



Craft your own Generative AI Chatbot with Amazon Bedrock



AGENDA

PART 01

Intro

What is
Generative AI?

PART 02

Amazon Bedrock

Building a GenAI
Chatbot

PART 03

Best Practices and Tips

Conclusion and Q&A



INTRODUCTION

Théo Lebrun



lebruntheo



Falydoor



@Falydoor



<https://theolebrun.com/>



Data Engineer and Technical Manager
at Ippon Technologies



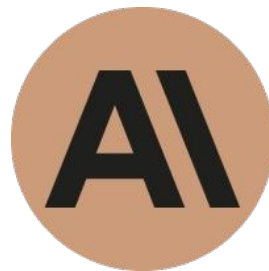
What is Generative AI?

- **Generate Human-like Text**
 - Based on an input (or prompt)
 - Can be enhanced with a context (RAG)
- **Tailored**
 - Model
 - Temperature
- **Use Cases**
 - Customer support
 - Content creation
 - Virtual assistants

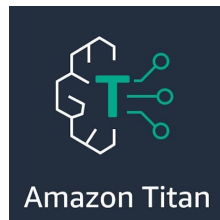


ChatGPT

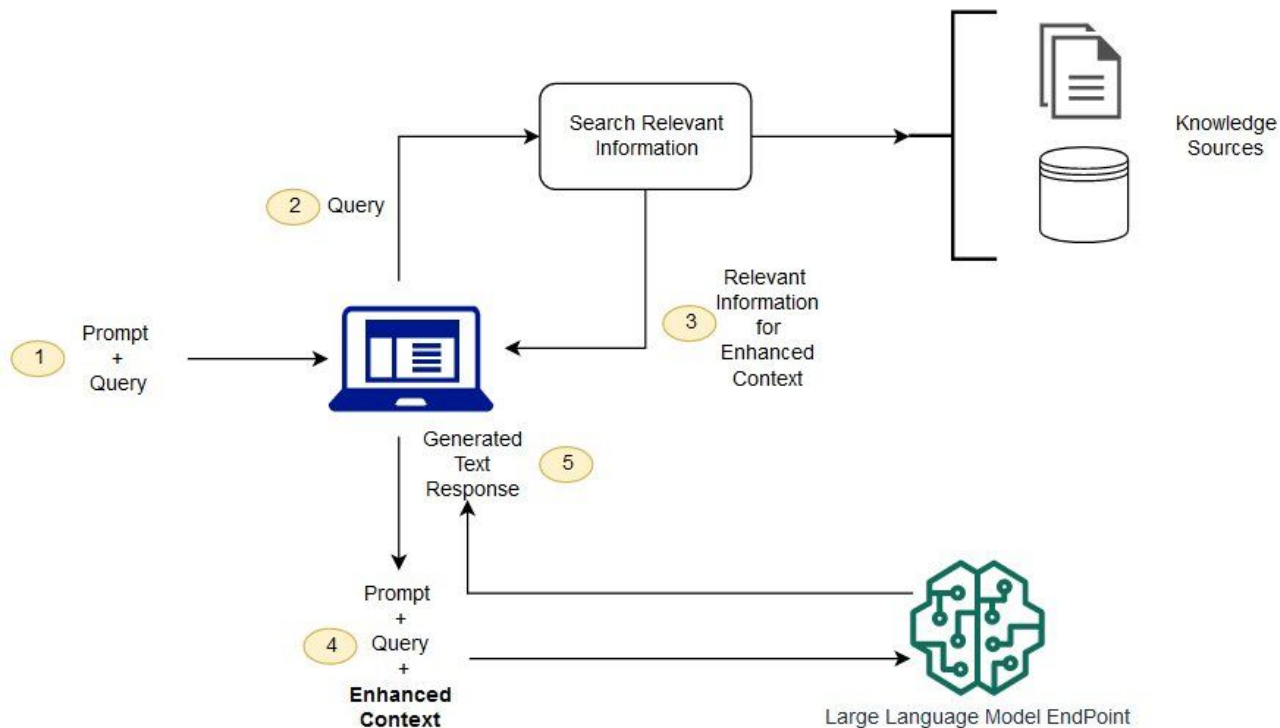
 Meta



 Gemini



Retrieval-Augmented Generation (RAG)



Amazon Bedrock

- **Fully Managed Serverless Service**
 - High-performing foundation models (Llama, Claude, Titan, etc)
 - Interactive playgrounds
 - Bring your own custom models
- **Knowledge Bases**
 - S3 and third party (Salesforces, Confluence, etc)
- **Builder Tools**
 - Prompt flows and management
 - Agents
- **Data Protection and Privacy**
 - Encrypted in transit and at rest
 - Not shared with model providers
 - Monitoring and logging capabilities



Amazon Bedrock



Developing your Chatbot

- **Picking a LLM Framework**
- **Embeddings**
 - Vectorstore
 - Model
- **Choosing the Right Model**
- **Bot Platform**
 - Gradio
 - Slack, Teams, etc
- **Deployment and Scaling**

