



with tensorflow

구현을 위한 딥러닝

■ 고려대학교 물리학과 한승희

■ mod96@naver.com

Contents

1

Introduction

2

Deep Learning Models

3

Tensorflow vs PyTorch

4

H.W.

Contents

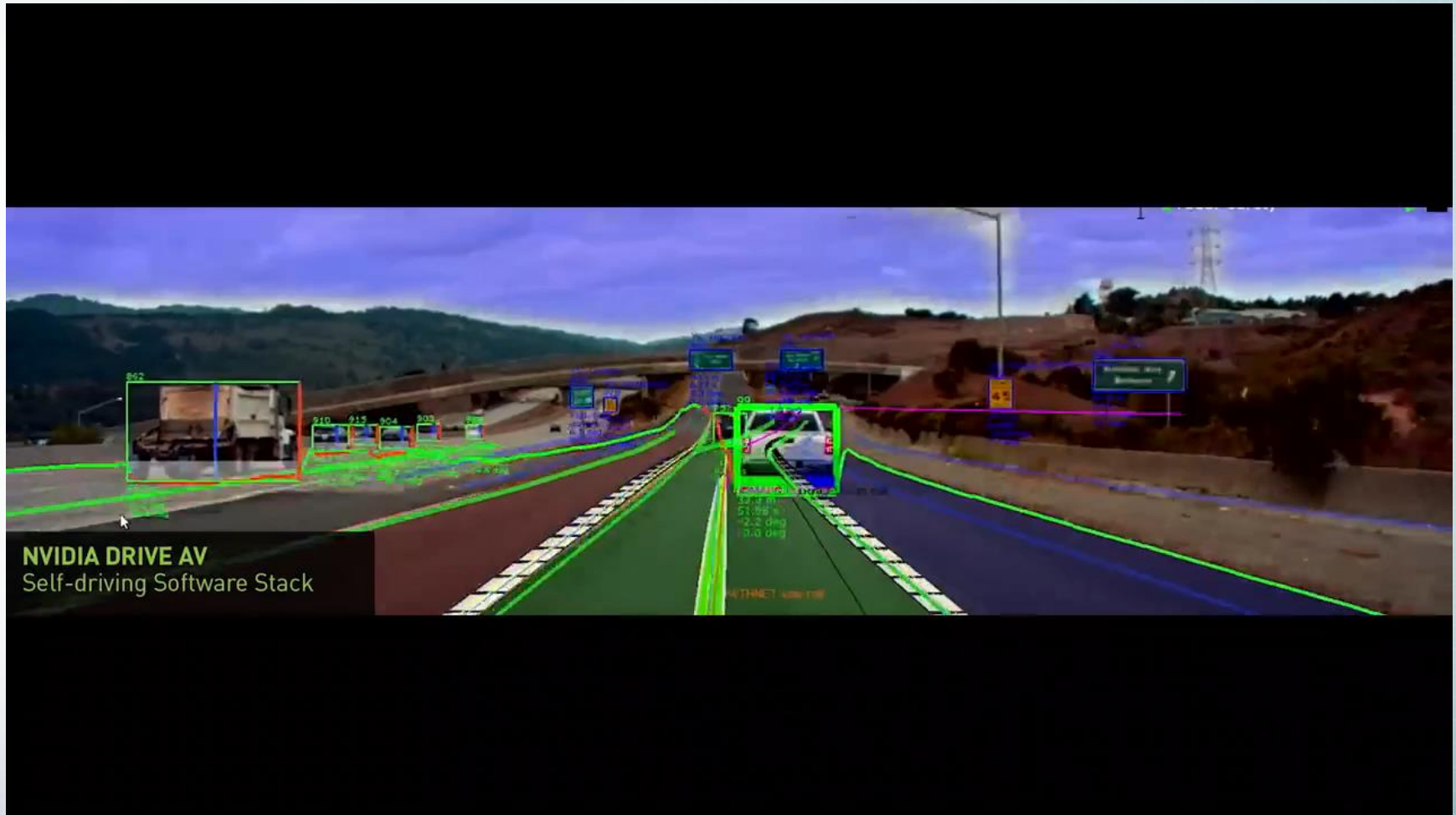
1

Introduction

Self Driving Car



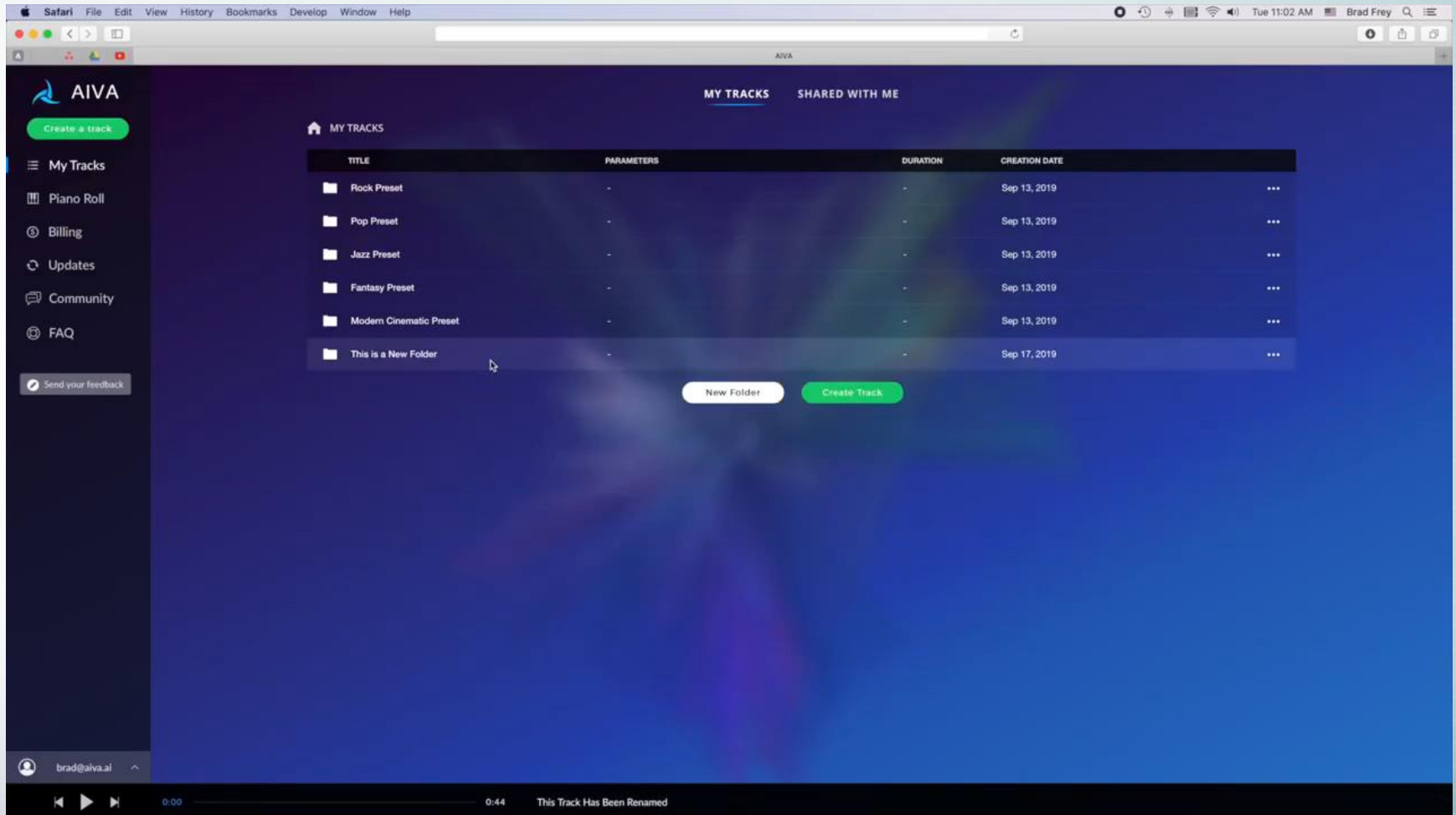
Self Driving Car



Amazon Prime Air



AIVA (AI Virtual Artist)



AI Speech (2018)

Audio samples from "Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis"

Paper: [arXiv](#)

Authors: Ye Jia *, Yu Zhang *, Ron J. Weiss *, Quan Wang, Jonathan Shen, Fei Ren, Zhifeng Chen, Patrick Nguyen, Ruoming Pang, Ignacio Lopez Moreno, Yonghui Wu. (*: equal contribution.)

