SAS® Retrieval Agent Manager - Generative AI for Enterprise Knowledge Base

Powered by SAS Event Stream Processing

Table of Content

Retrieval Augmented Generation

RAG connects Unstructured data to Outcomes

Retrieval Augmented Generation

- Faster Access to Institutional Knowledge
- Scalable Customer & Employee Support
- Reduced Manual Document Review Effort
- Operational Efficiency
- Stronger Compliance & Risk Mitigation

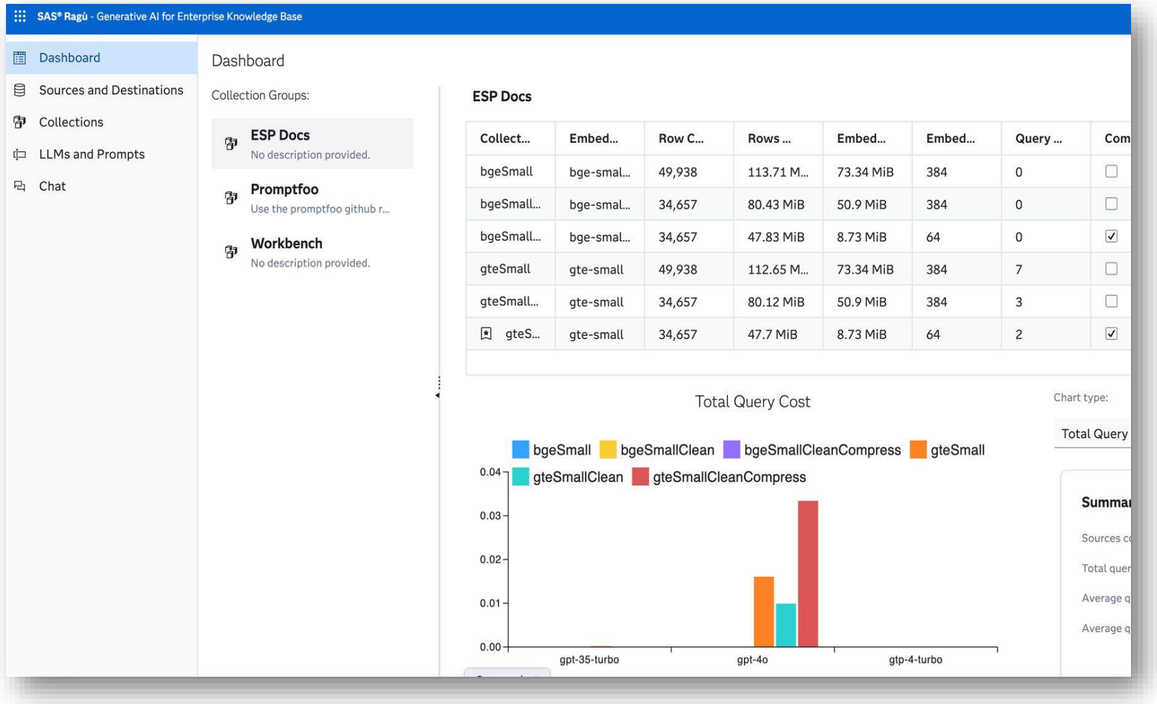
Benefits of RAG

- Uses information from unstructured sources to generate contextual responses
- Overcomes LLM's limitations of static training data
- Separation of key components supports agility in using latest innovation

What is SAS® Retrieval Agent Manager (RAM)?

Introducing Retrieval Agent Manager

Simplifies and accelerates the development of Generative AI solutions by streamlining retrieval-augmented generation (RAG) workflows



Capabilities

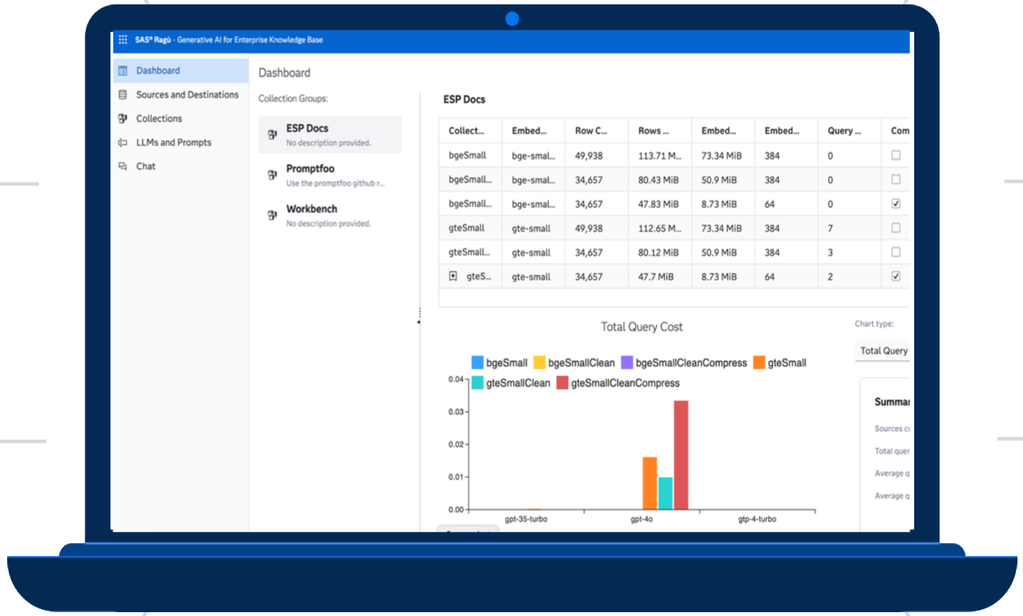
Enables secure, intelligent knowledge retrieval from proprietary documents without compromising sensitive data

