Technology Topic Summary

intel ai

What Is Retrieval-Augmented Generation (RAG)?

RAG is an artificial intelligence (AI) framework that allows organizations to customize large language models (LLMs) without fine-tuning, leading to faster deployment and lower costs.

RAG at a Glance

LLM challenges:

- LLMs need fresh and specialized data to provide more reliable and relevant responses.
- Updating LLM data requires fine-tuning, which takes time and can be costly.
- Sharing private data with third-party LLMs presents data security, privacy, and compliance risks.

Solution:

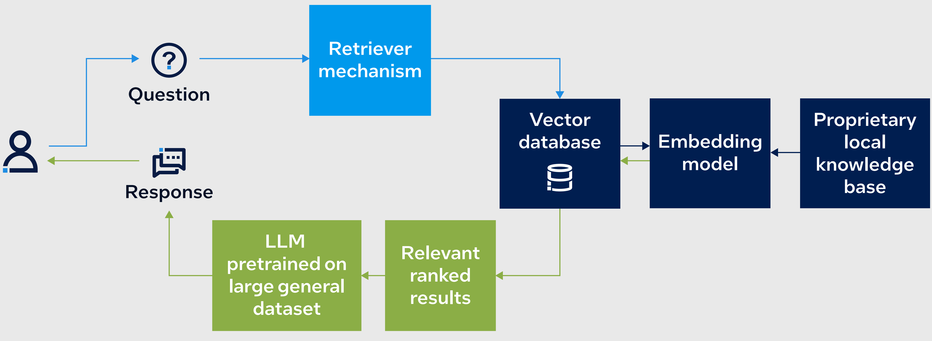
- RAG connects LLMs to proprietary local databases, which helps keep data safe.
- RAG allows frequent database updates without requiring LLM fine-tuning.
- RAG helps launch customized AI apps faster and cheaper.

Benefits of Using a RAG Framework

Providing LLMs with access to fresh, specialized data helps drive business outcomes and customer satisfaction.

- **Reduce time and cost:** Launch GenAI apps faster without fine-tuning.
- **Improve outputs:** Deliver more accurate, relevant responses based on custom content.
- **Shorten time to action:** Respond faster to insights based on continuously updated data.
- **Enhance data privacy:** Improve data security with protected local access.

The RAG Process Simplified



intel ai

Dive into the details

Learn more at intel.com/ai

Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.
© Intel Corporation