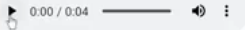
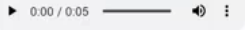
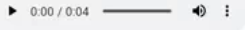
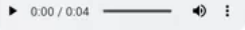
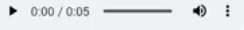
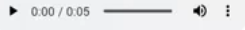
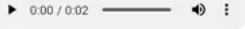
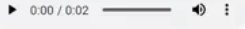
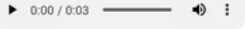
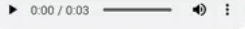
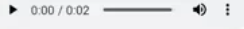
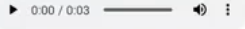
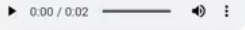
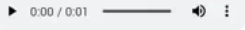
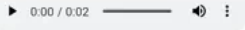
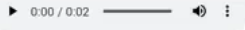
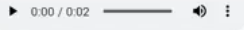
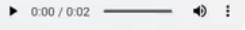


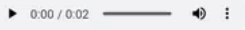



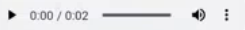
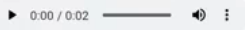
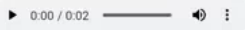
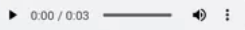
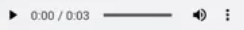
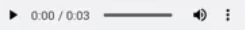
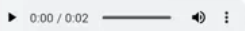
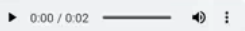
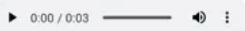
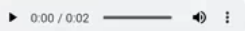
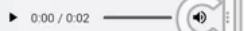

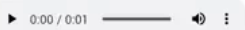

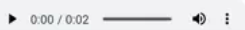
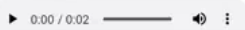
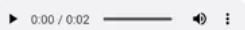

Abstract: We describe a neural network-based system for text-to-speech (TTS) synthesis that is able to generate speech audio in the voice of many different speakers, including those unseen during training. Our system consists of three independently trained components: (1) a speaker encoder network, trained on a speaker verification task using an independent dataset of noisy speech from thousands of speakers without transcripts, to generate a fixed-dimensional embedding vector from seconds of reference speech from a target speaker; (2) a sequence-to-sequence synthesis network based on Tacotron 2, which generates a mel spectrogram from text, conditioned on the speaker embedding; (3) an auto-regressive WaveNet-based vocoder that converts the mel spectrogram into a sequence of time domain waveform samples. We demonstrate that the proposed model is able to transfer the knowledge of speaker variability learned by the discriminatively-trained speaker encoder to the new task, and is able to synthesize natural speech from speakers that were not seen during training. We quantify the importance of training the speaker encoder on a large and diverse speaker set in order to obtain the best generalization performance. Finally, we show that randomly sampled speaker embeddings can be used to synthesize speech in the voice of novel speakers dissimilar from those used in training, indicating that the model has learned a high quality speaker representation.

[Click here for more from the Tacotron team.](#)

Speaker Adaptation for Unseen Speakers

Each column corresponds to a single speaker. The speaker name is in "Dataset SpeakerID" format. All speakers are unseen during training. The first row is the reference audio used to compute the speaker embedding. The rows below that are synthesized by our model using that speaker embedding.

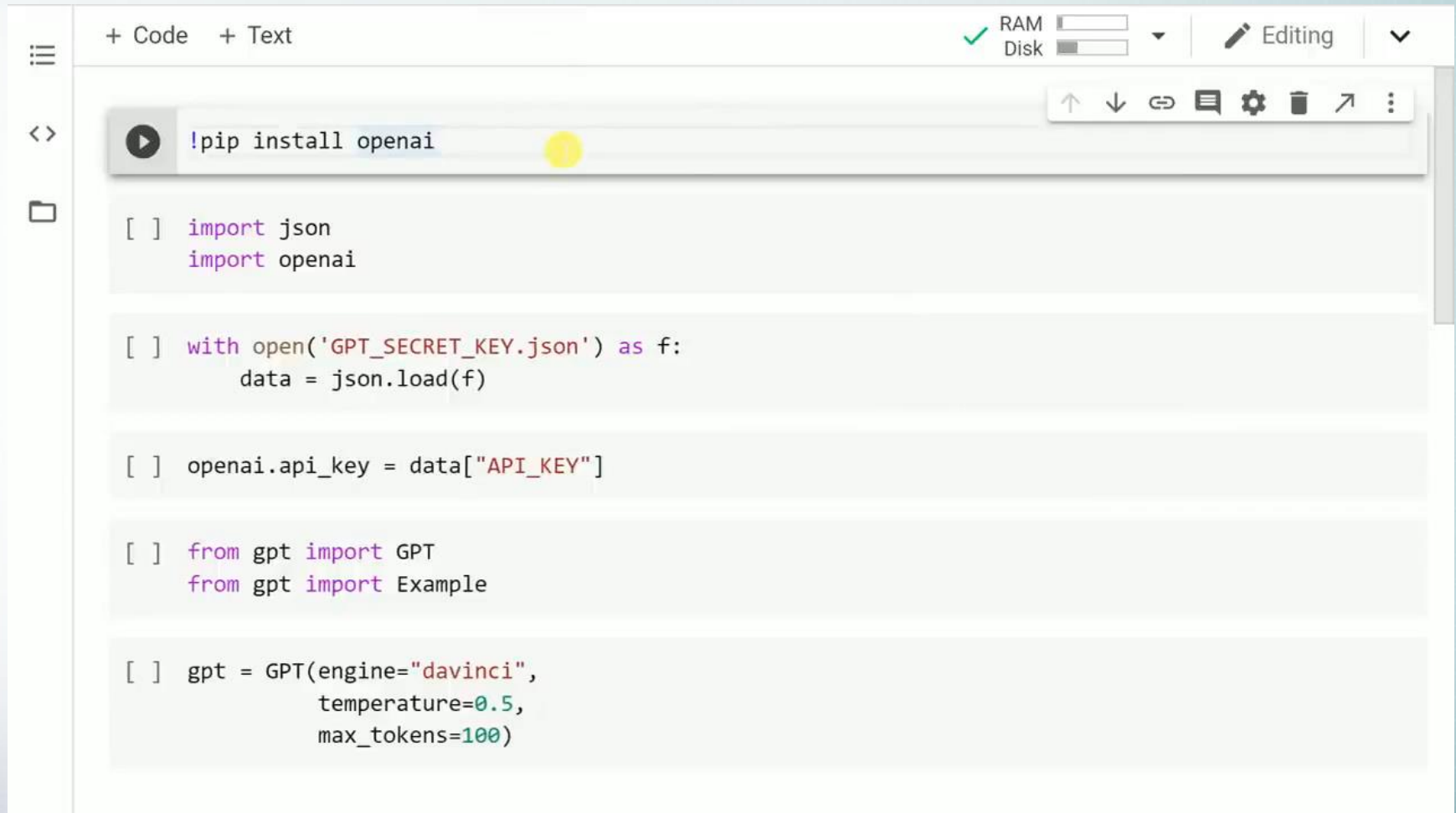
These examples are sampled from the evaluation set for Table 1 and Table 2 in the paper.

