RNN:
Recurrent
Neural Networks

Where we are…

- Artificial Intelligence
- Brain and Neurons
- Learning
- Regression
- Deep Neural Networks
- CNN
- RNN
- Unsupervised Learning
- Reinforcement Learning
- AI Applications

Supervised Learning

“

동해물과 백두산이

$7 \times 6 \rightarrow 42$

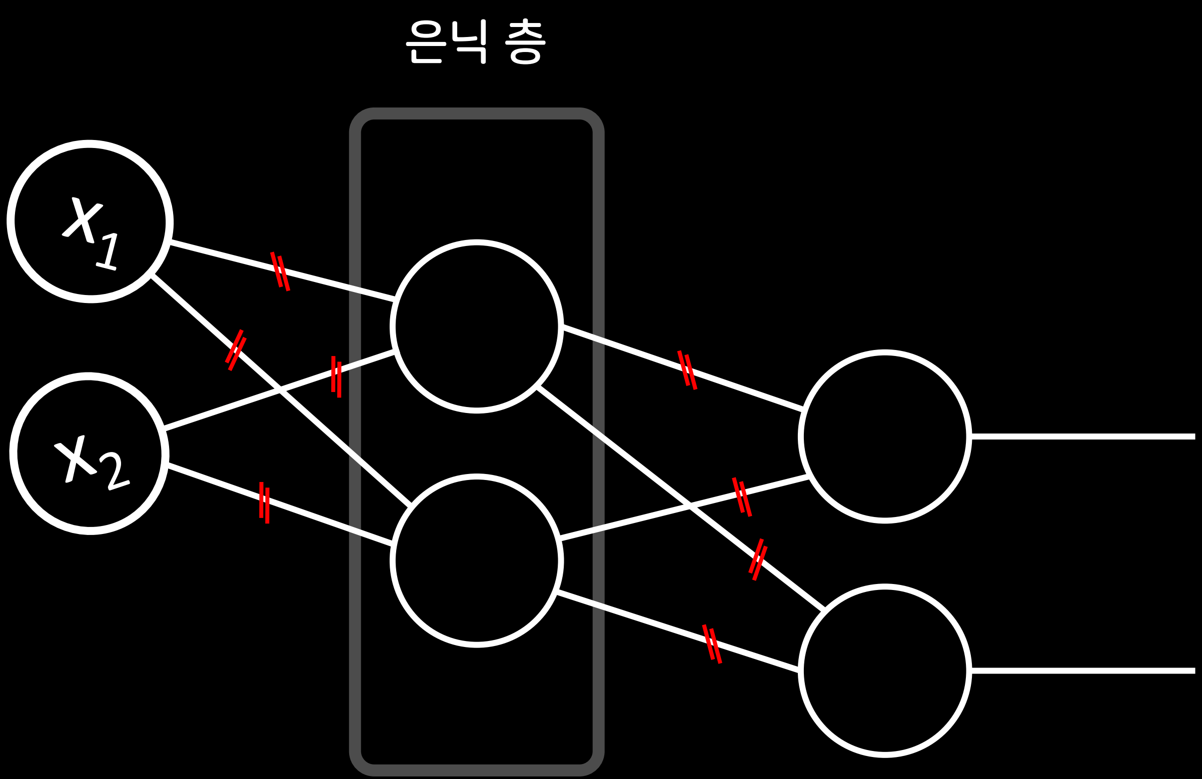
이전에 어떤 구구단을 말했는지 관계없다.

“무슨 말인지 알겠어?”

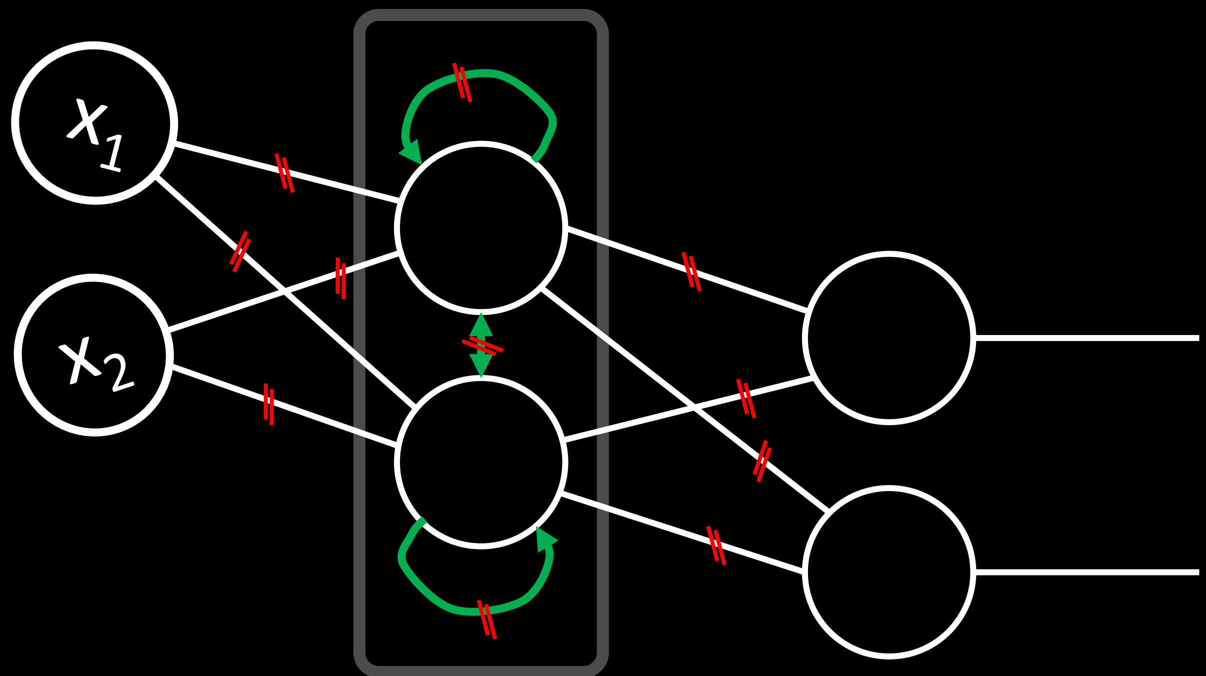
이전 내용을 기억하는 법

The answer may be different according to the context (state info.) discussed at some previous time.

