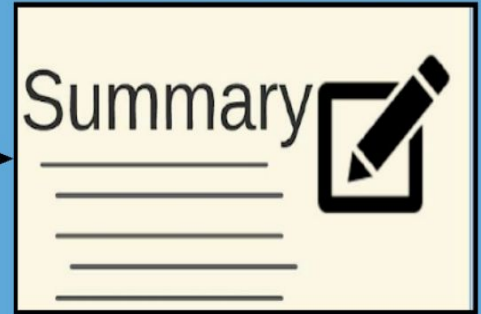


Text Summarization

- Akanksha Rawat
- Karishma Kuria
- Rajvi Shah



PROJECT OVERVIEW

- Generates succinct, accurate, easy to read, and comprehensive summaries of lengthy Text and Audio files.
- Different text summarization methods/model are implemented to get best condensed text summary.
- Reading a condensed summary can give anyone a sense of what audio or text files focuses on or help catch up on hours' worth of content in minutes.

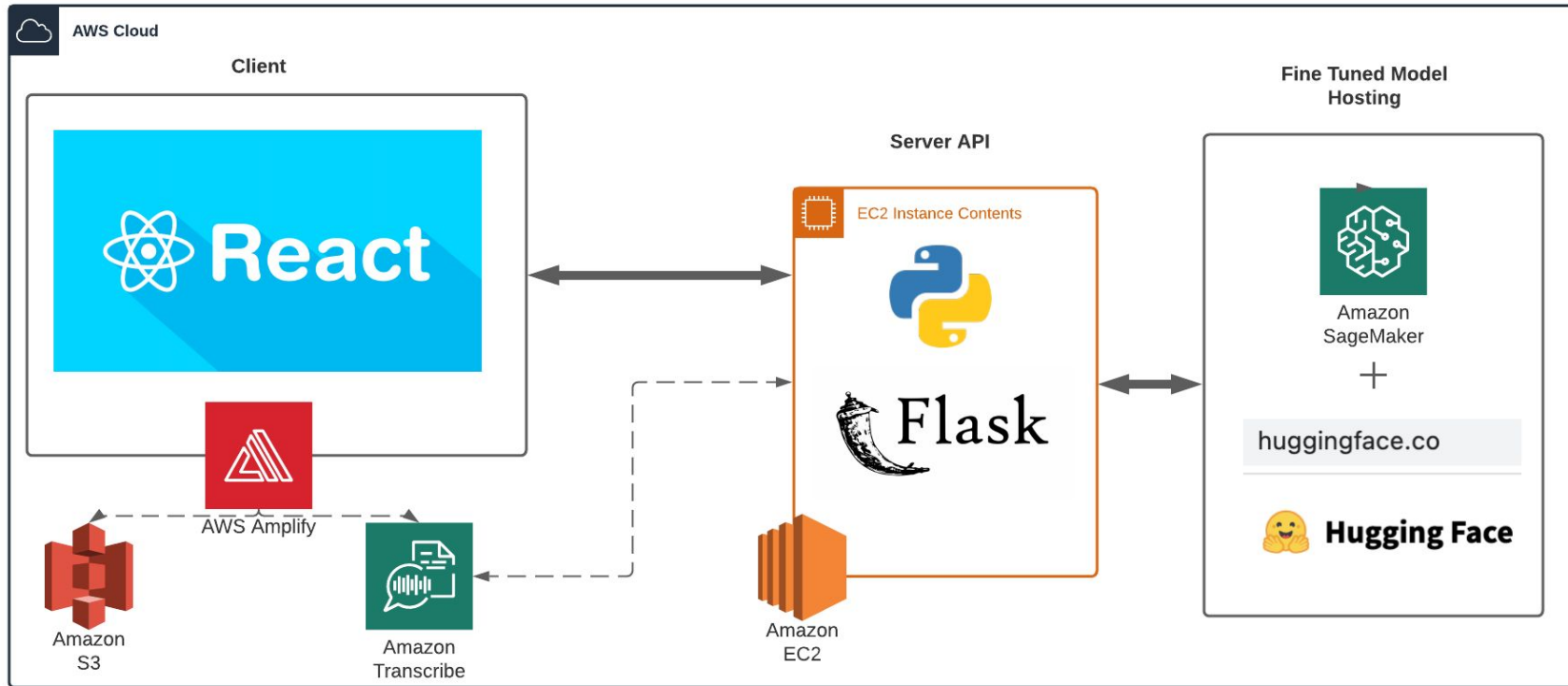
PIPELINE OF THE PROJECT



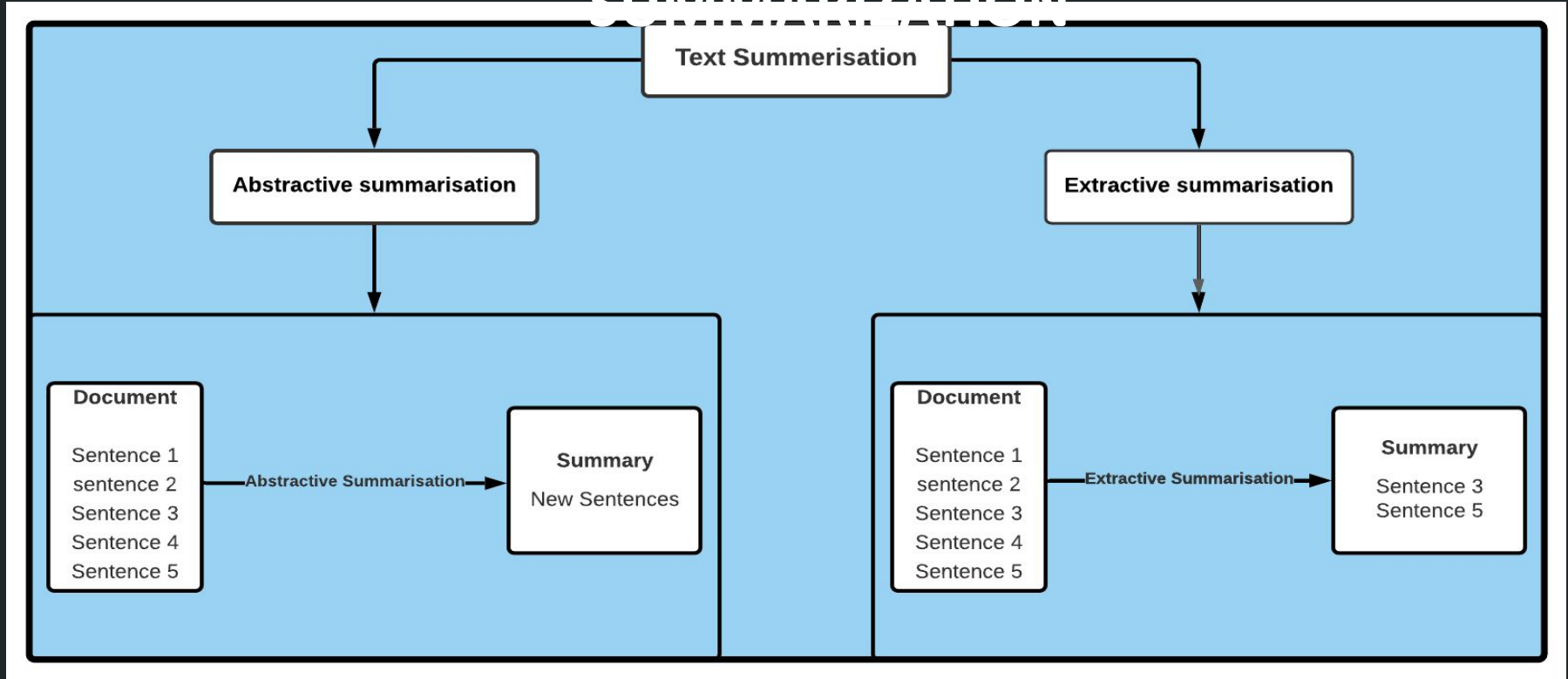
Two phases in the pipeline:

- Input- take input as audio or text file, convert Speech to text, display input as text.
- Output- Text-to-Text Summarization- Summary.