	VCTK p240	VCTK p260	LibriSpeech 1320	LibriSpeech 3575	LibriSpeech 6829	LibriSpeech 8230
Reference:						
Synthesized:						
0: Take a look at these pages for crooked creek drive.						
1: There are several listings for gas station.						
2: Here's the forecast for the next four days.						
3: Here is some information about the Gospel of John.						
4: His motives were more pragmatic and political.						
5: She had three brothers and two sisters.						

clideo.com

Source

GPT – 3 (OpenAI, 2020)



The screenshot shows a Jupyter Notebook interface. At the top, there are tabs for '+ Code' and '+ Text'. On the right, there are status indicators for 'RAM' and 'Disk' (both with green checkmarks and progress bars), and a button labeled 'Editing' with a dropdown arrow. Below the tabs, there is a toolbar with icons for undo, redo, link, comment, settings, delete, and a refresh icon. The main area contains a code cell with the following Python code:

```
!pip install openai

[ ] import json
    import openai

[ ] with open('GPT_SECRET_KEY.json') as f:
    data = json.load(f)

[ ] openai.api_key = data["API_KEY"]

[ ] from gpt import GPT
    from gpt import Example

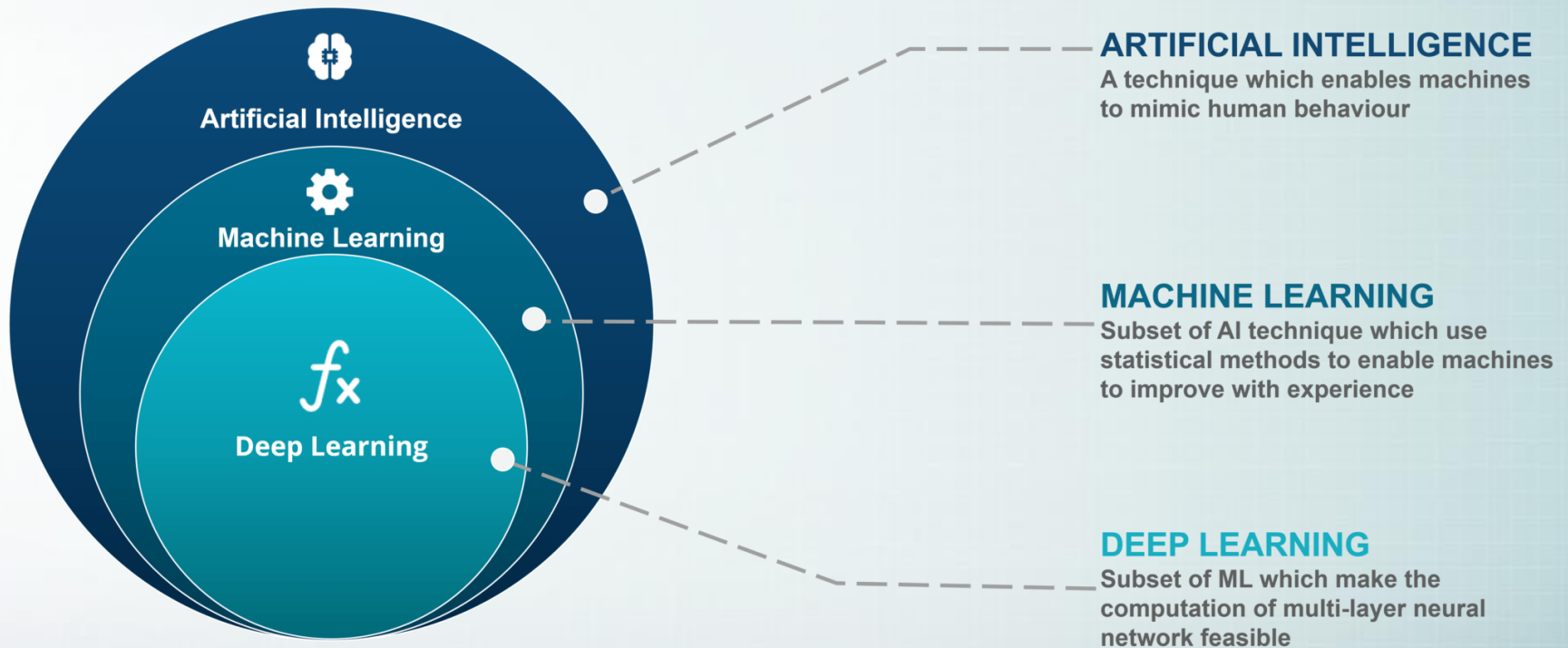
[ ] gpt = GPT(engine="davinci",
              temperature=0.5,
              max_tokens=100)
```

AI Gaming

Gen 25 species 10 genome 7 (30%)
Fitness: 2365 Max Fitness: 2365



AI vs ML vs DL

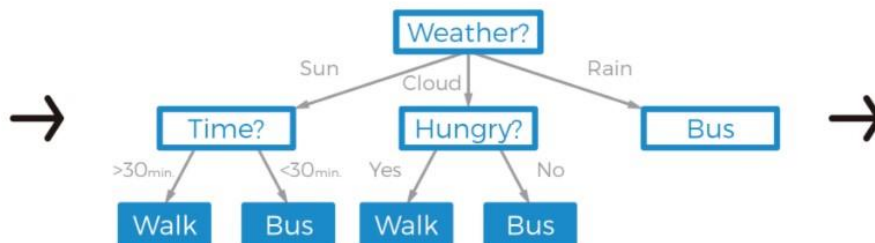


Deep Learning?

