Overview (very simple & dev-friendly)

We build one big local RAG library with:

Federal/U.S. universal content

Contract datasets (NDA/MSA)

Litigation & investigation datasets

Multi-doc case datasets

New York–specific law + cases

Then each brain pulls only what it needs.

Brains:

NDA BRAIN
MSA BRAIN
INVESTIGATION BRAIN
CASE INTELLIGENCE BRAIN
NEW YORK STATE MODULE

Everything below is free, legal, downloadable, and suitable for local RAG.

1) NDA BRAIN
Purpose:

Understands NDAs, detects deviations, generates safe redlines.

Dev Explanation:

NDAs are simple contracts. 90% of clauses follow templates.
This brain needs lots of real NDAs and clean clause libraries.

✅ Core Free Data Links (NDA Brain)
SEC EDGAR — NDAs from public companies

Direct link to search EDGAR:
https://www.sec.gov/edgar/search-and-access

Query example for dev:
https://www.sec.gov/edgar/search/#/q=nda&filter_forms=10-K

Contract Standards Clause Library (1.2M clauses)

https://github.com/ContractStandards/Contract-Clauses

OneNDA (Open-source standard NDA)

https://onenda.org/

CUAD Dataset (lawyer-labeled contracts including NDAs)

https://huggingface.co/datasets/atticus-project/cuad

Open-source NDA collections

https://github.com/jamesacampbell/contracts

https://github.com/alangrafu/legal-docs

⭐ NEW EXTRA LINKS (high value NDAs)
U.S. Public Agency NDAs (procurement)

https://www.dol.gov/agencies/oasam/site-closures/nda

(Often includes NDA templates used by U.S. agencies.)

Internet Archive – NDA Collection

Direct NDA search:
https://archive.org/search?query=non-disclosure+agreement

2) MSA BRAIN
Purpose:

Understands MSAs (Master Service Agreements) used for B2B services.

Dev Explanation (simple):

An MSA = main rules two companies follow when working together.
Hard parts include:

Who pays if something goes wrong (indemnity)

Limits on financial damages (liability caps)

Does the SOW match the main contract (SOW alignment)

Termination, warranties, service levels

Needs lots of SaaS/MSA/vendor contract examples.

✅ Core Free Data Links (MSA Brain)
SEC EDGAR — MSAs / SaaS / Services Agreements

https://www.sec.gov/edgar/search-and-access

Query example for SaaS/MSA:
https://www.sec.gov/edgar/search/#/q=%22master%20service%20agreement%22

NVCA Model Legal Documents (MSA families)

https://nvca.org/model-legal-documents/

Tech Contracts Handbook (free sample MSAs)

https://www.techcontracts.com/resources/

LEDGAR (100k+ contracts w/ clause labels)

https://huggingface.co/datasets/lex_glue

(Search inside for “ledgar”)

Practical Law — Free Sample Clauses

https://content.next.westlaw.com/practical-law

(Some samples are free without account)

Public Agency Contract Libraries (MSAs)

Miami-Dade County contract search:
https://www.miamidade.gov/Apps/ContractSearch/

U.S. open data portal (contracts):
https://www.data.gov/search?q=contract

⭐ NEW EXTRA LINKS (high value MSAs)
SAM.gov — Federal contract attachments (lots of MSAs + SOWs)

Direct data source:
https://sam.gov/content/opportunities

Many opportunities include PDFs of MSAs, SOWs, and IT service agreements.

Internet Archive — MSA & Service Agreement collection

https://archive.org/search?query=%22service+agreement%22

3) INVESTIGATION BRAIN
Purpose:

Understands evidence: emails, chats, witness statements, transcripts.
Finds contradictions, exposure, timelines.

Dev Explanation:

This is the “detective” brain.
Needs real emails, statements, investigation reports, government case files.

