Generative AI Applications

Ram N Sangwan

- Text Based Applications
- Image-based Applications
- Video Generation
- Audio Applications
- Generative AI Ecosystem

Code Models
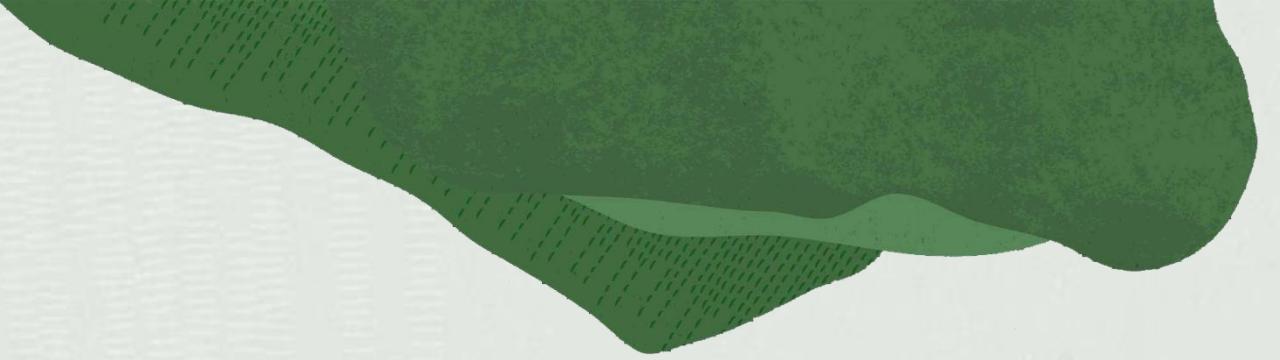
- Instead of training on written language, train on code and comments

- Cline with MCP, Co-pilot, Codex, Code Llama

- Complete partly written functions, synthesize programs from docstrings, debugging

- Largely successful: >85% of people using co-pilot feel more productive

- Great fit between training data (code + comments) and test-time tasks (write code + comments). Also, code is structured ➔ easier to learn

This is unlike LLMs, which are trained on a wide variety of internet text and used for many purposes (other than generating internet text); code models have (arguably) narrower scope

Multi-Modal

- These are models trained on multiple modalities, e.g., language and images

- Models can be autoregressive, e.g., DALL-E or diffusion-based e.g., Stable Diffusion

- Diffusion-models can produce a complex output simultaneously, rather than token-by-token
  - Difficult to apply to text because text is categorical
  - Some attempts have been made; still not very popular

- These models can perform either image-to-text, text-to-image tasks (or both), video generation, audio generation

- Recent retrieval-augmentation extensions

Text Based Applications

Text Based Applications

Customer Operations

- Automated customer service based on the customer's product suite, experience, and language

Marketing

- Content generation for e-commerce, articles optimized for SEO etc.

Sales
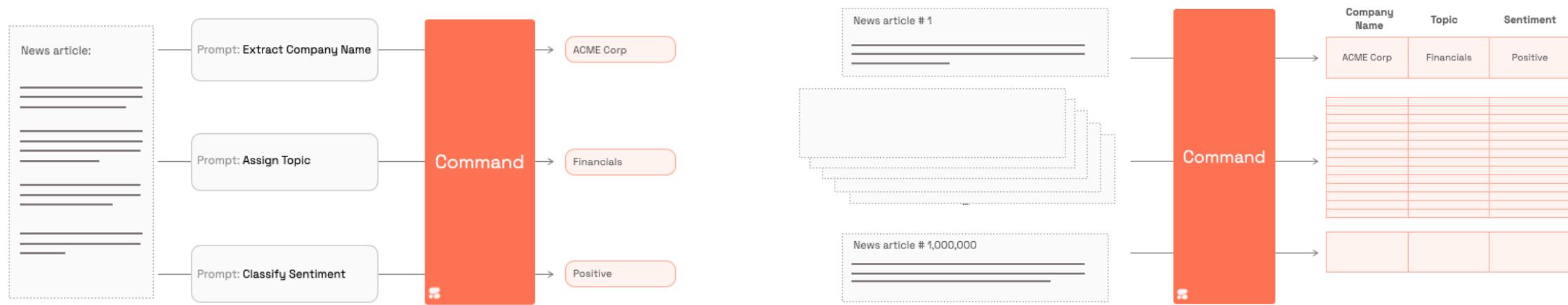
- Custom sales outreach based on interaction history and prospect profile to free up sales representative's time

Product Development

- Analysis, cleaning, and labelling of large volumes of data, such as user feedback, market trends, logs

Cohere Command-A

In-depth analysis of documents

- Command A can write product descriptions, help draft emails, suggest example press release copy, and much more.