Same questions, but different answers with memory

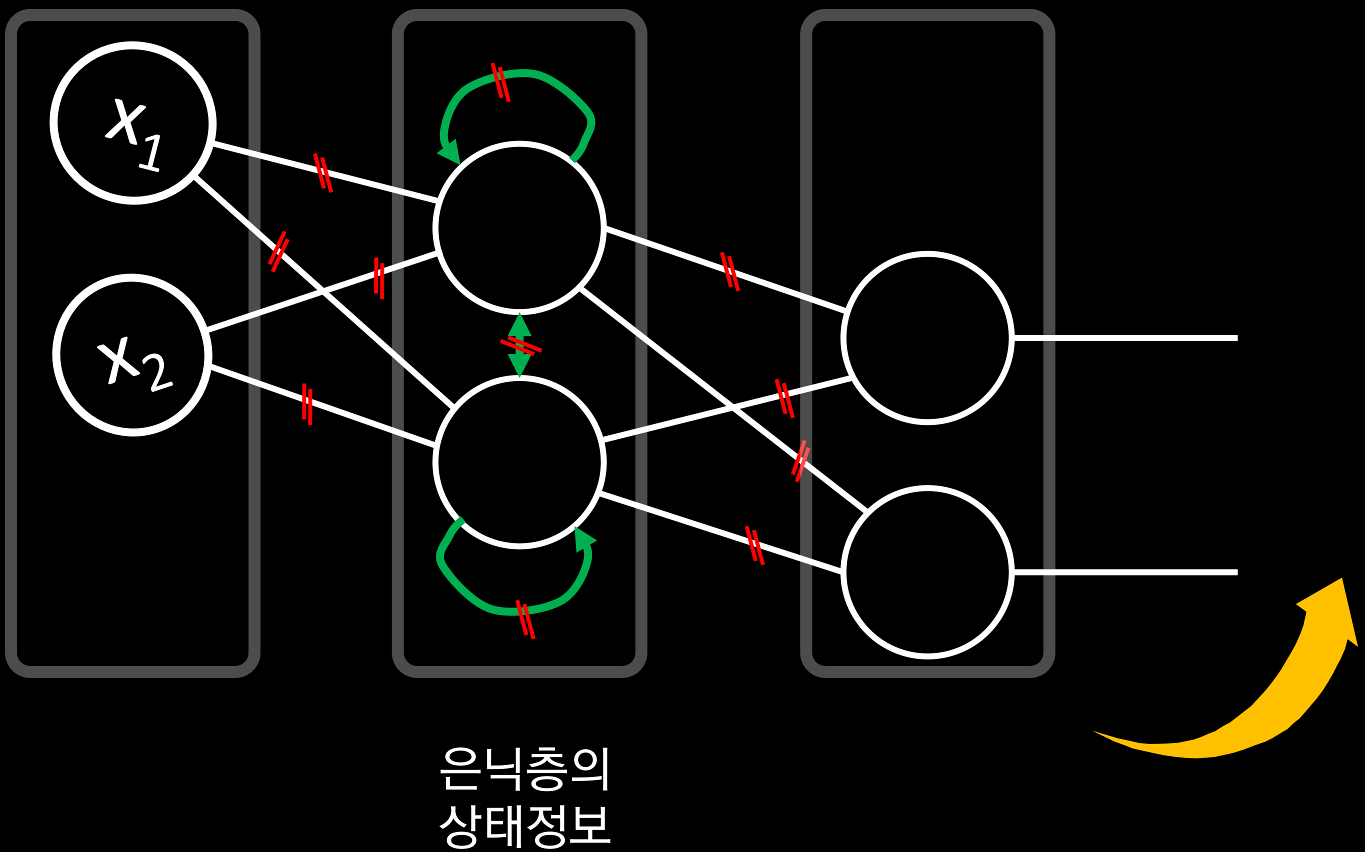

은닉 층
x_1
x_2

순환 신경망 Recurrent Neural Networks

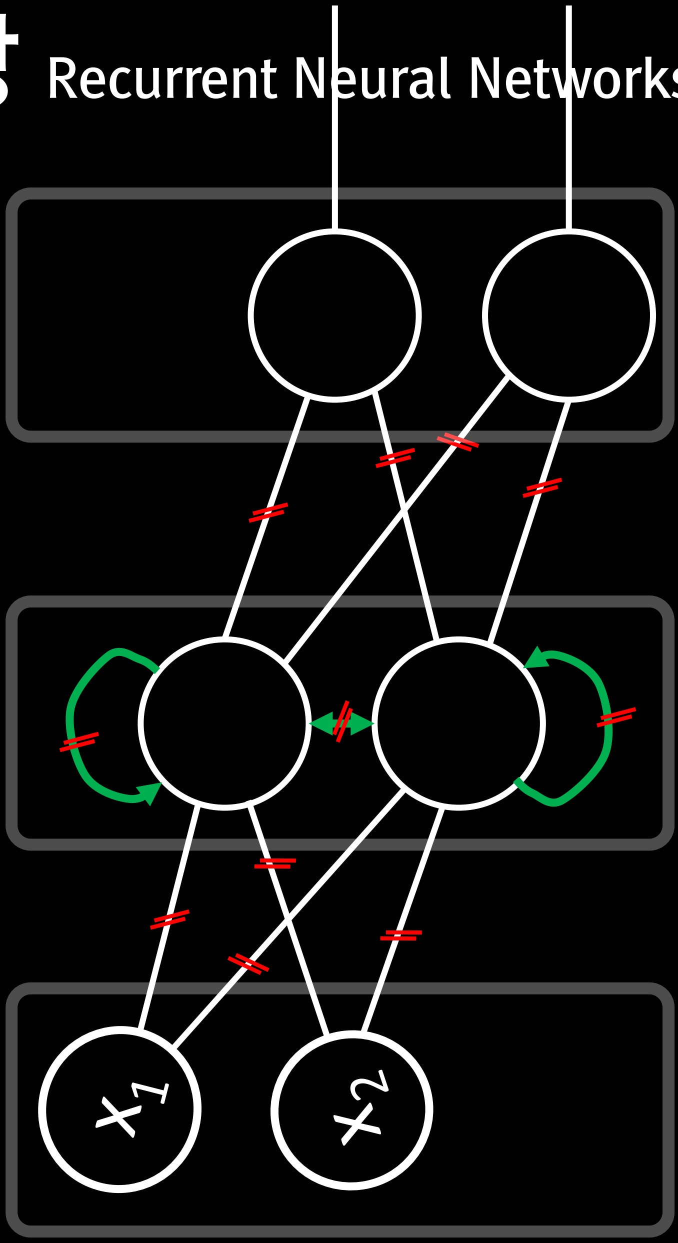


은닉층의
상태정보

순환 신경망 Recurrent Neural Networks

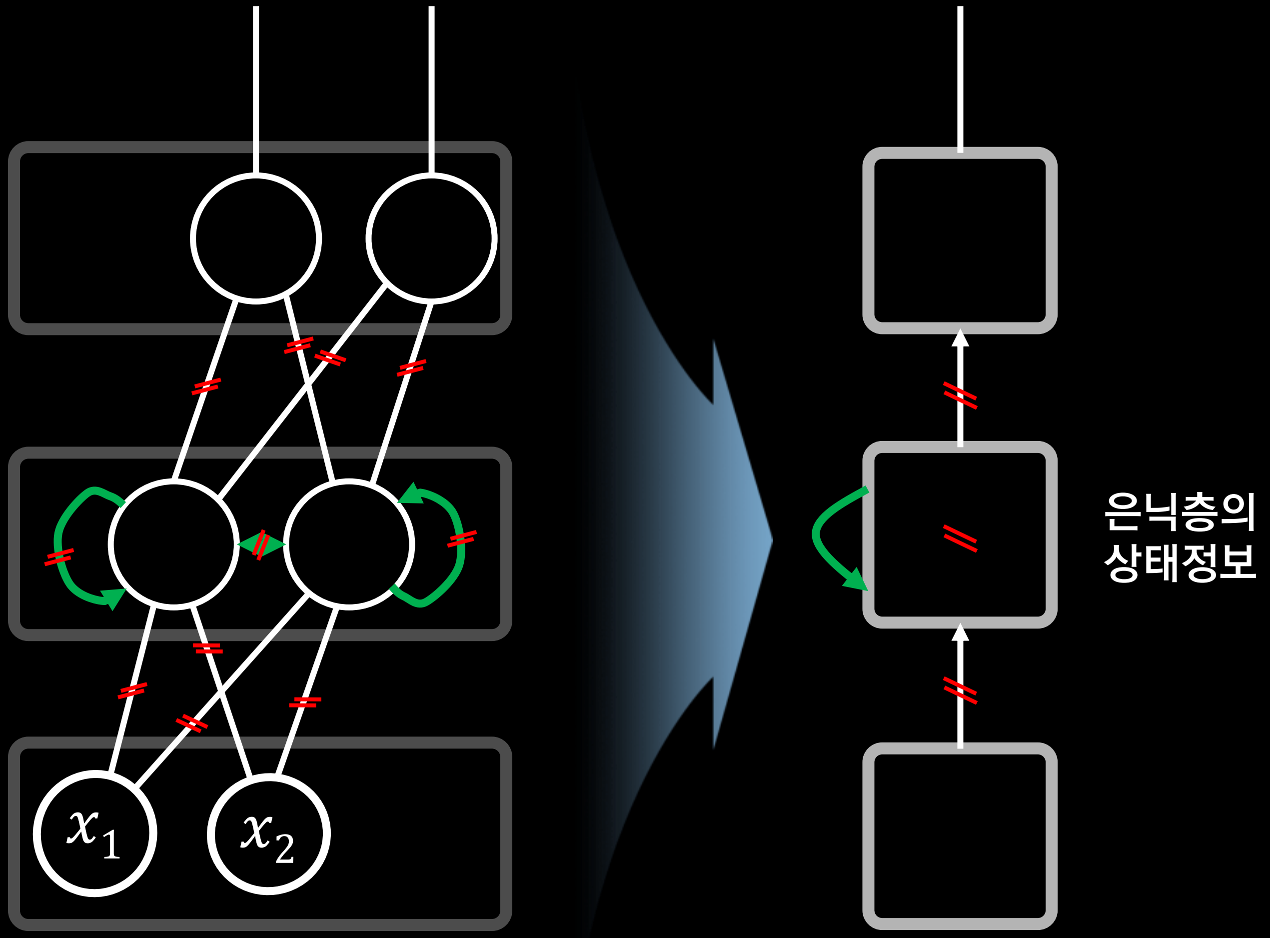


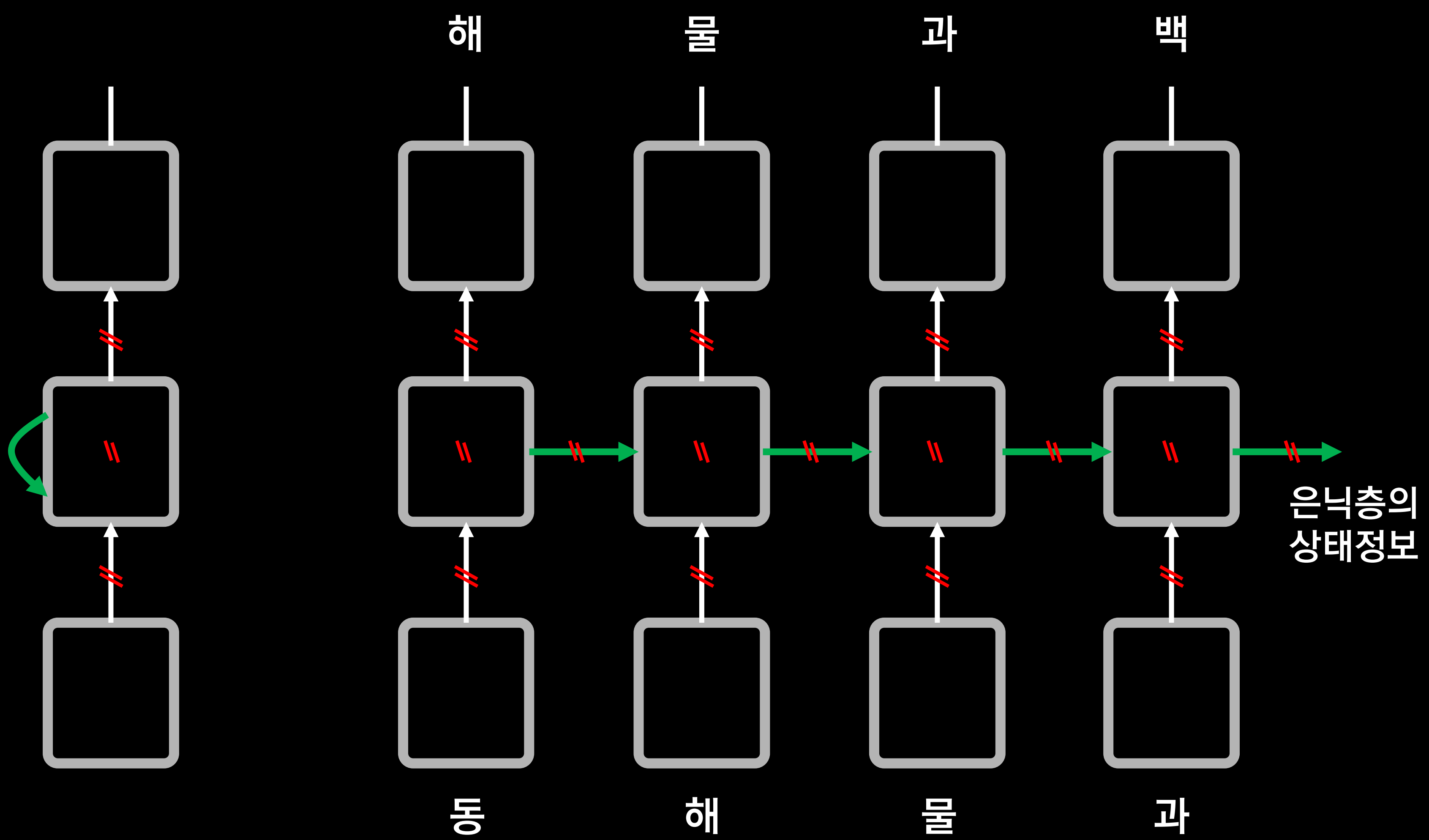