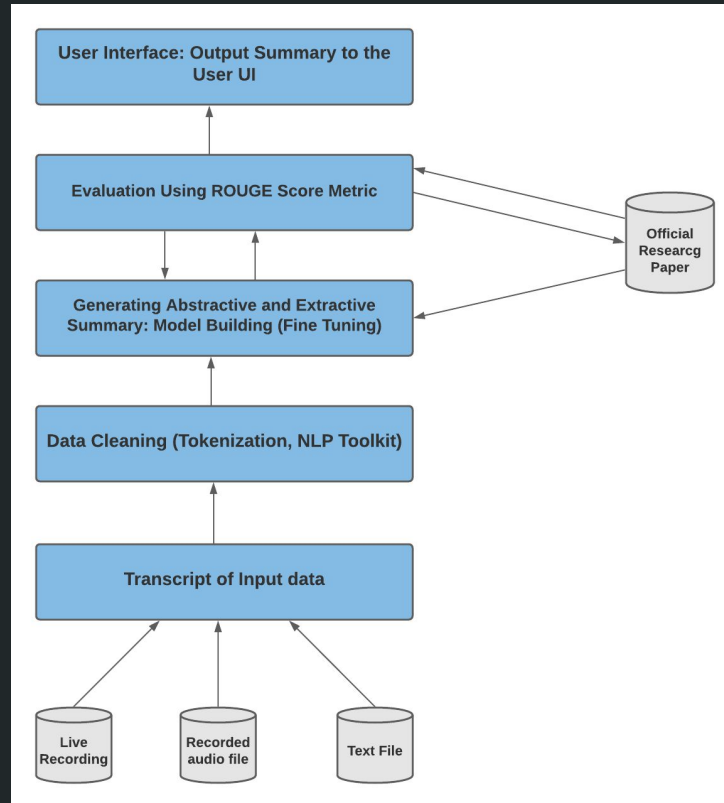
COMPONENTS



METHODS USED FOR TEXT SUMMARIZATION



HIGH-LEVEL ARCHITECTURE DESIGN AND DATA FLOW DIAGRAM



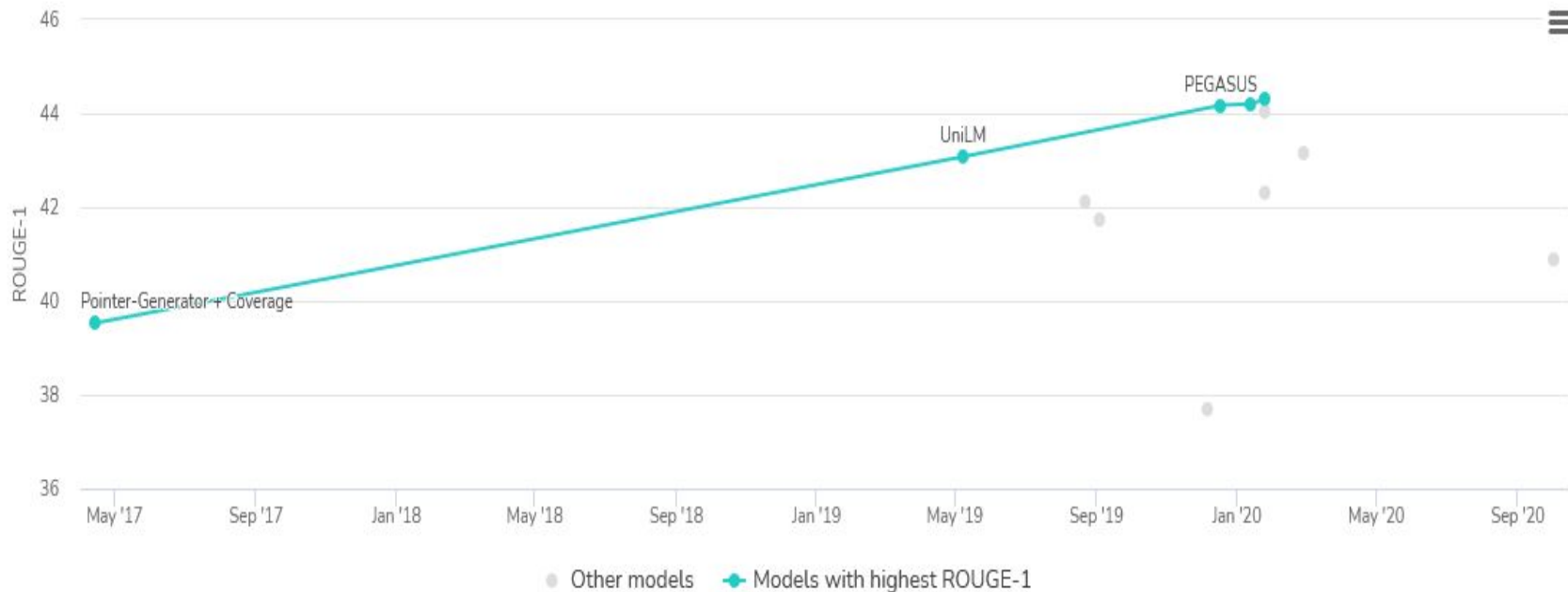
WHY PEGASUS ?

- Google's PEGASUS further improved the state-of-the-art (SOTA) results for abstractive summarization in particular with less resources.
- Uses Encoder-decoder model for sequence to sequence learning.
- Gap sentence Generation
- Masked Language Model
- ROUGE Score: to select important sentence.

