

The background features a dark blue gradient with large, overlapping, semi-transparent shapes in shades of purple and magenta. Two thin, light blue lines cross the scene diagonally. The text is positioned on the left side.

AWS re:Invent

DECEMBER 2 - 6, 2024 | LAS VEGAS, NV

AIM363

Advanced strategies for prompt engineering in gen AI application development

Ahmed Raafat

(he/him)

Principal Solutions Architect

AWS

Amit Lulla

(he/him)

Principal Solutions Architect

AWS



Agenda

- Introduction to Amazon Bedrock & prompt engineering
- Whiteboarding for real-world scenario
- Architecture overview
- Amazon Bedrock Prompt management & Flows
- Best practices & advanced strategies
- Recap & key takeaways

AWS Generative AI Stack

APPLICATIONS TO BOOST PRODUCTIVITY



Amazon Q Business
INSIGHTS AND AUTOMATION



Amazon Q Developer
SOFTWARE DEVELOPMENT LIFECYCLE

MODELS AND TOOLS TO BUILD GENERATIVE AI APPS

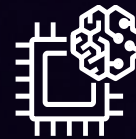


Amazon Bedrock
AMAZON MODELS | PARTNER MODELS

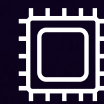
INFRASTRUCTURE TO BUILD AND TRAIN AI MODELS



Amazon SageMaker
MANAGED INFRASTRUCTURE



AWS Trainium
AWS Inferentia



GPUs

HIGH PERFORMANCE COMPUTE



Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models (FMs)

Choice of leading FMs through a single API

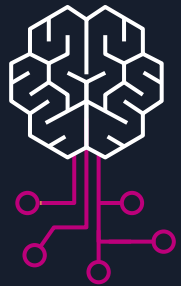
Model customization

Retrieval Augmented Generation (RAG)

