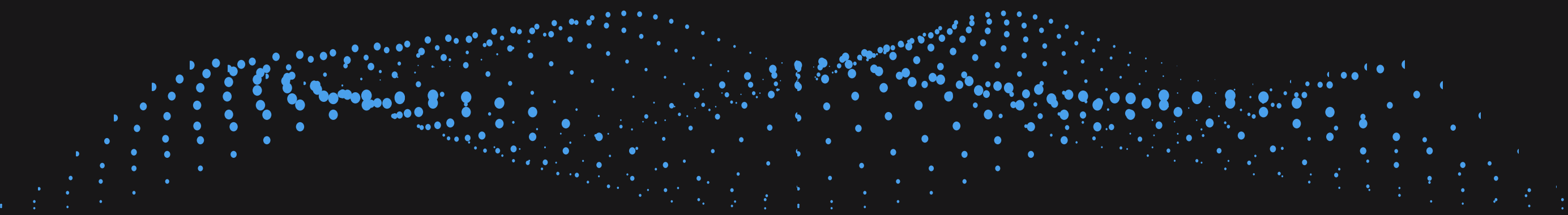



# VRiN — AI Deep Search for Your Data

HybridRAG retrieval with multi-hop  
reasoning and citations



 Vedant Patel

 [vedant@vrin.cloud](mailto:vedant@vrin.cloud)

 +1 (530)-204-8133

## PROBLEM

VRiN

# 20-30%

### Time Sink

Employees lose 20–30% of their week just finding information. McKinsey/IDC show **~1.8–2.5 hours/day** loss searching or re-creating info

[Read More ↗](#)

# \$3.1 T/yr

### Economy Wide Losses

IBM/IDC report that the cost of “bad or inaccessible” data is massive. Gartner pegs poor data quality at **\$12.9–\$15 M/year per enterprise**

[Read More ↗](#)

# 20+

### SaaS tools & private stores

Teams use dozens of SaaS apps; answers require **cross-document, permission-aware, explainable reasoning** — today’s generic chatbots and **vanilla RAG fall short.**

WHY NOW

VRiN

78%

### Mainstream Adoption

Companies report using AI in at least one function (up from 72% in 2024)

[Read More ↗](#)

92%

### Budget Rise

Executives expect to increase AI spend over the next 3 years; pressure to prove ROI is high

[Read More ↗](#)

60%

### SMB Momentum

Will increase spending in 2025 (faster than many other IT categories)

[Read More ↗](#)

100%

### Market Signal

Market leaders center messaging around permission-aware search, grounded answers —demand for cited, context-rich systems is durable

# SOLUTION

VRiN

From documents to decisions.

Reasoned answers, permission-aware, grounded in company data

## 1. Facts-First Approach

🌐 Extracted 15 facts and 7 vector chunks

- Knowledge Graphs + Vector Chunks to ground every response in company facts
- NO Hallucinations

## 2. Multi-hop Reasoning

🌐 Explored 3 Source Documents

- Cross-document synthesis
- Traces the chain of facts
- Produces citations and confidence

