

Deep Learning : Transformer Model



राष्ट्रीय प्रौद्योगिकी संस्थान सिक्किम
NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM

Course Instructor:
Dr. Bam Bahadur Sinha
Assistant Professor
Computer Science & Engineering
National Institute of Technology
Sikkim

Limitations of RNN

RNN-Generated Poem

The sun is shining bright,
In the sky so high, it takes flight.
The birds are singing, the trees stand tall,
But something seems off, no sense at all.

The moon is rising, the stars do glow,
But the day is here, it doesn't show.
The grass is green, the water is blue,
Yet everything feels old, and somehow new.

RNN-Generated Output - Repetition

The cat is on the table. The cat is on the table. The cat sat on table.

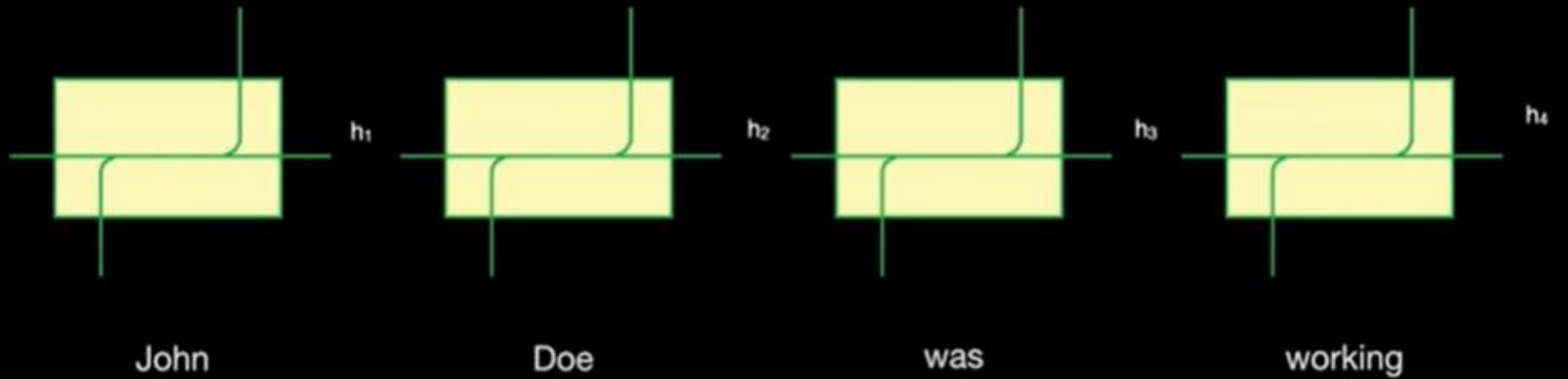
RNN-Generated Output - Inappropriate

The boy went to the store to buy some apples. He paid the cashier.
The sun was shining brightly, and he ate the cake at the park.

RNN-Generated Output - Grammatically incorrect

he group of people is walking to the park. They was excited about the picnic.