Machine Learning



Input



Decision tree

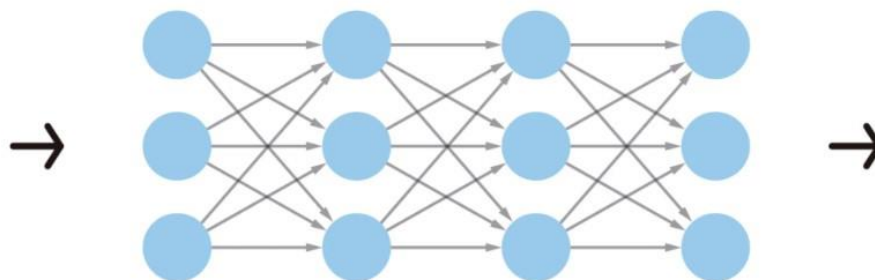


Output

Deep Learning



Input



Feature extraction + Classification



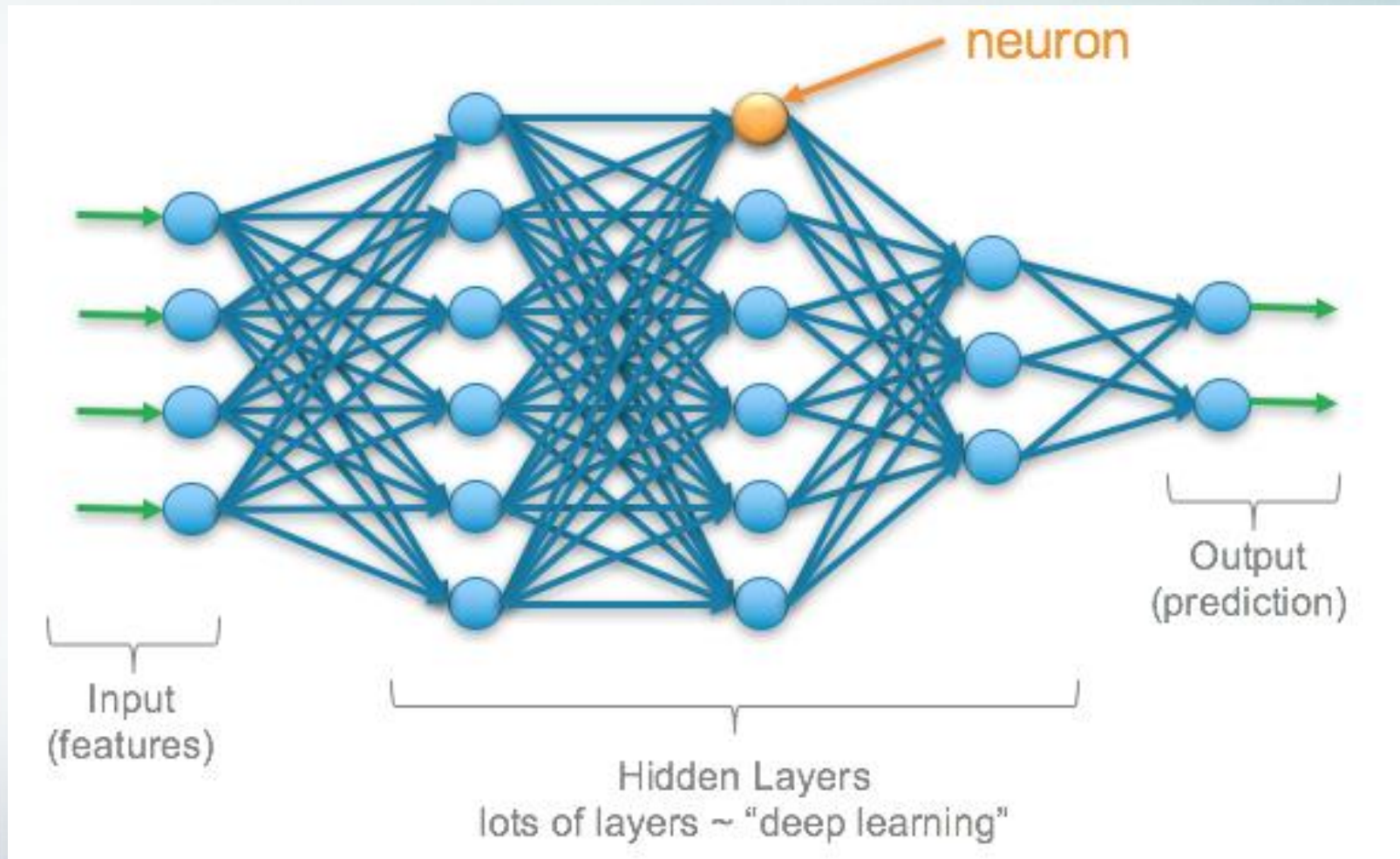
Output

