

# Deploy a secure Generative AI using your own data in a few simple steps

How to safely accelerate your Generative AI journey with an on-premises Proof-of-Concept.



How can I **deploy my own Generative AI**?

**Great question!**

Here are four approaches for building generative AI. I recommend starting model augmentation.



**Simple inferencing with a pre-trained model**  
ChatGPT-like tools where a user asks "prompt engineering" questions to get answers from a pre-trained model.



**Model augmentation**  
Combining simple inferencing with domainspecific data by using retrieval-augmented generation, or RAG.



**Fine-tuning models**  
Altering model weighting and making adjustments using your own data delivers better results.



**Model training**  
Building a highly bespoke model and training it with specific data can help solve complex problems, but typically involves the most effort and investment.





**RAG models sounds interesting.**


Can you tell me a little bit more about how this works?


**RAG models provide LLMs with domain-specific knowledge**

RAG is used in LLMs (Large Language Models) to retrieve knowledge relevant to user queries, combining the user's prompt with domain-specific data. You can deploy powerful LLMs like LLaMA 2 and boost capabilities with propriety data while keeping it secure and protected. Data types include:

  
PDFs

  
Text files

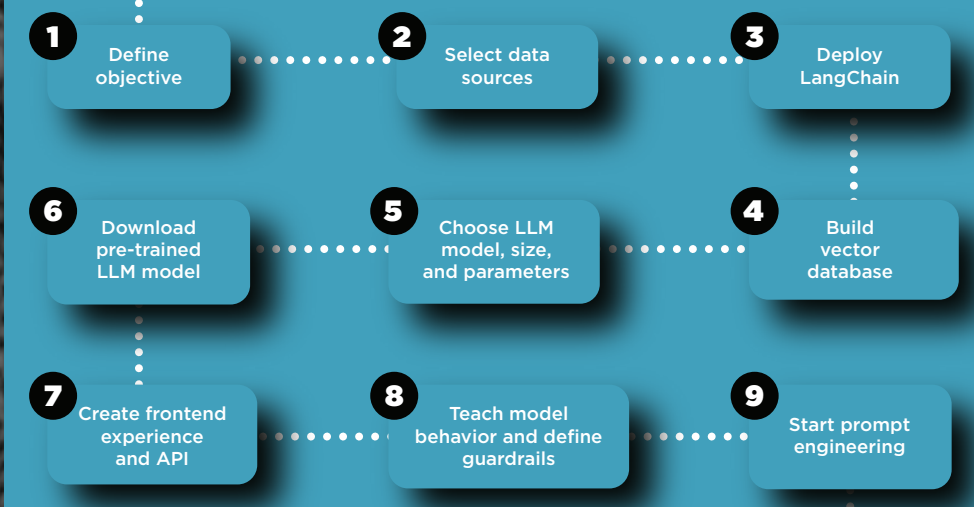
  
Databases

  
Images



Interesting. So, how do I start **building my own RAG model**?

Here are the **9 steps to building your own RAG model**:



You can do RAG anywhere, so **why should I do it on premises?**

There are many **advantages of running Generative AI on premises**. Here are some of them:

  
Enhance security controls and data access

  
Reduce risk and create guardrails

  
Capitalize on real-time data

  
Create cost efficiencies

  
Drive energy efficiency



Now that I have my LLM connected to my data, **how do I use it?**

**Start with the data you want to use** and then "prompt" the model on what data to use and what it should produce for you.

**Need tips on good prompting? Here's a prompt primer:**

- Tell the model what role it should portray itself as
- Indicate which data sets the model should use
- Define the specific output you want generated



Can you provide an **example of prompts** to try?

**Here are three prompts you can copy and paste:**

RAG can empower any team in your organization depending on the data you want to use.

**1 Sample marketing prompt:**  
[You are a B2B marketer]. Using our list of [customer reference stories], write an [email detailing the three most common challenges our product can solve].

**2 Sample job hiring prompt:**  
[You are a hiring manager] analyzing your [internal role 2 descriptions], to write a [public description for our open positions].

**3 Sample technical document prompt:**  
[You are a product manager]. Using [our internal technical specification documents], produce [release notes for our customers about our latest update].\*

\*For best results, specifying document and data location may help.



How can I **accelerate** my enterprise

**Learn from Generative AI experts.**

Reach out about Generative AI transformation and learn the difference between traditional AI and Generative AI. Start accelerating your AI journey today.

[Contact Us](#)