## 3. BYOC/BYOK

- Respect ACLs
- Deploy in customer cloud
- Complete audit trail

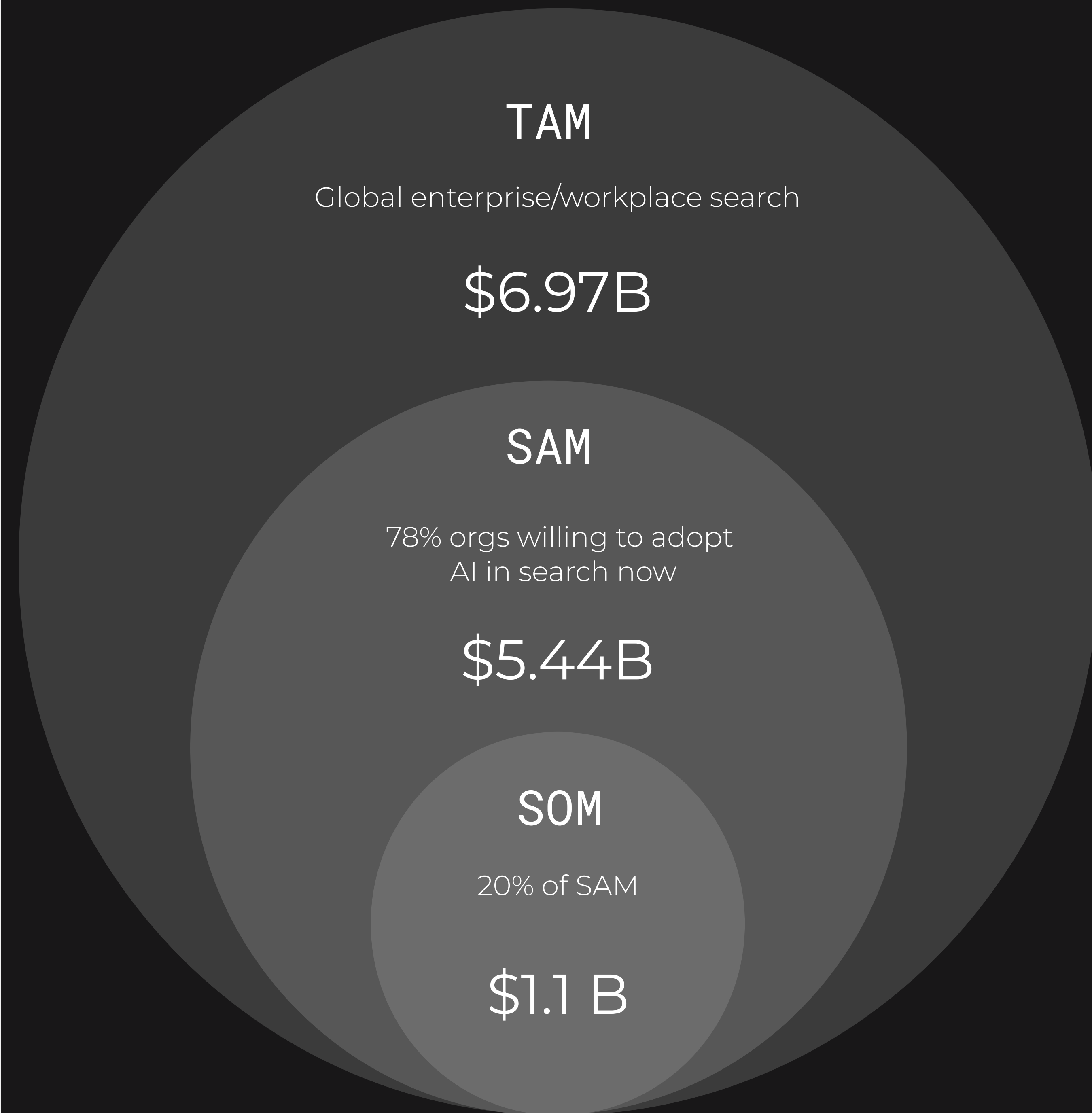
## 4. Expert Specialization

- Custom experts for Support, Engineering, Legal, Finance & other industries

## MARKET OPPORTUNITY

A wedge into a  
\$7B core —  
expanding into  
the broader  
GenAI for Work  
stack

[Read More ↗](#)