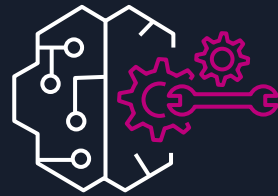
Agents that execute multistep tasks

Security, privacy, and safety

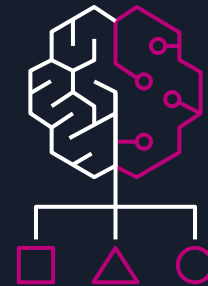
Amazon Bedrock **simplifies**



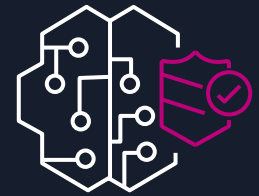
Choice



Customization

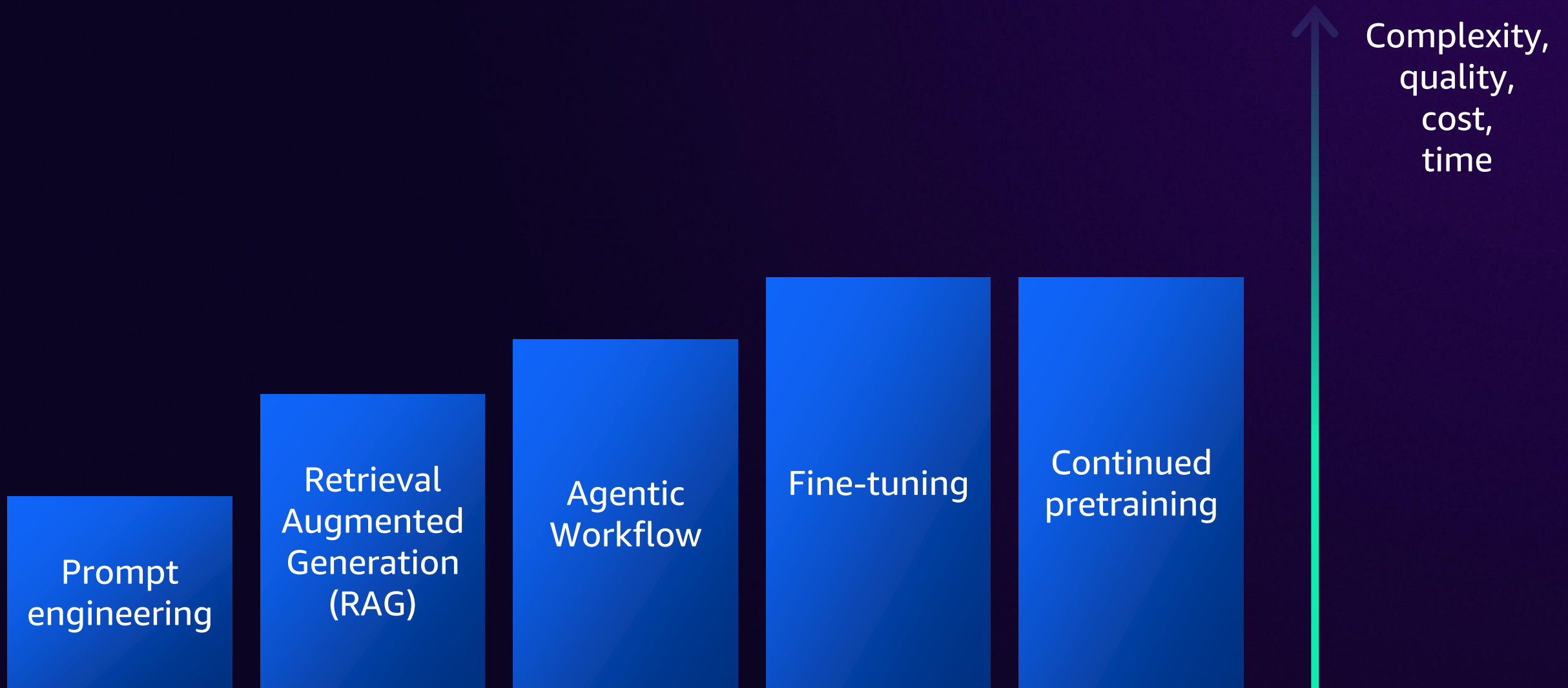


Integration



**Security &
governance**

Common approaches for customizing FMs



Prompt engineering – Introduction

- What is a **prompt**?
 - ✓ Text input provided to an AI system to elicit a response
- What is **prompt engineering**?
 - ✓ Using NLP techniques to craft prompts that steer FMs/LLMs towards desired responses
- Why is this important?
 - ✓ Enables fine-grained and strategic control over models' behavior
 - ✓ Targets desired capabilities
 - ✓ Mitigates risks



**NLP = Natural Language Processing
FM = Foundation Model
LLM = Large Language Model*

(Typical) prompt structure

The following is text from a restaurant review:

"I finally got to check out Alessandro's Brilliant Pizza and it is now one of my favorite restaurants in Seattle. The dining room has a beautiful view over the Puget Sound but it was surprisingly not crowded. I ordered the fried Castelvetro olives, a spicy Neapolitan-style pizza and a gnocchi dish. The olives were absolutely decadent, and the pizza came with a smoked mozzarella, which was delicious. The gnocchi was fresh and wonderful. The waitstaff were attentive, and overall the experience was lovely. I hope to return soon."

Summarize the above restaurant review in one sentence.

Contextual information about the task

Reference text for the task

Simple, clear and complete instructions

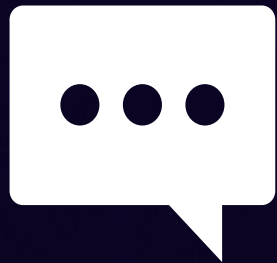
Instructions placed at the end

The form of output is specified

Prompt engineering philosophy

Empirical science: always test your prompts & iterate often!