Generative AI - Language Translation

Code snippet using Hugging Face's **transformers** library to perform a simple translation from English to French:

```
from transformers import pipeline
translator = pipeline('translation_en_to_fr', model='Helsinki-NLP/opus-mt-en-fr')
english_text = "Hello, how are you?"
french_translation = translator(english_text)

print(french_translation[0]['translation_text'])
```

```python
import warnings
warnings.filterwarnings('ignore', category=UserWarning, message='TypedStorage is deprecated')

from transformers import pipeline

# Initialize the translation pipeline
translator = pipeline('translation_en_to_fr', model='Helsinki-NLP/opus-mt-en-fr')

# Translate text from English to French
english_text = "Hello, how are you?"
french_translation = translator(english_text)

print(french_translation[0]['translation_text'])
```

```
Bonjour, comment allez-vous ?
```

- You can also specify the model you want to use for translation.
- **'Helsinki-NLP/opus-mt-en-fr'**, is a popular choice for English to French translation.
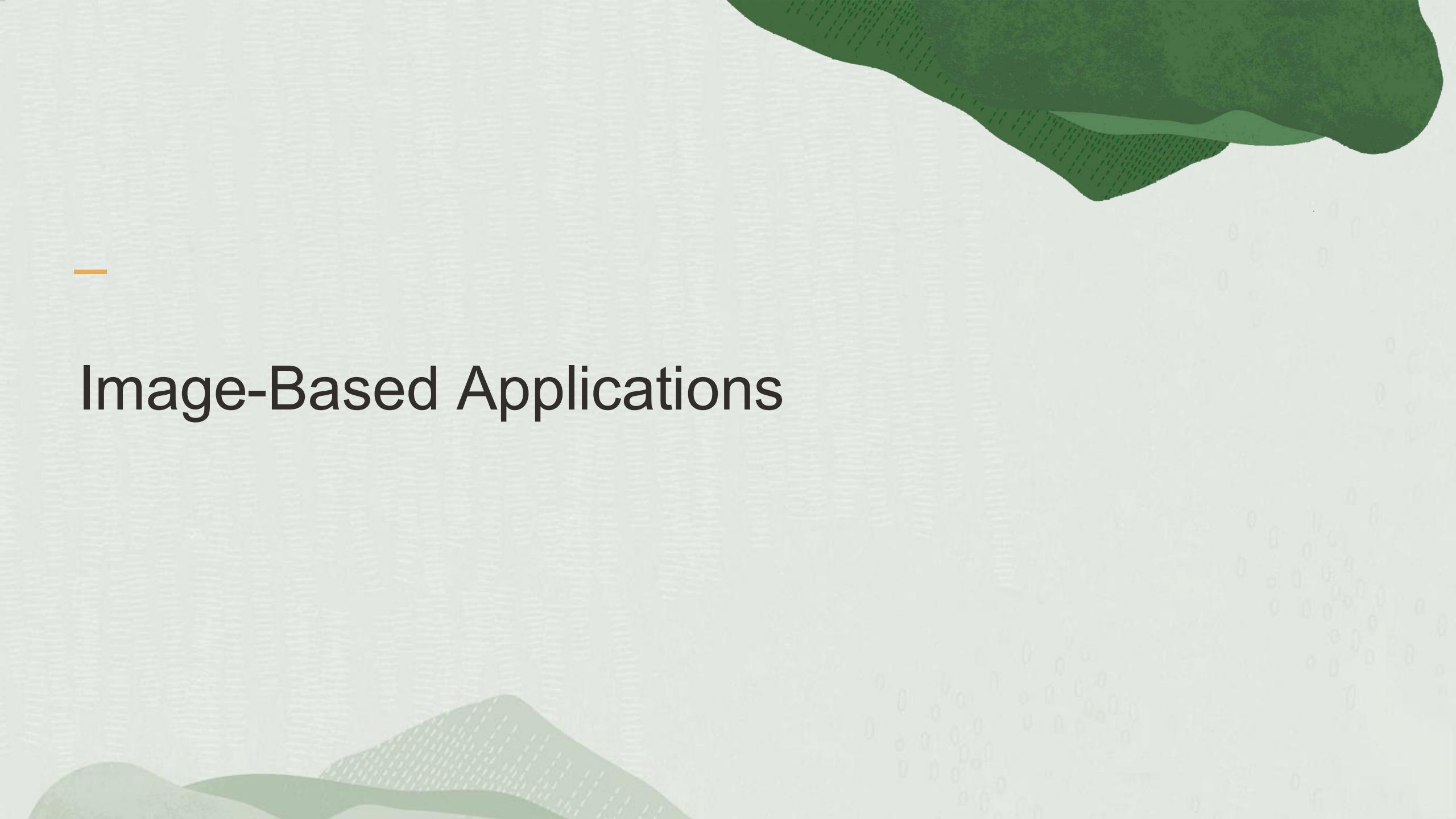
Image-Based Applications