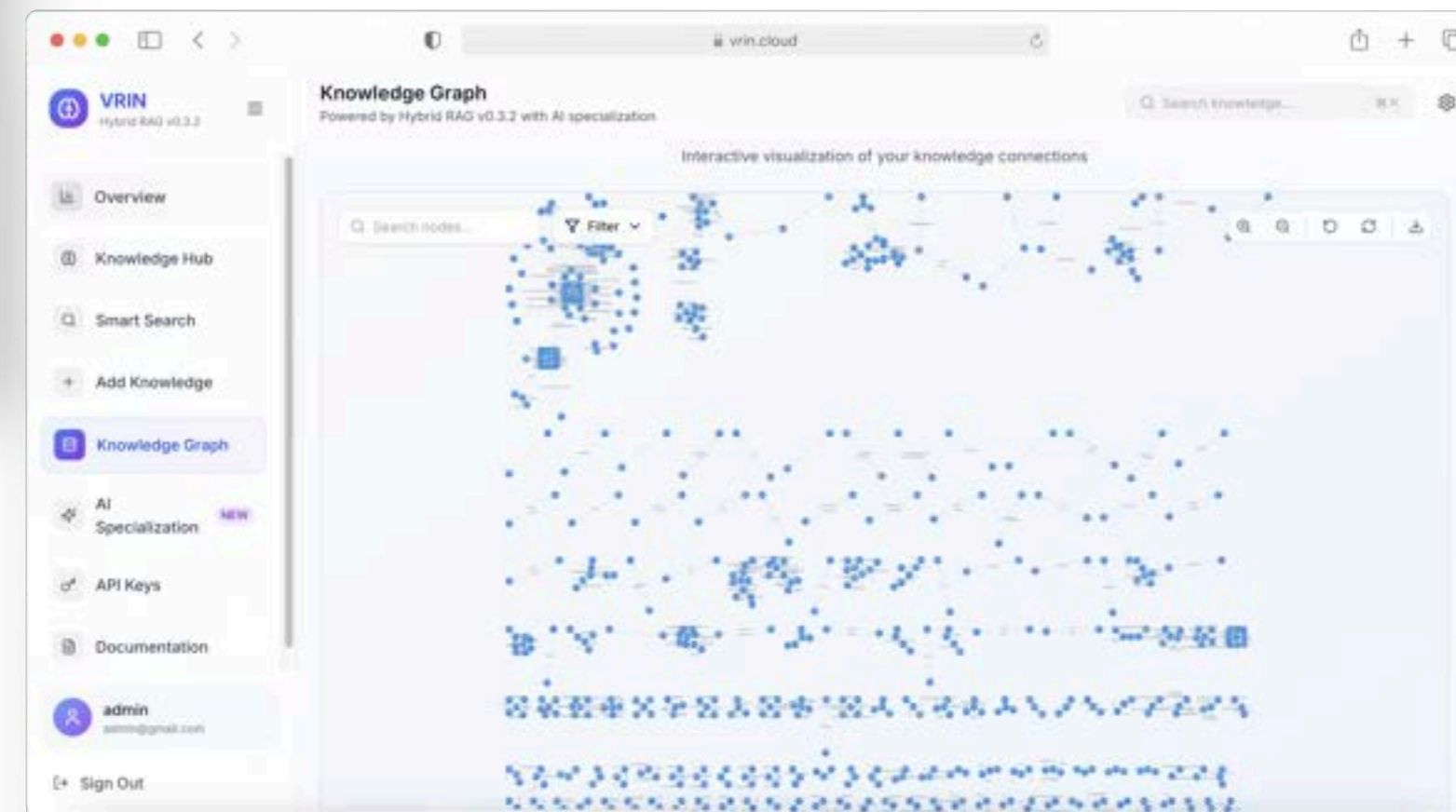
```
from vrin import VRINClient

client = VRINClient(api_key=os.getenv("API_KEY"))

insert_doc = client.insert(
    content=tech_acquisition_doc,
    title="TechGlobal Strategic M&A Assessment"
)
```

## Ingest

Connectors → de-dup →  
extract entities/relations  
and embeddings



## Structure

Store as knowledge  
graph + vector index.