Allows flexible experimentation with different embedding models and LLM's using automated and user-defined evaluation frameworks

Supports development of advanced AI agents (Agentic AI) for precise, context-aware interactions across cloud and on-prem environments

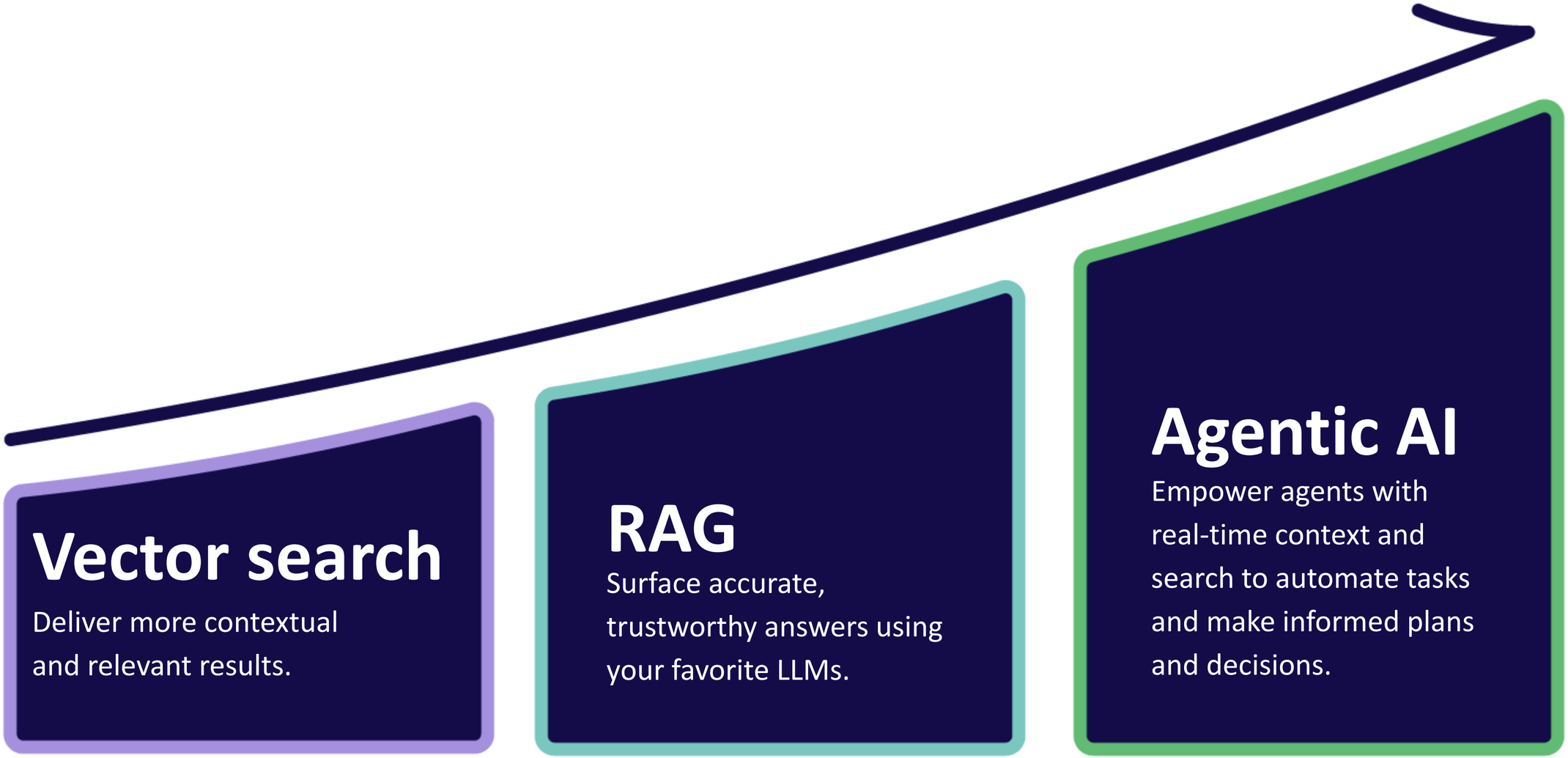
Enterprise grade features

- No Code UX
- Plug & Play GenAI Services
- Cloud & On-Prem Support
- Compact, Optimized Architecture
- Granular Security
- Agentic Support
- API Layer
- Built-in Evaluation & Citations