Image Synthesis
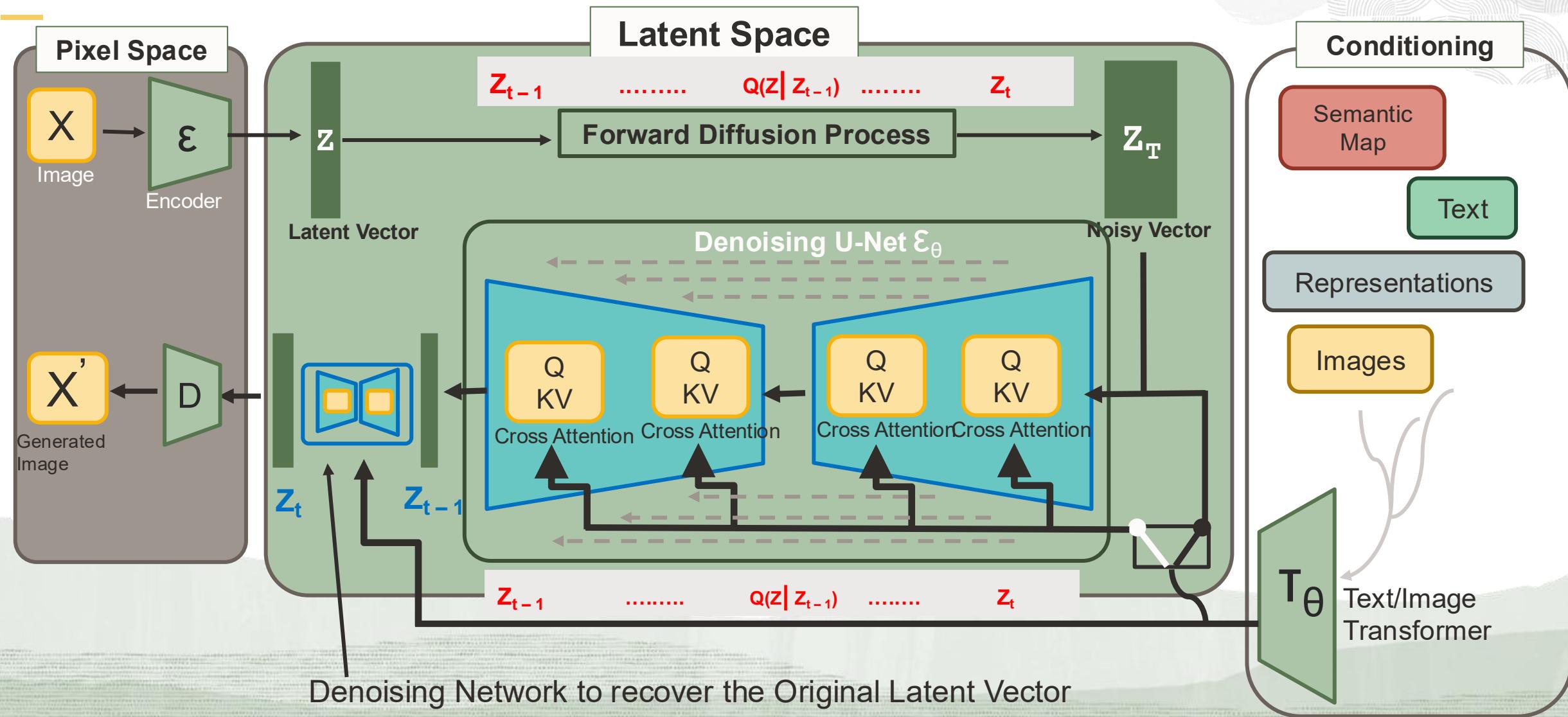
- DALL-E 3 and Stable Diffusion are two of the most advanced AI systems designed for image synthesis from textual descriptions.