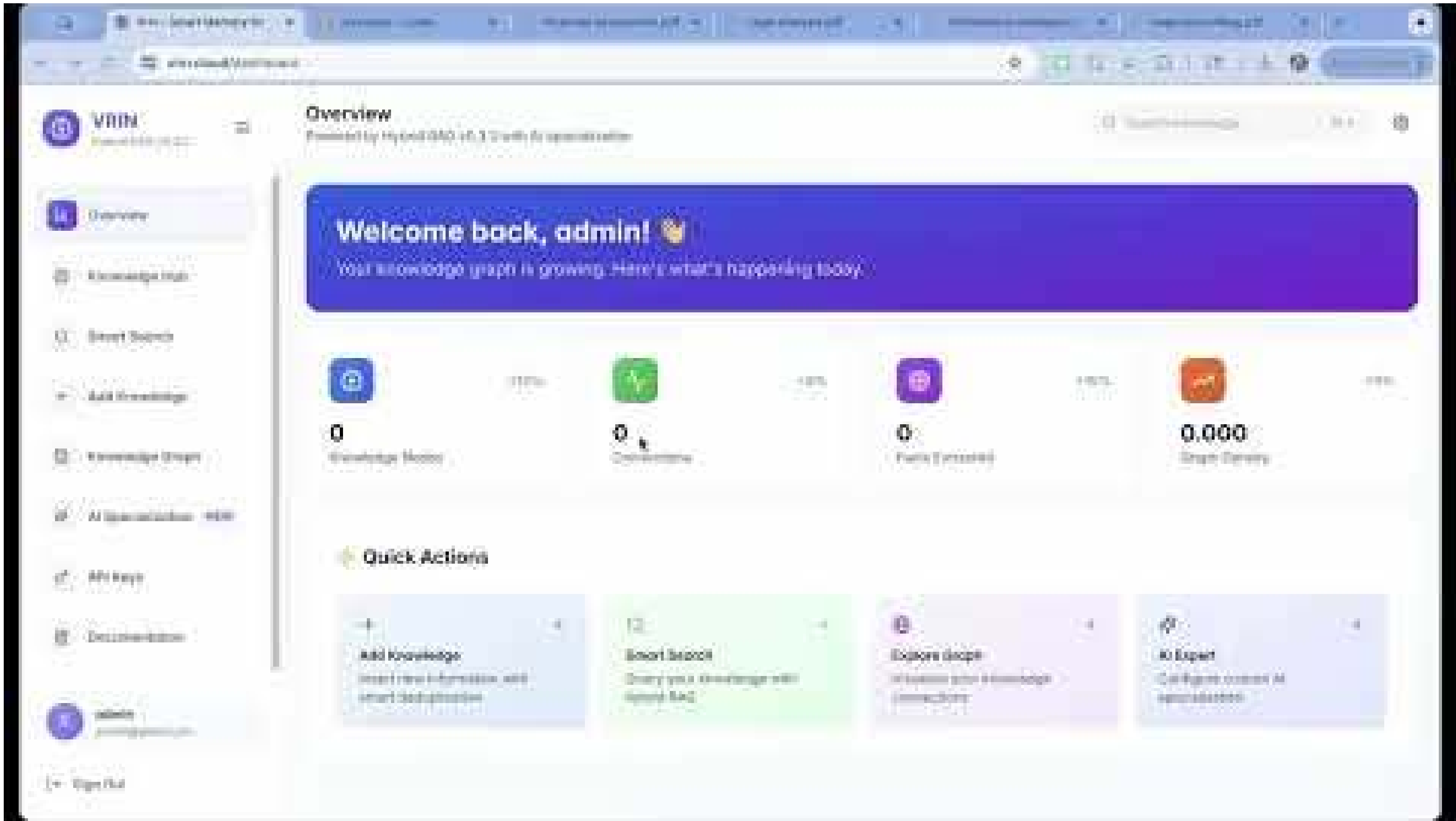
```
query = """
Analyze the TechGlobal-EuroManufacturing acquisition. Find
me conclusive evidence that justifies the $12B price tag,
and how does this
acquisition give TechGlobal competitive advantages over
QuantumCorp?
"""

response = client.query(query)
summary = response['summary']

# Other performance metrics
total_facts = response['total_facts']
total_chunks = response['total_chunks']
entities = response['entities']
multi_hop_chains = response['multi_hop_chains']
cross_doc_patterns = response['cross_document_patterns']
reasoning_confidence = response['reasoning_confidence']
```

## Retrieve & Reason

Entity + semantic retrieval → result  
fusion → multi-hop chains →  
answers with citation

