기계 번역과 attention

ZP 32기 김동영

Most Popular AI Research July 2022

Top 10 List Based On Total Likes On Twitter

designed & made by: @bycloudai

The below metrix are all taken from labml.ai, Twitter and GitHub. This infographic is not affiliated with labml.ai.
Please note that this top 10 list is not intending to define and categorize what is a successful or a revolutional research.
This top 10 list is purely to entertain the idea of what is research is popular.
The research selected for this top 10 list was merely available on labml.ai, in no way it is intending to exclude other research.

RETWEETS
TWITTER LIKES
GITHUB STARS

1 Formal Algorithms for Transformers

Mary Phuong, Marcus Hutter

N/A 2.2K 374

2 Confident Adaptive Language Modeling

Tal Schuster, Adam Fisch, Jai Gupta, Mostafa Dehghani, Dara Bahri, Vinh Q. Tran, Yi Tay, Donald Metzler

N/A 1.7K 308

3 Language Models (Mostly) Know What They Know

Kadavath et al.

N/A 1.7K 271

4 YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao

3.8K 1.6K 316

5 Language Model Cascades

David Dohan, Winnie Xu, Aitor Lewkowycz, Jacob Austin, David Bieber, Raphael Gontijo Lopes, Yuhuai Wu, Henryk Michalewski, Rif A. Saurous, Jascha Sohl-dickstein, Kevin Murphy, Charles Sutton

N/A 1.6K 221

6 Collaborative Neural Rendering using Anime Character Sheets

Zuzeng Lin, Ailin Huang, Zhewei Huang, Chen Hu, Shuchang Zhou

N/A 1.3K 373

7 Neural Networks and the Chomsky Hierarchy

Grégoire Delétang, Anian Ruoss, Jordi Grau-Moya, Tim Genewein, Li Kevin Wenliang, Elliot Catt, Marcus Hutter, Shane Legg, Pedro A. Ortega

50 1.3K 265

8 Language Modelling with Pixels

Phillip Rust, Jonas F. Lotz, Emanuele Bugliarello, Elizabeth Salesky, Miryam de Lhoneux, Desmond Elliott

140 1.3K 233

9 On the Principles of Parsimony and Self-Consistency for the Emergence of Intelligence

Yi Ma, Doris Tsao, Heung-Yeung Shum

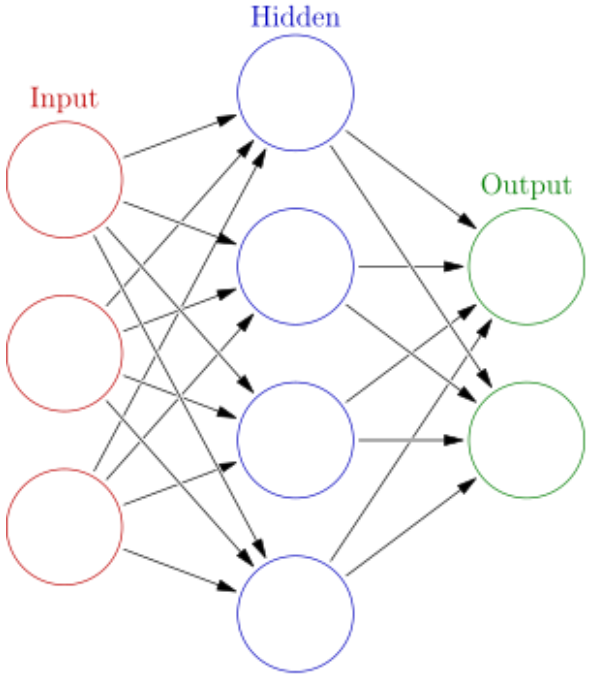
N/A 1.4K 204

10 Scaling Laws vs Model Architectures: How does Inductive Bias Influence Scaling?

Yi Tay, Mostafa Dehghani, Samira Abnar, Hyung Won Chung, William Fedus, Jinfeng Rao, Sharan Narang, Vinh Q. Tran, Dani Yogatama, Donald Metzler