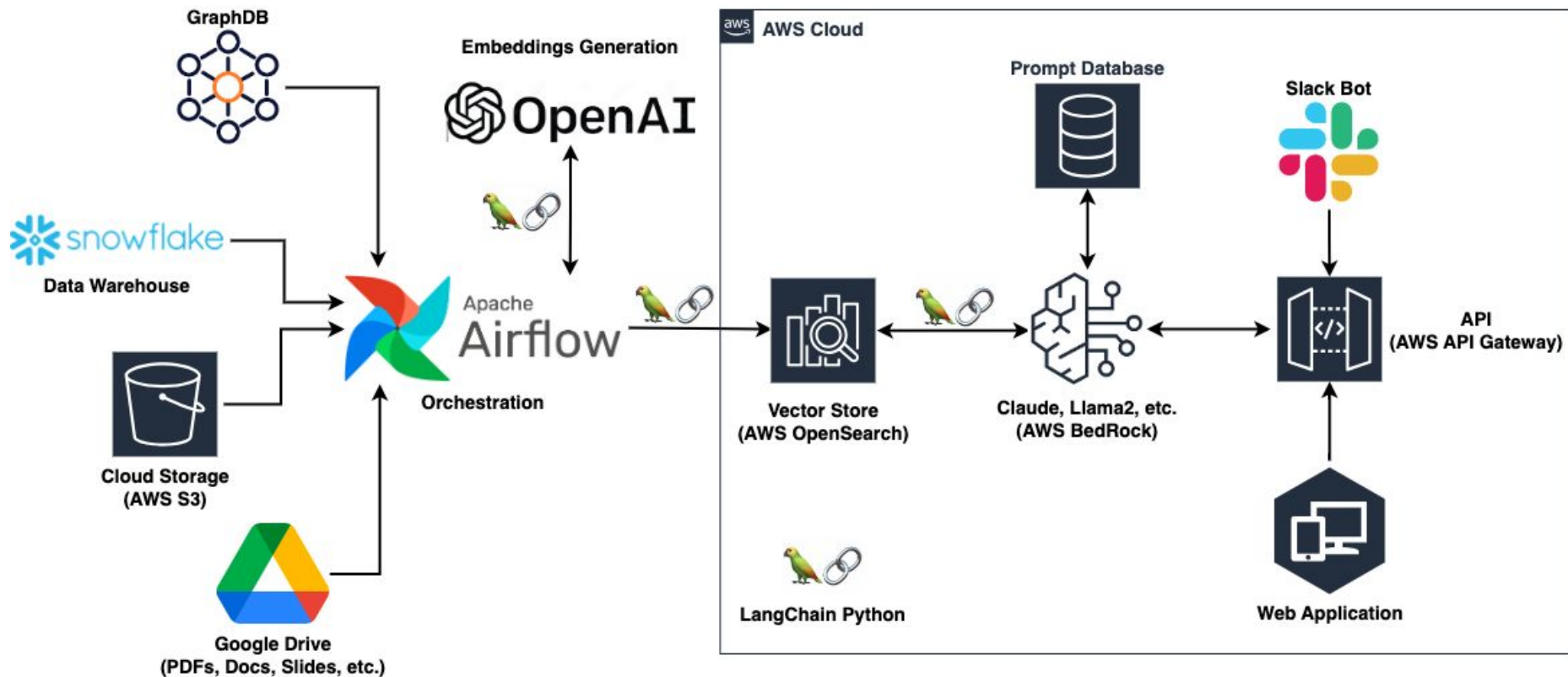


Preparing Your Data

- **Sources**
 - File based
 - Database
 - Third party
- **Data Requirements**
 - How to collect and clean
 - Preprocessing
 - Update frequency
- **Document Integration**
 - Type of documents (doc, pdf, txt, etc)
 - Chunking strategies



Architecture



Best Practices

- **Security Considerations**
 - User's data not shared
 - Amazon Bedrock Guardrails
- **Ongoing Maintenance**
 - Monitoring
 - Feedback system
- **Agents for Amazon Bedrock**
 - Automate repetitive tasks
 - Improve generated response



Tips and tricks

- **Conversational Retrieval**

- Timestamp, who is speaking, etc
- History
- Include source documents
- Multi index and prompt (reduce LLM hallucination)

- **Performance Optimization**

- Streaming
- Max tokens
- Model type (turbo, express, lite, etc)



Useful links

- <https://aws.amazon.com/bedrock/>
- <https://www.langchain.com/>
- [Build a Chatbot with Langchain and Amazon Bedrock](#)
- <https://github.com/Falydoor/talks>



Thank you for your time!

I would love to chat more with you about building a successful Gen AI Chatbot using Amazon Bedrock or any other technology.

You can email me at tlebrun@ipponusa.com or chat with me after the talk.





Q&A

en.ippon.tech

contact@ipponusa.com — +1 844-477-6687 — @ipponUSA