

● RAG FOR AP & FINANCE TEAMS

From document pile to cited answer.

A grounded RAG system that answers questions over your vendor and contract documents in seconds — cited to the source, or an honest “not found.” Here’s the everyday problem it solves, how it works, and who it fits.

[See the live engine →](#)[See how it works ↓](#)

Tons of documents

Contracts, policies, scans — scattered everywhere.



We index it all

OCR + a searchable RAG index over every file.



Ask, get cited answers

Retrieve → rerank → answer → cite — or an honest “not found.”



Nothing invented

If it isn't in your documents, it says so.

THE PROBLEM WE SET OUT TO SOLVE

Answers buried in paperwork.

This is the scenario we hear on almost every call with a document-heavy finance or AP team.

01

Finance & AP teams — hours lost hunting for one clause

The everyday problem. Once a month, someone on the AP or finance team has to go through hundreds of vendor contracts and invoices by hand to answer a question like “do we have a signed NDA with this vendor?” or “what are the payment terms in this specific contract?” It takes hours, and if something gets missed, nobody finds out until it’s already a problem.

What we built. A system where you type the question in plain English — “which vendor contracts have net-30 payment terms?” — and it searches every document, gives you the answer with a link to the exact contract and clause, or tells you honestly it isn’t there. No guessing.

Fit for. Finance and AP teams at financial-services organizations (lending, insurance, factoring) with a lot of vendor and compliance paperwork — the kind of team currently doing this search by hand across shared drives and email, not on some abstract “digital transformation journey.”

WHAT WE ENGINEERED

A pipeline built to refuse rather than guess.

Four real engineering decisions, in order — this is the actual architecture running on the live demo below.

1

OCR everything

Scanned and image-only PDFs are recognized with OCR (OpenAI vision, or a fully local engine for on-prem) and turned into searchable text alongside native files.

2

Retrieve, then rerank

We embed the question alone (not diluted by a system prompt), pull a wide candidate set, then rerank by a hybrid of semantic score and literal keyword overlap — the clause that actually matches the question wins.

3

Guardrail before generation

If the best retrieved match scores below a set relevance threshold, we refuse before the model ever runs — no low-confidence guess reaches the user.

4

Grounded synthesis, cited

The model answers only from the retrieved excerpts, with a strict system prompt: extract facts, quote the source, and say “not found” rather than invent.

⚙ Built with

REAL STACK, NOT VAPORWARE

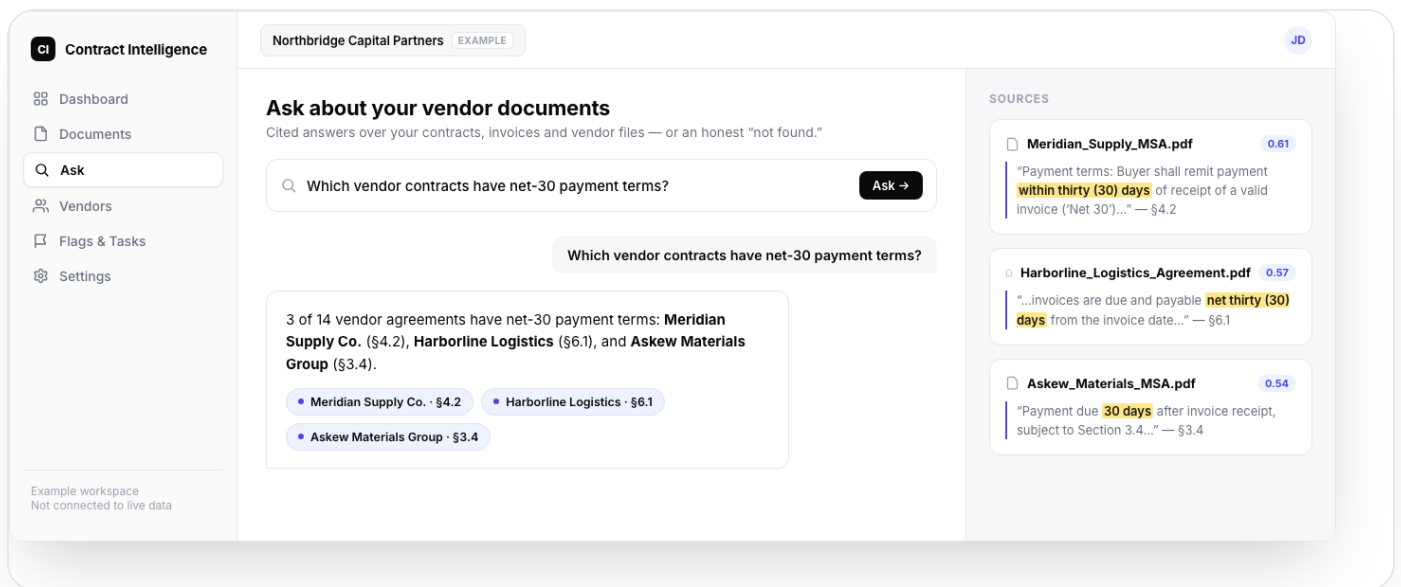
OpenAI + LlamaIndex.js + Tesseract.js

A separate, isolated vector index per tenant — not a metadata filter — so cross-tenant leakage is structurally impossible, not just policy. Deployed on Vercel, Node runtime.

