles and Policies

- **Ingestion Agent.** Monitors sources (EDGAR live feed / schedules), respects robots/rate limits, tags doc types. Securities and Exchange Commission
- **Layout Agent.** Produces tokens with (x, y) positions; picks OCR-free vs OCR-based path dynamically (Donut vs OCR+LayoutLMv3). ECVA+1
- **XBRL Agent.** Resolves concepts/units/periods using XBRL API; aligns facts to FIBO classes (e.g., DebtInstrument, FinancialStatement). XBRL US+1
- **Reasoner Agent.** Uses ReAct to plan calculations (e.g., Net debt = total debt – cash) and call tools (pandas, unit converters). arXiv
- **Verifier Agent.** Runs cross-checks (unit consistency, subtotals = sum of parts), compares narrative to table facts, flags inconsistencies.
- **Compliance Agent.** Enforces redaction and export policies; attaches provenance and versioning for audits (SOX/GDPR policies are organization-specific, not cited).

Coordination. We use AutoGen-style multi-turn conversations with role prompts and termination conditions (e.g., all checks passed + citations compiled). arXiv+1

IV. EXTRACTION AND NORMALIZATION

Layout-Aware Parsing

We parse page geometry, detect blocks (title, table, figure, footnote), and generate structure-aware spans. LayoutLMv3 provides joint text-image embeddings; Donut enables OCR-free parsing for born-digital PDFs and high-quality scans. arXiv+1

Table Understanding

Tables in 10-K filings often include multi-row headers, hierarchical labels, and footnote markers. We apply header detection, spanning-cell reconstruction, and semantic typing (e.g., CurrentAssets, Liabilities), informed by public layout datasets (DocLayNet, FUNSD) for generalization. arXiv+2ACM Digital Library+2

XBRL Alignment

When XBRL is available, the XBRL Agent requests facts and contexts (entity, period, unit) via the SEC or XBRL US endpoints, then aligns them to the KG using FIBO identifiers. This yields source-of-truth numeric values and reduces OCR risk. EDM Council+3Securities and Exchange Commission+3Securities and Exchange Commission+3

Ontology-First Schema

We model issuers, instruments, statements, and time periods using FIBO, enabling SPARQL queries like “total long-term debt for CIK X in FY2024” and consistent semantics across issuers. SpecEDM Council

V. QUERY EXECUTION

Figure 2 depicts a typical sequence:

1. **Intent detection & routing.** The Query Router classifies the request (fact lookup, computed metric, narrative synthesis).
2. **Retriever.** Pulls (a) KG facts, (b) nearest text/table chunks, and/or (c) XBRL facts for the period/unit.
3. **IE Agent.** Extracts/normalizes missing spans (e.g., from a PDF where XBRL is unavailable).
4. **Validator.** Cross-checks arithmetic (e.g., operating income + other income – interest = pre-tax income), compares against alternate disclosures.
5. **Aggregator.** Composes the final answer with citations to page images or XBRL concept URIs; adds a compact derivation trace.
6. **Orchestrator.** Applies ReAct steps to iterate until all checks pass or ambiguity is flagged for the user. arXiv

VI. EVALUATION PROTOCOL

Datasets and Tasks

- **FinQA** (numeric reasoning over financial reports): accuracy of final answers and program traces. arXiv
- **TAT-QA** (hybrid table+text): exact match, F1, and operation-level correctness. ACL Anthology
- **FinTextQA** (long-form finance/regulatory QA): factuality and source attribution. arXiv

Metrics

- **Answer quality:** EM/F1 (TAT-QA), numeric error tolerance (FinQA).
- **Faithfulness:** citation-support rate (answer sentences trace to retrieved spans/XBRL facts).
- **Extraction quality:** table structure F1; entity/relationship F1 (layout-labeling tasks, leveraging DocLayNet/FUNSD annotations). arXiv+1
- **Auditability:** proportion of answers with complete derivation traces.
- **Latency & cost:** wall-clock per query, tokens/tool calls per answer.

