Amazon Bedrock and Prompt Engineering

✦ **Key Concepts** ⌃

- **Prompt Engineering:** Prompt engineering is the process of crafting effective inputs for large language models to improve response quality.
- **Amazon Titan Models:** Amazon Titan models are specific foundation models in Bedrock, each with unique capabilities and use cases.
- **Foundation Models:** Foundation models are pre-trained models available in Amazon Bedrock that simplify the development of generative AI applications.
- **Generative AI:** Generative AI creates content such as text, images, or embeddings in response to a prompt.

This content was generated by AI, so please check for any mistakes.

Introduction to Amazon Bedrock

Amazon Bedrock is a managed service designed to build scalable generative AI applications using foundation models from leading providers. Bedrock simplifies access to FMs via APIs, abstracting away the complexities of infrastructure.

Generative AI creates content such as text, images, or embeddings in response to a prompt.

For instance:

- **In a travel app**: Generate travel recommendations or itineraries.
- **How?** FMs process vast datasets to generalize across various tasks. In this case, the App generates a prompt and submits to Amazon Bedrock, then it gets the answer and shows this to a user.

Foundation Models in Amazon Bedrock

Model	Max Tokens	Languages	Fine Tuning Supported	Use cases
Amazon Titan Text Premier	32K	English	Yes	Chat, chain of thought reasoning, open-ended text generation, brainstorming, summarization, code generation, table creation, data formatting, paraphrasing, rewriting, extraction, and Q&A.
Amazon Titan Text Express	8K	English (GA), 100+ (Preview)	Yes	Open-ended text generation, brainstorming, summarization, code generation, table creation, data formatting, paraphrasing, chain of thought, rewriting, extraction, Q&A, and chat.
Amazon Titan Text Lite	4K	English	Yes	Summarization and copywriting.
Amazon Titan Text Embeddings	8K	25+ languages	No	Text retrieval, semantic similarity, and clustering.
Amazon Titan Text Embeddings V2	8K	100+ in pretraining	No	Semantic similarity searching to find documents (for example, detecting plagiarism), classifying labels into data-based learned representations (for example, categorizing movies into genres), and improving quality and relevance of retrieved or generated search results.
Amazon Titan Multimodal Embeddings	128	English	Yes	Search, recommendation, and personalization.
Amazon Titan Image Generator	512	English	Yes	Text-to-image generation, image editing, and image variations.

Amazon Titan Image Generator v2	512	English	Yes	Text-to-image generation, image editing, image variation, image conditioning using a reference image, subject consistency using fine tuning (preserve specific subjects in generated images), and automated background removal.

More information at: [Foundation Model for Generative AI - Amazon Titan - AWS](#) ⧉

Introduction to Prompt Engineering

Prompt engineering is a critical skill for leveraging Large Language Models (LLMs) effectively. A **prompt** is the input or query that defines the task for the LLM. In Amazon Bedrock, foundational models (FMs) are pre-trained and available for use, sparing you from the high cost of training models yourself. Crafting effective prompts can dramatically improve the quality of LLM responses!

Examples of LLM Tasks

1. Classification

- **Objective**: Categorize a restaurant review based on tourism type.
- **Approach**: Provide a list of categories for the LLM to choose from and structure the prompt clearly.

2. Question-Answering

- **Objective**: Extract free-form answers from a given context.
- **Approach**: Present the context first, followed by the question.

3. Text Generation

- **Objective**: Generate ideas for a website front page.
- **Approach**: Include specific instructions about tone and style.

Conclusion

Prompt engineering is an iterative process requiring experimentation and refinement, much like writing code. Amazon Bedrock simplifies generative AI adoption by offering pre-trained models and scalable infrastructure. By combining thoughtful prompt design and the appropriate FM you can unlock powerful capabilities in applications, like the prompt performed in the Travel App.

Mark as completed

👍 Like   👎 Dislike   🚩 Report an issue

Module 1 Quiz
Graded Assignment • 20 min

Go to next item →