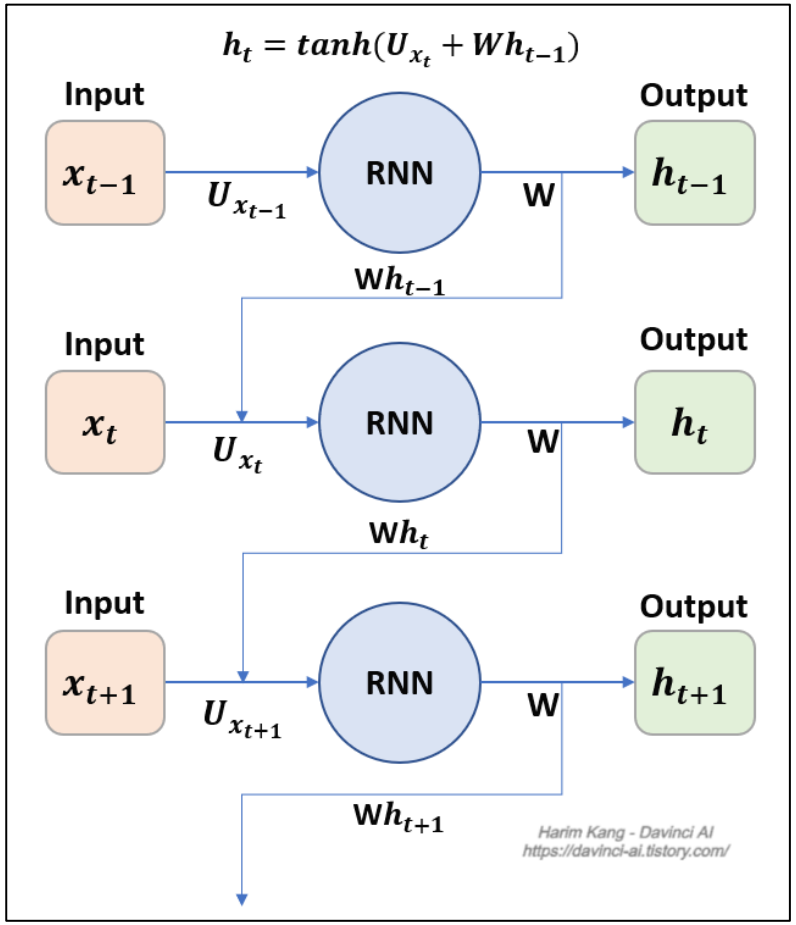
N/A 1.2K 245

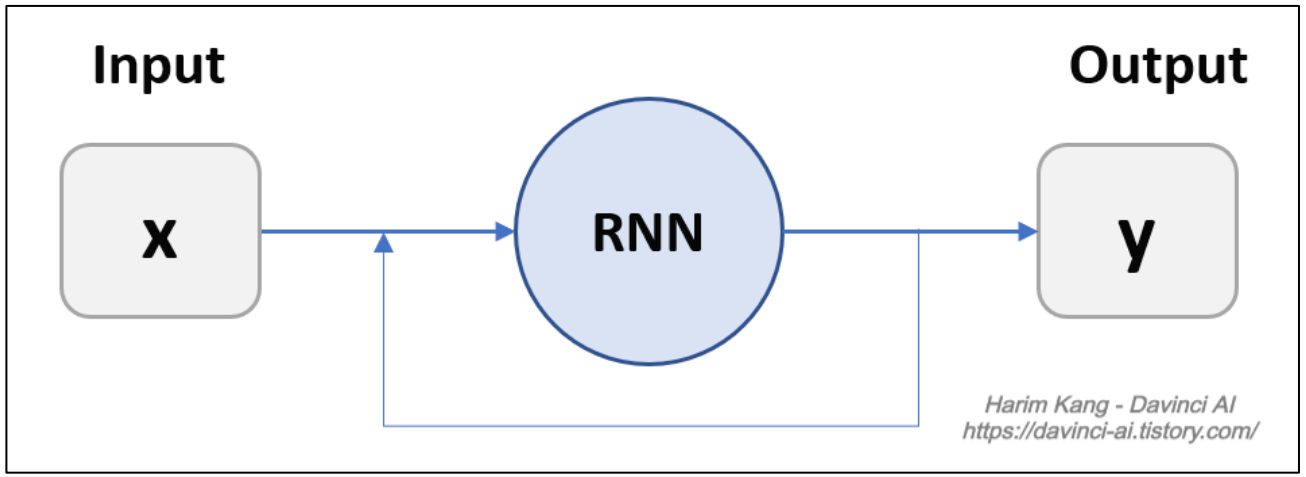
ffn(dnn)

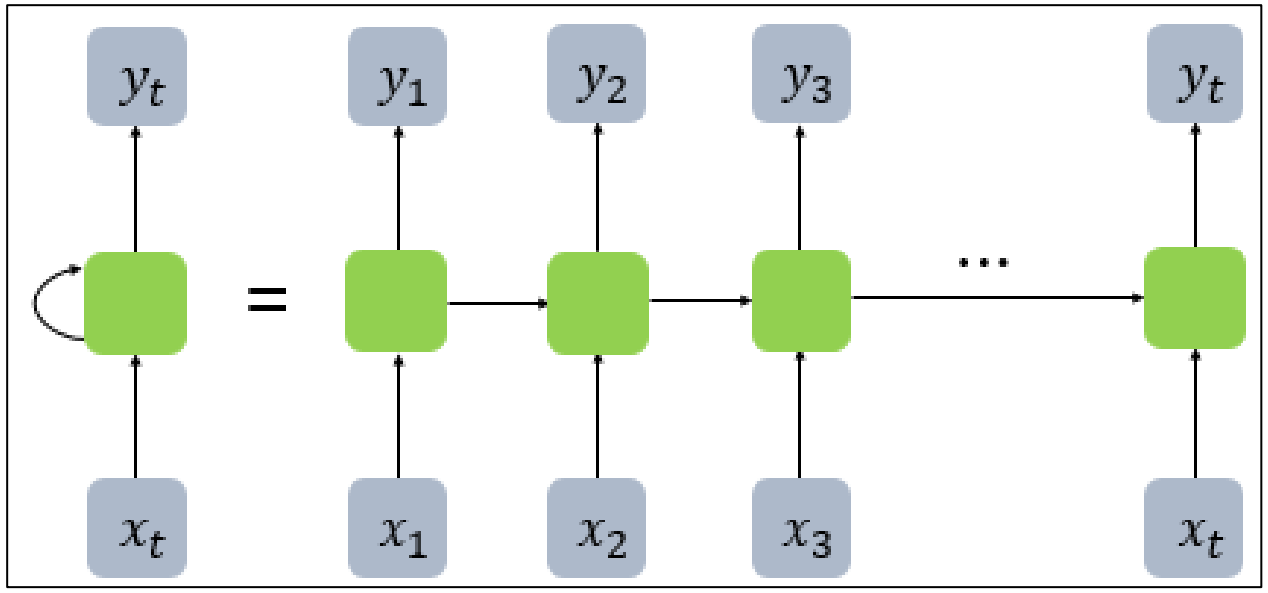


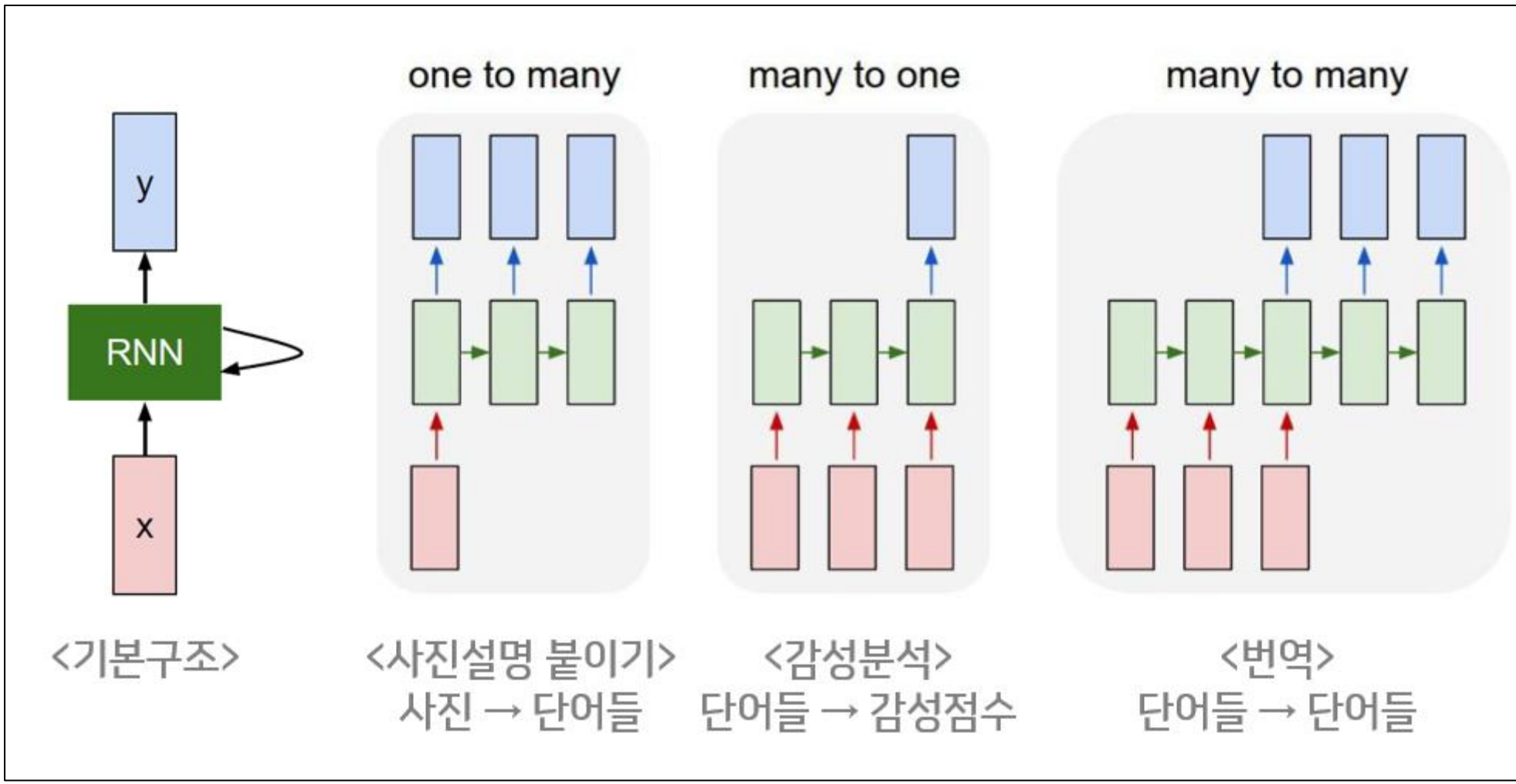
Cat!!