Contents

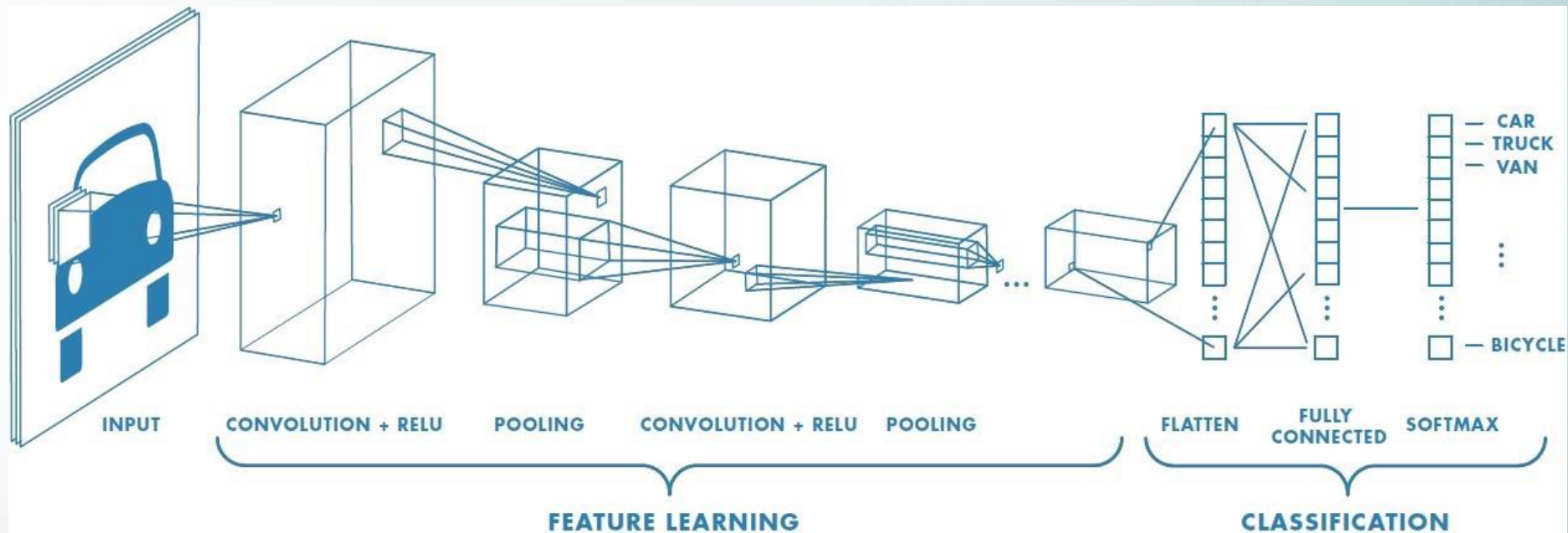
2

Deep Learning Models

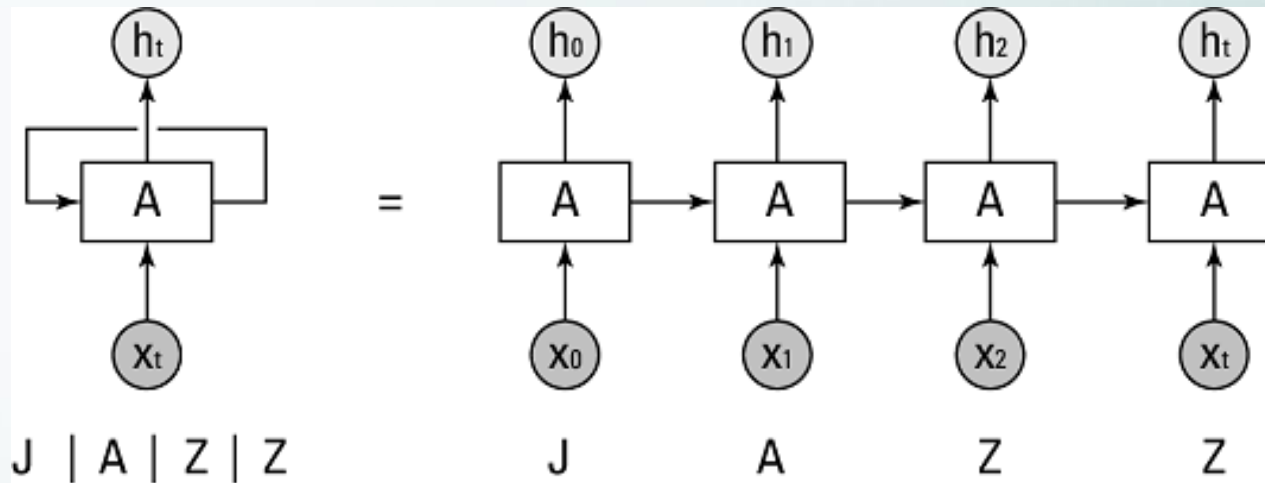
Idea of Neural Networks (NN)



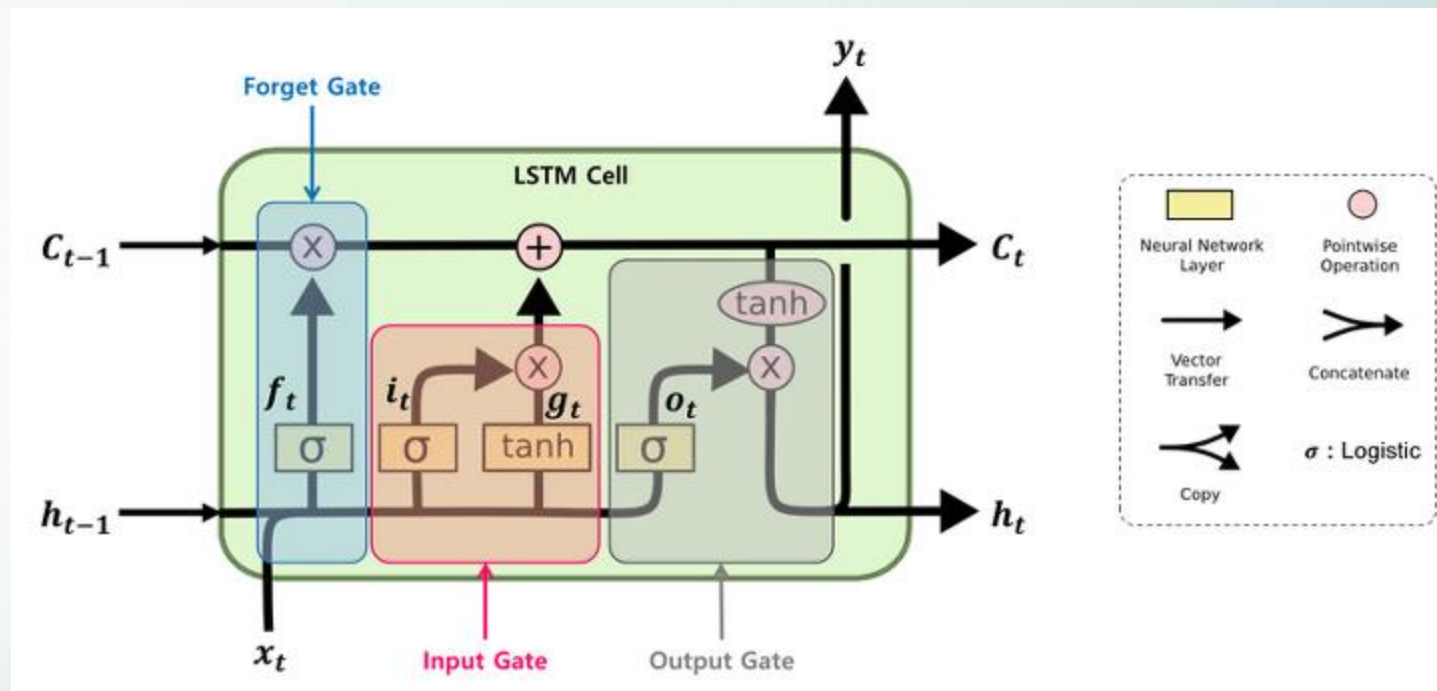
Convolutional NN (CNN)