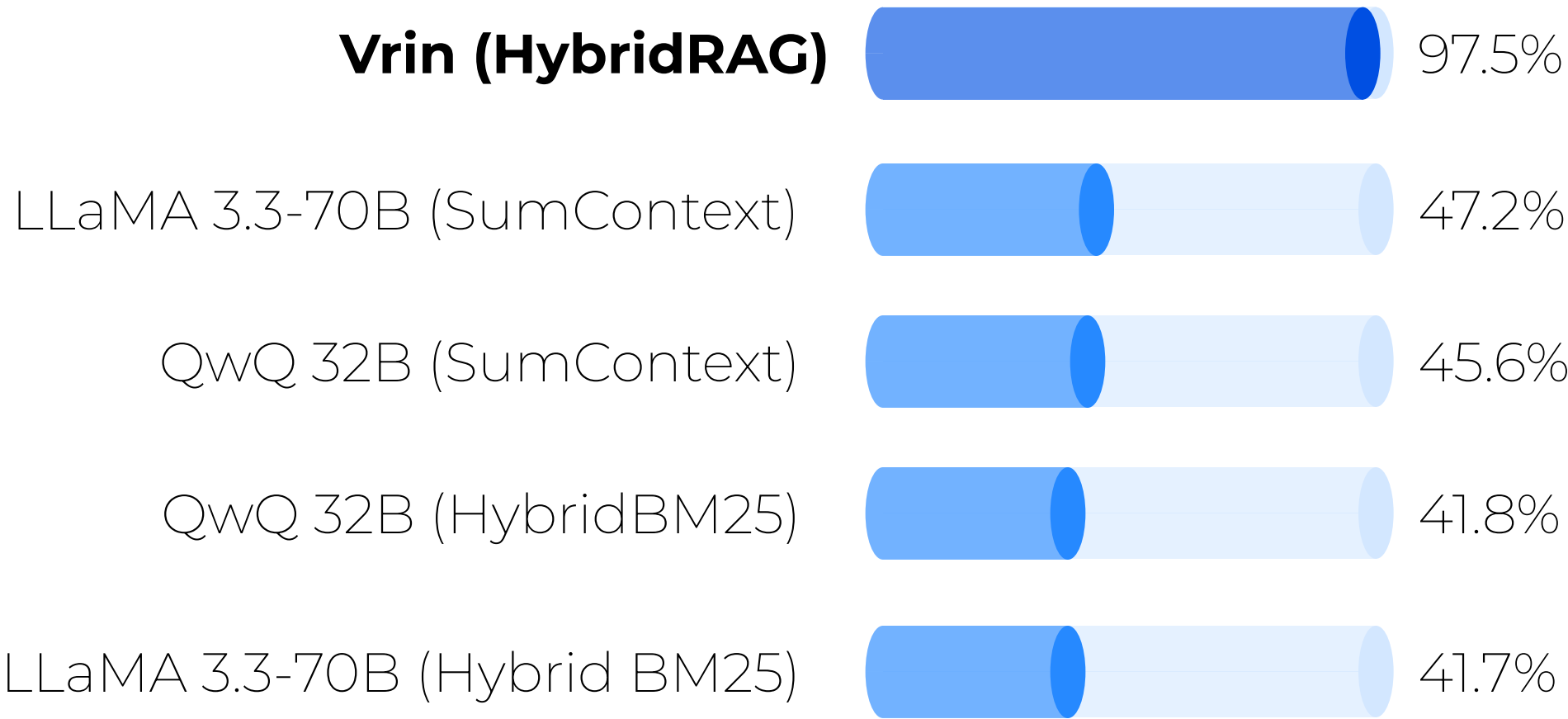


[Watch ↗](#)

BENCHMARKS

RAGBench FinQA

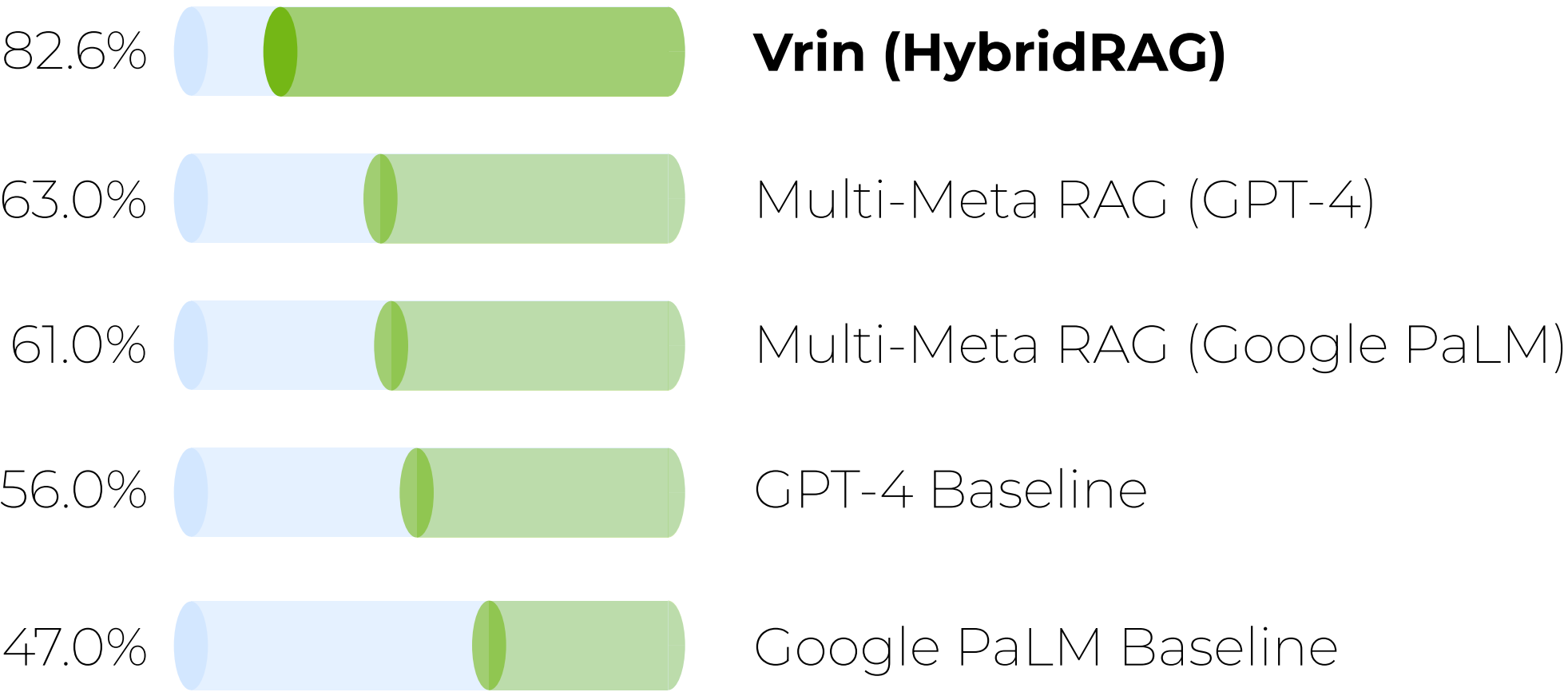
Metric: Number Match (NM)