rnn

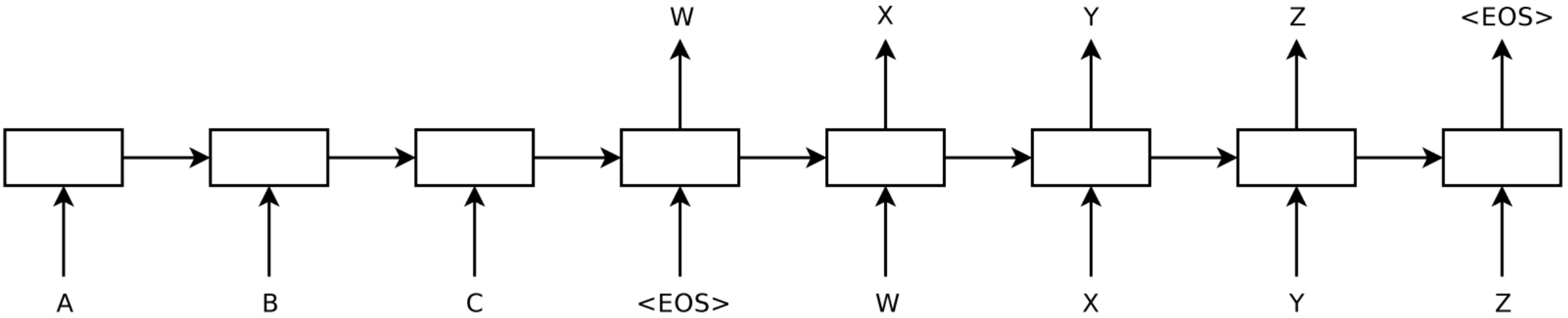

$h_t = tanh(U_{x_t} + Wh_{t-1})$
Input
Output
x_{t-1}
$U_{x_{t-1}}$
RNN
W
h_{t-1}
Wh_{t-1}
Input
Output
x_t
U_{x_t}
RNN
W
h_t
Wh_t
Input
Output
x_{t+1}
$U_{x_{t+1}}$
RNN
W
h_{t+1}
Wh_{t+1}
Harim Kang - Davinci AI
https://davinci-ai.tistory.com/


Input
Output
x
RNN
y
Harim Kang - Davinci AI
https://davinci-ai.tistory.com/


y_t
y_1
y_2
y_3
y_t
=
...
x_t
x_1
x_2
x_3
x_t

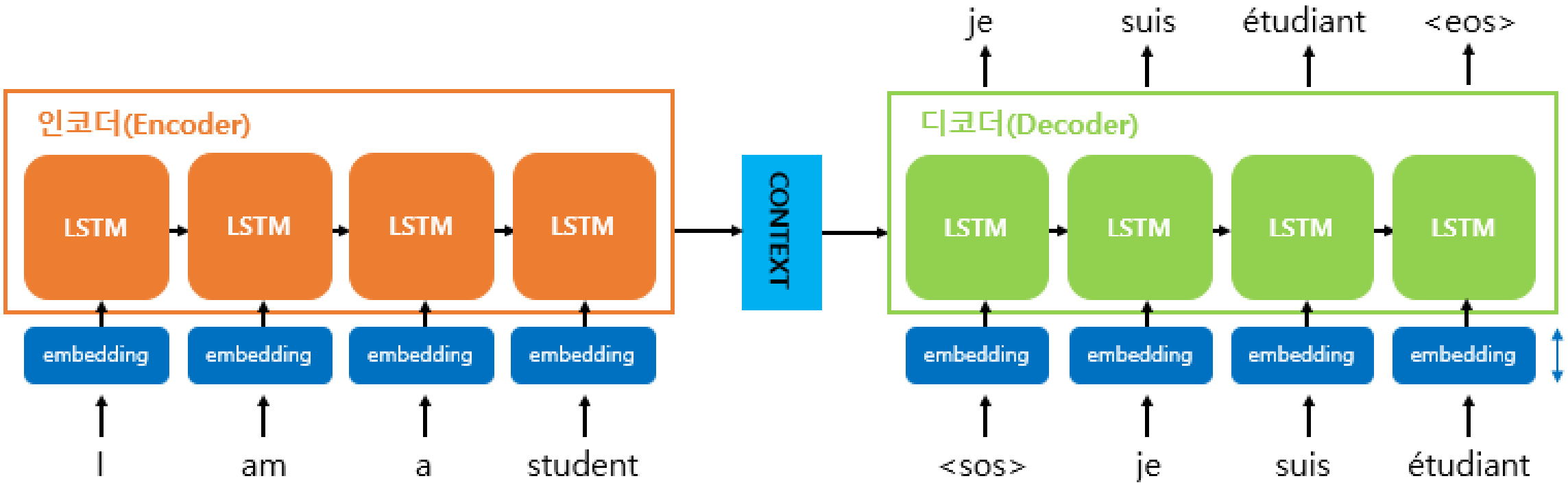

y
RNN
x
one to many
many to one
many to many
<기본구조>
<사진설명 붙이기>
사진 → 단어들
<감성분석>
단어들 → 감성점수
<번역>
단어들 → 단어들

seq2seq

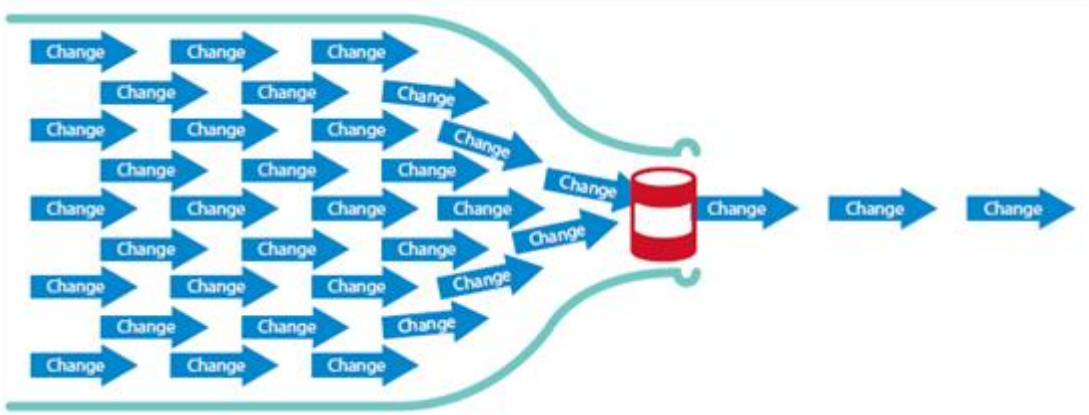


Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences.

Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems, pages 3104–3112, 2014.


je
suis
étudiant
<eos>
인코더(Encoder)
디코더(Decoder)
LSTM
LSTM
LSTM
LSTM
CONTEXT
LSTM
LSTM
LSTM
LSTM
embedding
embedding
embedding
embedding
embedding
embedding
embedding
embedding
I
am
a
student
<sos>
je
suis
étudiant

Problem 1

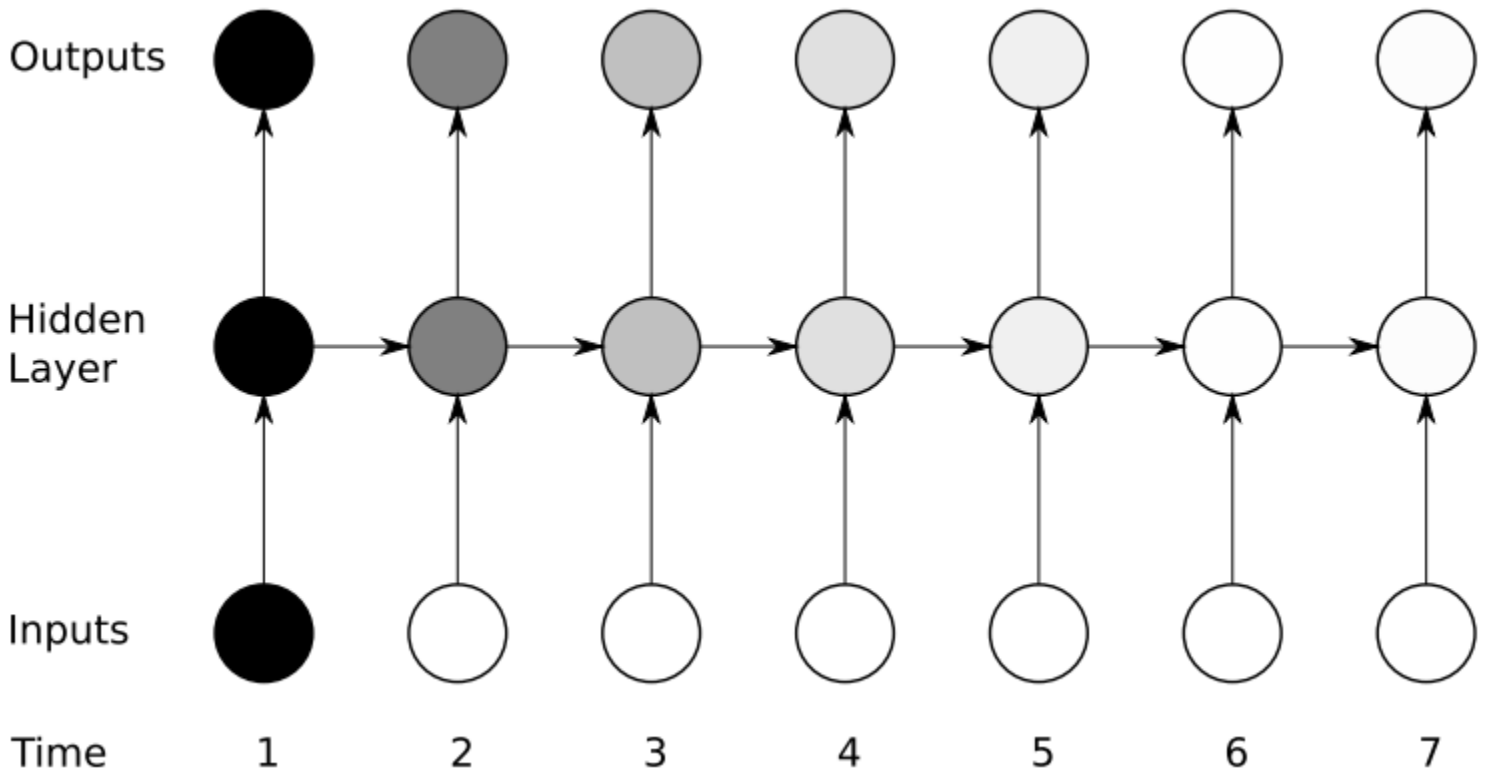


- 문맥 벡터가 고정된 크기이다!

- potential issue with this encoder–decoder approach is that a neural network needs to be able to compress all the necessary information of a source sentence into a fixed-length vector.

D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

Problem 2



summary

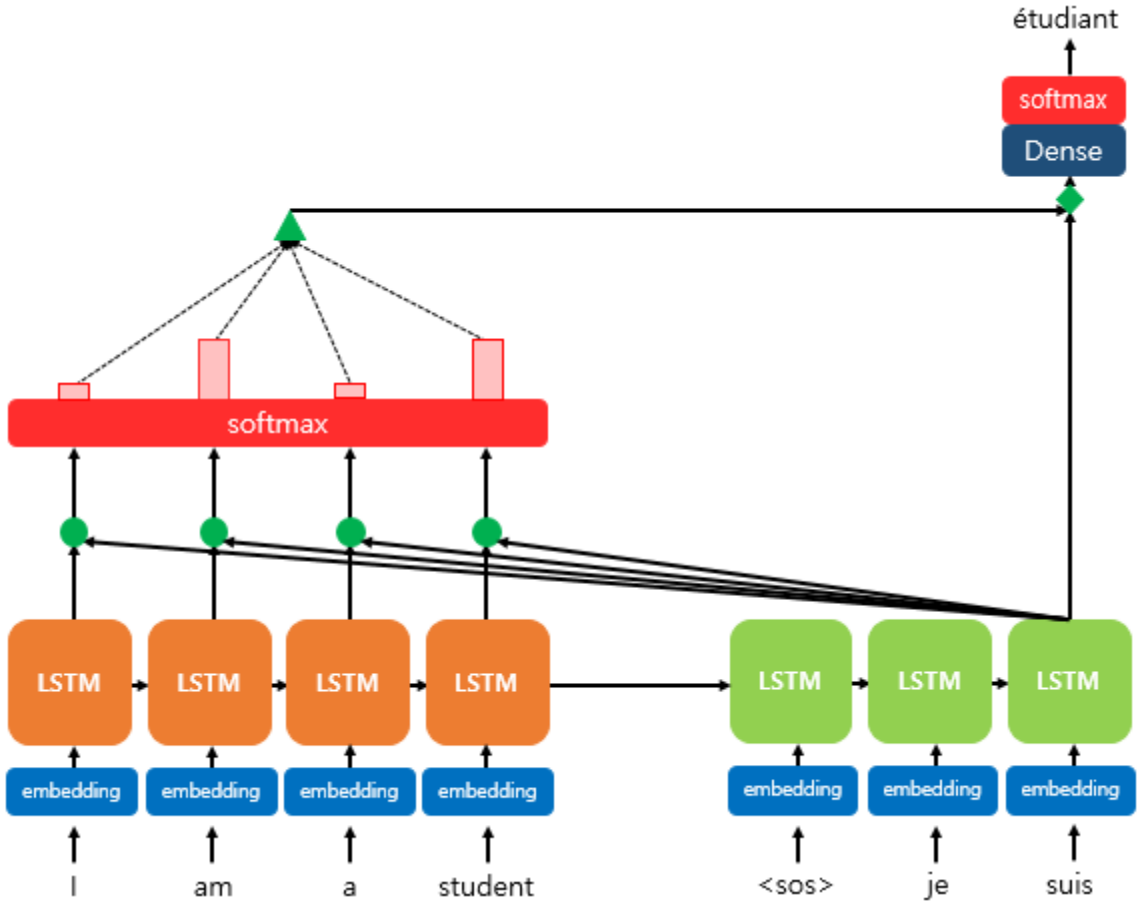
1. Mnt에서는 seq2seq를 많이 썼음. (dnn은 성능이 좋지 않으니까!)