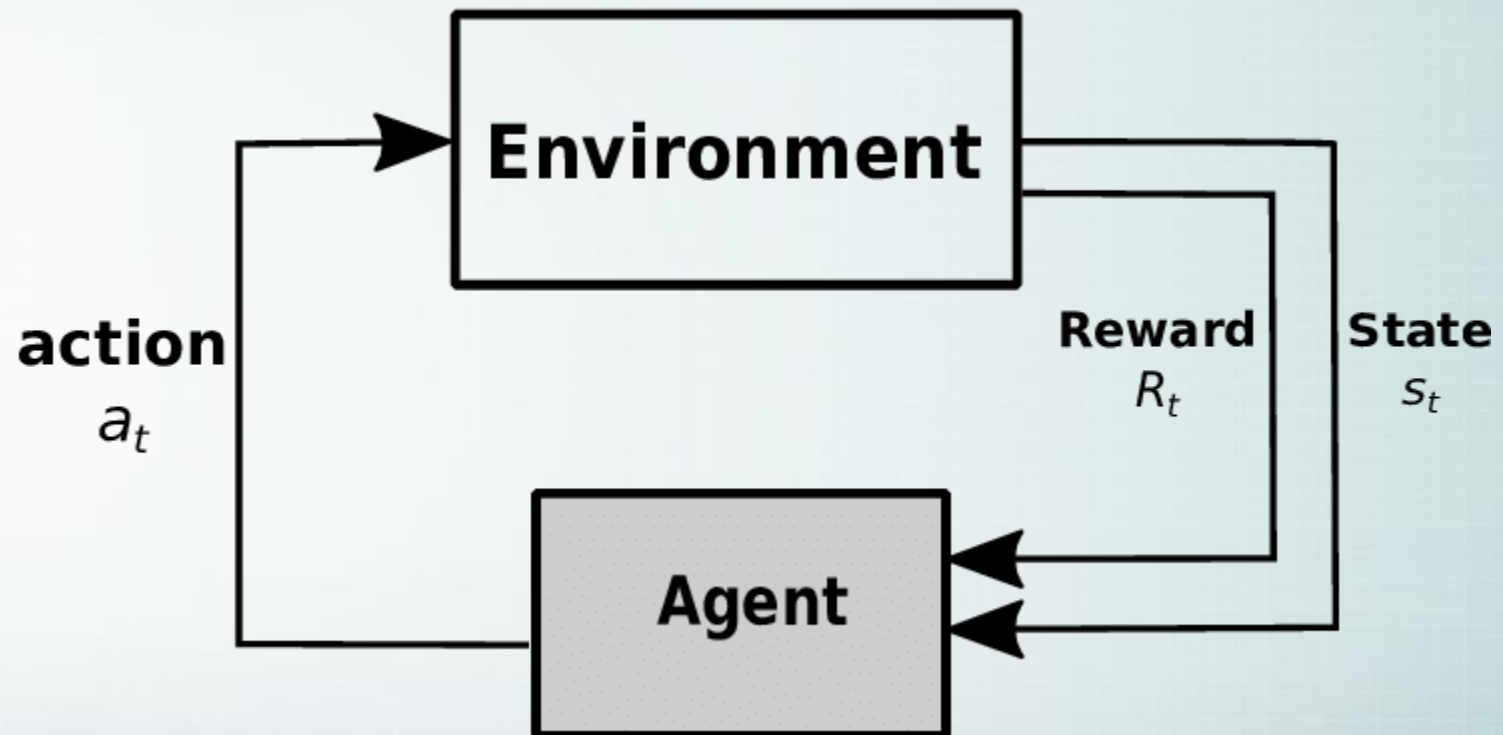
RNN



RNN - LSTM



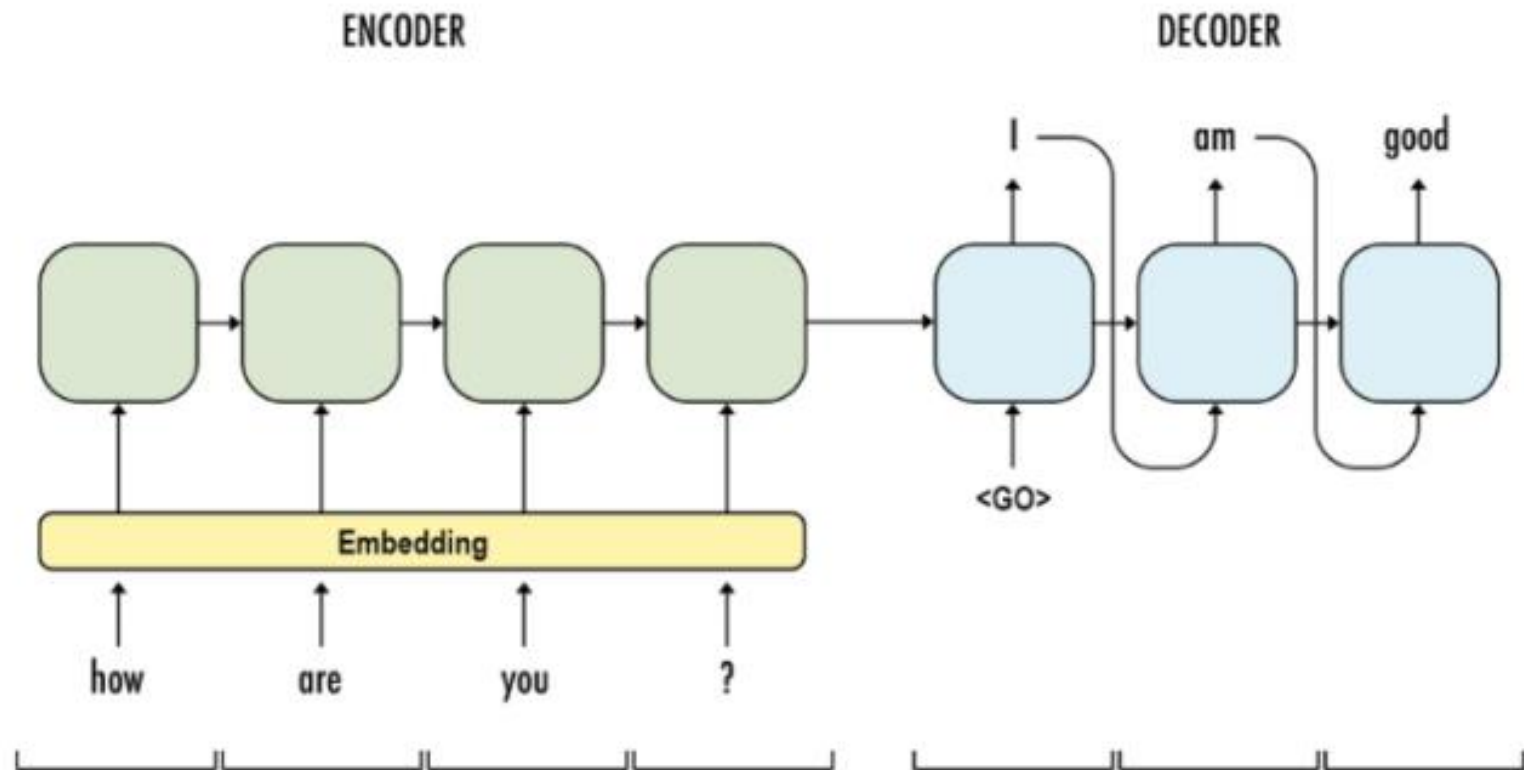
Reinforcement Learning



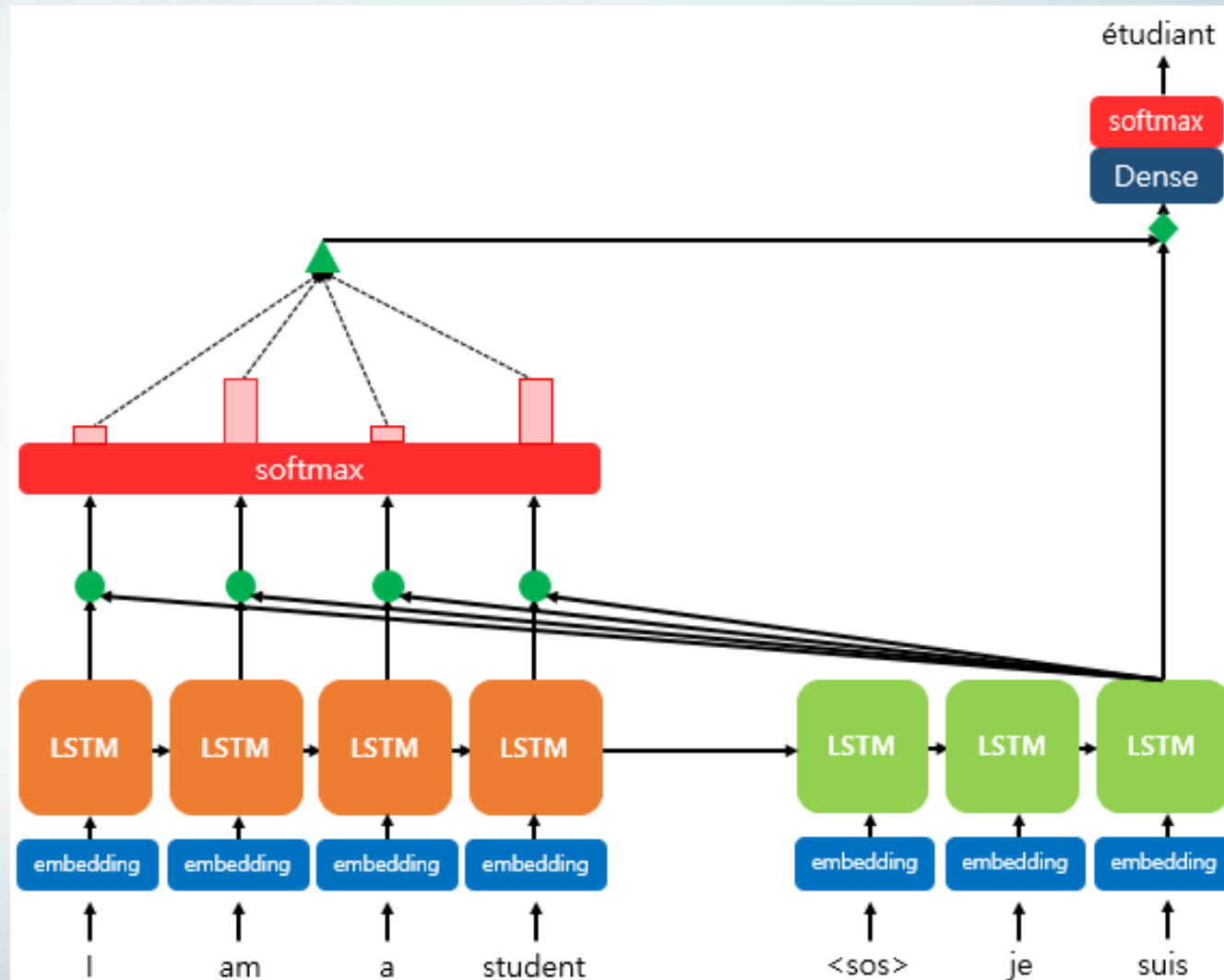
Some Important Models...

- ❖ Resnet (CNN)
 - ❖ VAE (CNN + probability)
 - ❖ GAN (CNN + classification)
 - ❖ Seq2Seq (RNN)
 - ❖ Image Captioning (CNN + RNN)
-
- ❖ BERT (Transformer)
 - ❖ Image Captioning (Visual Attention)
 - ❖ XLNet (Transformer - XL)
 - ❖ GPT - 3

