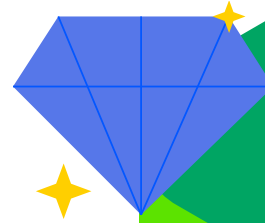


Unstructured Data & AI

- **80% - 90%** of world's data is unstructured.
- Industry-specific solutions in **Healthcare, Ed-tech, Finance, and Retail** and more are witnessing significant development, addressing regulatory compliance and personalized customer engagement needs.
- Emerging technologies such as **IoT** and edge computing are integrating with **unstructured data** solutions, fostering smarter, more connected enterprise ecosystems.
- Integration of **advanced AI and machine learning** algorithms continues to revolutionize unstructured data processing, enabling predictive analytics and automated insights.



Generative AI / LLM Concerns



Hallucinations

Generative AI confidently state / provide incorrect data 15% to 20% of the time.¹



Data Biases

Possibility of producing results that reflect biases encoded in the training data.



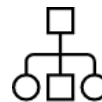
Understanding

Lack of human reasoning and understanding, leading to stereotypical and incorrect answers.



Data Cutoffs

Data cutoffs and missing information on recent events and developments.



Explainability

Lack of reasoning and reference to the answers they provide.



Robustness

Robustness against bad inputs and struggle with specific tasks.

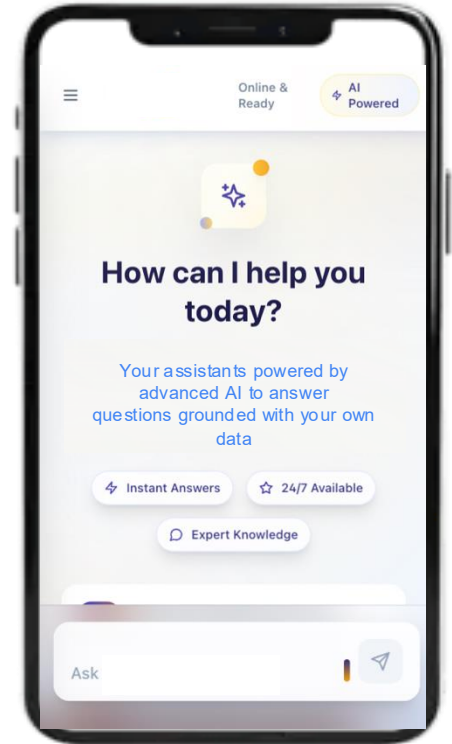
What is Retrieval Augmented Generation (RAG)?



Definition: RAG combines the power of large language models (LLMs) with your **own proprietary data** to generate answers. Instead of relying solely on the LLM's training (which can produce hallucinations), RAG **retrieves factual content from your knowledge base** and uses it to craft accurate, relevant responses

Why **RAG** is Game-Changing

RAG ensures **answers are trustworthy and verifiable**, with citations or links to source content. It's like a supercharged intelligent search – users get direct answers, not just links, and those answers are grounded rather than generic internet information.