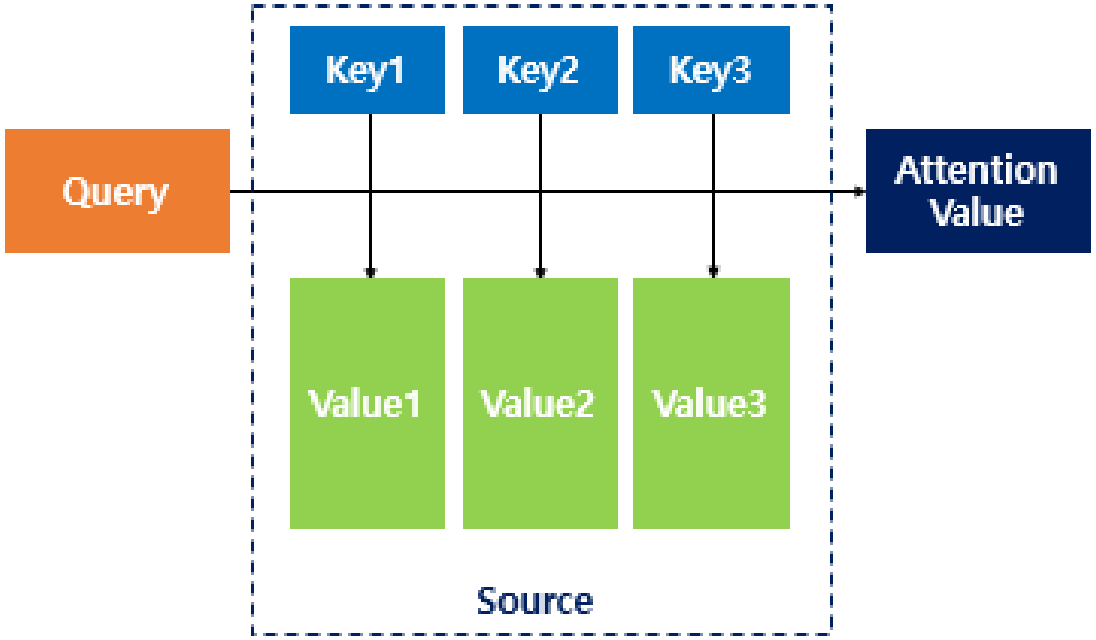
2. 기존 seq2seq는 인코더에서 입력이 문맥 벡터로 압축되고, 디코더에서 이 문맥 벡터를 사용해 출력을 만드는 방식으로 작동함

3. 하지만 이 방식은 문제가 있음
   - long-term dependecy (vanishing gradient)
   - compress all the necessary information into a fixed-length vector.
   - 병렬 처리도 어려움

4. 따라서 seq2seq는 입력 문장의 길이가 길어질수록 성능이 떨어짐

Attention mechanism





transfromer

Object detection

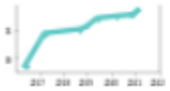
COCO minival

DINO (Swin-L, multi-scale)

Machine translatic

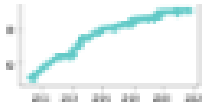
WMT201

WMT201



IWSLT2014 German-English

BiBERT



ImageNet

CoCa (finetuned)

CIFAR-10

ViT-H/14

Image classification

ADE20K

FD-SwinV2-G

Cityscapes test

ViT-Adapter-L (Mask2Former, BEiT pretrain, Mapillary)

ADE20K val

FD-SwinV2-G

segmentation

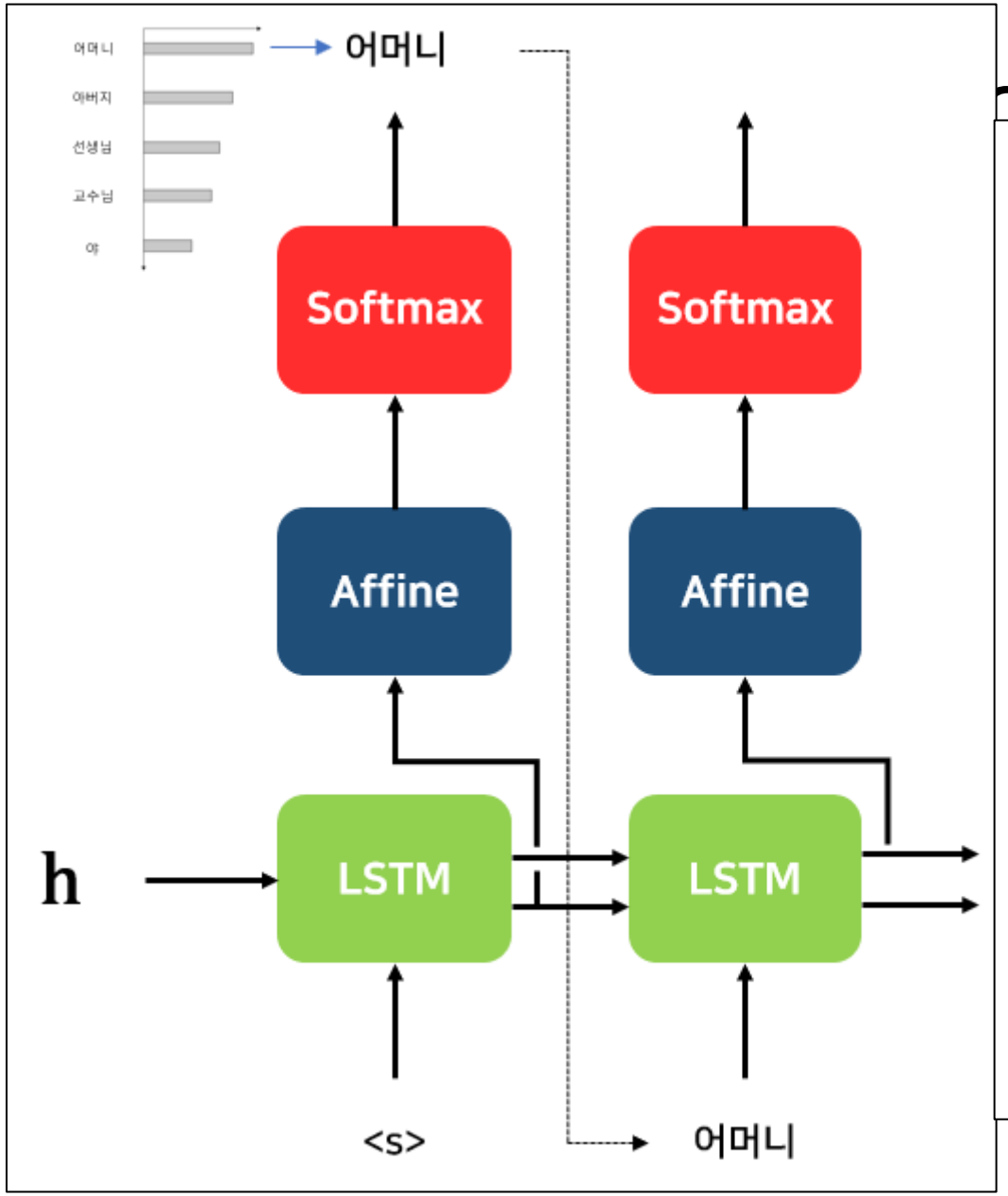
Transformer?