[Learn More](#) ↗

MultiHop-RAG

Metric: Semantic Accuracy (SA)

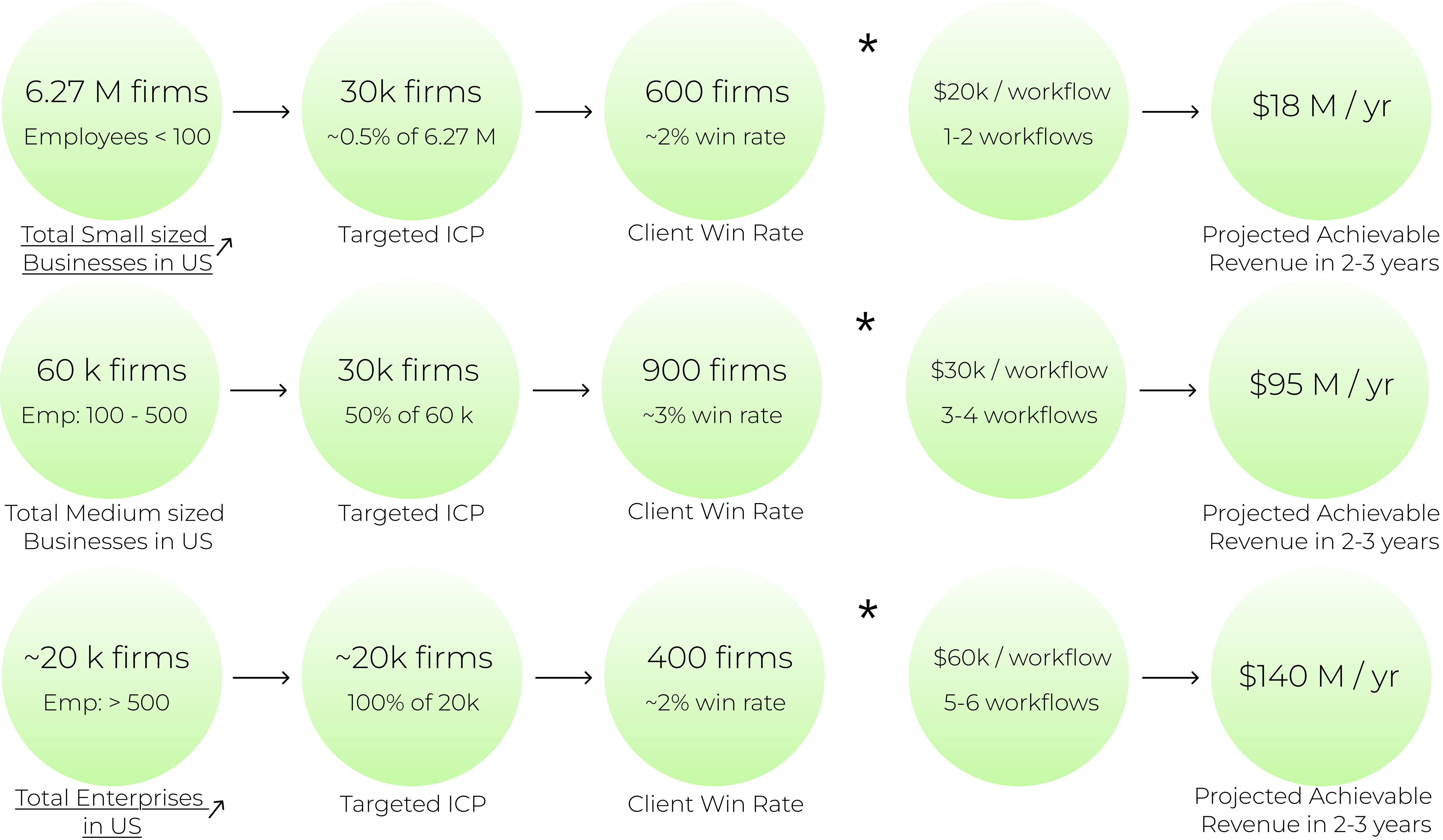


[Learn More](#) ↗



# REVENUE MODEL

VRiN



# BUSINESS MODEL

VRiN

Features	Builder (Free)	Team (\$2–4k/mo)	Business (\$8–15k/mo)	Enterprise (\$60–\$250k+/yr + usage)
Data scope	up to 200k chunks / 200k edges	2–5M chunks / 3–7.5M edges	10–25M chunks / 15–35M edges	Custom (100M+ chunks; 150M+ edges)
Queries/mo	10k	100–250k	500k–1M	Custom (SLA'd)
HybridRAG	Shared infra	Dedicated indices	Dedicated + VPC peering	Private/VPC or on-prem
Memory & CBOM	Basic	Full CBOM & TTL	Full + exports	Full + auditor packs
Security	API keys	Basic RBAC	SSO/SAML + SCIM	SSO/SAML, SCIM, data residency
Connectors	CSV/S3	+ Postgres/Drive	+ Slack/Jira/Confluence/SharePoint	All + custom