Develop test cases



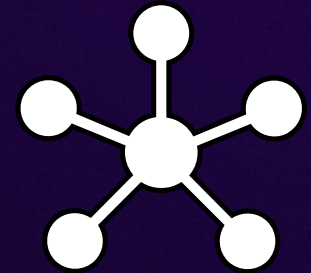
Test prompt against cases



Refine prompt

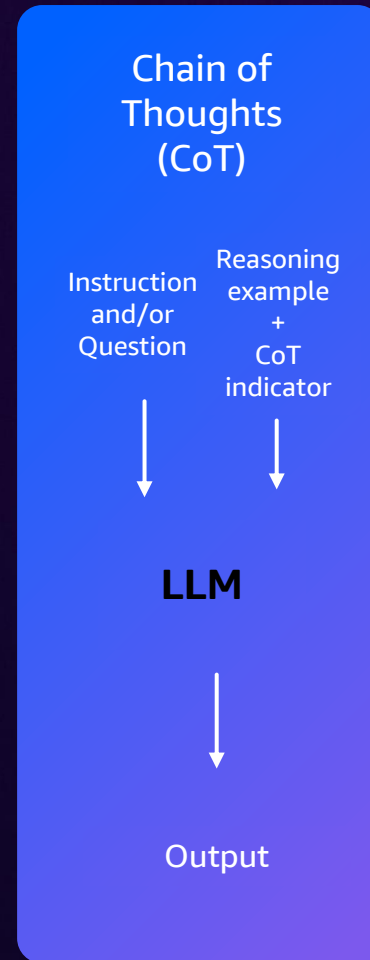
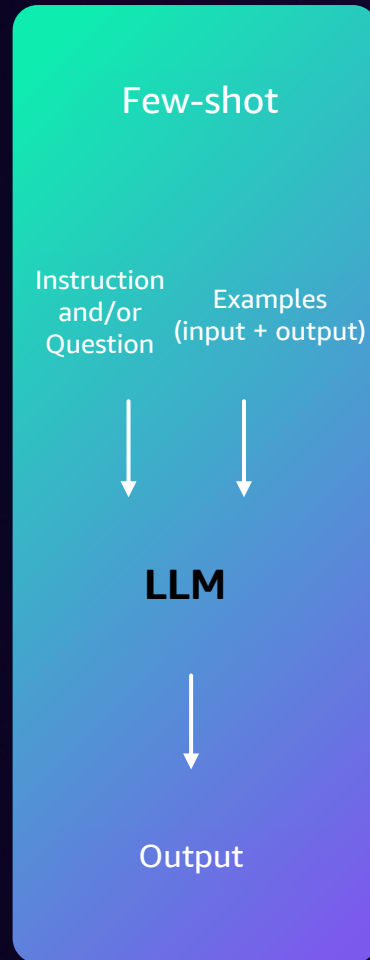
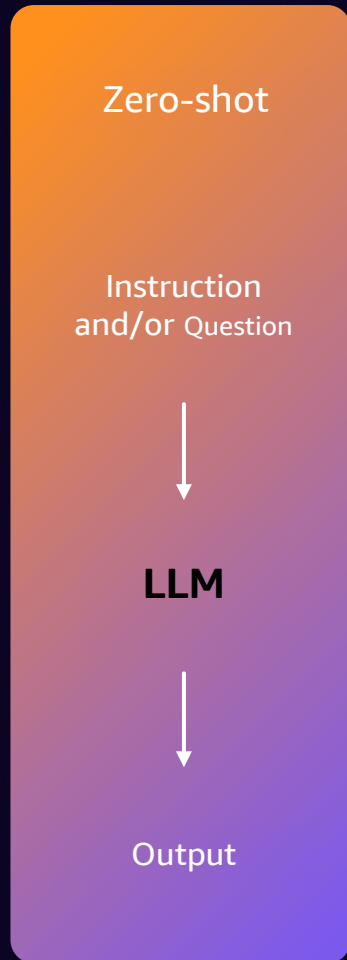


Share polished prompt



Don't forget edge cases!