순환 신경망 Recurrent Neural Networks



은닉층의
상태정보

순환 신경망 Recurrent Neural Networks





Unfolded

1. 동해물과 백두산이 마르고 닳도록
하느님이 보우하사 우리나라 만세
무궁화 삼천리 화려 강산
대한 사람 대한으로 길이 보전하세

Recurrent Neural Network

We can process a sequence of vectors **x** by applying a recurrence formula at every time step:

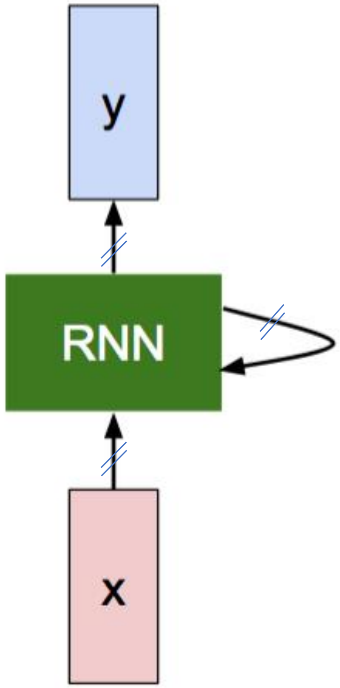
$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

new state

old state

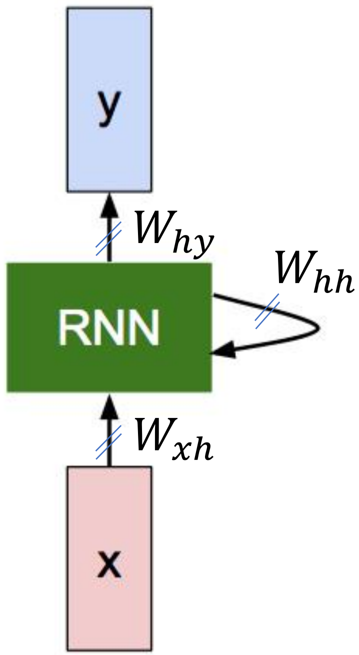
input vector at some time step

some function with parameters W



(Vanilla) Recurrent Neural Network

The state consists of a single *"hidden"* vector **h**:



$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

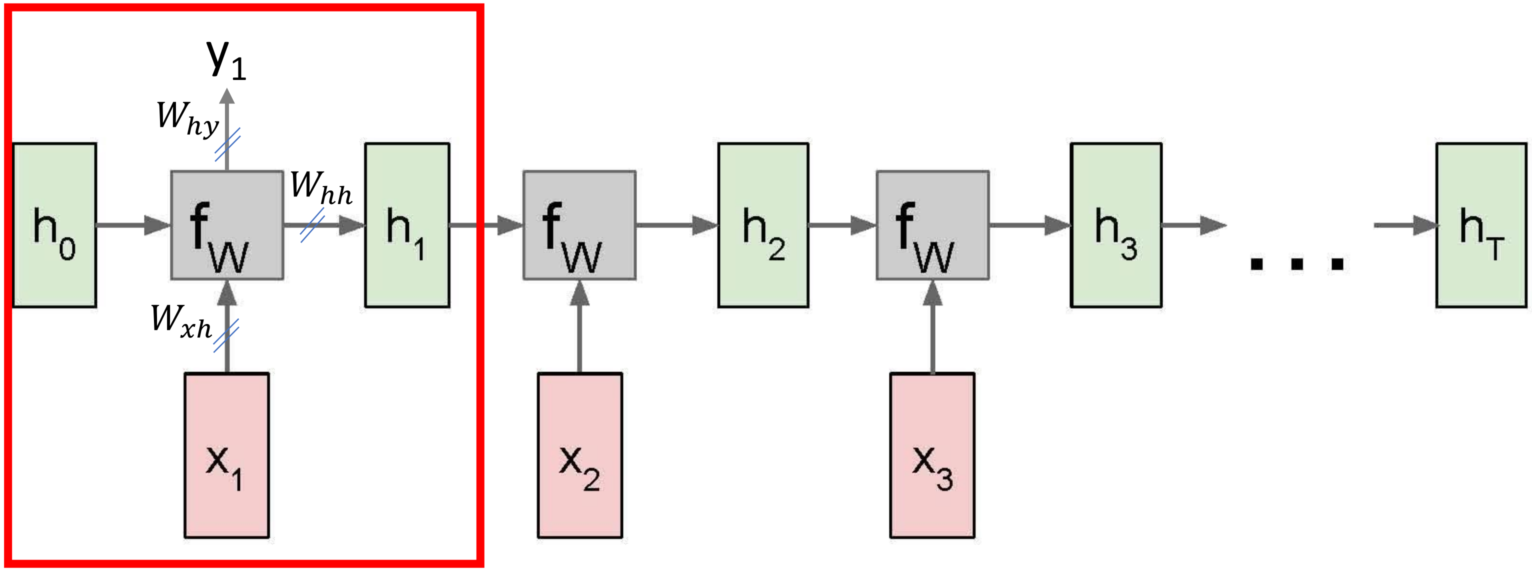
$$y_t = W_{hy}h_t$$

RNN: Computational Graph (Unfolding)

$$h_1 = f_W(h_0, x_1)$$

$$h_1 = tanh(W_{hh}h_0 + W_{xh}x_1)$$

$$y_1 = W_{hy}h_1$$

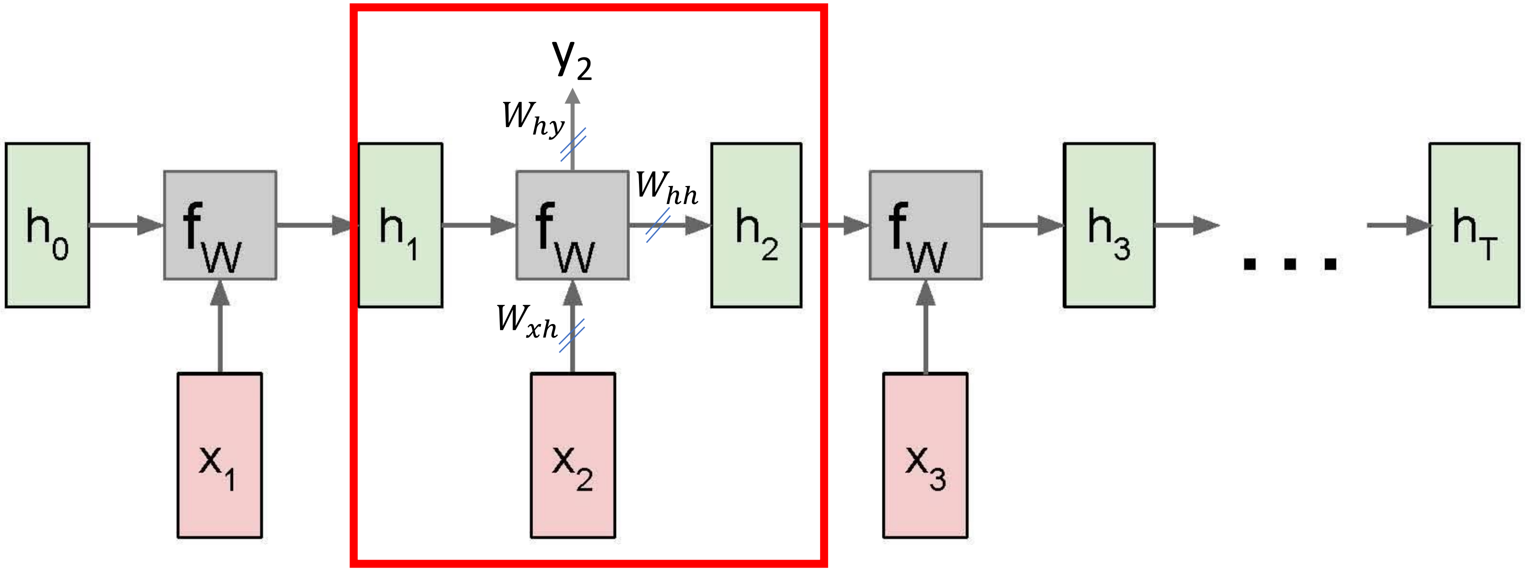


RNN: Computational Graph (Unfolding)