- They use different architectures but share the common goal of converting text prompts into detailed images.
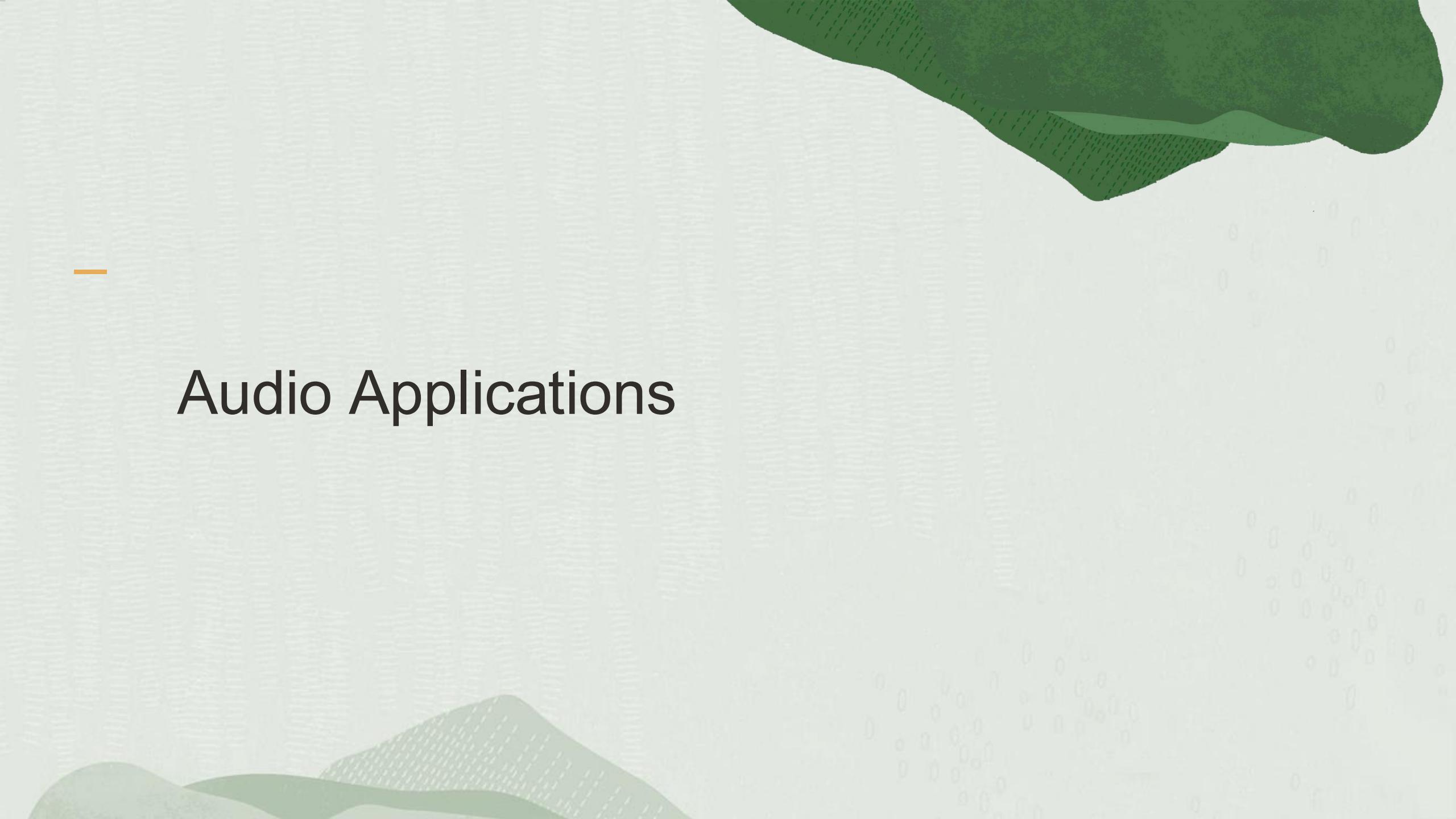
Stable Diffusion

Known for its efficiency and ability to generate high-quality images quickly.

- **Latent Diffusion Process**:
    - Operates on concept of latent space, where images are represented in a compressed form.
    - The model diffuses this latent representation starting from noise, guided by the textual description.
- **Text Conditioning**:
    - The model is conditioned on the text through an encoder that represents the textual description in a form that guides the generation process.
- **Iterative Refinement**:
    - Stable Diffusion also refines the image over several steps.

Diffusion Models – How they Work

Denoising Network to recover the Original Latent Vector

Audio Applications

Speech synthesis and Music generation

- Text-to-speech (TTS) technologies like Google WaveNet and Amazon Polly are examples of advanced speech synthesis systems.
- They convert text into lifelike spoken audio, using deep learning techniques to produce natural-sounding voices that closely mimic human speech patterns.

https://cloud.google.com/text-to-speech?hl=en#demo

https://ai-service-demos.go-aws.com/polly

Try at skillpedia.ai

Video Generation - OpenAI Sora

Try it at https://openai.com/index/sora/

- OpenAI Sora is able to generate complex scenes with multiple characters, specific types of motion, and accurate details of the subject and background.

- The model understands not only what the user has asked for in the prompt, but also how those things exist in the physical world.

