# Auditable financial knowledge graphs explained: schema-driven LLM extraction with provenance

24 November 2026 · 8 min read · blogs.iictp.org/auditable-financial-knowledge-graphs-llm-explained

Short answer: an auditable financial knowledge graph is a structured map of entities, figures and relationships extracted from financial documents by a language model, with every element validated against a schema and traceable to its source page. An arXiv paper submitted on 20 November 2026, accepted at the ICAIF 2026 conference, proposes a method for building one and a seven-dimension way to score its quality. It matters because plain chunk-based retrieval struggles with time-dependent facts and corporate structure. This is an arXiv preprint. The authors say it was accepted at ICAIF 2026, so check the proceedings for the reviewed version, and its numbers are the authors' own claims.

**TL;DR The short version**

- An auditable financial knowledge graph maps entities, figures and relationships extracted by a language model, each validated against a schema and traced to its source page.
- An arXiv paper accepted at ICAIF 2026 proposes a construction method and a seven-dimension quality evaluation.
- Define the schema first and refuse any fact without a source page and quote, because audit needs to trace every number.

Read online, listen to it, or change the text size: https://blogs.iictp.org/auditable-financial-knowledge-graphs-llm-explained

## What is the method?

The paper, "Auditable Financial Knowledge Graphs: Schema-Driven Construction and Intrinsic Quality Evaluation", starts from the observation that financial documents mix narrative disclosures, accounting metrics, corporate structures, management commentary and facts that are only true for a period. Conventional chunk-based retrieval-augmented generation (RAG) represents these poorly. The authors propose building a temporal knowledge graph instead.

## How it works

- **Parsing.** Documents are parsed page by page by a general-purpose document-intelligence service.
- **Phased LLM extraction.** Overlapping windows of pages go through phased extraction of structural entities, quantitative observations and domain-specific relationships.
- **Schema validation.** Candidate graph elements are checked against predefined schemas.

- **Entity resolution.** Elements are consolidated with scope-aware exact, normalised, fuzzy, embedding-based and LLM-assisted matching.
- **Temporal and provenance rules.** Merge constraints, deterministic edge de-duplication and conflict-aware properties keep issuer identity and reporting-period validity, and fine-grained provenance links facts back to source evidence.

## What is new compared with earlier approaches

The paper's contribution is threefold, as stated: the construction method, an evidence-grounded temporal graph built from heterogeneous financial documents, and an intrinsic framework to evaluate the graph's quality. That framework scores seven weighted dimensions, applies two hard validation gates and reports structural diagnostics, and it separates graph quality from downstream retrieval and answer quality so problems can be located.

## Results reported by the source

The abstract describes the methodology and the evaluation framework but gives no headline accuracy figure, so none is reported here. Read the paper's tables for the measured graph-quality scores and for which documents were used.

## Limitations and open questions

- We read the abstract and submission details, not the full text, so the experimental scale and failure cases are not covered here.
- LLM-assisted entity resolution can merge entities that are different, such as similarly named subsidiaries. The hard gates and provenance exist to catch that, and their effectiveness needs checking against the paper.
- A graph is only as good as its parsing; the method depends on a document-intelligence service for page structure.

## How to apply the ideas

1. Define the schema first: issuers, periods, line items, units and relationships, and what a valid fact must contain.
2. Store a source page and quote for every extracted fact, and refuse facts without them.
3. Resolve entities within a scope, such as one issuer, before attempting cross-company merges.
4. Score the graph on its own, before any question answering, so extraction faults are not confused with retrieval faults.

## What it means for accountants and auditors

Audit and review need to trace a number to its source. A provenance-first graph is a design for that, and it lines up with structured filings, where inline XBRL tags already give each fact a concept, period and unit. A related benchmark is described in Filing-Evidence.

## Frequently asked questions

- **What is an auditable financial knowledge graph?** A graph of entities, figures and relationships from financial documents where every element is schema-validated and traceable to source evidence.
- **Why not just use RAG?** The authors say chunk-based retrieval represents time-dependent facts and corporate structure poorly.
- **What does the quality framework measure?** Seven weighted dimensions, two hard validation gates and structural diagnostics, per the abstract.
- **Where was it published?** It was accepted at the ICAIF 2026 conference, according to the arXiv comments.

## Related reading

- Filing-Evidence: small LLMs reading SEC filings
- What is a vector database?
- RAG explained
- AI Practitioner

## Sources and currency

Sources were read on 24 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail. This post is based on the abstract and arXiv details.

> Related credential: AI Practitioner. Explore the exam facts and syllabus.

### Sources and further reading


1. Auditable Financial Knowledge Graphs: Schema-Driven Construction and Intrinsic Quality Evaluation. arXiv 2611.27835, 20 November 2026; ICAIF 2026. arxiv.org/abs/2611.27835.



Published by IICTP, an independent technical credential body, on 24 November 2026. The current version of this post is at https://blogs.iictp.org/auditable-financial-knowledge-graphs-llm-explained. More guides: blogs.iictp.org.