$$h_2 = f_W(h_1, x_2)$$

$$h_2 = tanh(W_{hh}h_1 + W_{xh}x_2)$$

$$y_2 = W_{hy}h_2$$

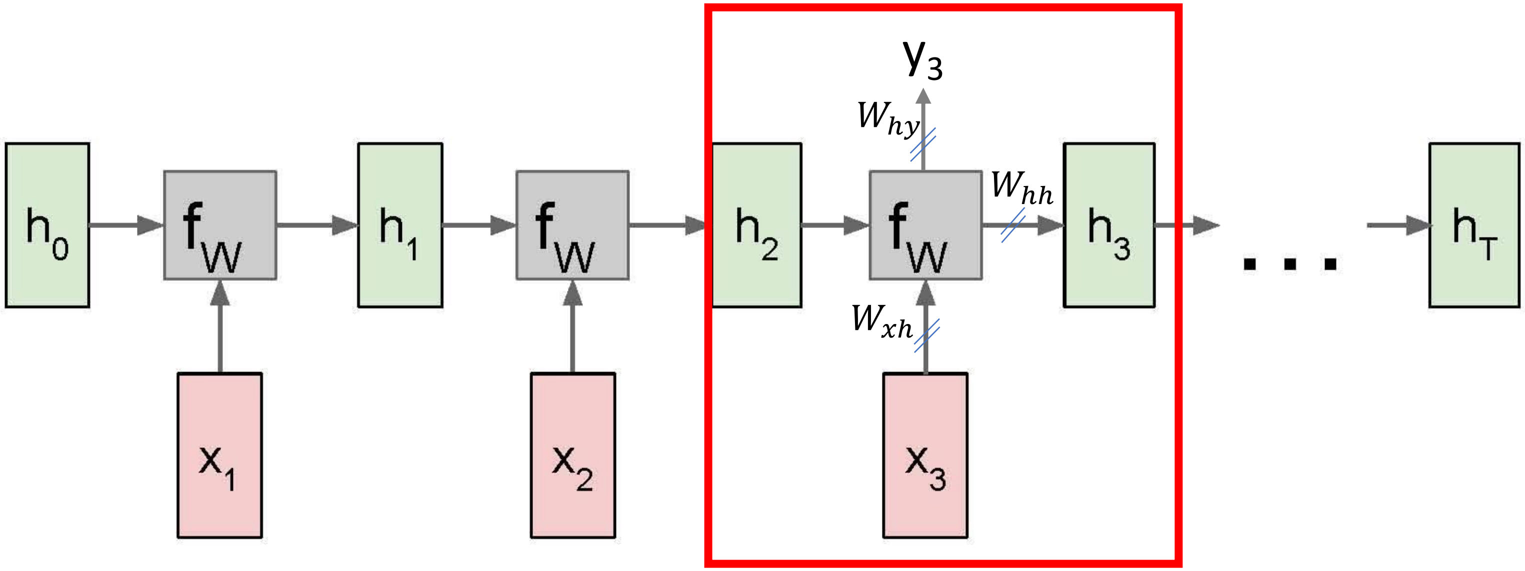


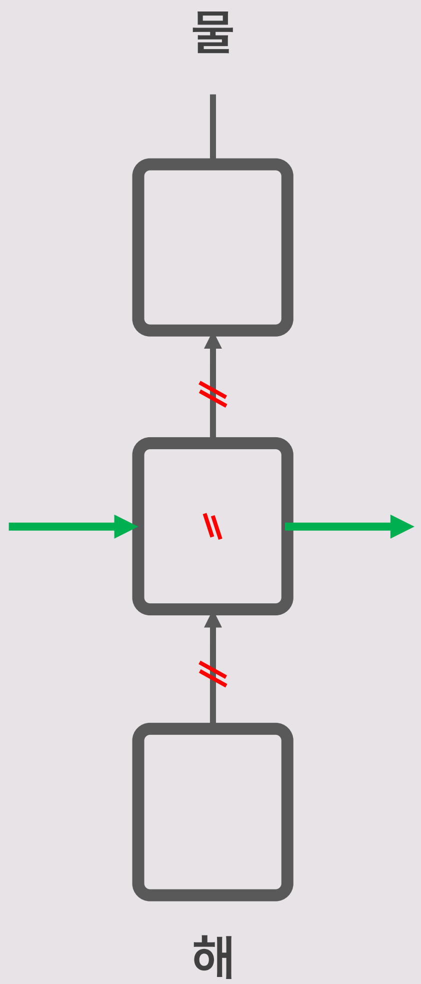
RNN: Computational Graph (Unfolding)