Abstracting the complexity associated with RAG, without losing the flexibility

Use Case Maturity Curve



Agentic AI Use Cases

Across Industry

Manufacturing

RAM enhances production control, monitors supply chains, enables predictive maintenance, and improves quality using insights from equipment and process data.

Healthcare

Retrieves and reasons over patient records and literature to support treatment planning, drug discovery, medical reviews, and patient monitoring.

Finance / Banking

Applies transaction-aware retrieval for fraud detection, risk analysis, and real-time support through compliant, document-grounded chatbots.

Government

Powers virtual agents, automates inventory and procurement tasks, and informs pricing decisions using regulatory and operational data.

sas

Use case Scenario: Predictive Maintenance

Industrial Setting

Accelerating the time from Detection to Correction

Problem:

Alert Response Management in Operational Setting …

- Too Many Alerts
- Often Requires Tradeoffs
- Specialized Knowledge

Solution:

Generative AI enables “Contextualized Insights” …

- Converting alerts to insights
- Aligned to corrective action plans
- Upskills Workforce

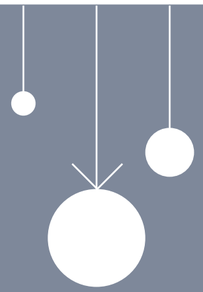
Why This Is Challenging

A typical gas turbine has over 20,000 components—each with its own maintenance procedures and dependencies.

The OEM manual for a GE Frame 9F turbine runs over 6,000 pages—not including separate documents for the generator, auxiliary systems, or site-specific integrations.

Field service bulletins can number in the hundreds per year for major OEMs like Siemens or GE, requiring constant monitoring to stay current.

Average time to identify the root cause of a turbine trip (using traditional methods): 6–10 hours.



Number of unstructured documents consulted per event:

- **15–30+ PDFs** across systems
- **Inspection reports from 5+ years of logs**
- **Vendor manuals** with non-standard terminology
- Emails and handwritten field notes from previous crews

Proposed solution

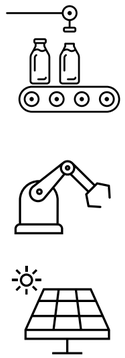
Integrate GenAI:

RAG in existing ML-based
Predictive Maintenance workflow

Operationalize ML – based pipelines
for Predictive Maintenance

Industry Copilot - Maintenance

FACTORY



MACHINE DATA

SENSOR DATA

VIBRATION | PRESSURE | ENERGY

DOCUMENTATION

KNOWLEDGE BASE

EQUIPMENT MANUALS | MAINTENANCE RECORDS | WORK INSTRUCTIONS

EVENT STREAM PROCESSING

STREAMING AI

DATA ACQUISITION | ENRICHMENT /QUALITY | ANOMALY DETECTION

RAW ALERT + DATA

RETRIEVAL AGENT MANAGER

INDUSTRY AGENT

MACHINE STATE | ADVANCED ANALYTICS | AUTOPROMPT GENERATION

LLM/RAG

NO CODE INTERFACE | MODEL AGNOSTIC | TRUSTWORTHY RESPONSES

API

DECISION SUPPORT

CONTEXTUAL ALERT

ROOT CAUSE

CORRECTIVE ACTION

WORKFORCE EFFICIENCY

Predictive Maintenance: Industrial Fan Demo

Industrial Fan Overall Dashboar

Device VFAN00

Last 2 minutes

Refresh 5s

Industrial Fan Monitoring

Individual Fan Air Pressure

Detected Anomalies

Snapshot

Current Value Readings (VFAN00)

Speed (RPS)

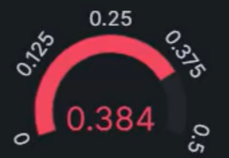
Current (ARMS)

Torque (Nm)