LEADERBOARD FOR ABSTRACTIVE SUMMARIZATION

Leaderboard

Dataset



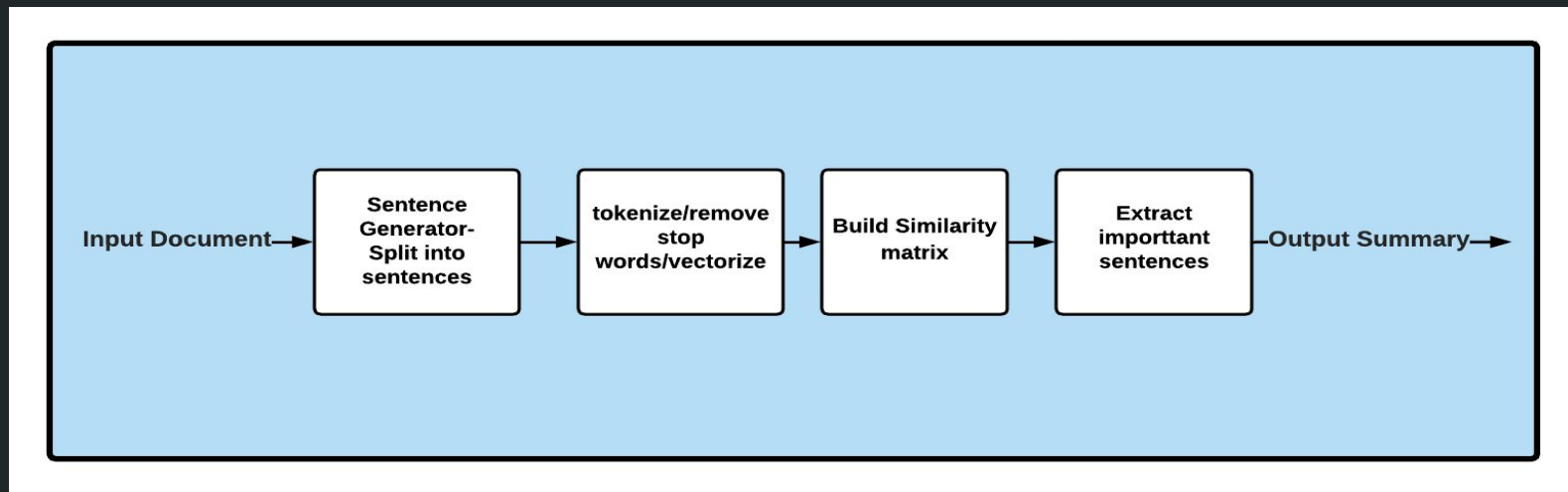
Implementation Insights

- Unsupervised Extractive Summarization

Input article → split into sentences → remove stop words → build a similarity matrix → generate rank based on matrix → pick top N sentences for summary.