RNN Problems

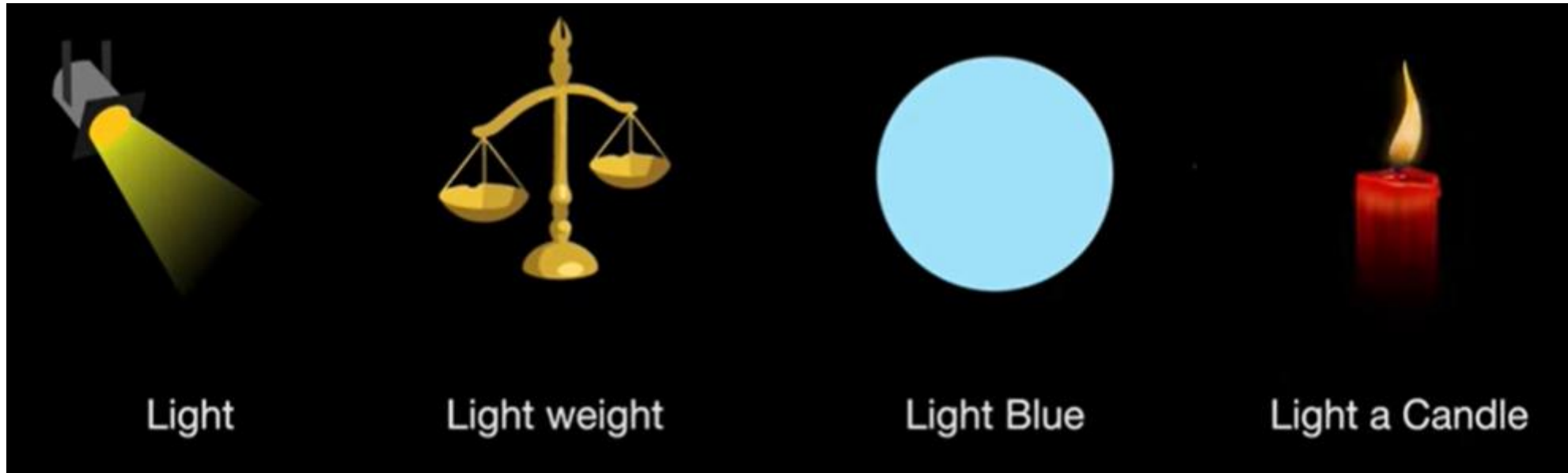


RNN Problems



When Jake first laid eyes on the Apple iPhone, he wasn't expecting much. It felt like holding the future. His old phone had served him well, and he wasn't one to chase the latest tech trends. But the moment he held the sleek, glassy device in his hand, there was something fundamentally different about it.

He pressed the home button, and the phone lit up with a vibrant, colorful display, unlike anything he had seen before. The icons were sharp, and the touchscreen was so fluid that he couldn't resist swiping through the screens just to watch them glide by. The polished metal edges and smooth, responsive screen drew him in.



Different Context

“I love Apple phones”



isTech	0.2
canCall	0.1
isSmart	0.19
.	.
.	.
.	.
isFruit	0.3
isFood	0.18

Apple

“I love Apple Juice”



Word Embeddings

isTech	0.01
canCall	0.005
isSmart	0.005
.	.
.	.
isFruit	0.75
isFood	0.20

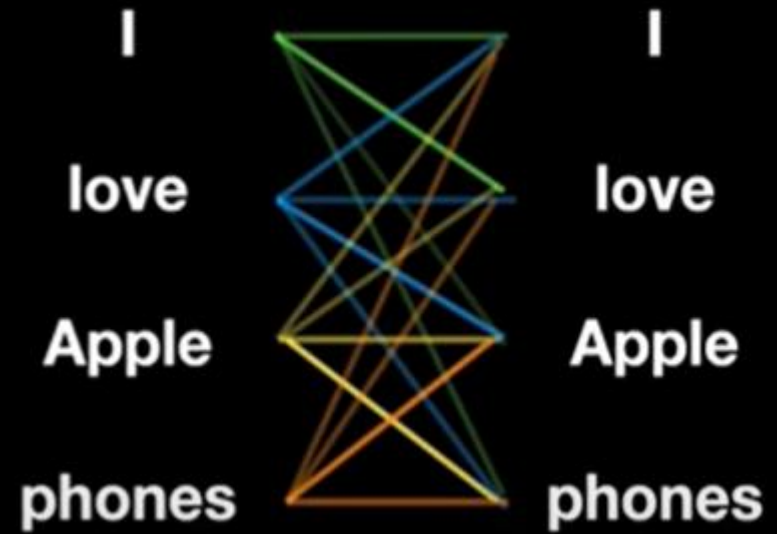
Apple

“I love Apple Juice”