✅ Core Free Data Links (Investigation Brain)
Enron Email Dataset (classic investigation corpus)

https://www.cs.cmu.edu/~enron/

SEC Enforcement Actions (investigation files)

https://www.sec.gov/enforcement

DOJ Case Docs + Press Releases

https://www.justice.gov/news

(Contains statements of fact, indictments, evidence summaries)

CourtListener (federal & state filings)

https://www.courtlistener.com/

DOL Inspector General Reports (investigations)

https://www.oig.dol.gov/reports.htm

FTC Cases & Complaints

https://www.ftc.gov/legal-library/browse/cases-proceedings

Congressional Investigations (huge)

https://www.govinfo.gov/app/collection/chrg

Police Interviews / Interrogation transcripts (OpenJustice CA)

https://openjustice.doj.ca.gov/

⭐ NEW EXTRA LINKS (high value investigations)
FBI Vault — Full investigative files (PDF)

https://vault.fbi.gov/

NIST Incident Reports (technical investigations)

https://www.nist.gov/publications

4) CASE INTELLIGENCE BRAIN
Purpose:

Builds narratives and timelines across 50–500+ documents.
Understands dockets, motions, evidence sets.

Dev Explanation:

This is the “litigation overview brain.”
It needs multi-document cases, filings, opinions, court records.

✅ Core Free Data Links (Case Intelligence Brain)
CourtListener Bulk Data (massive federal + state)

https://www.courtlistener.com/api/bulk-data/

RECAP Archive (federal dockets + PDFs)

https://www.recapthelaw.org/

Supreme Court Opinions

https://www.supremecourt.gov/opinions/opinions.aspx

U.S. Courts — Federal Opinions

https://www.uscourts.gov/court-records

GovInfo — federal filings & case materials

https://www.govinfo.gov/app/collection/uscourts

FBI Vault (narrative investigation files)

https://vault.fbi.gov/

⭑ NEW EXTRA LINKS (multi-doc litigation)
Harvard Caselaw Access Project (6M+ decisions)

https://case.law/

OpenJustice NYC — litigation-related datasets

https://opendata.cityofnewyork.us/

5) NEW YORK STATE MODULE (MUST BE IN RAG FROM DAY 1)
Purpose:

Adds New York–specific laws & court opinions.

Dev Explanation:

New York is the #1 state for:

contract law

finance

litigation

commercial MSAs

white-collar investigations

Must be included from day one.

✓ Core Free Data Links (New York Module)
New York Consolidated Laws (Full text)

https://public.leginfo.state.ny.us/laws/

NY Court of Appeals Decisions

https://www.nycourts.gov/courts/appeals/decisions/

NY Appellate Division Decisions

https://nycourts.gov/courts/ad1/Decisions.shtml

NY Attorney General Enforcement Actions

https://ag.ny.gov/press-releases

NYC Open Data (investigations + agency cases)

https://opendata.cityofnewyork.us/

NY State & City Contracts

NY Senate transparency contracts:
https://www.nysenate.gov/transparency/contracts

NY UCC (Commercial Code rules)

https://www.nycourts.gov/courthelp/UGC/UCC.shtml

⭐ NEW EXTRA LINKS (high value NY data)
NYC Law Department Publications

https://www.nyc.gov/site/law/publications/publications.page

NY State Comptroller Contract Finder

https://wwe1.osc.state.ny.us/transparency/contracts/contractsearch.cfm

FINAL STRUCTURE (DON'T CHANGE)
Dev Download Plan (simple):

Download federal + universal contract datasets

Download New York laws + cases (mandatory for v1)

Build indices:

contract_index

evidence_index

case_index

ny_state_index

Run RAG with the 4 brains:

NDA BRAIN → universal contracts + NY corrections

MSA BRAIN → commercial contracts + NY corrections

INVESTIGATION BRAIN → emails, statements, investigations

CASE INTELLIGENCE BRAIN → multi-doc litigation

After pilot → Fine-tune 4 small adapters on Gradient.