Video Generation
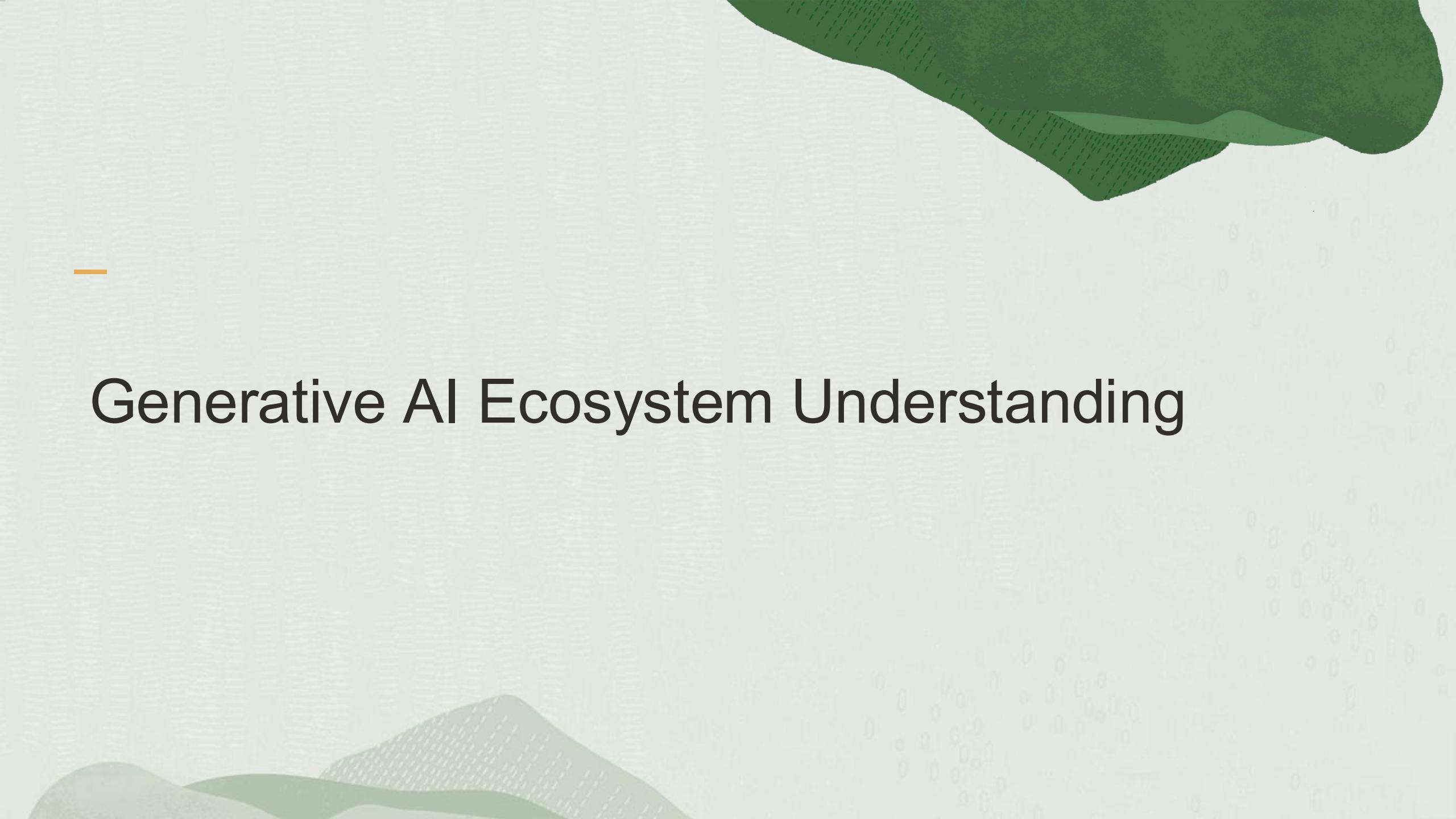
- Make-A-Video is a state-of-the-art AI system that generates videos from text.

- The system uses images with descriptions to learn what the world looks like and how it is often described.

- It also uses unlabeled videos to learn how the world moves.

- With this data, Make-A-Video lets you bring your imagination to life by generating whimsical, one-of-a-kind videos with just a few words or lines of text.

