Introductory Course on Transformer Models and Generative Artificial Intelligence Tools

➢ Popular Generative AI Tools - Application-wise

- Text-to-Text (T2T)
- Audio-to-Text (A2T)
- Text-to-Audio (T2A)
- Text-to-Image (T2I)
- Image-to-Text (I2T)
- Text-to-Video (T2V)
- Text-to-Motion (T2M)
- Text-to-Code (T2C)
- Brain-to-Text (B2T)
- Audio-to-Audio (A2A)

Text-to-Text (T2T)

- T2T Tools are NLP (Natural Language Processing) models that are designed to convert one type of text into another
- These tools are trained on large amounts of text data and can be fine-tuned for specific T2T task

Tasks T2T tools can perform:

- Text Summarization
- Text Generation
- Text Translation
- Text-to-speech
- Text-to-code

Text-to-Text (T2T)

One of the most advanced T2T Tools is OpenAI GPT-3 which can perform a wide range of T2T tasks with minimal fine-tuning

Other popular T2T tools include:

- Google Translate- Cloud-based machine translation service, T5, a text-to-text transfer model developed by Google AI
- BERT, a pre-trained language model developed by Google AI
- Seq2Seq, a neural network architecture for T2T tasks, and TTS (Text-to-Speech) engines like Google Text-to-Speech
- Amazon Polly
- IBM Watson Text-to-Speech

Some more popular tools:

Audio-to-Text (A2T)

- A2T tools are NLP models that are designed to convert speech or other audio inputs into written text
- These tools are trained on large amounts of audio data

Speech recognition

Speech-to-text

Transcription

Voice commands

Audio captioning

Audio-to-Text (A2T)

Popular T2A tools include:

- Google Speech-to-Text
- Amazon Transcribe
- Microsoft Azure Speech Services

Some more popular tools:

descript
AssemblyAI
Whisper
Otter
Temi

Text-to-Audio (T2A)

- **What are T2A Tools**
  - T2A tools are NLP models that are designed to convert written text into spoken words. These tools are also known as Text-to-Speech (TTS) systems. They are trained on large amounts of text data and can be fine-tuned for specific T2A tasks such as voice synthesis, voice-enabled applications, and speech-enabled devices.

Voice synthesis

Voice-enabled applications

Speech-enabled devices

Speech-enabled devices

E-book reading

Audio captioning

Text-to-Audio (T2A)

Popular T2A tools include:

- Google Text-to-Speech
- Amazon Polly
- IBM Watson Text-to-Speech
- Microsoft Azure Speech Services

Some more popular tools:

Play.ht MURF.AI RESEMBLE.AI

WELLSAID descript Aflorithmic

Synthesys MURF.AI

LOVO

Text-to-Image (T2I)

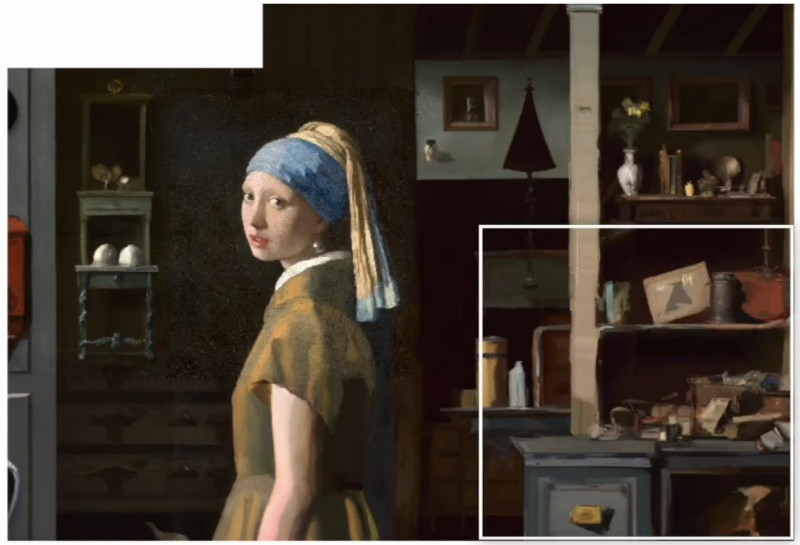
- T2I is an NLP task to generate images from a given text description.
- This task can be used to generate images of objects, scenes, or even abstract concepts that are described in natural language