Active Power

Reactive Power

Factor Power



Demo – Annual and Quarterly Reports of ALL S&P500 Companies

Source files, approx. 150.000 pages

```
=== 10-K DOCUMENT ===
aapl-20240928false2024FY0000320193P1YP1YP1YP1Yhttp://fasb.org/us-gaap/2024#MarketableSecuritiesCurrent
http://fasb.org/us-gaap/2024#MarketableSecuritiesNoncurrenthttp://fasb.org/us-
gaap/2024#MarketableSecuritiesCurrent http://fasb.org/us-
gaap/2024#MarketableSecuritiesNoncurrenthttp://fasb.org/us-gaap/2024#LongTermDebtCurrent http://fasb.org/us-
gaap/2024#LongTermDebtNoncurrenthttp://fasb.org/us-gaap/2024#LongTermDebtCurrent http://fasb.org/us-
gaap/2024#LongTermDebtNoncurrenthttp://fasb.org/us-gaap/2024#OtherAssetsNoncurrenthttp://fasb.org/us-
gaap/2024#OtherAssetsNoncurrenthttp://fasb.org/us-gaap/2024#PropertyPlantAndEquipmentNethttp://fasb.org/us-
gaap/2024#PropertyPlantAndEquipmentNethttp://fasb.org/us-gaap/2024#OtherLiabilitiesCurrenthttp://fasb.org/us-
gaap/2024#OtherLiabilitiesCurrenthttp://fasb.org/us-gaap/2024#OtherLiabilitiesNoncurrenthttp://fasb.org/us-
gaap/2024#OtherLiabilitiesNoncurrenthttp://fasb.org/us-gaap/2024#OtherLiabilitiesCurrenthttp://fasb.org/us-
gaap/2024#OtherLiabilitiesCurrenthttp://fasb.org/us-gaap/2024#OtherLiabilitiesNoncurrenthttp://fasb.org/us-
gaap/2024#OtherLiabilitiesNoncurrentP856DP473Diso4217:USDxbrli:sharesiso4217:USDxbrli:sharesxbrli:pureaapl:Vendora
apl:Subsidiaryiso4217:EUR000032019320231-10-012024-09-280000320193us-gaap:CommonStockMember2023-10-012024-09-
280000320193aapl:A0.000Notesdue2025Member2023-10-012024-09-280000320193aapl:A0.875NotesDue2025Member2023-10-
012024-09-280000320193aapl:A1.625NotesDue2026Member2023-10-012024-09-
280000320193aapl:A2.000NotesDue2027Member2023-10-012024-09-280000320193aapl:A1.375NotesDue2029Member2023-10-
012024-09-280000320193aapl:A3.050NotesDue2029Member2023-10-012024-09-
280000320193aapl:A0.500Notesdue2031Member2023-10-012024-09-280000320193aapl:A3.600NotesDue2042Member2023-10-
012024-09-2800003201932024-03-2900003201932024-10-180000320193us-gaap:ProductMember2023-10-012024-09-
280000320193us-gaap:ProductMember2022-09-252023-09-300000320193us-gaap:ProductMember2021-09-262022-09-
240000320193us-gaap:ServiceMember2023-10-012024-09-280000320193us-gaap:ServiceMember2022-09-252023-09-
300000320193us-gaap:ServiceMember2021-09-262022-09-2400003201932022-09-252023-09-3000003201932021-09-262022-09-
2400003201932024-09-2800003201932023-09-3000003201932022-09-2400003201932021-09-250000320193us-
gaap:CommonStockIncludingAdditionalPaidInCapitalMember2023-09-300000320193us-
gaap:CommonStockIncludingAdditionalPaidInCapitalMember2022-09-240000320193us-
```

Prompt ideas from the audience?

Beyond KPI lists

- You are the world's best value / quality / growth investor. You have access to all 10-Ks and the latest 10-Qs for all SP500 companies. Create a checklist what you want to research to find stocks that are normally overlooked. Go broad first, create a shortlist, then go deep. Give me a shortlist of stocks.
- Take a look at the 10-K and 10-Q forms - which filings are signed by founders or CEOs with >10% ownership?
- Find companies that segment reporting shows undervalued divisions with higher margins or growth rates than the consolidated business.
  - Scan segment footnotes and MD&A for: Segments with operating or gross margins 500–1,000+ bps above consolidated.
  - Find Segments growing double-digits while consolidated growth is mid-single digit or less.
- Which companies will profit exponentially from AI / artificial intelligence but are currently not seen as AI stocks? Go deep in the 10-Ks and 10-Qs.
  - Which of these companies already profit exponentially from AI but their multiples dont reflect it yet? screen them also for solid financials and value investment risks. give me a shortlist of 3 solid companies with high growth potential and solid fundamentals.
- Find language in the forms that could be interpreted as willingness to buy back shares and give me the respective companies and excerpts of the 10-Ks and 10-Qs.
- Identify stocks with risk factors where management admits the issue is unlikely or immaterial, but the market might overreact (e.g., 'we depend on a single supplier, but we have alternatives'), give me a list of these stocks. These probably represent currently undervalued stocks.

Thanks

Saurabh.Mishra@sas.com
sas.com