Support	Community	Email (48h SLA)	Priority (8–12h SLA)	TAM & DSE
Add-ons	—	Extra storage/queries	Compliance exports, private LLM	On-prem, managed upgrades

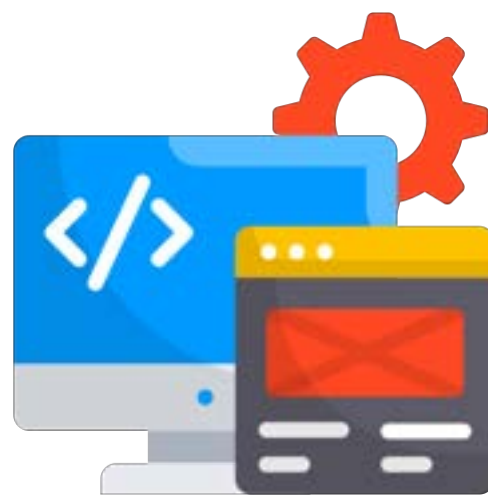
# GO-TO-MARKET

VRiN



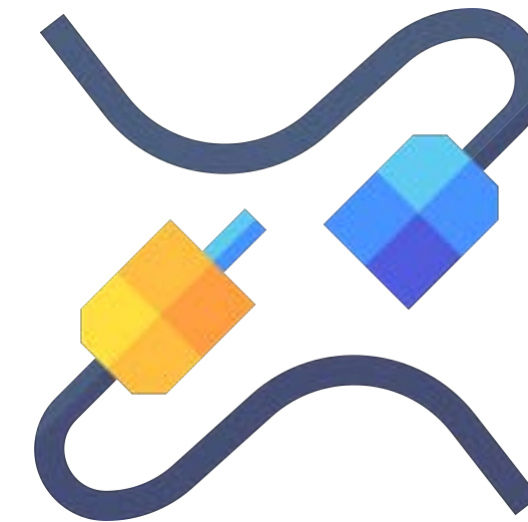
## Founder-Led Enterprise Sales

Target **CIO/CDO/Head of Platform** at data-rich, compliance-bound orgs (engineering, financial services, healthcare, industrial). Lead with pilot that measures time-to-answer and reduction in duplicate work.



## Self Serve Developer Motion

SDKs  
Guides  
Learning Modules  
Tutorials  
Community



## SaaS tools & private stores

iPaaS & data integration partners for connectors. Connectors are commodities, reasoning & analysis are the real value