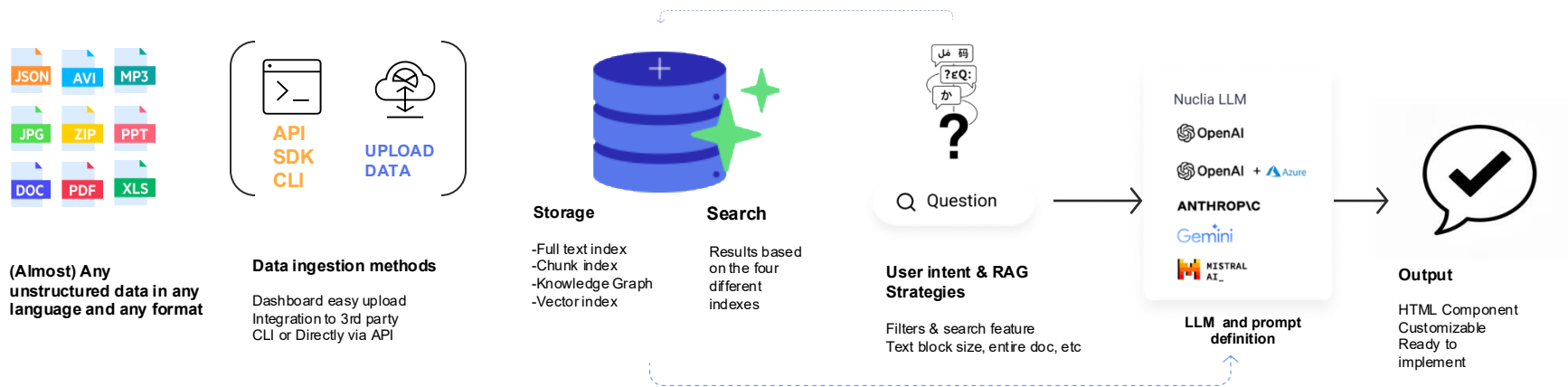


...but building a RAG platform is not an easy task

- **Expertise Requirement:** Building RAG programmatically requires strong expertise in NLP, machine learning, and software development.
- **Development Time:** Building RAG involves an iterative process of data collection, pre-processing, model training, evaluation, and refinement. Each iteration require significant time to achieve satisfactory results.
- **Complexity of Tasks:** Training RAG models for tasks such as passage retrieval and response generation can be complex, requiring meticulous tuning of hyperparameters and model architectures.
- **Resource Intensiveness:** Training and deploying RAG models require significant computational resources, including high-performance hardware and storage infrastructure.
- **Maintenance and Updates:** Continuous maintenance and updates are necessary to keep the RAG model up-to-date with evolving language patterns and domain-specific knowledge.

Progress Agentic RAG

An end-to-end RAG platform that transforms enterprise knowledge into accurate, actionable answers.



AI Extraction Smart chunking & Embeddings & Ingest Agents

Choose the best embeddings for specific use cases: Nuclia, OpenAI, Google, Hugging Face

NucliaDB

Multi-index database 100% focused on unstructured data storage with integration of various vector indexes.

Retrieval Strategy & Retrieval Agents

How to retrieve the context based on the desired output thanks to different retrieval strategies

Agentic & LLM

Select the best LLM for specific use cases Fine-tune the prompt the behaviour, and the agentic

Quality & Metrics

Get to know the quality of the output based on REMI

Progress Agentic RAG's advantages over “just an AI model”



How **Progress Agentic RAG** fits within enterprise technology stacks, providing an intelligent layer driven by **your knowledge**.