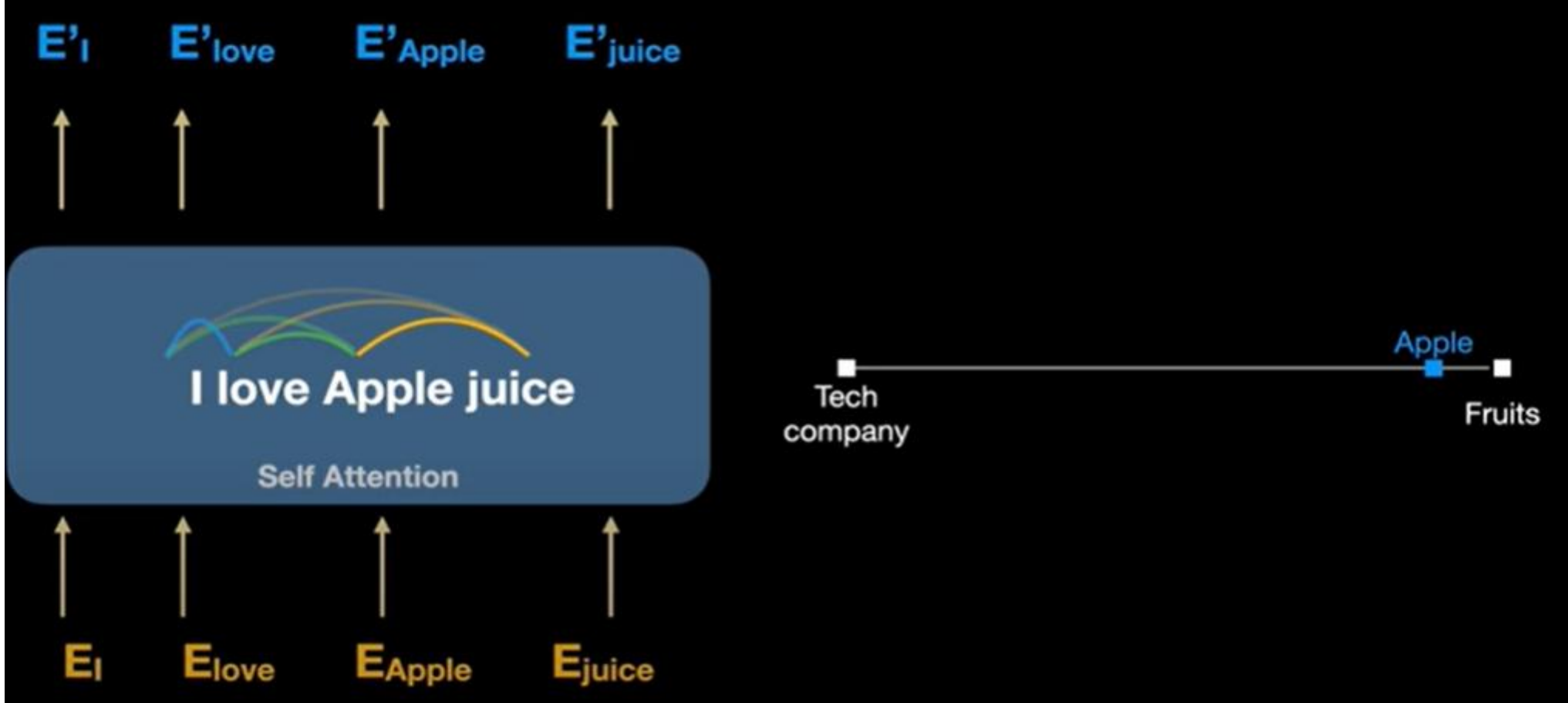


Dynamic Word Embeddings

Self Attention



Self Attention



Self Attention

Generate an image of a **leafy cloud**



Cloud

+



Leafy

=



Leafy cloud

Power of Transformers

She saw the man with the telescope

1. She saw the man who was holding the telescope



2. She used a telescope to see the man

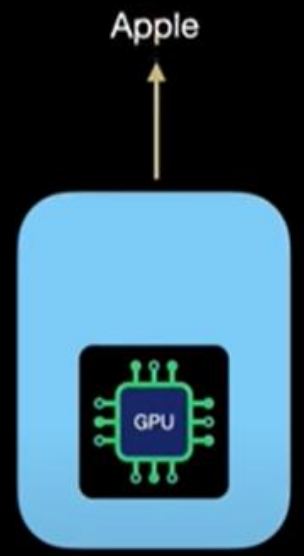
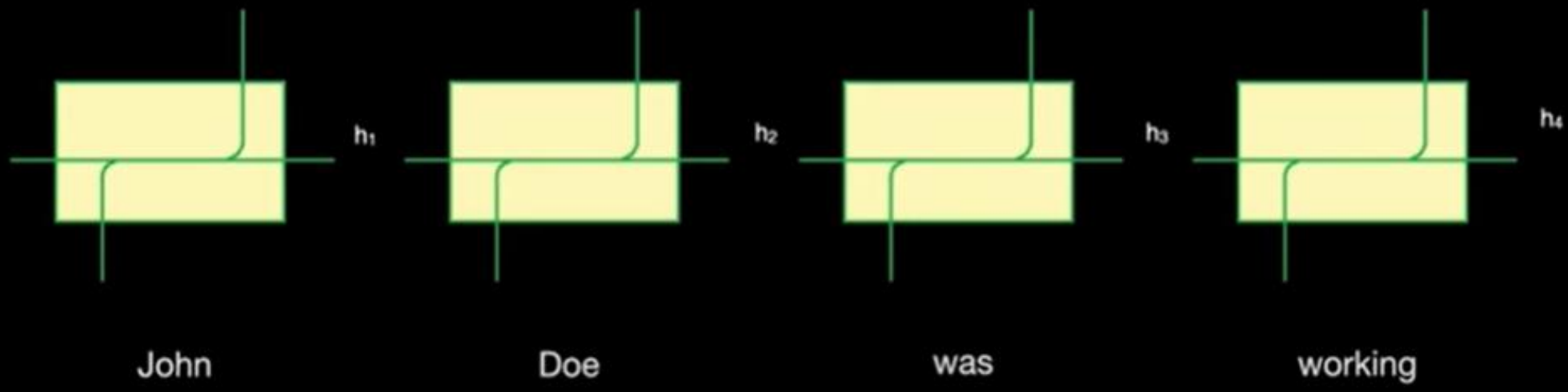


