

# Generative AI

A builder's guide.



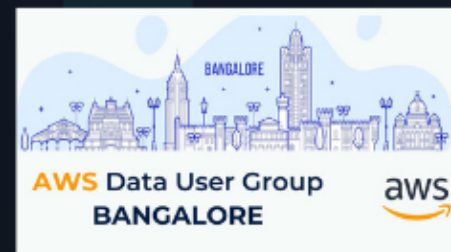
7TH WORLD MACHINE LEARNING SUMMIT

08 - 09 June, 2023 | Bangalore | Venue : Sterlings Mac Hotel Bengaluru



Ayyanar Jeyakrishnan

ORGANIZER & LEADER



CloudnLoud Community

**AWS**  
community  
builder

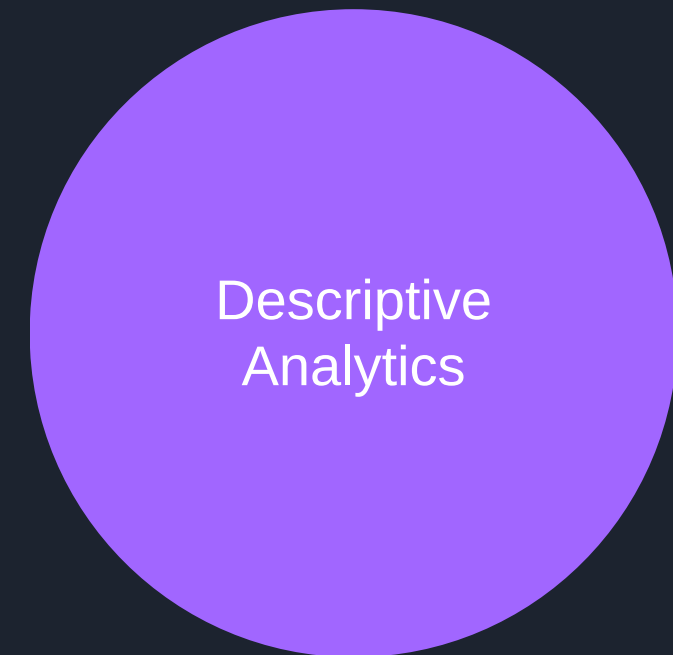
Machine Learning



# Agenda

- Path generative AI
- Highlevel Overview of Transformers
- AWS Foundation Models Overview
- Demos

# Path to Gen AI



# Transformers - Attention Is All You Need - 2017

## The Transformer model architecture.

- “Attention” enabled models to scale the understanding of relationships between words.
- Efficiently use parallel computing.
- Encoder: To convert the input sequence into a vector representation.
- Decoder: To generate the output sequence from the vector representation and To predict the next word in the output sequence