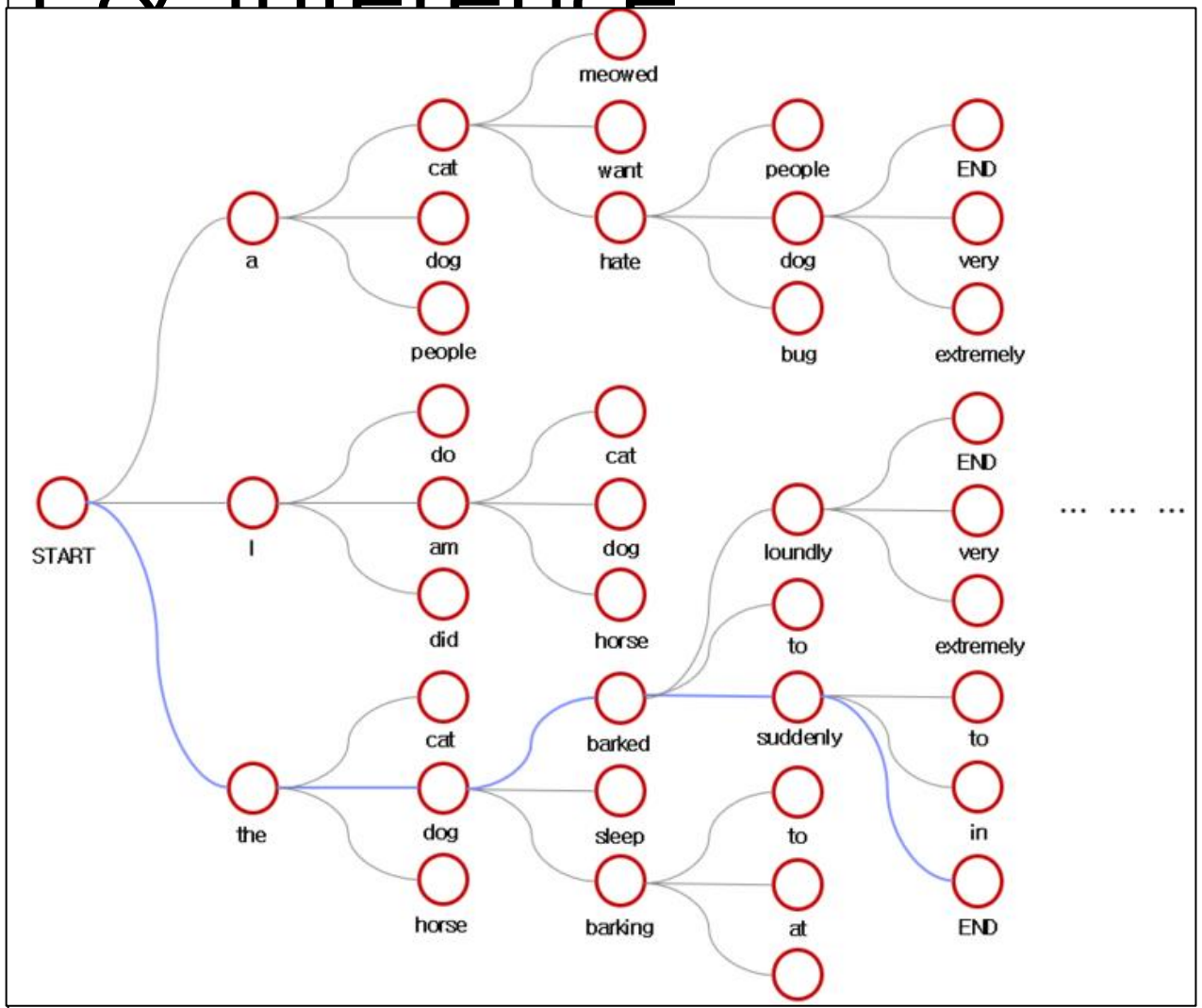
- Seq2seq = rnn + attention
- Seq2seq는 rnn 기반이기에 병렬 처리가 쉽지 않음

  → rnn을 빼자!!
- Transformer는 rnn 없이 attention 만을 사용함

  → attention is all you need (2017, google)
- Seq2seq에서는 input sequence와의 attention score을 구하기 위해 input lengt만큼의 시간이 필요했음
- Transformer는 self-attention을 사용해 attention score 계산의 시간복잡도를 O(1)으로 줄여 긴 문장도 효과적으로 학습할 수 있도록 함

& inference


어머니
아버지
선생님
교수님
어머니
Softmax
Softmax
Affine
Affine
h
LSTM
LSTM
<s>
어머니


START
I
the
a
do
am
did
cat
dog
horse
cat
dog
sleep
barking
barked
meowed
want
hate
people
dog
bug
loundly
to
suddenly
to
at
cat
dog
horse
people
very
extremely
END
very
extremely
END
to
in
END
... ...

X-formers

- improve the vanilla Transformer from different perspectives
- Model Efficiency

  - self-attention이 비효율적임 → lightweight attention (sparse attention variants) and Divide-and-conquer methods (recurrent and hierarchical mechanism).

- Model Generalization

  - transformer는 inductive bias가 부족하기 때문에, 데이터셋이 작을 때는 훈련하기 어려움 → pre-training, introducing structural bias