She saw the man with the telescope

She saw the man with the telescope

Power of Transformers (Multi-headed Attention)

RNN

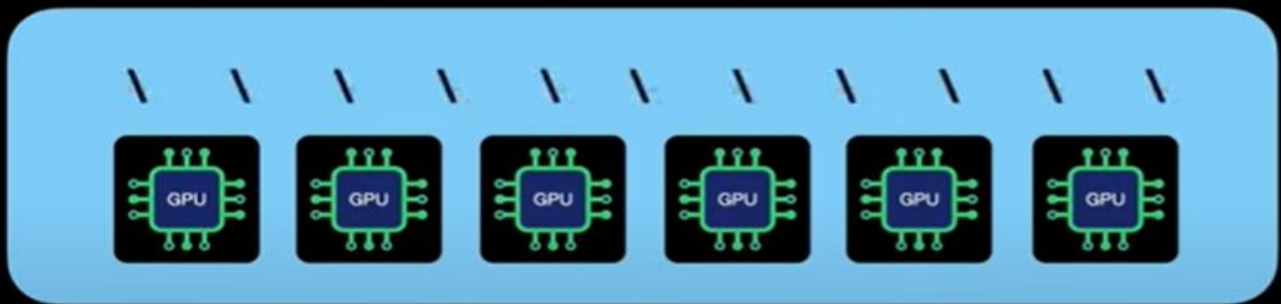


Transformer

E'_I E'_{love} E'_{Apple} E'_{juice}



When Jake first laid eyes on the Apple iPhone, he wasn't expecting much. His old phone....



When Jake first laid eyes on the Apple iPhone, he wasn't expecting much. His old phone....