Image Synthesis	Data Augmentation	Audio Captioning
By Generating New Images From Text Descriptions T2I Can Be Used To Increase The Size Of The Training Dataset And Improve The Performance Of Image Recognition Models	Advertising And E-commerce	Medical Imaging

Text-to-Image (T2I)

Popular T2I tools include:

- DALL-E, a pre-trained T2I model that can generate images from textual descriptions
- GPT-3, a pre-trained model that can generate images of objects, scenes, and abstract concepts
- Google's Show and Tell, a T2I model that can generate images of objects, scenes, and people



DALL·E 2 · Stable Diffusion · craiyon · Lexica · Midjourney · Imagen · WOMBO · NightCafé · GauGAN2 · DeepAI · Jasper · artbreeder · Wonder · pixray-text2image · neural.love · Omneky · alpaca · mage.space · KREA · Nyx.gallery · ROSEBUD.AI · PhotoRoom

Image-to-Text (I2T)

What are I2T Tools

- I2T is a challenging task that involves generating text descriptions from images
- I2T is an NLP task that involves generating a text description from an image

Generate captions

Label

Stories

Image-to-Text (I2T)

- Combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)
- Extraction of the feature representations of an image using a CNN; and then use feature representations as input to an RNN
- Using attention mechanisms
- I2T tools automate the task of deciphering the image to describe them in natural sentences, improving workflow and efficiency
- Learn continuously from past experiences and adapt to changes- making it the most suitable for generating meaningful captions eventually

Popular I2T tools include:

neural .love

OpenAI GPT-3 x Image Captions

srijan:

Text-to-Video (T2V)

T2V is an NLP task that involves generating a video from a given text description. This task can be used to generate videos of objects, scenes, or even abstract concepts that are described in natural language. There are several techniques and approaches that have been proposed to tackle the T2V problem.

- **Some possible use cases of T2V includes:**
- AI video creator uses advanced natural language processing (NLP) and machine learning algorithms to create high-quality videos from text in over 50 languages
- AI text-to-video generator can analyze the content of your blog posts, news articles, and web pages to generate relevant and engaging videos
- Tool leverages AI to generate natural-looking lip movement based on the words of your text
- Uses AI to find musical visuals that will further enhance your message

Text-to-Video (T2V)

Popular T2V tools include

Video-GPT and DALL-E 4

Fliki synthesia Meta AI Google AI Phenaki CONTENDA

runway WOMBO descript

OPUS >ROSEBUD.AI

Text-to-Motion (T2M)

- Text-to-Motion (T2M) is an NLP task that involves generating animation or motion from a given text description. This task can be used to generate animations of objects, characters, or even abstract concepts that are described in natural language.

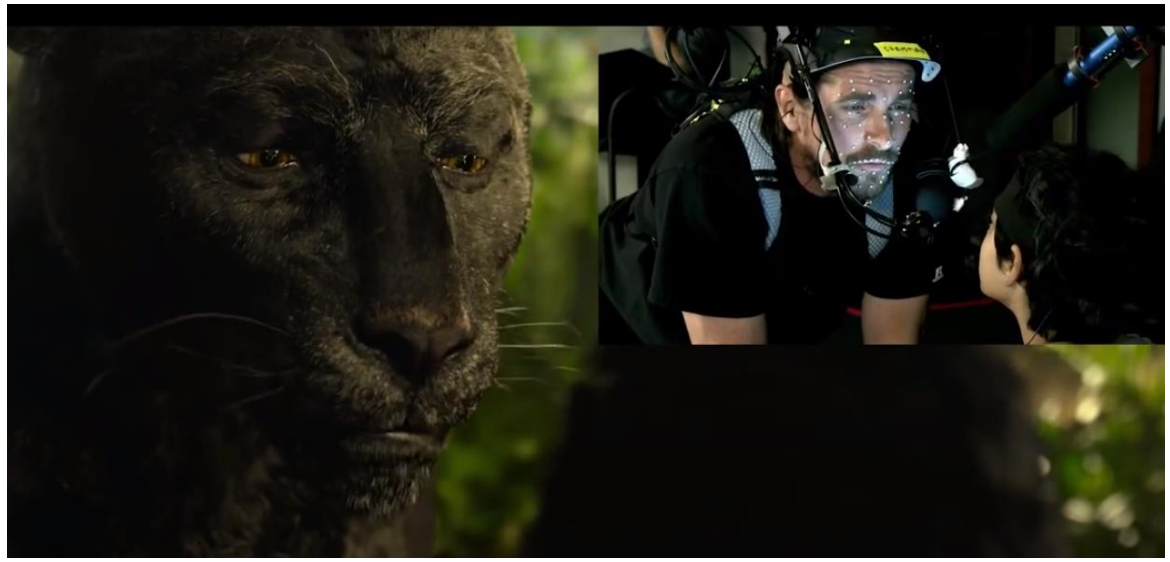
Motion Capture	Motion Synthesis	Motion Analysis	Interpret Complex Text