Implementation Insights

- Unsupervised Extractive Summarization



Implementation Insights

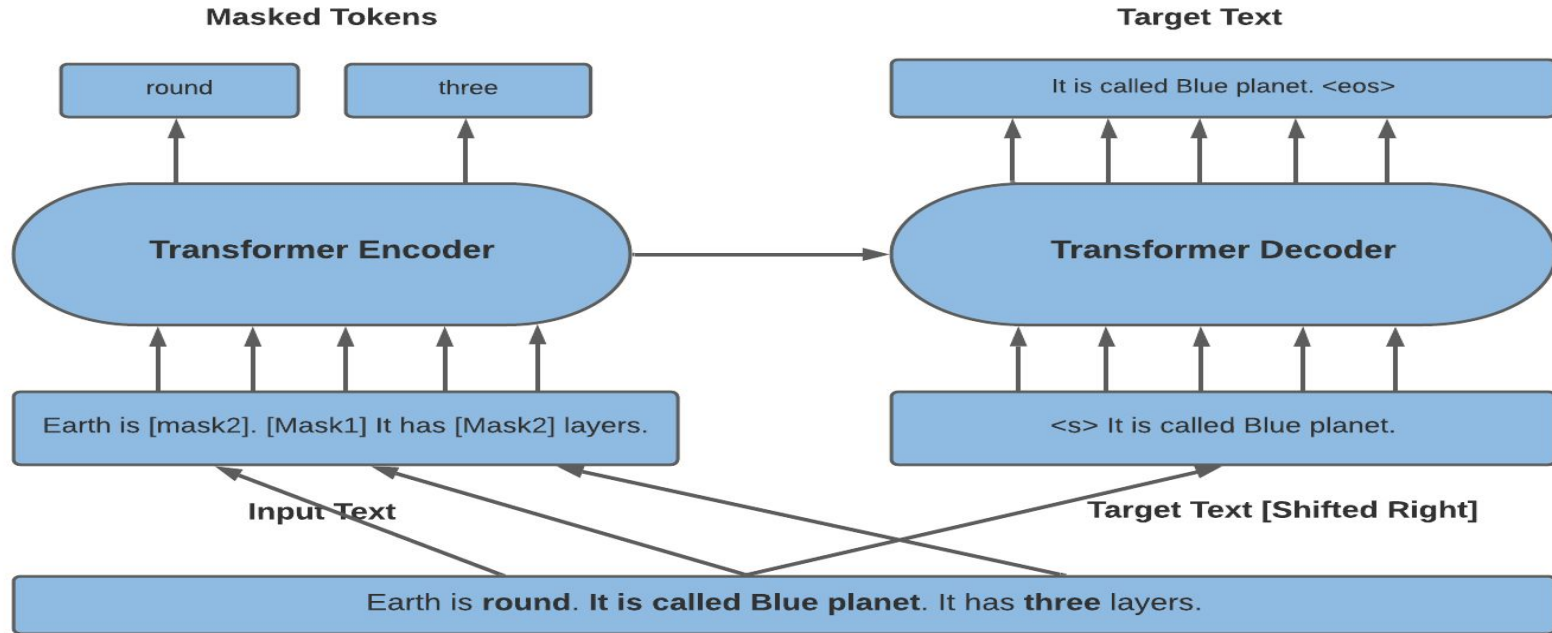
- Abstractive Summarization

Fine-tuned CUSTOM_MODEL_NAME = 'arawat/pegasus-custom-xsum'

Pretrained models-

1. google/pegasus-billsum
2. google/pegasus-cnn_dailymail
3. google/pegasus-multi_news

PEGASUS ARCHITECTURE



FINE-TUNE MODEL

- Custom- fine-tuned google/pegasus-large model on XSUM dataset
- Deployed on Huggingface Library/AWS Sagemaker

FINE-TUNE MODEL

The screenshot shows the Hugging Face interface for the model `arawat/pegasus-custom-xsum`. The page includes a navigation bar with links to Models, Datasets, Spaces, Resources, Solutions, and Pricing. The model's name is displayed with a copy icon and a like count of 0. Below the name are tags for Text2Text Generation, PyTorch, Transformers, pegasus, seq2seq, generated_from_trainer, model-index, and AutoNLP Compatible. A tabbed interface shows the Model card, Files and versions, and Settings. The Model card tab is active, displaying the results section which states the model is a fine-tuned version of `google/pegasus-large` on an unknown dataset. The model description section is also visible. On the right, there is a 'Train' button with a dropdown, a 'Deploy' button, and a 'Use in Transformers' button. A 'NEW' badge is present next to a note about selecting AutoNLP in the 'Train' menu. A bar chart shows 4 downloads last month. A 'Hosted inference API' section is also visible.

← → ↻ <https://huggingface.co/arawat/pegasus-custom-xsum> ☆ 🛡️ ⬇️ ☰