  https://makeavideo.studio/

Video Generation –Kling AI
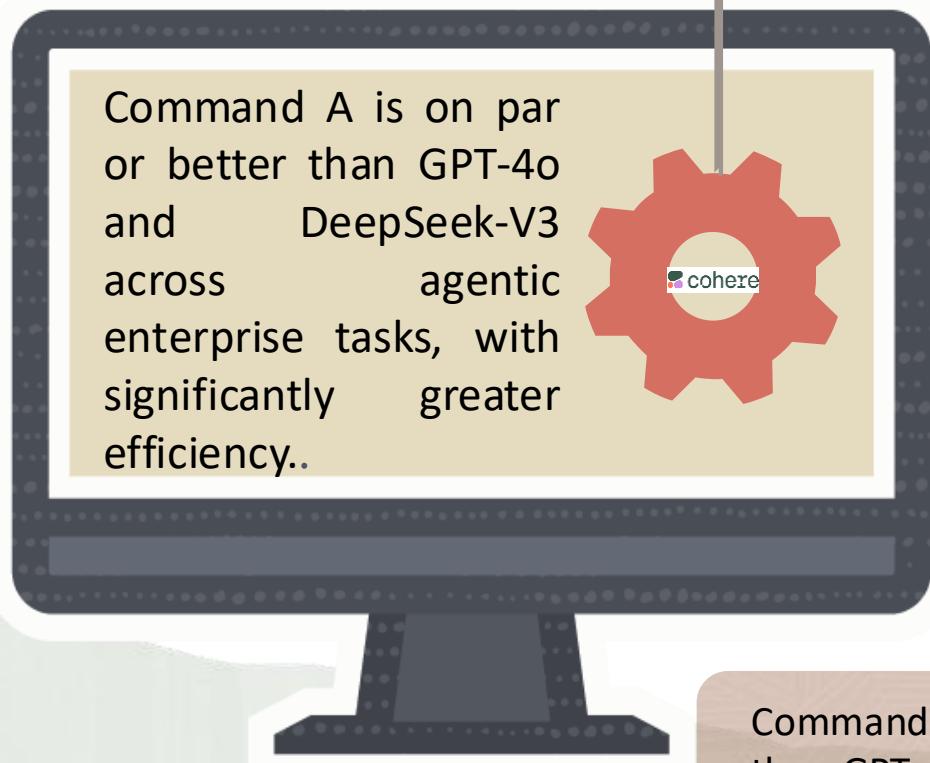
Kling AI in action here

- Generate videos up to 2 minutes long at 30fps.

- Deep understanding of text-video semantics and Diffusion Transformer architecture.
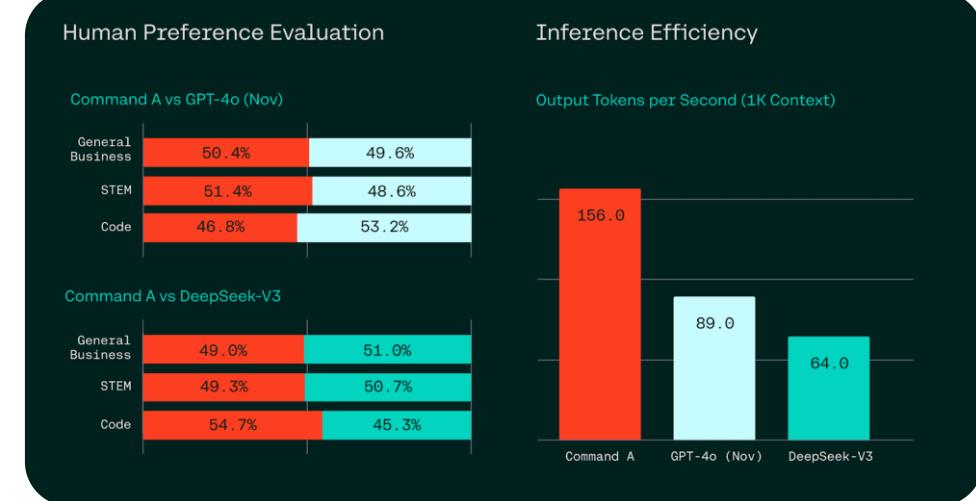
Generative AI Ecosystem Understanding

Cohere Models – Command-A
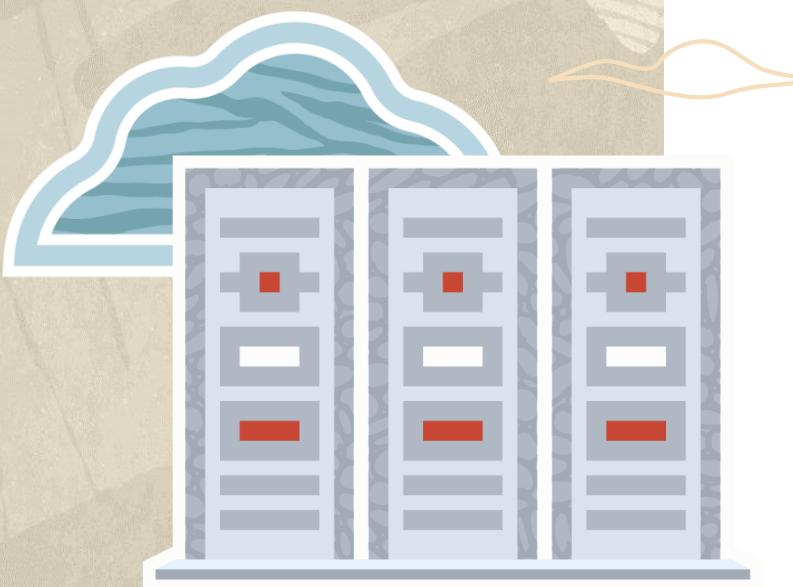
Command-A

Across a range of standard benchmarks Command A provides strong performance on instruction following, SQL, agentic, and tool tasks.

Command A is on par or better than GPT-4o and DeepSeek-V3 across agentic enterprise tasks, with significantly greater efficiency..