Prompt engineering techniques



Prompt engineering

- Prompt engineering – **C**ontext **O**bjective **S**tyle **T**one **A**udience **R**esponse
- Prompt templates
- Chaining prompts for complex tasks
- Dynamic prompt generation based on content type
- A/B testing different prompt versions
- Incorporating user feedback for continuous improvement

Whiteboarding



Amazon Bedrock Prompt management



Prompt management

- Simplifies prompt creation, evaluation, and versioning
- Facilitates prompt sharing and collaboration
- Helps optimize responses for specific use cases
- Demo: Creating and versioning prompts for content moderation

Amazon Bedrock Prompt management

STREAMLINE YOUR GENERATIVE AI APPLICATION DEVELOPMENT

Built-in prompt builder in Amazon Bedrock console and SDK APIs

Prompt library for easy cataloging and management

Seamless integration with Prompt flows and Amazon Bedrock Studio

The screenshot shows the 'Prompt management' page in the Amazon Bedrock console. At the top, there's a breadcrumb trail: 'Amazon Bedrock > Prompt management'. The main heading is 'Prompt management' with an 'info' link. Below this, a 'How it works' section is divided into three columns: 'Create a prompt', 'Test the prompt', and 'Use the prompt'. Each column has an icon and a brief description. The 'Create a prompt' section explains that prompts are inputs to guide FMs. The 'Test the prompt' section describes inputting sample values for variables. The 'Use the prompt' section explains publishing a snapshot for production. Below this is a 'Prompts (3)' section with a search bar and a table of prompts. The table has columns for Name, Description, and Last updated. Three prompts are listed: 'MyTestPrompt-20240628-135832', 'CustomerService', and 'MyTestPrompt-001'. Action buttons for 'Delete', 'Edit', and 'Create prompt' are visible.

The screenshot shows the 'Prompt builder: CustomerServiceSummary' page. It has a breadcrumb trail: 'Amazon Bedrock > Prompt management > 87VLP47NYL > Prompt Builder'. The main heading is 'Prompt builder: CustomerServiceSummary' with 'Delete', 'Create version', and 'Save draft' buttons. The page is split into several sections. The 'Prompt' section contains a text area with a placeholder: 'You're a customer service analyst. Always respond in {{language}} of the user and in {{format}} format. Consider the following context: {{context}} Summarize the following conversation in a single sentence: {{conversation}}'. Below this is a note about valid characters for variables. The 'Configurations - Optional' section includes a 'Model' dropdown set to 'Claude 3 Haiku v1', 'Inference parameters' with a 'Reset' button, and 'System prompts' and 'Randomness and diversity' sliders for Temperature (set to 1), Top P (set to 0.999), and Top K (set to 250). The 'Test variables (4)' section shows a table of variables: language (English), format (JSON), context (The customer has an Android phone and has complained about the device 3 times this month), and conversation (Customer: My phone is not charging, can you please help me? Agent: Sure, let me retrieve your details. Customer: Thanks. Agent: I see you have multiple issues wit...). A 'Test window' on the right shows a chat interface with a system prompt and a user response. A 'Run' button is at the bottom right.



Amazon Bedrock Flows



Amazon Bedrock Flows

ACCELERATE USER-DEFINED GENERATIVE AI WORKFLOWS

- Low-code visual builder for user-defined generative AI workflows
- Native integration of FMs, Amazon Bedrock Prompts, Amazon Bedrock Agents, Amazon Bedrock Knowledge Bases, Amazon Bedrock Guardrails, and other AWS services
- Accelerates creation, testing, and deployment

Amazon Bedrock Flows

ACCELERATE USER-DEFINED GEN AI WORKFLOWS

The screenshot displays the Amazon Bedrock Flow Builder interface. The main workspace shows a workflow diagram with the following nodes: a 'Flow Input' node (output: document, String), a 'Prompt' node (Prompt_1, input: input, String; output: modelCompletion, String), a 'Condition' node (ConditionNode_1, input: ConditionInput, String; condition: ConditionInput == "octank"; go to node: KnowledgeBaseNode_1; if all conditions are false, go to node: Prompt_2), a 'Knowledge base' node (KnowledgeBaseNode_1, input: retrievalQuery, String; output: outputText, String), and two 'Flow Output' nodes (FlowOutputNode and FlowOutputNode_1, both with input: document, String). A left sidebar contains a 'Flow Builder Info' panel with categories: Logic (Collector, Condition, Iterator), Orchestration (Agents, Prompts), Code (Lambda function), Data (Knowledge base, S3 Retrieval, S3 Storage), AI Services (Lex), and Other (Flow Output). At the top right, there are 'Save' and 'Save and exit' buttons. On the right, a 'Test Flow' panel shows a test question: 'What's the content for Octank in the topic of Marketing? answer briefly'. Below it, a search result snippet is displayed: 'The search results indicate that Octank's marketing team is involved in the following activities: - Organizing a trade show exhibit to showcase their renewable energy products (source 2) - Launching a new website to showcase their biomass energy solutions (source 2) - Launching a new marketing campaign to highlight the cost-savings of their smart grid technology (source 1) - Launching a social media campaign to increase awareness about their biofuel products (source 2, source 3) - Organizing a webinar to promote their waste-to-energy solutions (source 3)'. Below the snippet is another test question: 'What are the types of energy sources in the world?' and a corresponding answer snippet: '{ "answer": "The main types of energy sources in the world include fossil fuels (such as oil, natural gas, and coal), renewable sources (such as solar, wind, hydroelectric, geothermal, and biomass), and nuclear energy." }'. At the bottom of the test panel is a 'Test your flow' input field and a 'Run' button.