Hugging Face 🔍 Models Datasets Spaces Resources Solutions Pricing

arawat/pegasus-custom-xsum 📄 👍 like 0

Text2Text Generation PyTorch Transformers pegasus seq2seq generated_from_trainer model-index

AutoNLP Compatible

Model card Files and versions Settings Train Deploy Use in Transformers

Edit model card

results

This model is a fine-tuned version of `google/pegasus-large` on an unknown dataset.

Model description

NEW Select `AutoNLP` in the “Train” menu to fine-tune this model automatically.

Downloads last month
4

Hosted inference API ⓘ

Text2Text Generation

FINE-TUNE MODEL

AWS

Services

Search for services, features, blogs, docs, and more

[Option+5]

Console Home & Tag Editor

Amazon SageMaker

Dashboard

Search

SageMaker Domain

- Studio
- RStudio

Images

- Ground Truth
- Notebook
- Processing
- Training
- Inference
 - Compilation jobs
 - Model packages
 - Models
- Endpoint configurations
- Endpoints
- Batch transform jobs

Edge Manager

Augmented AI

AWS Marketplace

Amazon SageMaker > Models > huggingface-pytorch-inference-2021-11-29-21-55-51-399

huggingface-pytorch-inference-2021-11-29-21-55-51-399

Actions Create batch transform job Create endpoint

Model settings

Name	ARN	Creation time	IAM role ARN
huggingface-pytorch-inference-2021-11-29-21-55-51-399	arn:aws:sagemaker:us-east-1:559976405953:model/huggingface-pytorch-inference-2021-11-29-21-55-51-399	Nov 29, 2021 21:55 UTC	arn:aws:iam::559976405953:role/service-role/AmazonSageMaker-ExecutionRole-20211129T120132

