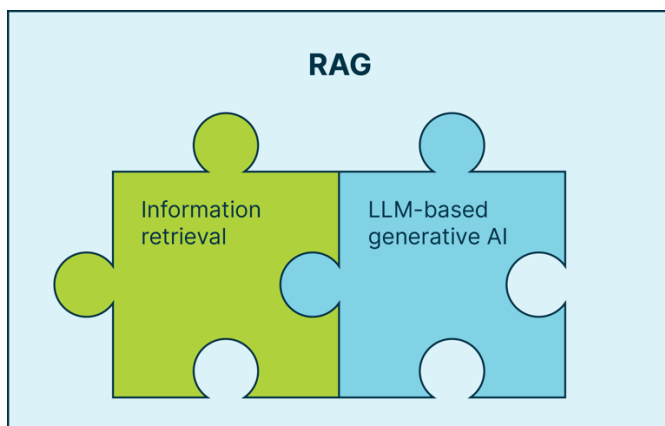


RAG: How to Get All the Right Answers

With retrieval-augmented generation (RAG), you can take the capabilities of generative AI to the next level. RAG helps large language models (LLMs) tap into additional data sources so that employees and customers can intuitively gather context-related and up-to-date knowledge.

LLMs are only as good as the data they were trained on. RAG is the key to avoiding obsolete and overly generalized answers from an LLM: with this technology, users can search across a wide range of both internal and external data sources. The «basic knowledge» of the LLM is enriched with the most recent company-specific information on products and services. This greatly speeds up and improves the search for relevant data from various distributed sources.



Our offer

We help you define a target architecture for your use cases. We fulfill your operational requirements and needs for data protection, data security, and scalability while ensuring a sustainable implementation of your RAG system. And we prepare your data and set up an LLM tailored to your purposes.

Our designs for search inputs and result listings follow a user-centered approach: we use prompt engineering to consistently align the persona of the virtual assistant and its tonality with the needs and goals of the users.

Your benefits

- RAG makes the replies of generative AI faster, up to date, more contextual, and more accurate.
- A variety of internal and external systems can be integrated as additional data sources.
- The data sources used for the search comply with the respective user authorization.
- Multi-level inquiries in natural language are supported.
- The search processes can be monitored and traced.
- Full transparency: In contrast to LLMs, RAG systems can cite the sources of their answers.

Integration and use

- A practical example of integrating and using RAG is an advanced chatbot to support your customer service. Such a chatbot understands queries, collects relevant information from a company database, and generates precise answers tailored to the user by combining retrieval and generative models.

Contact

Adnovum AG
Stéphane Mingot

Badenerstrasse 170
8004 Zurich

T +41 79 306 08 58
stephane.mingot@adnovum.ch

www.adnovum.com



adnovum

And your digital business works