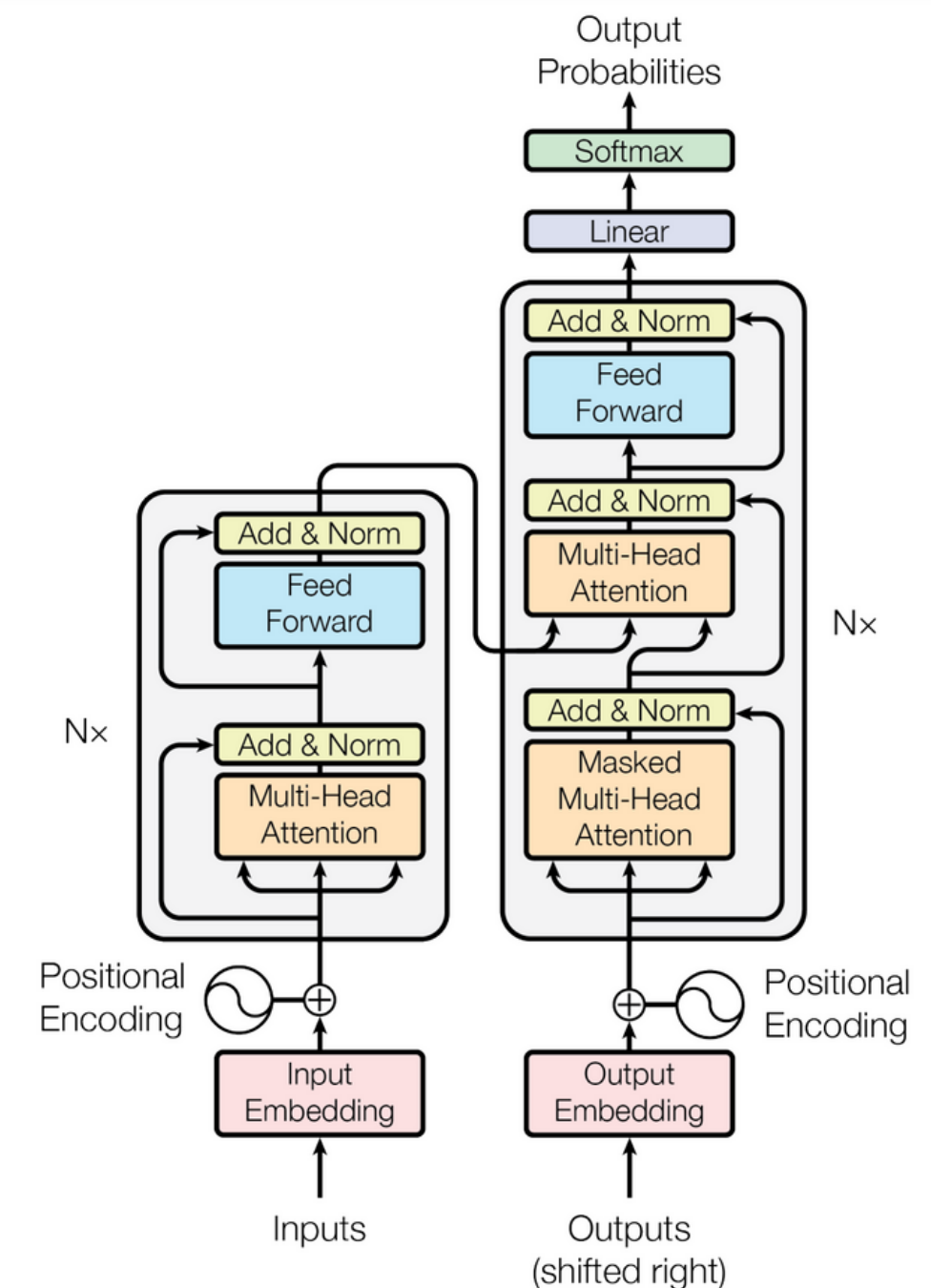


Figure 1: The Transformer - model architecture.

# Foundation models available on SageMaker JumpStart for self-managed access

## Publicly available

stability.ai

### Models

Text2Image  
Upscaling

### Tasks

Generate  
photo-realistic  
images from  
text input  
Improve quality  
of generated  
images

### Features

Fine-tuning on  
SD 2.1 model



### Models

AlexaTM  
20B

### Tasks

Machine  
translation  
Question  
answering  
Summarization  
Annotation  
Data generation



### Models

Flan T-5 models  
(8 variants)  
DistilGPT2, GPT2  
Bloom models  
(3 variants)

### Tasks

Machine  
translation  
Question  
answering  
Summarization  
Annotation  
Data generation

## Proprietary models

?

co:here

### Models

Cohere  
generate-med

### Tasks

Text generation  
Information  
extraction  
Question  
answering  
Summarization

Light<sup>star</sup>

### Models

Lyra-Fr  
10B

### Tasks

Text generation  
Keyword  
extraction  
Information  
extraction  
Question  
answering  
Summarization  
Sentiment  
analysis  
Classification

AI21labs

### Models

Jurassic-1  
Grande 17B

### Tasks

Text generation  
Long-form  
generation  
Summarization  
Paraphrasing  
Chat  
Information  
extraction  
Question  
answering  
Classification

# Demo - GenerativeAI using AWS Foundation Models

1. Text generation - Question and Answering, Summarization using AI21 Jurassic-2 Jumbo.
2. AWS code whisperer - Demo
3. Image Generation using Stability.AI 2.0 Model



# Connect with me for future Session or free workshop -

## <https://www.linkedin.com/in/jayyanar/>

- Sagemaker 101
- MLOps with Sagemaker Workshop Level 1 - 3 Hrs
- MLOps with Sagemaker Workshop Level 2 - 3 Hrs
- Generative 101
- Generative AI Workshop - 3 Hrs
- DevOps with AWS Workshop - 3 Hrs



**Thank you!**