Container 1

Container Name	Model data location
Container 1	-
Image	Mode
763104351884.dkr.ecr.us-east-1.amazonaws.com/huggingface-pytorch-inference:1.7.1-transformers4.6.1-cpu-py36-ubuntu18.04	Single model
Training job	
-	
Environment variables	
Key	Value
HF_MODEL_ID	arawat/pegasus-custom-xsum
HF_TASK	summarization
SAGEMAKER_CONTAINER_LOG_LEVEL	20
SAGEMAKER_REGION	us-east-1

Network

Enable network isolation
False

Project-Demo

Text Summarisation

Audio File Input <input type="button" value="Browse..."/> No file selected. Upload your audio file	Text Input <input type="button" value="Browse..."/> msft.txt Upload your text file	Voice Input <input type="button" value="Start recording"/> Press 'start recording' to begin your transcription. Press STOP recording once you finish speaking.
Input Text In an attempt to build an AI-ready workforce, Microsoft announced Intelligent Cloud Hub which has been launched to empower the next generation of students with AI-ready skills. Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services. As part of the program, the Redmond giant which wants to expand its reach and is planning to build a strong developer ecosystem in India with the program will set up the core AI infrastructure and IoT Hub for the selected campuses. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset. This will require more collaborations and training and working with AI. That's why it has become more critical than ever for educational institutions to integrate new cloud and AI technologies." The program aims to build up the cognitive skills and in-depth understanding of developing intelligent cloud connected solutions for applications across industry. Earlier in April this year, the company announced Microsoft Professional Program in AI as a learning track open to the public. The program was developed to provide job ready skills to programmers who wanted to hone their skills in AI and data science with a series of online courses which featured hands-on labs and expert instructors as well. This program also included developer-focused AI school that provided a bunch of assets to help build AI skills.		
Generate Extractive Summary	This program also included developer-focused AI school that provided a bunch of assets to help build AI skills. Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset. Earlier in April this year, the company announced Microsoft Professional Program in AI as a learning track open to the public. As part of the program, the Redmond giant which wants to expand its reach and is planning to build a strong developer ecosystem in India with the program will set up the core AI infrastructure and IoT Hub for the selected campuses	<input type="button" value="Show Rogue Score Board"/> <input type="button" value="Print Score Card"/> Extractive/Model Rogue =====> Value google/pegasus-bllsum =====> 0.3103243697478992 google/pegasus-cnn_dailymail =====> 0.1818181818181818 google/pegasus-multi_news =====> 0.4129076923076923 arawat/pegasus-custom-xsum =====> 0.5104602510460251 Best Score Model Name: arawat/pegasus-custom-xsum Score: 0.5104602510460251
Generate 'google/pegasus-bllsum' Abstractive Summary	In an attempt to build an AI-ready workforce, Microsoft announced Intelligent Hub which has been launched to the next generation of students with AI-ready skills as a three-year collaborative program, Intelligent Hub will support around 100 institutions with AI infrastructure, course content and curriculum developer support, development tools, and give students access to cloud and AI services.	
Generate 'google/pegasus-cnn_dailymail' Abstractive Summary	Microsoft Intelligent Cloud Hub will support around 100 institutions. -n=Program will provide AI infrastructure, course content and curriculum -n=Aims to build up cognitive skills and understand developing AI solutions.	
Generate 'google/pegasus-multi_news' Abstractive Summary	- "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset," Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, says in a press release. "That's why it has become more critical than ever for educational institutions to integrate new cloud and AI technologies." To that end, Microsoft has announced Intelligent Cloud Hub, a three-year collaborative program that will support around 100 institutions in India with AI infrastructure, course content, and curriculum, developer support, development tools, and give students access to cloud and AI services, the Verge reports. Microsoft will also provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services, and Azure Machine Learning. The Intelligent Cloud Hub is part of Microsoft's Professional Program in AI, which is open to the public.	
Generate 'arawat/pegasus-custom-xsum' Abstractive Summary	The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset.	

Any Questions...



Thank you!