End-User Experience



Websites



Employee Hubs



Personalized Apps



Learning Portals

Application & Integration



Conversational AI



Intelligent Search



Custom Applications



CMS/DXP/LMS Plugins

Progress Agentic RAG



AI Agents



RAG Engine



Multimodal Processing



Knowledge Base



Security

Data Sources



CMS/DXP



LMS



CRM



Video/Audio



Documents



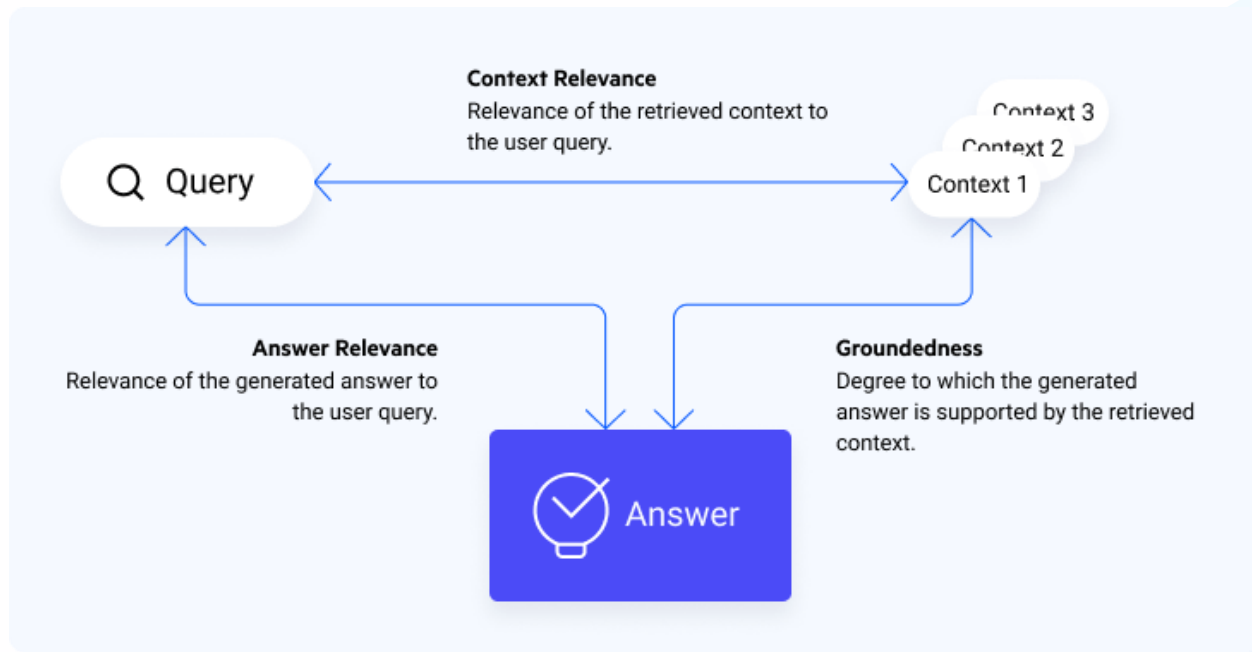
Databases

Embedding Models

Model Name	Size	Threshold	Max Tokens	Matryoshka	Multilingual	External
en-2024-04-24	768	0.47	2048	No	No	No
multilingual-2023-08-16	1024	0.7	512	No	Yes	No
multilingual-2024-05-06	1024	0.4	2048	No	Yes	No
Open AI small	1536	0.5	8192	Yes	No	No
Open AI large	3072	0.5	8192	Yes	No	Yes
Google multilingual Gecko	768	0.55	3072	Yes	Yes	Yes
Hugging Face	N/A	N/A	N/A	N/A	N/A	Yes

What can REMi do?

It computes the metrics defined by the RAG triad



ARAG's secret sauce: Built-in RAG quality metrics

We built REMi, the best open source model delivering measurable quality metrics to ensure **accurate, reliable, and trustworthy** results.

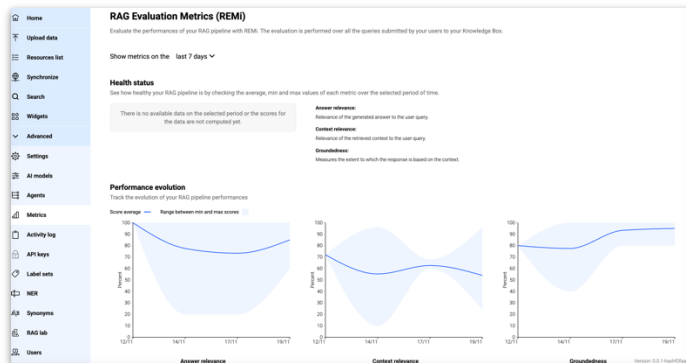
REMi: Redefining RAG Evaluation

What is REMi?

REMi is a specialized evaluation model developed by Progress Agentic RAG to assess the quality and performance of Retrieval Augmented Generation (RAG) pipelines.

Unlike traditional LLM evaluations, REMi is fine-tuned specifically for RAG, providing targeted metrics such as **Answer Relevance**, **Context Relevance**, and **Groundedness**.

[More about Quality Metrics](#)
[More about REMi](#)



Quality Metrics for RAG

REMi provides detailed metrics for every query submitted to your Knowledge Box, empowering you with insights into pipeline performance.

Context Relevance: Assesses the relevance of retrieved content to the original query.

Answer Relevance: Measures the relevance of AI-generated responses to user queries.

Groundedness: Indicates how well the AI's responses are based on provided context.

Latest release: Out of the Box AI Agents for RAG

Task-Specific Agents

Ingestion agents

Ingestion agents offer additional analysis during the data ingestion process, enabling labeling, knowledge graph extraction, synthetic data generation, and FAQ creation

These agents, available during ingestion or as batch processes, facilitate **cleaning** and **transforming** user-extracted data into actionable information for use with large language models (LLMs) or as metadata accessible directly from the client side.



Retrieval Agents

Progress Agentic RAG offers over 30 retrieval strategies tailored to various use cases, each delivering unique experiences when generating answers. These agents operate during the retrieval process to refine and optimize the information needed to address complex questions.

Multi-Step Agents

Combine multiple tasks, like classifying documents and summarizing them, to handle complex workflows efficiently.