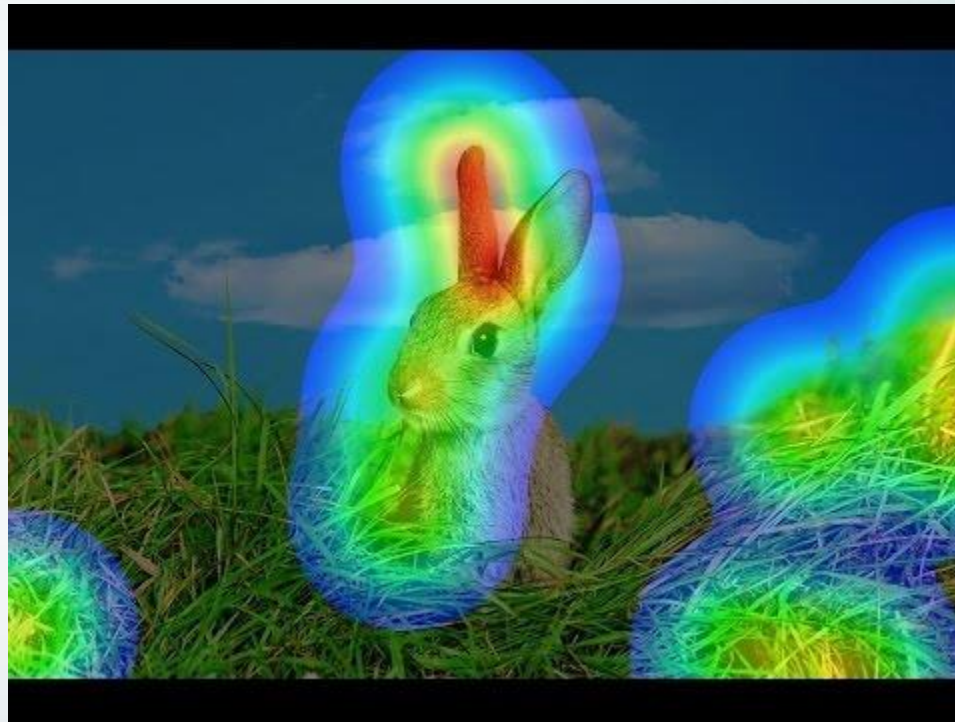
Seq2Seq before Attention



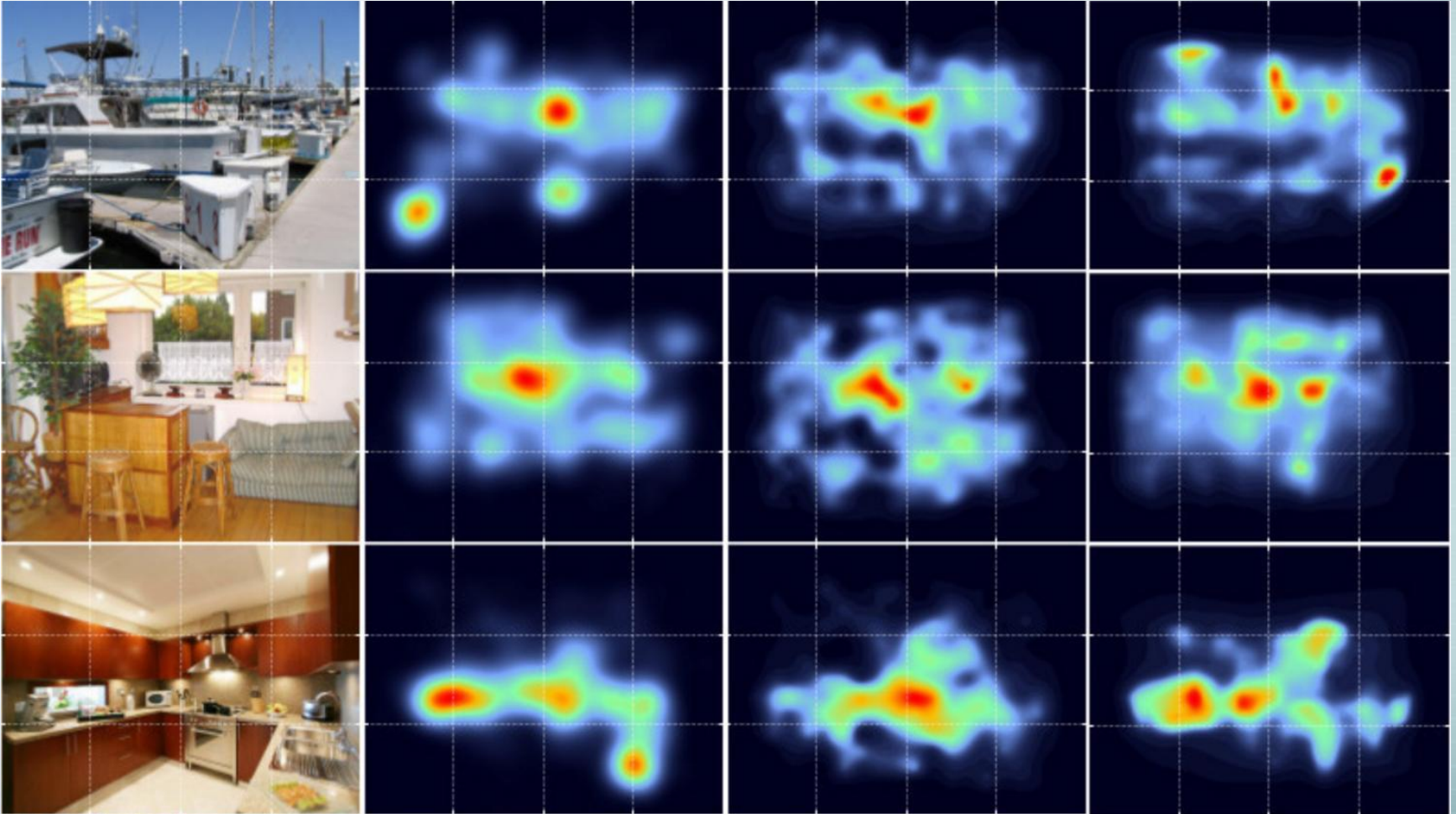
Seq2Seq with Attention



Visual Attention



Visual Attention



Transformer

- ❖ **Attention Is All You Need (2017)** - Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

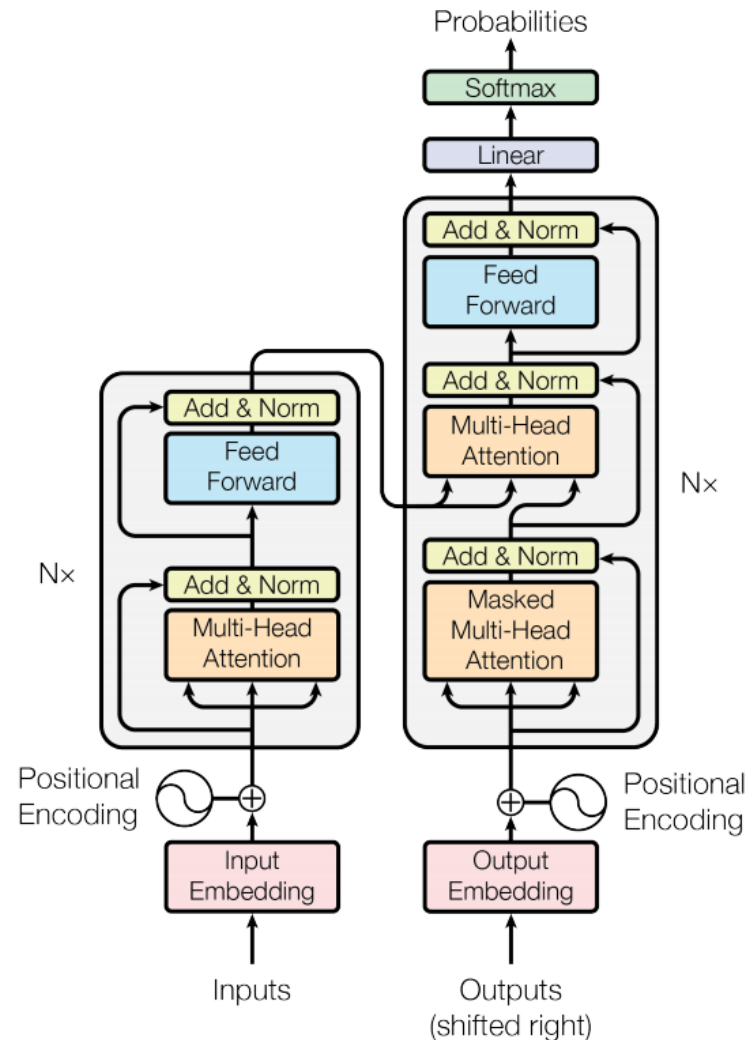


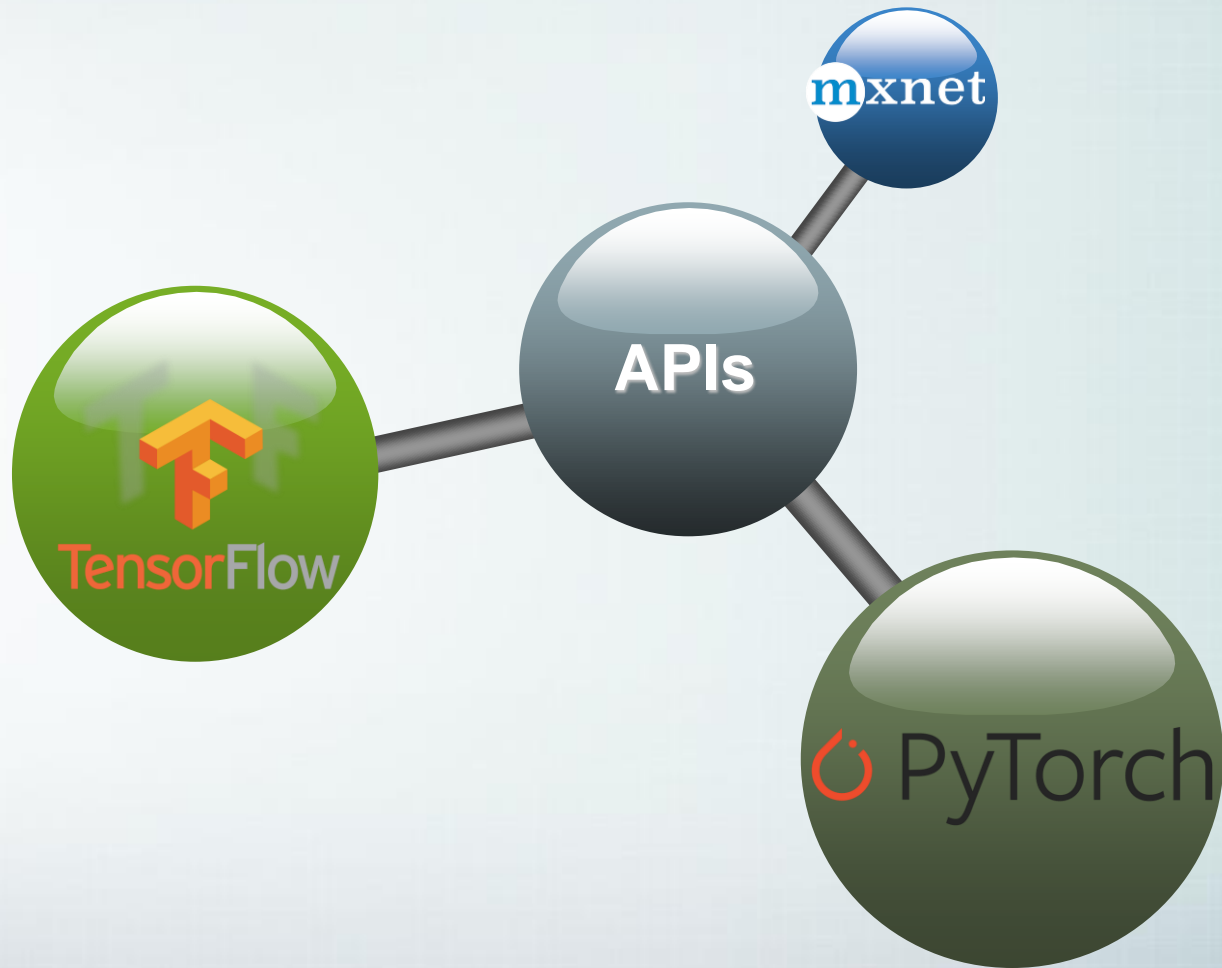
Figure 1: The Transformer - model architecture.