Scale of Transformers

Transfer Learning



Pre-training
CNN

Flower

Fine Tuning

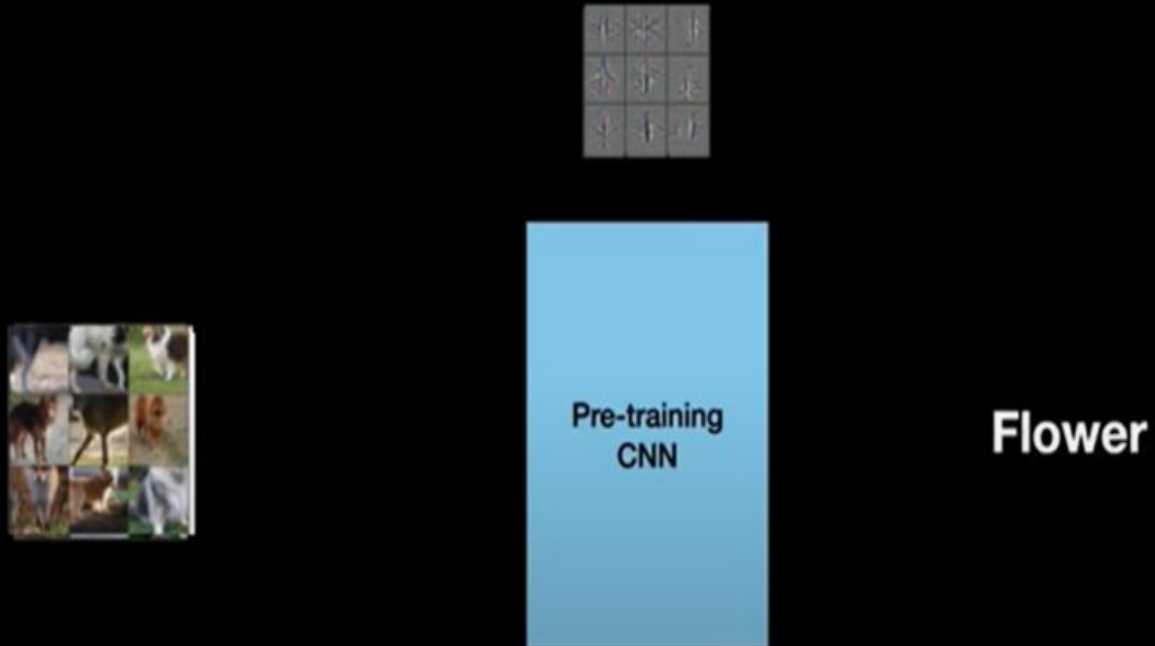


Cat or Dog

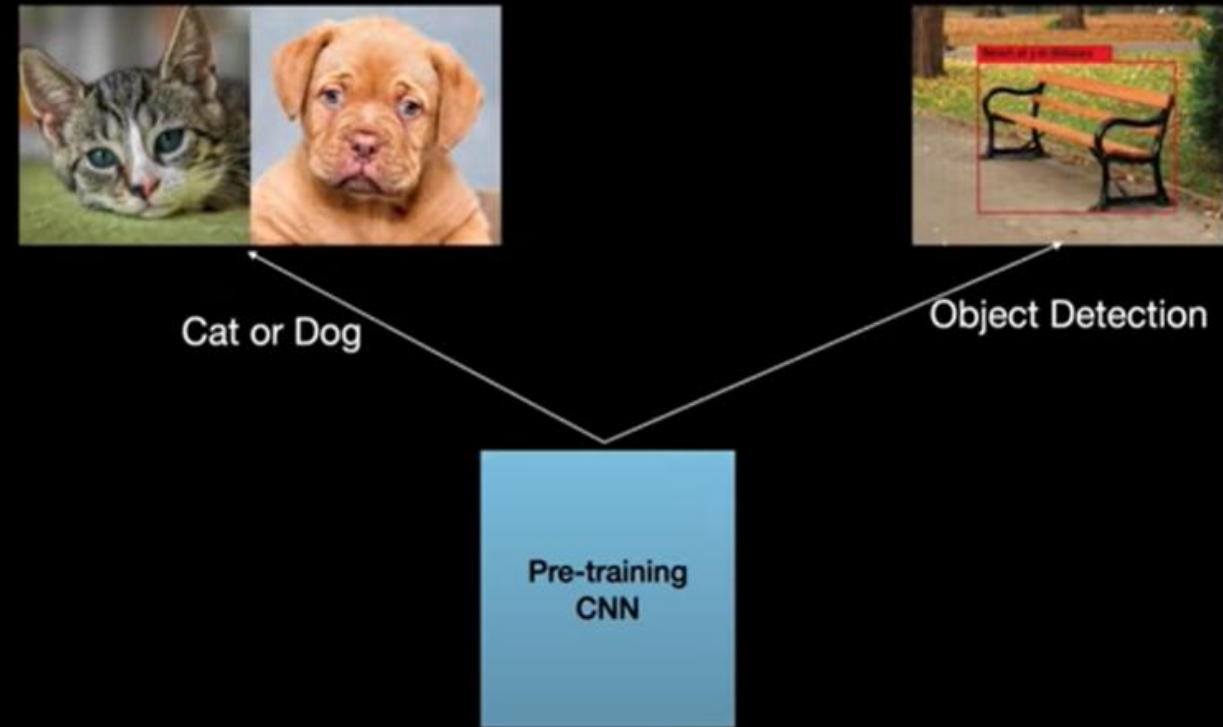
Pre-training
CNN

Transfer Learning

Transfer Learning



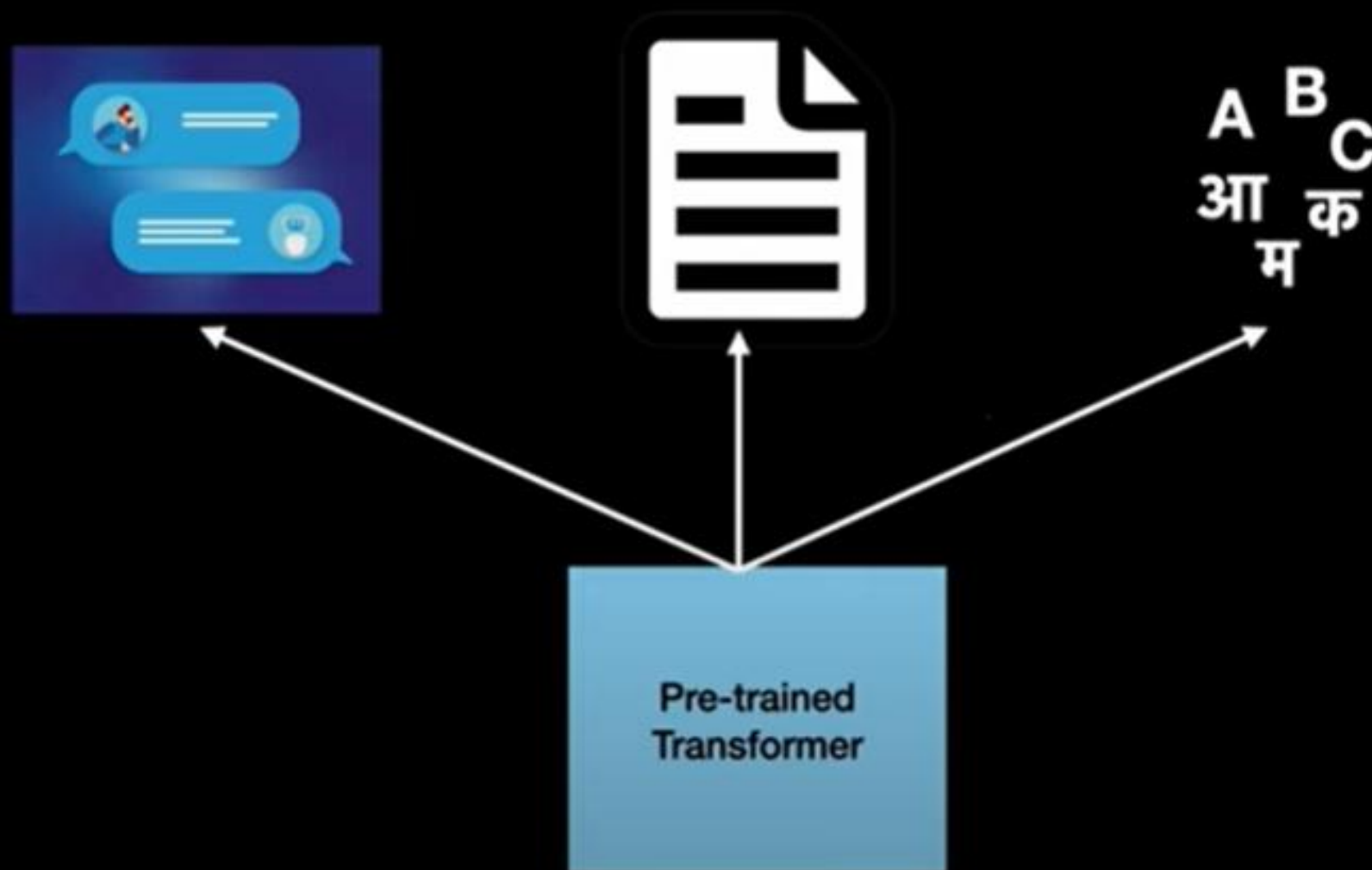
Fine Tuning




Transfer Learning

Pretrained transformer and Fine-Tuning


Fine-tuning Transformer



 **Hugging Face**

Search models, datasets, users...

ModelsDatasetsSpacesPostsDocsEnterprisePricingLog InSign Up



The AI community building the future.

The platform where the machine learning community collaborates on models, datasets, and applications.

Explore AI Appsor Browse 1M+ models

TasksLibrariesDatasetsLanguagesLicensesOther

Filter Tasks by name

Multimodal

- Text-to-ImageImage-to-Text
- Text-to-VideoVisual Question Answering