Ablations



- LayoutLMv3 vs Donut vs OCR+heuristics for extraction;
- With/without XBRL alignment;
- Single-agent vs AutoGen multi-agent orchestration (dialogue turns, success rate). arXiv

VII. IMPLEMENTATION NOTES

Agent framework. AutoGen-style agents with role prompts; CAMEL role-playing to stabilize collaboration; ReAct action steps for tool use (retrievers, parsers, calculators). Microsoft GitHub+2arXiv+2

Tools and stores.

- **Retrieval:** hybrid BM25 + dense embeddings over text and tables; vector store with chunk-level provenance (doc, page, bbox).
- **KG:** RDF store (FIBO vocabulary) with SPARQL endpoints. SpecEDM Council
- **External data:** SEC EDGAR APIs for filings; XBRL facts via SEC or XBRL US. Securities and Exchange Commission+1
- **Standards mapping:** for payment documents (wires, statements), map to ISO 20022 message items to improve cross-bank interoperability. ISO20022+1

Routing heuristics. If a requested metric is definable via XBRL (e.g., us-gaap:LongTermDebtNoncurrent), prefer XBRL over PDF extraction; otherwise route to IE Agent and verify against totals.

Provable lineage. Store: (i) exact page images, (ii) bounding boxes, (iii) text spans, (iv) XBRL concept+context IDs. Answers link to these artifacts for audits.

VIII. GOVERNANCE, RISK, AND COMPLIANCE (GRC)

- Factuality & hallucinations. Enforce retrieve-then-read with mandatory citations; disallow generation when provenance fails.
- Data handling. Respect SEC terms/rate limits; cache only permitted content. Securities and Exchange Commission
- Standards alignment. FIBO ensures consistent semantics across issuers; ISO 20022 helps normalize banking/payment artifacts; both reduce downstream reconciliation risk. EDM Council+2SpecEDM Council+2
- Explainability. Publish derivation traces and arithmetic programs (as in FinQA) to satisfy internal model risk governance. arXiv

IX. LIMITATIONS AND FUTURE WORK

- Layout variability. Extremely irregular tables and embedded images still challenge automated parsing;

continued pretraining on DocLayNet-like diversity helps. arXiv

- Domain shifts. Non-US filings (IFRS), multilingual docs, and scanned legacy PDFs require locale-aware models and OCR fallback.
- Temporal reasoning. Rolling metrics (TTM) and restatements need time-aware contexts and versioned KGs.
- Agent safety. Multi-agent loops can be costly or unstable; guard with explicit termination rules and budget caps. arXiv

X. CONCLUSION

We present a multi-agent reference design that unifies layout-aware extraction, ontology-grounded normalization, and verifiable QA for financial documents. By combining AutoGen/CAMEL orchestration, ReAct tool-use, LayoutLMv3/Donut extraction, and FIBO/XBRL/ISO 20022 standards, organizations can obtain auditable, accurate, and timely insights from complex filings and statements. The accompanying evaluation plan (FinQA, TAT-QA, FinTextQA) emphasizes numeric and hybrid reasoning with citation fidelity, making the system ready for regulated environments.

REFERENCES

1. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework (arXiv + OpenReview + MSR page). arXiv+2OpenReview+2
2. CAMEL: Communicative Agents for “Mind” Exploration of LLM Society (arXiv/OpenReview/PDF). arXiv+2OpenReview+2
3. ReAct: Synergizing Reasoning and Acting in Language Models (arXiv/PDF). arXiv+1
4. LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking (arXiv/PDF/ACM). arXiv+2arXiv+2
5. Donut: OCR-Free Document Understanding Transformer (ECCV/ACM + GitHub). ECVA+2ACM Digital Library+2
6. FUNSD dataset (arXiv + project). arXiv+1
7. DocLayNet dataset (arXiv/ACM/GitHub). arXiv+2ACM Digital Library+2
8. FinQA (arXiv/ACL). arXiv+1
9. TAT-QA (ACL/arXiv/GitHub). ACL Anthology+2arXiv+2
10. FinTextQA (arXiv). arXiv
11. SEC EDGAR APIs / developer resources and live filings feed. Securities and Exchange Commission+2Securities and Exchange Commission+2
12. XBRL US API. XBRL US
13. FIBO (EDM Council spec + overview + GitHub). SpecEDM Council+2EDM Council+2
14. ISO 20022 (official site + SWIFT + Fed guides + industry explainer). StoneX+4ISO20022+4ISO20022+4.