[More about Agents](#)

Competitive Landscape

	Progress Agentic RAG	Competitors			
		Indexing pipelines	Vector DBs	RAG providers	LLM providers
Modular RAG Multiple strategies, multi-modal and retrieval agents	●	●	●	●	●
Quality metrics tracking	●	●	●	●	●
Data Augmentation / Ingestion & Retrieval Agents	●	●	●	●	●
RAG Lab / Fine tuning cycle Fast experimentation for best RAG output	●	●	●	●	●
API - Dashboard - Widget	●	●	●	●	●
Any type file/link/text processing	●	●	●	●	●
On-prem support & no third party dependencies	●	●	●	●	●
Multiple Vector chunks	●	●	●	●	●
Multiple indexes full text, text blocks, graph and vector index	●	●	●	●	●

Value drivers for our customers

- Accelerate and scale deployments of new AI-based customized and personalized end solutions tailored to the specific business domain, products, services, and customer needs
= **ACCELERATED INNOVATION**
- Improve and automate workflows, enable real time data driven decisions
= **OPERATIONAL EFFICIENCY & SCALE**
- Delight their users by providing a state of the art conversational experience
= **CUSTOMER RETENTION & UPSELL**
- Reduced hallucination, improved accuracy, relevance of responses, auditability and transparency, traceability
= **TRUST & COMPLIANCE**

The Value Agentic RAG Delivers

AI-Powered Knowledge, Delivered Without Disruption



AI Search + Conversations

Conversational access to organizational knowledge grounded in real documents



Enrichment at Scale

Auto-organizes and enhances PDFs, videos, transcripts, and more



Zero Disruption

Integrates with existing systems — no migration required



Transparent AI

All answers are sourced, cited, and defensible



No-Code Automation

Create AI agents in plain English — no dev time needed



Time to Value in 30 Days

Live in 4 weeks with measurable outcomes

Use Case	Data Needed	Pilot Time	Success KPIs	ROI Levers	Natural Expansion
AI Knowledge Hub for SOPs & Manuals	PDFs of SOPs, tech manuals, shift guides (ENAR/TR)	2 weeks	Search time ↓ 70–90%; first-contact resolution ↑	Fewer escalations; faster incident handling	Add maintenance logs; mobile access airside
Passenger FAQ Copilot (Multilingual)	Existing FAQs, service pages, terminal maps	2–3 weeks	Web/call deflection 20–40%; CSAT ↑	Lower AHT; 24/7 service without headcount	Kiosk/app embed; voice; airline partners
Contract & RFP Summarizer	Concession/IT contracts, tenders	1–2 weeks	Review time ↓ 60–80%; risk clause detection rate	Legal/PM time saved; fewer missed obligations	Clause library; renewal alerts
Incident Investigation Assistant	Safety reports, shift logs, email PDFs	3 weeks	Time-to-root-cause ↓; repeat incident rate ↓	Faster CAPA; less downtime	Add CCTV transcripts, SCADA exports
Policy Diff & Compliance Check	Current/prev policies, ICAO/CAA regs	2 weeks	Diff cycle time ↓ 80%; audit findings ↓	Faster compliance updates; audit prep time ↓	Continuous monitoring; attestations
Training Video/Audio Search	LMS exports, training videos (auto-transcribed)	2–3 weeks	Time-to-answer in training ↓ 70%	Reduced retraining; faster onboarding	Auto-quiz generation; skills gap insights
Retail & Concession Intelligence (Docs)	Sales reports, proposals, emails as PDFs	3 weeks	Contract insight time ↓; promo speed ↑	Better rev-share terms; targeted promo	Blend POS/footfall later for uplift analysis
ORAT/Readiness Summarizer	Checklists, punch lists, commissioning docs	2 weeks	Open items surfaced in minutes; slippage ↓	Fewer delays; exec visibility ↑	Milestone trend analytics
Service Desk Assist (Internal IT/Facilities)	Knowledge articles, known errors, runbooks	2–3 weeks	Ticket handle time ↓ 30–50%	Less L2 load; higher self-serve	Chat-in-ticket; auto draft resolutions
Disruption Comms Playbook Copilot	Past comms, templates, SOPs	1–2 weeks	Draft time ↓ 80%; approval time ↓	Faster, consistent comms; brand risk ↓	Live data hooks (flights, gates) later

Resources

- Progress blog:
 - [REMi v0 + general explanation](#)
 - [REMi v1](#)
 - [REMi v1 vs IBM Granite-Guardian 3.1](#)
- Agentic RAG docs: [Evaluate your RAG Experience using REMi](#)
- Agentic RAG Understanding API: [REMi endpoint](#)
- Agentic RAG Python SDK: [Predict REMi](#)