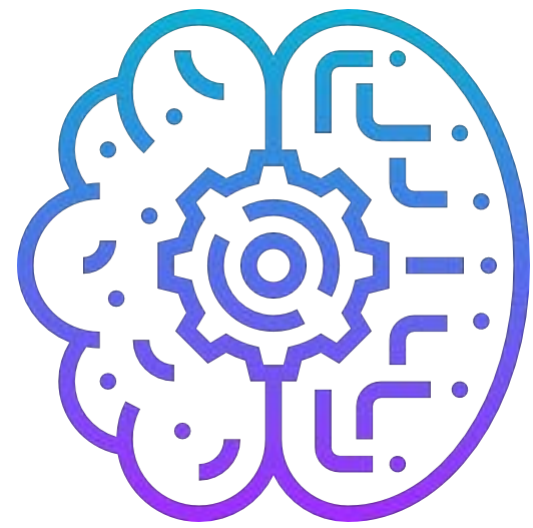
# COMPETITION





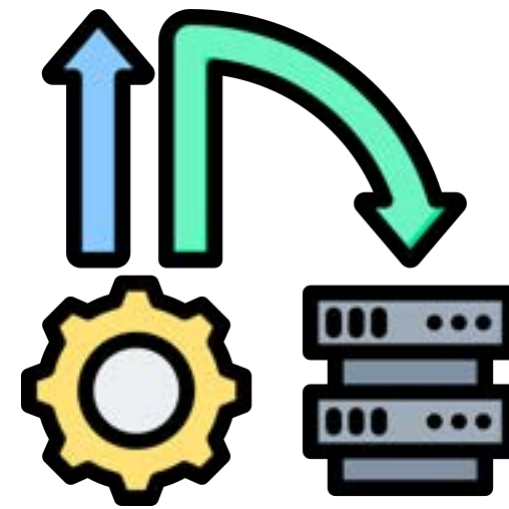
# COMPETITIVE ADVANTAGE

VRiN



## Deeply reasoned Explainable responses

Knowledge Graphs + Vectors to support multi-hop reasoning while retaining permission awareness to address industry's need for grounded outputs and lower hallucinations.



## Governance & deployment control as first-class

Similar governance posture as hyper-scalers (AWS Kendra, Google Vertex) but deeper cross-source reasoning chains and domain specializations instead of generic retrieval.



## Faster Path to ROI

We optimize for answer quality with citations and time-to-value.  
“Trust + Measurable Impact” wins budget in 2025.  
Ship deep search/answers inside existing tools via clean APIs & SDKs for rapid integration





## TEAM

I built a **production AI assistant system** for internal developers at **Microsoft** and saw vanilla RAG break under real workflows—fragile pipelines, context bloat, and no provenance. At UC Davis, my research pushed TinyML ECG models to extend battery life from ~2 weeks to 1+ month and used cloud offload for heavier multi-label analysis (**research published in IEEE ISCAS**). I also led a multi-agent Health LLM system that had to fetch the right medical signals and choose the right model for each question—exposing **RAG’s gaps in multi-hop reasoning, cross-document synthesis, and time awareness**. Vrin is the system I wished I’d had: a neutral control plane that makes AI answers accurate, auditable, and enterprise-ready.

[Learn More ↗](#)

# TRACTION



## The Challenge

A leading research lab at UC Davis is standing up a multi-agent Health LLM to answer patient questions using continuously recorded signals plus clinical context. The existing approach struggled with reasoning — vector-only RAG missed cross-signal links and couldn't explain answers; and integration velocity—DIY pipelines across silos were slow to deploy and hard to maintain.

## The Solution

We integrated Vrin as the memory & context layer inside the multi-agent orchestration, providing HybridRAG with permission-aware retrieval, per-claim citations, and BYOK/VPC deployment patterns aligned to the health system's AI/privacy guidance.



## The Challenge

A multi-brand hospitality & retail operator needed a smarter website assistant. Knowledge was spread across a proprietary customer-insights database plus wikis, PDFs, and ticket notes — each with strict permissions. The team wanted trustable responses that referenced their own data, without standing up a costly DIY RAG stack or rewriting existing systems.

## The Solution

We deployed Vrin via SDK into their existing web assistant. Customer-insights tables and content were ingested with entity/relationship extraction, then fused during retrieval to support multi-hop, cited answers. Row-level permissions and VPC/BYOK patterns enforced data access. The assistant calls Vrin's API for grounded responses.



**Pipeline: Engaged with & Actively scoping multiple enterprise and mid-market buyers; pilot statements of work in progress to embed Vrin in priority workflows.**

## THE ASK

We're raising 12 months financing for GTM, Security/Compliance, iPaaS/connector partners, Reasoning & Governance R&D to land pilots → and convert to scalable motion.



**\$500K**



# VRiN

Thank you

 Vedant Patel  vedant@vrin.cloud  +1 (530)-204-8133

<https://www.vrin.cloud>