Drag-and-drop interface

Direct testing, deployment, & traceability

Version control and aliasing

SDK APIs for code automation



Demo





Machine Learning

Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models (FMs)

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Prompt flow builder: Octank-QA-Assistant

Prompt flow builder [Info](#)

Nodes [Configure](#)

Prompts [Info](#)

Prompts allow you to create a library of configurable commands handled by LLMs.

[Go to prompts](#)

Node name

Q_A

Valid characters are a-z, A-Z, 0-9 and _ (underscore). The name can have up to 50 characters.

Use a prompt from your Prompt Management

Define in node

Select model

Claude 3 Sonnet v1 | [ODT](#)

[Change](#)

Message

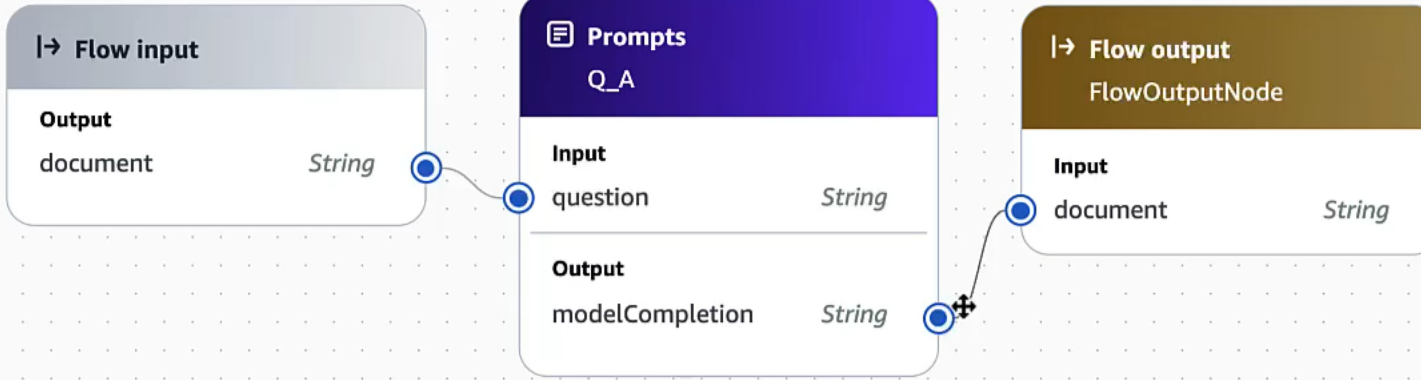
You are a Q&A expert to answer employees' questions `{{question}}`

► Inference configurations

► Input

▼ Output

Name



CloudShell Feedback

Changes to Octank-QA-Assistant successfully saved. You can publish a version or create an alias for this prompt flow.

Notifications 0 1 1 0 0 0

Amazon Bedrock > Prompt flows > Octank-QA-Assistant > Working Draft

Prompt flow builder: Octank-QA-Assistant

Save

Save and exit

Knowledge base Info

Provide FMs and agents with contextual information from your company's private data sources for RAG.