- Model Adaptation

  - 여러 downstream task에 적용시키는 것

- Taxonomy of Transformers

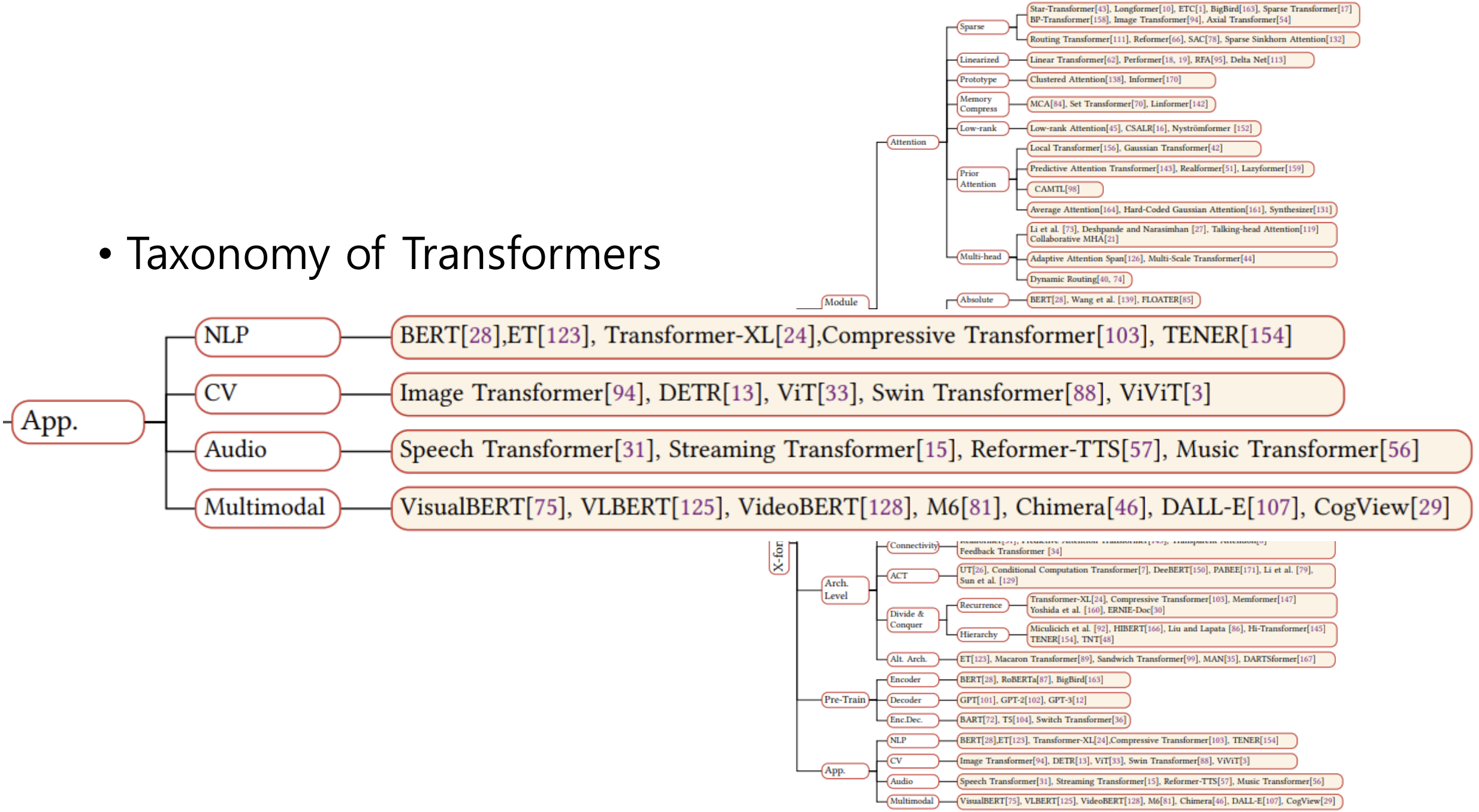
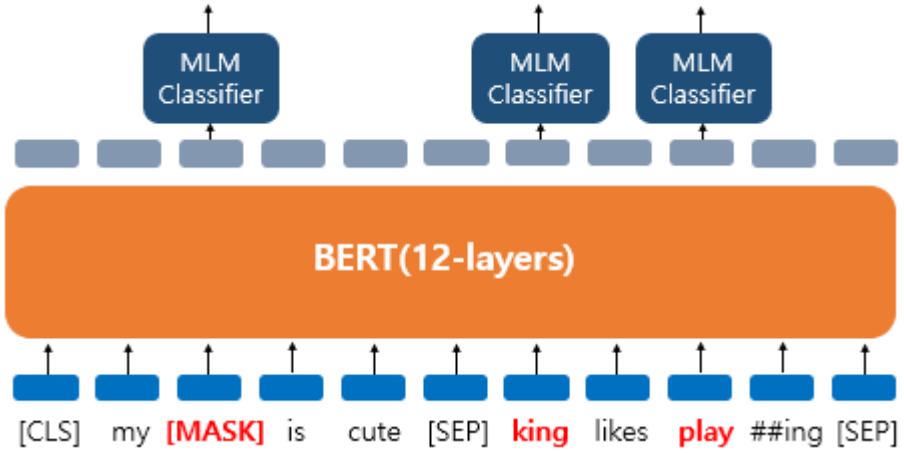


Fig. 3. Taxonomy of Transformers

bert

- Bidirectional Encoder Representations from Transformers
- 2018년 등장하자마자 다양한 nlp task에서 sota를 갈아치움
- 위키피디아(30억 단어)와 같은 label이 없는 텍스트 데이터로 훈련되었음
- 빈칸 뚫고 맞추기 방식으로 pre-trained



코드 유사성 판단 AI 경진대회

월간데이콘21 | AI프렌즈 | 대전 AI | Code NLP

₩ 상금 : 총 600만 원

2022.05.02 ~ 2022.06.10 16:59 + Google Calendar

769명 마감

Encoder is All you Need | 상하목장스누피팀	29	671
[Private 43등/0.95526] CodeBERT 사용	10	248
코드 텍스트 데이터 전처리 방법 공유	14	300
[Private 1등/0.9909] graphcodebert + codebert_mlm 앙상블	17	856
[Private 2nd, 0.9837], Part 1 of 2, Summary, Data creation	8	328
[Private 35 / 0.96088] CodeBERTa_small_v1	9	219
[Private 2nd, 0.9837], Part 2 of 2, GraphCodeBERT	7	237
Private 6위, codeBERT+codeT5, Random Sampling w/o duplicate	8	199
[Private 19등 / 0.9706] Random sampling + CodeBert	7	194
[Private 22등/0.96998] GraphCode Bert	8	217

쇼핑몰 리뷰 평점 분류 경진대회

데이콘 베이직 Basic | NLP | Accuracy

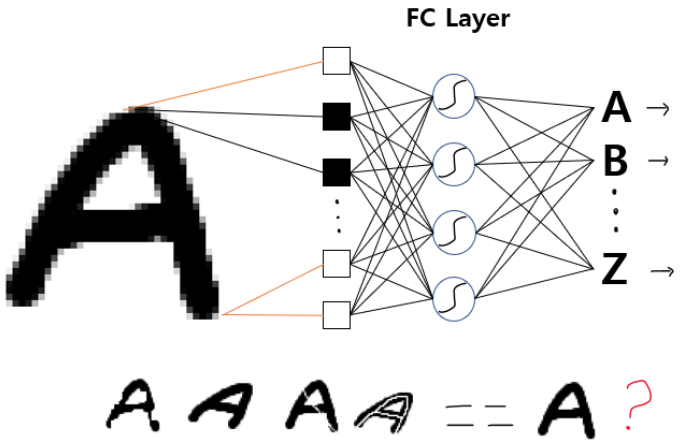
₩ 상금 : 인증서, 장학금, 스타벅스 기프티콘 등

2022.07.11 ~ 2022.08.05 17:59 + Google Calendar

524명 D-2

- BERT활용하여 텍스트 다중 분류 구현 with 허깅페이스
- 허깅페이스를 활용한 분류 모델 베이스라인 [roberta-large] score:0.6913
- Word2Vec를 이용한 평점 분석
- Pytorch Lightning을 사용한 kogpt2 classification구현 score: 0.64