- Document Question AnsweringGraph Machine Learning

Computer Vision

- Depth EstimationImage Classification
- Object DetectionImage Segmentation
- Image-to-ImageUnconditional Image Generation
- Video ClassificationZero-Shot Image Classification

Natural Language Processing

- Text ClassificationToken Classification
- Table Question AnsweringQuestion Answering
- Zero-Shot ClassificationTranslation
- SummarizationConversational
- Text GenerationText2Text Generation
- Sentence Similarity

Audio

- Text-to-SpeechAutomatic Speech Recognition
- Audio-to-AudioAudio Classification
- Voice Activity Detection

Models 469,541Filter by name

meta-llama/Llama-2-70b

- Text Generation • Updated 4 days ago • 25.2k • 64

stabilityai/stable-diffusion-xl-base-0.9

- Updated 6 days ago • 2.01k • 393

openchat/openchat

- Text Generation • Updated 2 days ago • 1.3k • 136

lillyasviel/ControlNet-v1-1

- Updated Apr 26 • 1.87k

cerspense/zeroscope_v2_XL

- Updated 3 days ago • 2.66k • 334

meta-llama/Llama-2-13b

- Text Generation • Updated 4 days ago • 328 • 64

tiiuae/falcon-40b-instruct

- Text Generation • Updated 27 days ago • 288k • 899

WizardLM/WizardCoder-15B-V1.0

- Text Generation • Updated 3 days ago • 12.5k • 332

CompVis/stable-diffusion-v1-4

- Text-to-Image • Updated about 17 hours ago • 448k • 5.72k

<https://huggingface.co/>

Multi-modality

Text



Image



Audio



Video

Multi-modality



Find me a similar
chair online

Transformer

Sure, here are
Some results

VLM (Vision Language Models)

3D render of a cute tropical fish in an aquarium on
a dark blue background, digital art



Open AI – DALL.E
(<https://labs.openai.com/>)