Human Preference Evaluation

Command A vs GPT-4o (Nov)

General Business	50.4%	49.6%
STEM	51.4%	48.6%
Code	46.8%	53.2%

Command A vs DeepSeek-V3

General Business	49.0%	51.0%
STEM	49.3%	50.7%
Code	54.7%	45.3%

Inference Efficiency

Output Tokens per Second (1K Context)

Command A	GPT-4o (Nov)	DeepSeek-V3
156.0	89.0	64.0

Command A can deliver tokens at a rate of up to 156 tokens/sec which is 1.75x higher than GPT-4o and 2.4x higher than DeepSeek-V3. Private deployments of Command A can be up to 50% cheaper than API-based access

Generative AI Ecosystem

cohere
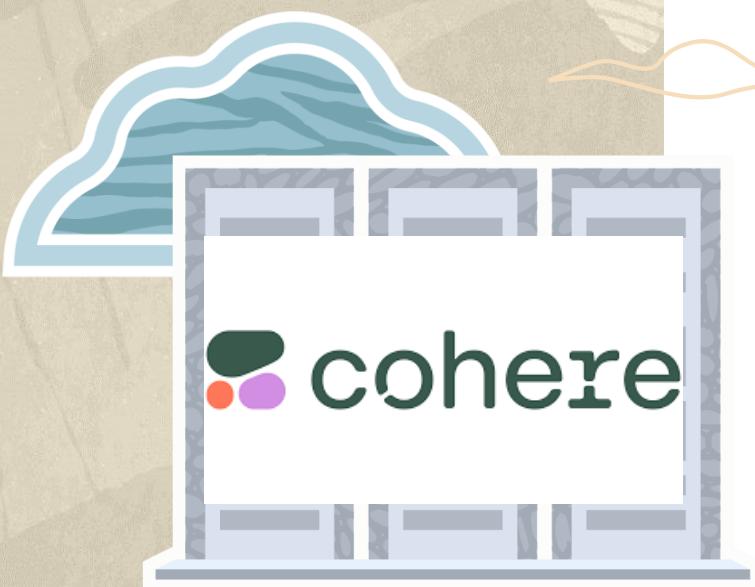
Command R:

- Instruction-following conversational model that performs language tasks at a higher quality, more reliably, and with a longer context than previous models. Used for code generation, RAG, tool use and agents.
- Support: 10 key languages
- MAX INPUT TOKENS: 128k
- MAX OUTPUT TOKENS : 4096
- API Endpoint: ***/chat***

Command R+ : **(New)**

- Command R+ is RAG-optimized model designed to tackle enterprise-grade workloads.
  - RAG with citation to reduce hallucinations
 - Multilingual coverage in 10 key languages.