[Go to Knowledge bases](#)

Node name

Octank_kb

Valid characters are a-z, A-Z, 0-9 and _ (underscore). The name can have up to 50 characters.

Knowledge base

Octank-knowledge-base

[Go to Knowledge bases](#)

- Return retrieved results
- Generate responses based on retrieved results

Select model

Claude 3.5 Sonnet v1 | ODT

[Change](#)

Input

Name

retrievalQuery

Type

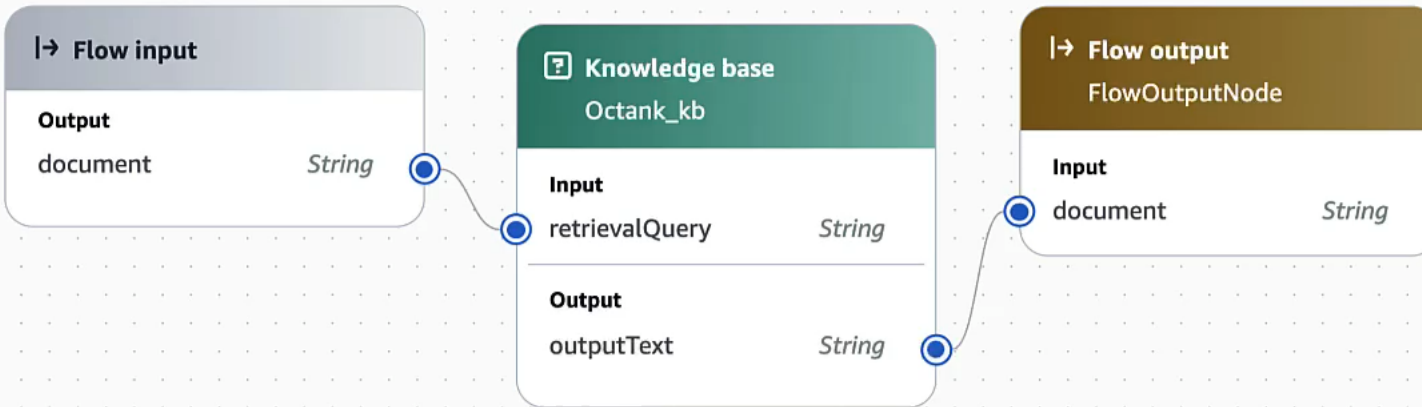
String

Expression

\$.data

Output

Name



Test Prompt flow Info

In one line, what do you know about Octank?

FlowOutputNode

Octank Financial is a leading financial services company founded in 2010 that provides comprehensive financial solutions to individuals, businesses, and institutions.

Enter your message here

Run

Advanced strategies for prompt engineering



Best practices



Best practices

- Keep prompts clear and concise
- Use context effectively
- Implement error handling and fallback mechanisms
- Regular evaluation and refinement of prompts

FLAWS framework

1

FOUNDATION MODEL SELECTION | 🤖 CLAUDE | 🐐 LLAMA 3 | ⚡ TITAN

2

LAYERED PROMPT DESIGN | 📄 SYSTEM | 👤 USER | 🔄 DYNAMIC

3

ORCHESTRATION PATTERN | 🔄 CHAIN | 📡 ROUTE | 🎯 DECIDE

4

WORKFLOW TESTING | 🔍 VALIDATE | 📊 MEASURE | 🔄 ITERATE

5

SAFETY AND SCALE | 🛡️ GUARDRAIL | 📊 MONITORING | 🚀 SCALE

Key takeaways



Recap & key takeaways

- Prompt engineering is crucial for effective generative AI applications
- Amazon Bedrock Flows enables visual workflow creation
- Prompt management streamlines the prompt lifecycle
- Advanced strategies can significantly improve AI performance
- Continuous iteration and improvement are key to success
- FLOWS framework

Content and workshop

Streamline generative
AI development in
Amazon Bedrock with
Prompt management
and Flows



Prompt Engineering
Academy for Generative
workshop



Amazon Bedrock Flows
demo



Q&A



Thank you!

Ahmed Raafat

arraafat@amazon.co.uk

Amit Lulla

alulla@amazon.co.uk



Please complete the session survey in the mobile app