

The Retrieval & Knowledge Landscape 2026

31

COMPANIES MAPPED

5

SUB-SECTORS

\$2bn

TOTAL CAPITAL RAISED*

~\$1bn

OF IT, VENTURE FUNDING

75%

US-HEADQUARTERED

The smallest and one of the earliest categories: **just 31 companies**, 75% still at Seed or Series A. RAG and retrieval infrastructure leads (10), but with Google, Microsoft and AWS all shipping native search and knowledge tooling, this is the layer most exposed to the hyperscalers absorbing it.

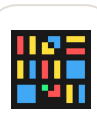
RAG & Retrieval Infrastructure 10



Caden AI



Formic



Instill AI
ALSO
MLOps & ML Lifecycle Platforms



LightOn



LlamaIndex
ALSO
Document Processing & Extraction (IDP)

ACQUIRED



Nuclia



Private GPT



Ragie



Unbody



Vectara
ALSO
Enterprise Knowledge & Context Layer

Web Search & Data APIs for Agents 7



Ceramic



Exa



LinkUp



Parallel



SID

ACQUIRED



Tavily



TinyFish

Enterprise Knowledge & Context Layer 7



Contextual AI
ALSO
RAG & Retrieval Infrastructure

ACQUIRED



Gathr
ALSO
Enterprise AI Platforms & Operating Systems



Glean
ALSO
Agent & Workflow Orchestration Frameworks

ACQUIRED



illumex
ALSO
Enterprise Data, BI & Analytics Platforms



Intelligible



Synapt



typedef

Vector & Search Infrastructure 4



Chroma

ACQUIRED



Jina



Neo4j

ACQUIRED



Voyage AI

Hyperscaler / SaaS Giant 3



Amazon Web Services



Google



Microsoft

*Funding from Crunchbase, 20 of 31 companies matched. "Total capital raised" includes all funding types (venture, debt, post-IPO, grants); "venture funding" excludes debt, private equity, post-IPO and grant rounds. Companies may appear in more than one sub-sector. Logos via logo.dev.