Generative AI Ecosystem

Cohere Embed-English:
- Generate embeddings from text based on various parameters.
- Embeddings can be used for estimating semantic similarity between two sentences, choosing a sentence which is most likely to follow another sentence, or categorizing user feedback.
- Outputs from the *Classify* endpoint can be used for any classification or analysis task.
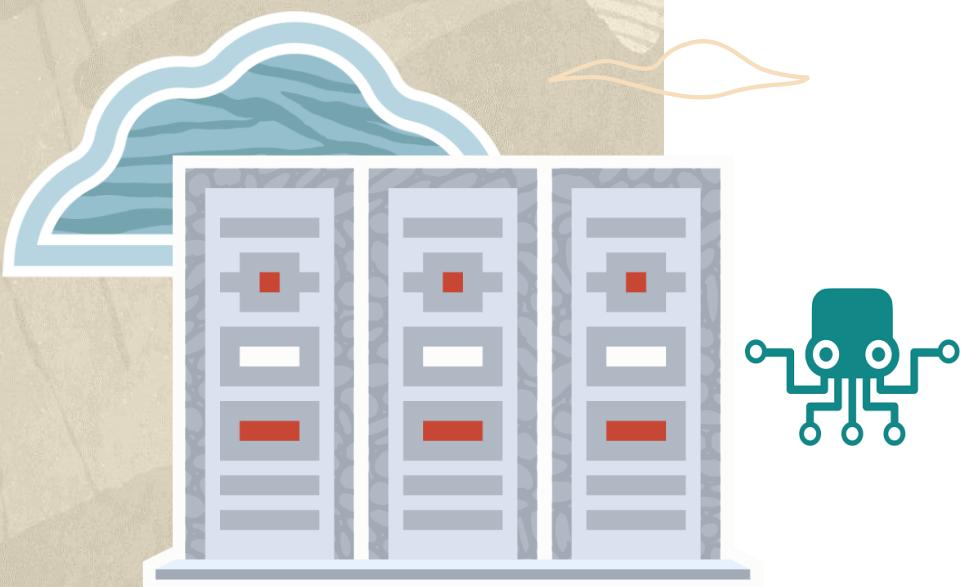
Generative AI Ecosystem
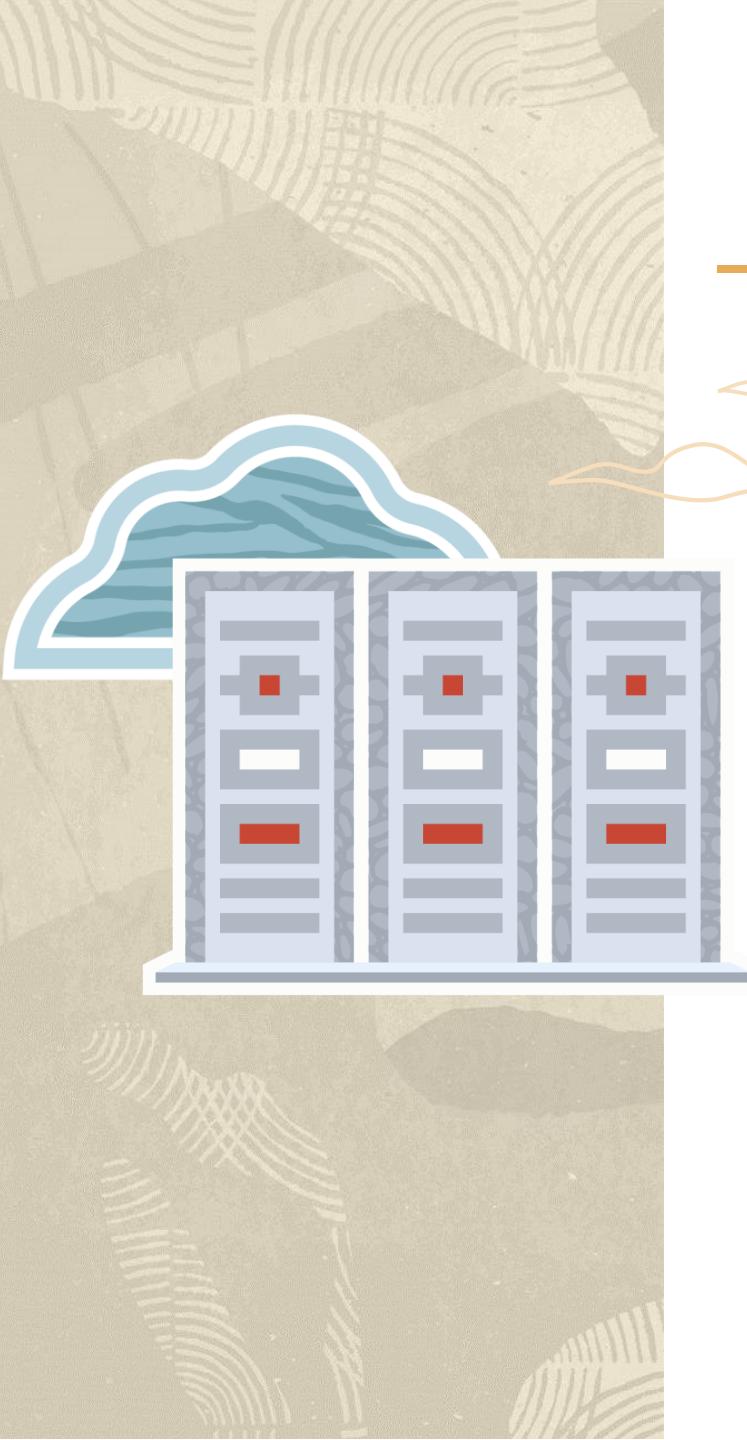
Cohere Aya:
- This is a multilingual research model from Cohere For AI (***c4ai-aya***)
- Support 101 Languages
- API Endpoint ***/generate***

AutoGPT:
- An open-source project developed by the community that aims to create AI Agents.
- Try AutoGPT in Action at https://theagenticai.io

Generative AI Ecosystem

Hugging Face: (Platform)

- Hugging Face develops and provides access to pre-trained language models.
- Models are trained on large amounts of text data and can be used for various NLP tasks such as sentiment analysis, named entity recognition, and machine translation.

LaMDA:

- LaMDA is a large language model developed by Alphabet (the parent company of Google) that can understand and generate human-like text based on user input.

Generative AI Ecosystem

LLaMA-3 and 4 Series:
- Refined post-training processes significantly lower false refusal rates, improve response alignment, and boost diversity in model answers.
- Elevates capabilities like reasoning, code generation, and instruction following.
- Meta Launched Llama-3 and Llama 4 Series of Models in multiple Sizes.

DALLE-3:
- DALL-E 3 is an AI program created by *OpenAI* that creates images from textual descriptions.
- Using more than 10-20 billion parameters, it interprets natural language inputs and generates the corresponding image.

An expressive oil painting of a basketball player dunking, depicted as an explosion of a nebula – created using DALLE 2

Thank You