ViT



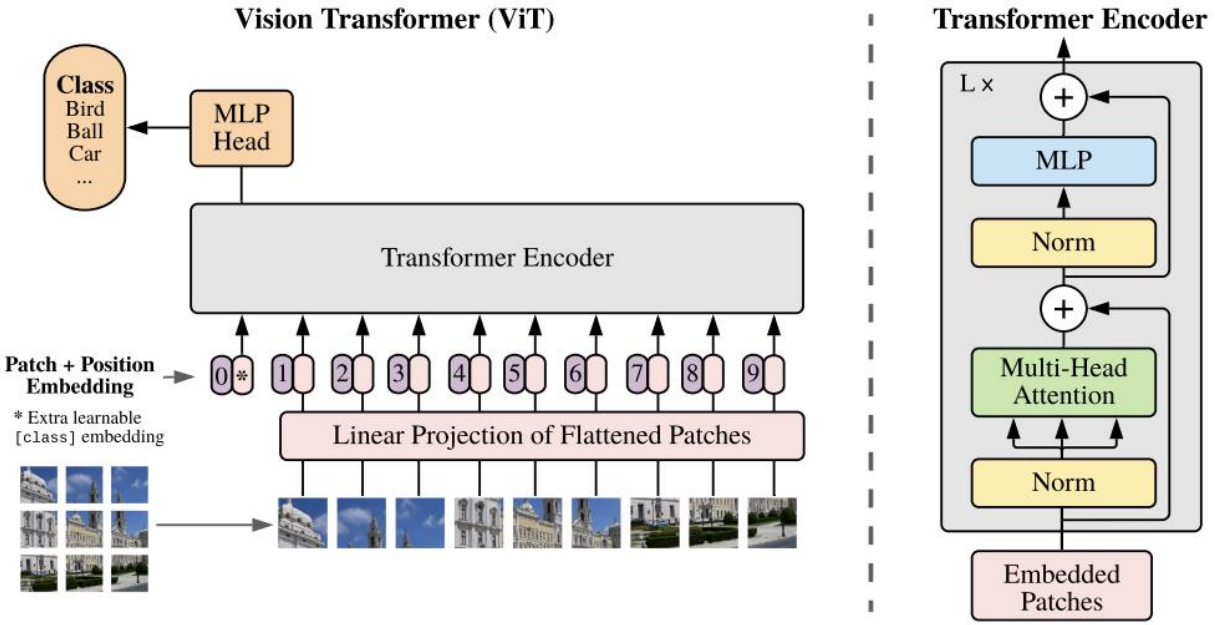


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

- Cnn을 사용하지 않음! Attention만 사용함

ViT(2)

- Thus, Vision Transformer matches or exceeds the state of the art on many image classification datasets, whilst being relatively cheap to pre-train.

- While these initial results are encouraging, many challenges remain. One is to apply ViT to other computer vision tasks, such as detection and segmentation.

LIFT: Language-Interfaced Fine-Tuning for Non-Language Machine Learning Tasks

Tuan Dinh*, Yuchen Zeng*, Ruisu Zhang, Ziqian Lin, Michael Gira, Shashank Rajput, Jy-yong Sohn, Dimitris Papailiopoulos, Kangwook Lee

University of Wisconsin-Madison, USA

Abstract

Fine-tuning pretrained language models (LMs) without making any architectural changes has become a norm for learning various language downstream tasks. However, for *non*-language downstream tasks, a common practice is to employ task-specific designs for input, output layers, and loss functions. For instance, it

Openai – DALL-E

Zero-Shot Text-to-Image Generation (arxiv.org)

(a) a tapir made of accordion. a tapir with the texture of an accordion.

(b) an illustration of a baby hedgehog in a christmas sweater walking a dog

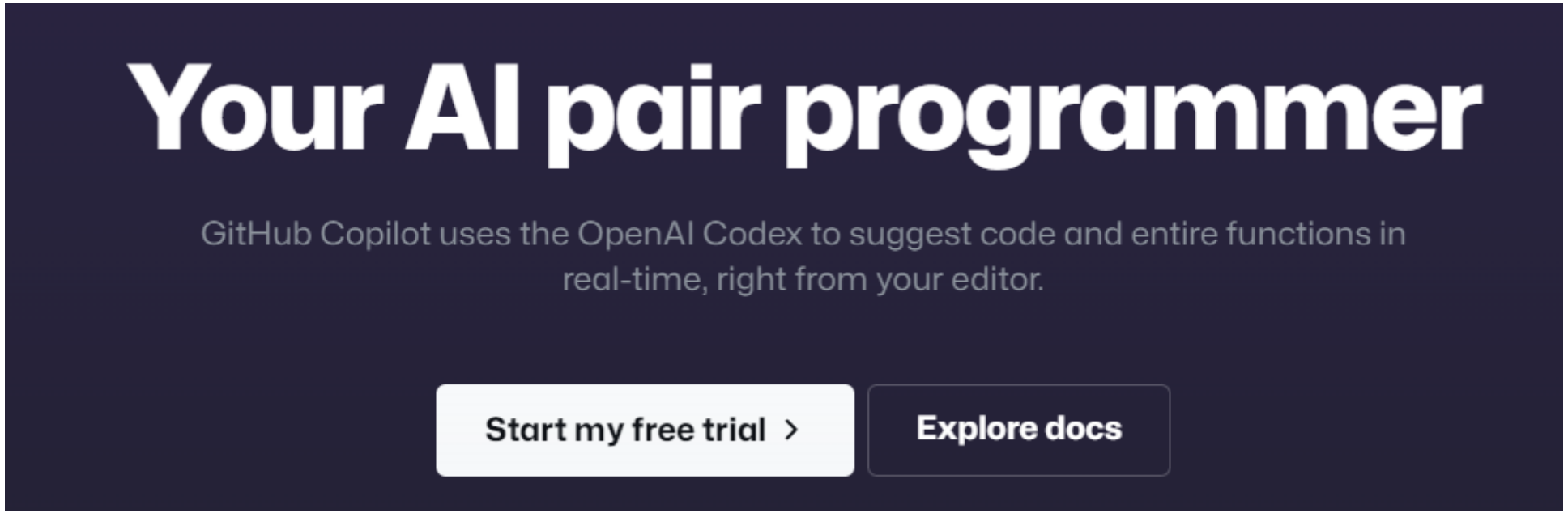
(c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign

(d) the exact same cat on the top as a sketch on the bottom

Figure 2. With varying degrees of reliability, our model appears to be able to combine distinct concepts in plausible ways, create anthropomorphized versions of animals, render text, and perform some types of image-to-image translation.

Openai – gpt-3

- copilot



GitHub Copilot · Your AI pair programmer

reference

- 고양이 주방 애완 동물 - Pixabay의 무료 사진
- 인공 신경망 - 위키백과, 우리 모두의 백과사전 (wikipedia.org)
- 1) 순환 신경망(Recurrent Neural Network, RNN) - 딥 러닝을 이용한 자연어 처리 입문 (wikidocs.net)
- 딥러닝 (7) - RNN(Recurrent Neural Network), LSTM, GRU (tistory.com)
- RNN과 LSTM을 이해해보자! · ratsgo's blog
- arXiv:1409.3215v3 [cs.CL] 14 Dec 2014
- 호다닥 공부해보는 CNN(Convolutional Neural Networks) - 호롤리한 하루 (gruuuuu.github.io)
- A Survey of Transformers (arxiv.org)
- [2206.06565] LIFT: Language-Interfaced Fine-Tuning for Non-Language Machine Learning Tasks (arxiv.org)
- 어텐션 메커니즘(Attention Mechanism) 간단히 이해하기 (tistory.com)
- 1) 어텐션 메커니즘 (Attention Mechanism) - 딥 러닝을 이용한 자연어 처리 입문 (wikidocs.net)
- 1409.0473.pdf (arxiv.org)
- Recurrent Neural Network | Fundamentals Of Deep Learning (analyticsvidhya.com)
- 1706.03762.pdf (arxiv.org)
- Long Short-Term Memory (LSTM) 이해하기 :: 개발새발로그 (tistory.com)
- The Unreasonable Effectiveness of Recurrent Neural Networks (karpathy.github.io)
- The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time. (jalammar.github.io)
- TIL/Neural machine translation by jointly learning to align and translate.md at master · strutive07/TIL (github.com)

감사합니다

- 감사합니다.