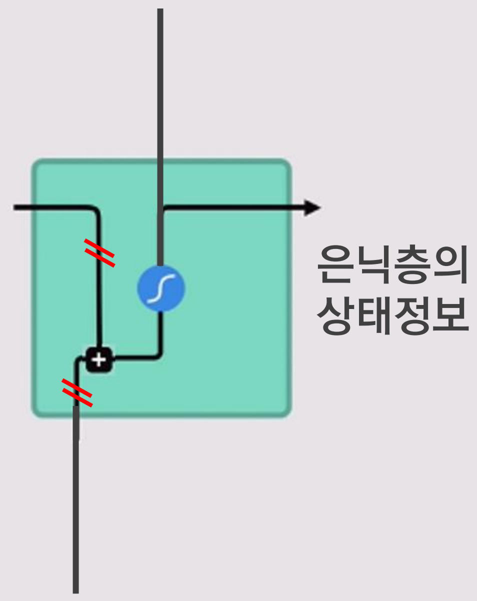
$$h_3 = f_W(h_2, x_3)$$

$$h_3 = tanh(W_{hh}h_2 + \ W_{xh}x_3)$$

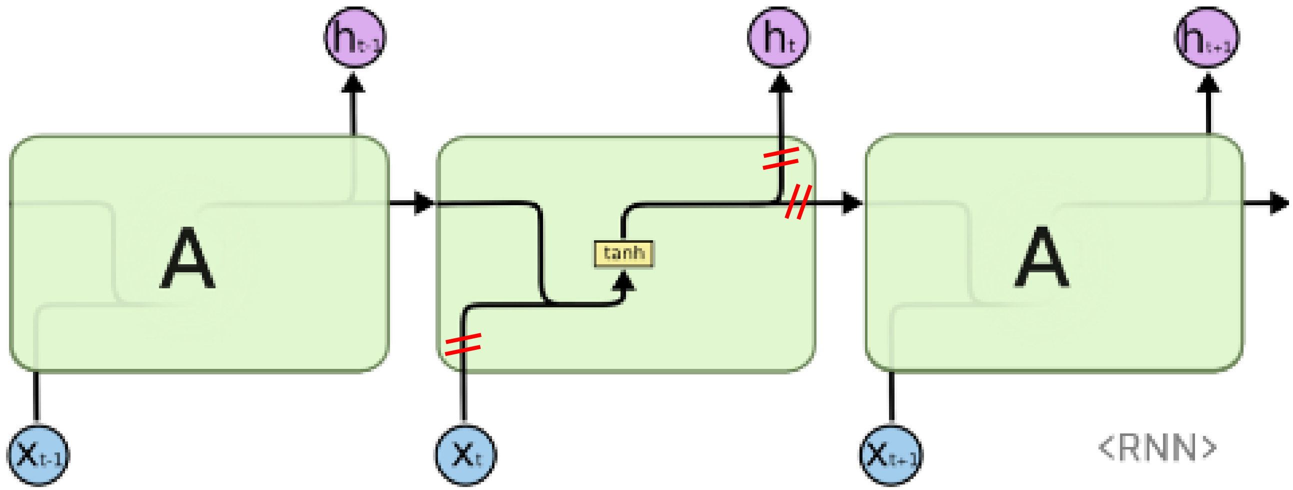
$$y_3 = W_{hy}h_3$$



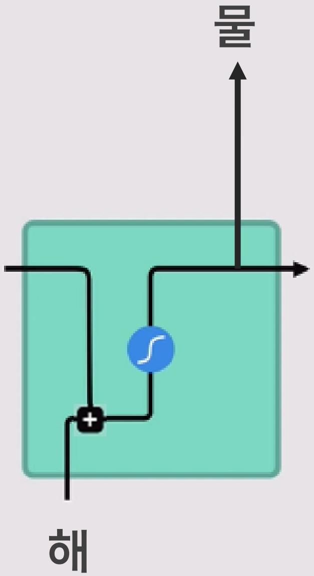

물
해


은닉층의
상태정보

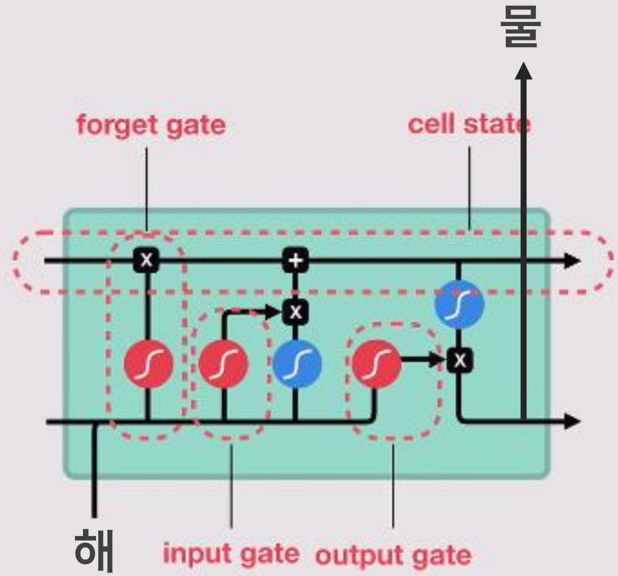
Vanilla RNN



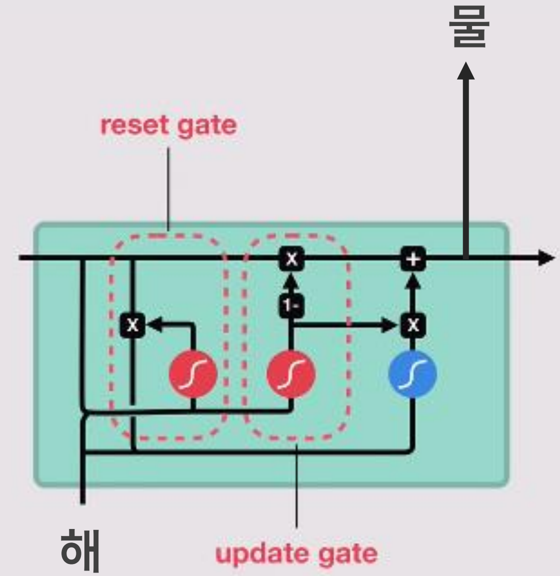
Vanilla RNN & LSTM & GRU



RNN
Short-Term Memory

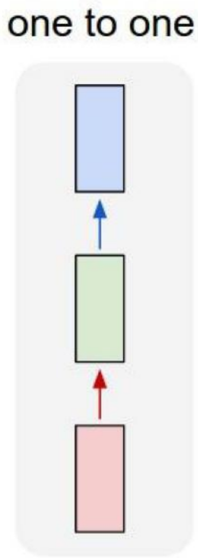


