## Módulo 02 - Modelos da Fundação

FMs - **(Foundation Models)**, são **modelos de inteligência artificial treinados em grandes quantidades de dados** para realizar uma ampla variedade de tarefas. Esses modelos servem como base para diferentes aplicações de IA, como geração de texto, código, imagens e muito mais. Exemplos:

* Amazon Titan
* Jurassic-2
* Jumba
* Claude 3
* Comando
* Incorporação
* Llama
* Mistral
* Mixtral
* Stable Diffusion
* Stable Diffusion XL (SDXL) 1.0

Parâmetros que controlam a geração de texto por modelos de linguagens:

Temperatura: Controla a aleatoriedade das respostas.

Top K: Limita a escolha às K palavras mais prováveis.

Top P: Corta palavras com baixa probabilidade cumulativa, tornando a resposta mais focada.

Comprimento de resposta: em modelos de linguagem define o número máximo de tokens (palavras ou caracteres) que podem ser gerados em uma resposta.

**Amazon Bedrock**

│

└── **Amazon Titan** (Modelos de Fundação da Amazon)

│

├── **Amazon Titan Text**

│ ├── Geração de Texto

│ ├── Escrita Criativa

│ ├── Assistência em Código

│ ├── Resumo de Texto

│

└── **Amazon Titan Embeddings**

├── Representação Vetorial de Texto

├── Busca Semântica

├── Agrupamento e Classificação de Dados

├── Aprimoramento de Recuperação em RAG

Jurassic 2 : são sistemas de inteligência artificial que geram texto automaticamente com base em um texto de entrada (prompt). Existem dois tipos:

* Jurassic-2 Mid: um modelo médio, rápido e eficiente para entender e responder comandos em linguagem natural.
* Jurassic-2 Ultra: um modelo maior e mais poderoso, ideal para criar textos criativos, resumos e responder perguntas complexas.

Para controlar esse modelo existem alguns parâmetros:

1. Comprimento de respostas:

**maxTokens**: Define o número máximo de palavras ou partes de palavras (tokens) que podem ser usadas na resposta.

**stopSequences**: Se o modelo encontrar essa sequência de palavras, ele para de escrever imediatamente.

1. Controle de repetições:

**presencePenalty**: Evita que o modelo repita palavras já usadas no texto.

**countPenalty**: Penaliza palavras que aparecem mais de uma vez.

**frequencyPenalty**: Reduz a chance de palavras repetidas com frequência.

Claude antropic: Claude é o nome de um modelo de inteligência artificial, como o ChatGPT, criado pela empresa Anthropic. Ele serve para gerar texto, responder perguntas, ajudar com tarefas, escrever códigos, entre outras funções.

### API Converse: Funciona com qualquer modelo que suporte mensagens.

* Usa o mesmo estilo de conversa com role (quem está falando, user ou assistant) e content (o conteúdo da fala), ideal para integrar Claude em aplicativos e assistentes.

Stability AI (SDXL) – Geração de Imagens: É um modelo que cria imagens a partir de texto.

Parâmetros importantes:

* cfg\_scale: define quão fiel a imagem será ao texto
* steps: mais etapas = imagem mais precisa
* seed: define o padrão de aleatoriedade (útil para repetir resultados)

Cohere Command – Geração de Texto: Modelo treinado para seguir instruções de forma prática e útil.

Parâmetros adicionais:

* return\_likelihoods: mostra (ou não) a probabilidade dos tokens gerados
* stream: se true, a resposta vem aos poucos, como streaming
* num\_generations: número de versões de resposta a serem geradas