Popular T2M tools include:

TREE Ind. MDM: Human Motion Diffusion Model

Approach:
One approach to T2M is to use a combination of deep learning and computer animation techniques to generate animations from text descriptions

Use motion capture data, where an actor's movements are recorded and then used to animate a digital character. The text input can be used to guide the selection of the appropriate motion capture data



Text-to-Code (T2C)

- T2C is an NLP task that involves generating computer programming codes from a given text description
- T2C is also known as "code generation" or "code synthesis"

T2C tools can perform:

write code in at least a dozen languages, including JavaScript, Go, Perl, PHP, Ruby, Swift and TypeScript, and even BASH. The model is trained on billions of lines of code available in the public domain

CodeT5 is an open source programming language model that can potentially bring three capabilities to software programming:

- **Text-to-code generation**
- **Code autocompletion**
- **Code summarization**

Text-to-Code (T2C)

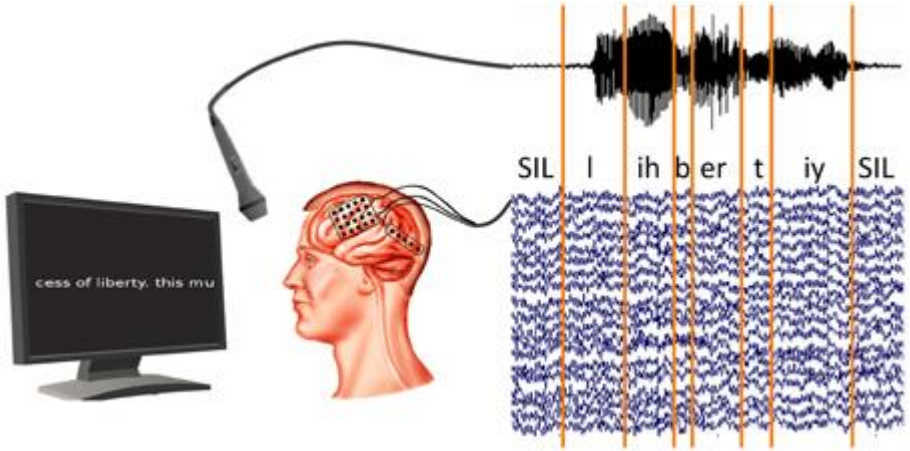
Popular T2C tools include:

There are also several pre-trained models like GPT-3 and GPT-4 which can be fine-tuned for T2C task.

Brain-to-Text (B2T)

- Brain-to-Text is a Generative AI tool that can decode brain signals into text
- B2T can be used for mind-reading, thought recognition, and emotion detection
- B2T is limited in its ability to accurately interpret brain signals

- B2T is a type of NLP task that involves transcribing thoughts or brain signals into text. This task is also known as "brain-computer interface" or "brain-machine interface".
- One approach to B2T is to use electroencephalography (EEG) to record brain activity, and then use machine learning algorithms to analyze the EEG data and transcribe it into text.



Popular B2T tools include:

Meta AI speech from brain — **non-invasive brain recordings**

Audio-to-Audio (A2A)

A2A is an NLP that involves transforming one type of audio signal into another. This can include tasks such as speech synthesis, speech enhancement, speech recognition, and audio style transfer. Each task addresses a specific need and can be used for a variety of applications in areas such as accessibility, virtual assistants and entertainment

A2A can used for:

Speech-enabled virtual assistants and accessibility for the visually impaired

Speech enhancement, which involves improving the quality of a noisy or degraded speech signal

Popular A2A tools include:

AudioLM

VOICEMOD

Synthetic Voice

SONANTIC, VOICEMOD, PolyAI, REPLICA, coqui, WELLSAID, readspeaker.ai, Fliki, podcast.ai, Listnr, RESEMBLE.AI, Dubverse

Music

soundful, USICO, SONIFY, Mubert, Aiva, SPLASH, Harmonai, Endel, INFINITE ALBUM, boomy

Thank You

For discussions/suggestions/queries email: ***isuw@isuw.in***

www.isuw.in

Links/References (If any)