Contents

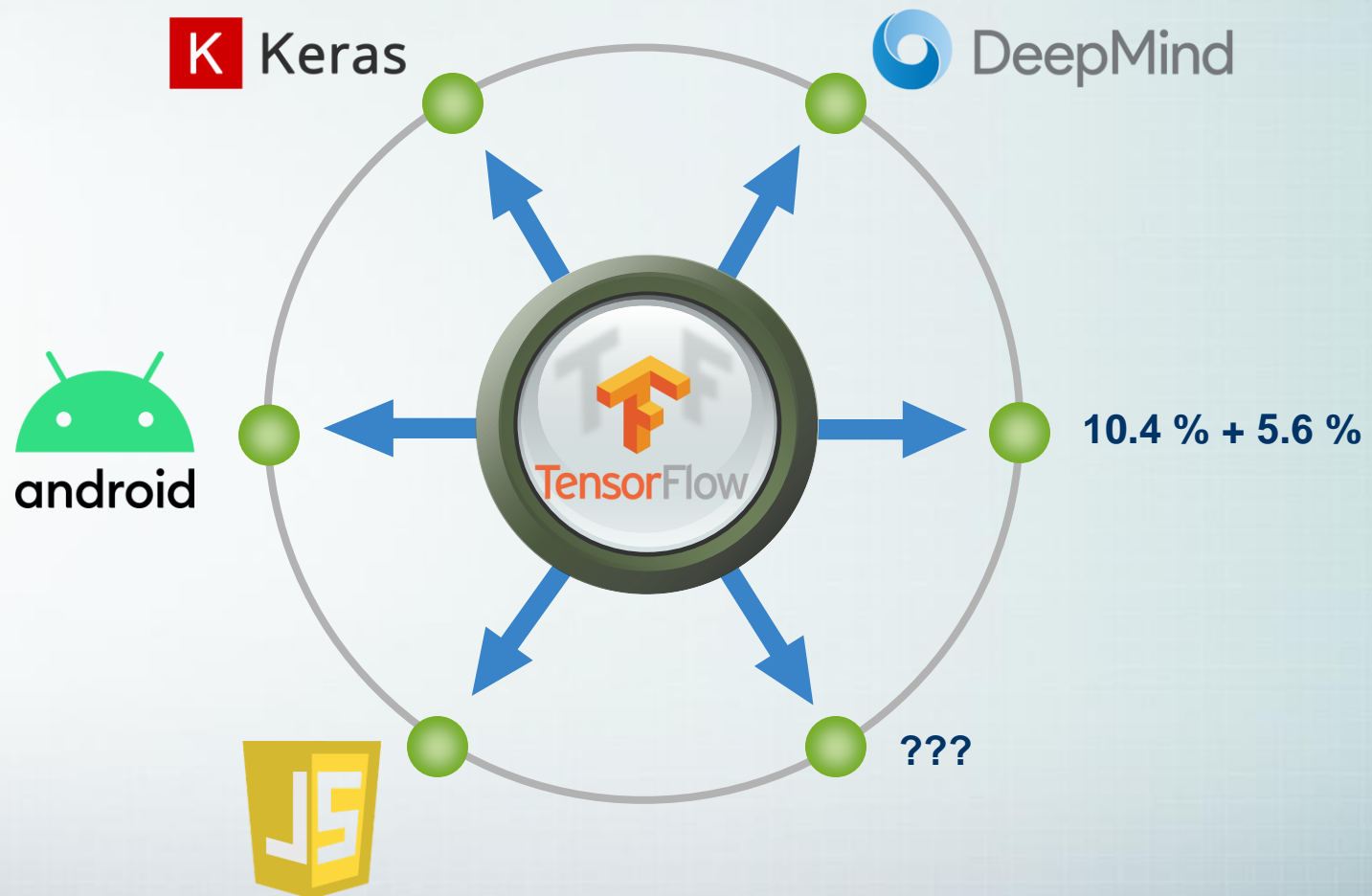
3

Tensorflow vs PyTorch

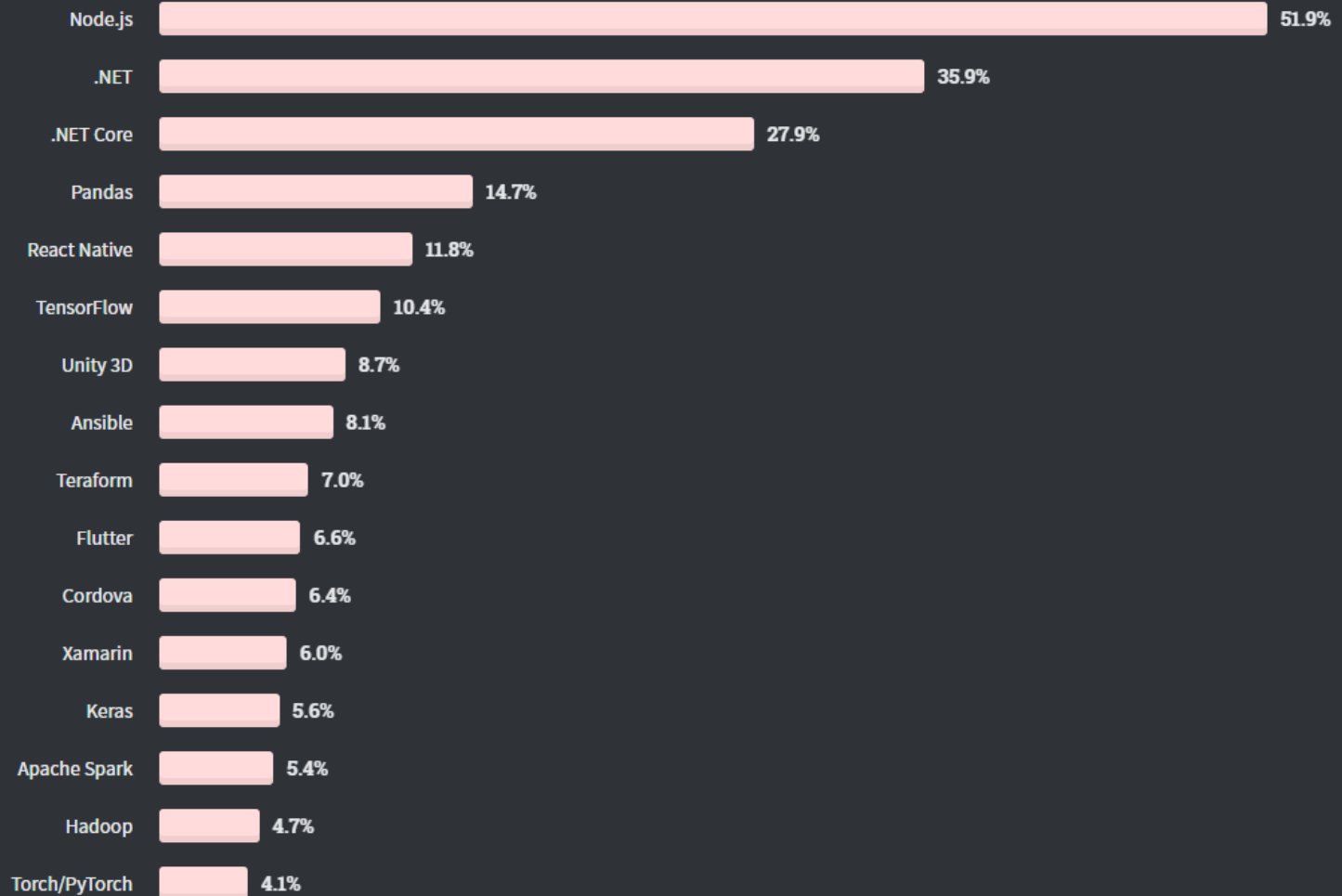
Tensorflow, PyTorch, MXNet



Why TensorFlow?



2020 Stack Overflow Developer Survey



H.W.

- ❖ **1. Run NumPy.html & Pandas.html on python (optional)**
- ❖ **2. from d2l.ai (Colab recommended)**
 - Run CH2.3.Linear Algebra
 - Run CH2.6.Probability
- ❖ **3. DLSeminar.pdf (optional)**
 - Read Lecture 1 ~ 2 (except Lecture2. Confusion Matrix and ...)



Thank You !

■ 고려대학교 물리학과 한승희