INTERFACE CONCEPT

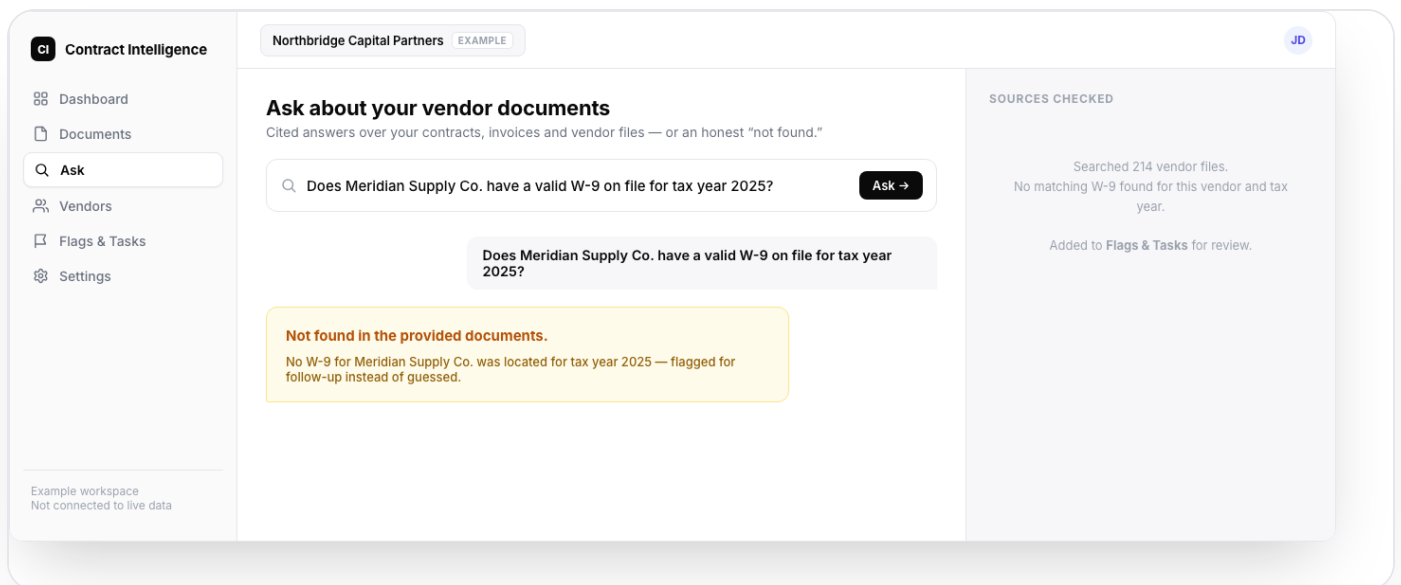
What the interface could look like.

A design concept for a workspace built around this engine — sidebar, vendor documents, and an Ask screen with sources cited alongside every answer. Not a live product screen.



A cited answer

The answer names the exact vendors and clauses, with the matching excerpt from each source document alongside it.



An honest refusal

When nothing in the documents answers the question, it says so — and hands the gap to a review queue instead of guessing.

WHAT IT PROVES

Verifiable properties, not claimed outcomes.

Here's what you can check for yourself on the live demo below, instead of taking our word for it.

Every answer is cited

Source document and a similarity score ship with every answer — verify it yourself, don't take our word for it.

It refuses on request

Ask something the documents don't contain — it says "not found," not a plausible-sounding guess.

Works on scans

Image-only PDFs go through OCR and come out searchable and citable, same as native files.

Per-tenant isolation

Built into the architecture, not bolted on — a query against one client's index cannot retrieve another's.

ESTIMATE THE IMPACT

What the time savings could look like.

This is a calculator, not a case study — plug in your own numbers. The default time-saved assumption is a starting point you should challenge, not a measured

result.

People who search documents regularly

4

Hours each spends searching, per week

5

Fully-loaded hourly cost (\$)

45

Assumed time saved

50%



\$46,800

Estimated annual cost of manual search time, at these numbers

\$23,400

Reclaimed per year at your assumed time-saved rate

Not a client result. Change any number above — this recalculates live.

Shown with example numbers — the calculator is interactive on the web version of this page.

SEE IT LIVE

Real retrieval, real citations.

Same retrieval, citation and refusal engine as the interface concept above — illustrated here with a vendor-contract question. Try your own on the live engine below.

SAME ENGINE

Which vendor contracts have net-30 payment terms?

3 of 14 vendor agreements have net-30 payment terms: **Meridian Supply Co.** (§4.2), **Harborline Logistics** (§6.1), and **Askew Materials Group** (§3.4). (Source: cited)

"Meridian_Supply_MSA.pdf" · §4.2 · score 0.61 →

Does Meridian Supply Co. have a valid W-9 on file for tax year 2025?

Not found in the provided documents.

Open the live demo

FAQ

Questions we'd ask too.

How accurate is it — can we trust it?

Every answer is grounded in the retrieved excerpts and cited to its source, so it's verifiable in seconds, not taken on faith. When the retrieved evidence is weak, it refuses instead of guessing — the confidence threshold for that is tunable per deployment.

How long to get this running on our documents?

The retrieval, citation and guardrail logic doesn't change — only the documents it's pointed at do. See the [legal](#) and [AP](#) pages for what that looks like for a specific team, typically a working pilot in 2–4 weeks.

Can this run on our documents instead?

Yes — that's the point of it. The retrieval, citation and guardrail logic doesn't change; only the documents it's pointed at do.

Is it safe for confidential documents?

Per-tenant isolation is structural (separate index per client, not a filter), and the same guardrail architecture supports on-prem/local OCR and a choice of LLM data-handling terms — see the security notes on the [legal](#) page for specifics.

GET STARTED

Want this pointed at your documents?

Tell us what you're searching for by hand today. We'll show the live engine and scope a pilot on your own files.

[Book a 20-min call →](#)

[See the live engine](#)

ONIX | Contract Intelligence

Grounded, citation-backed AI over your own documents. Built by
Onix-Systems.