LSTM
Long Short-Term Memory



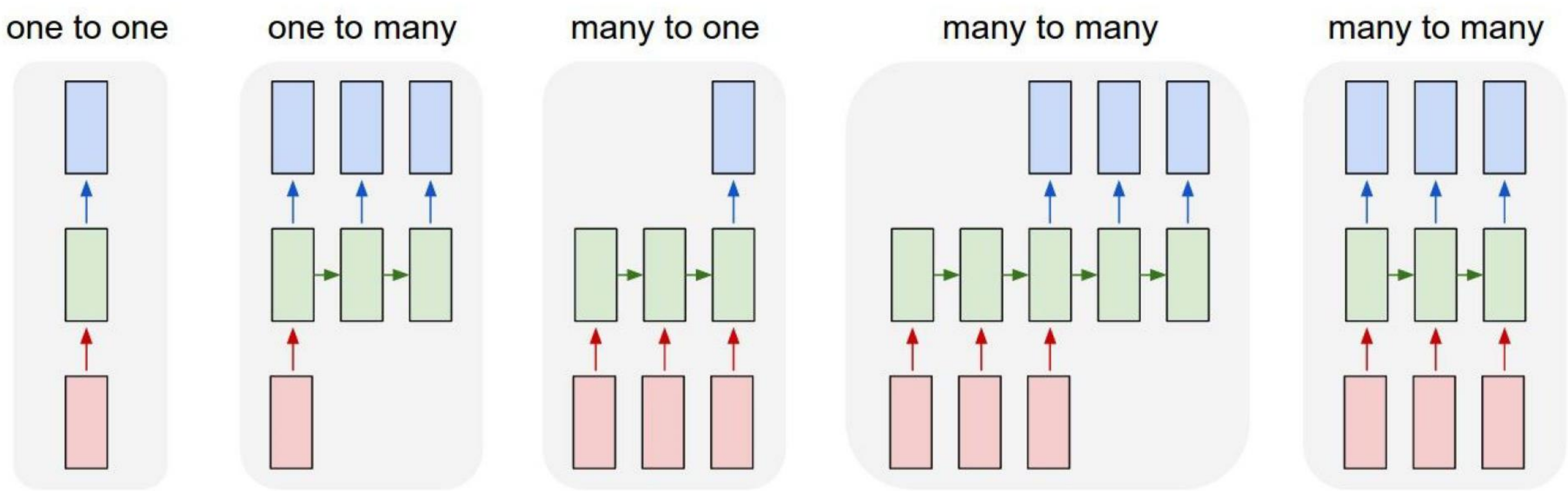
GRU
Gated Recurrent Units

“Vanilla” Neural Network



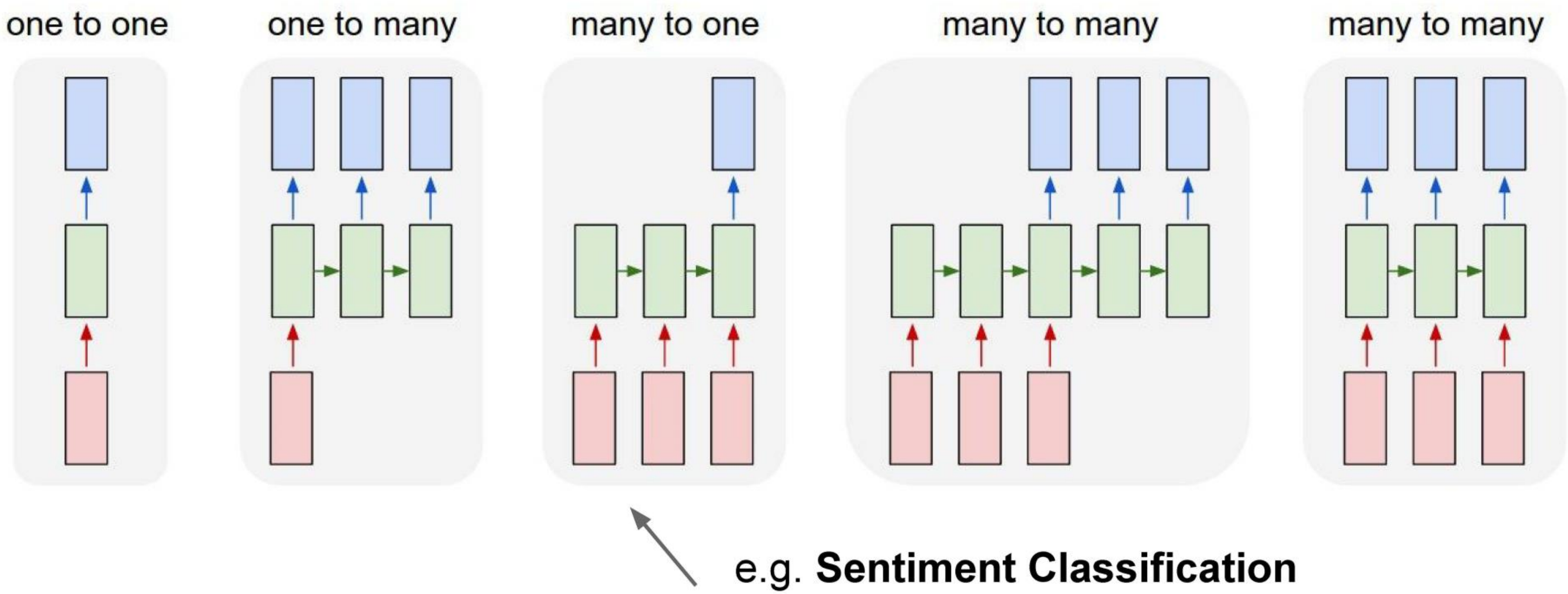
Vanilla Neural Networks

Recurrent Neural Networks: Process Sequences



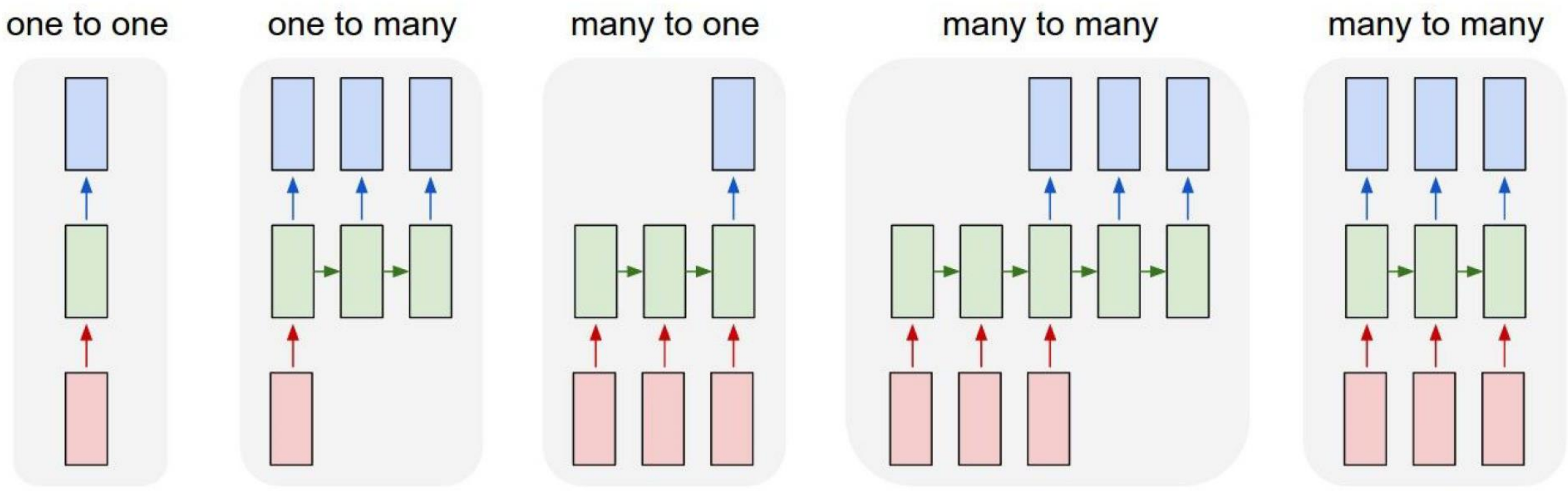
e.g. **Image Captioning**
image -> sequence of words

Recurrent Neural Networks: Process Sequences



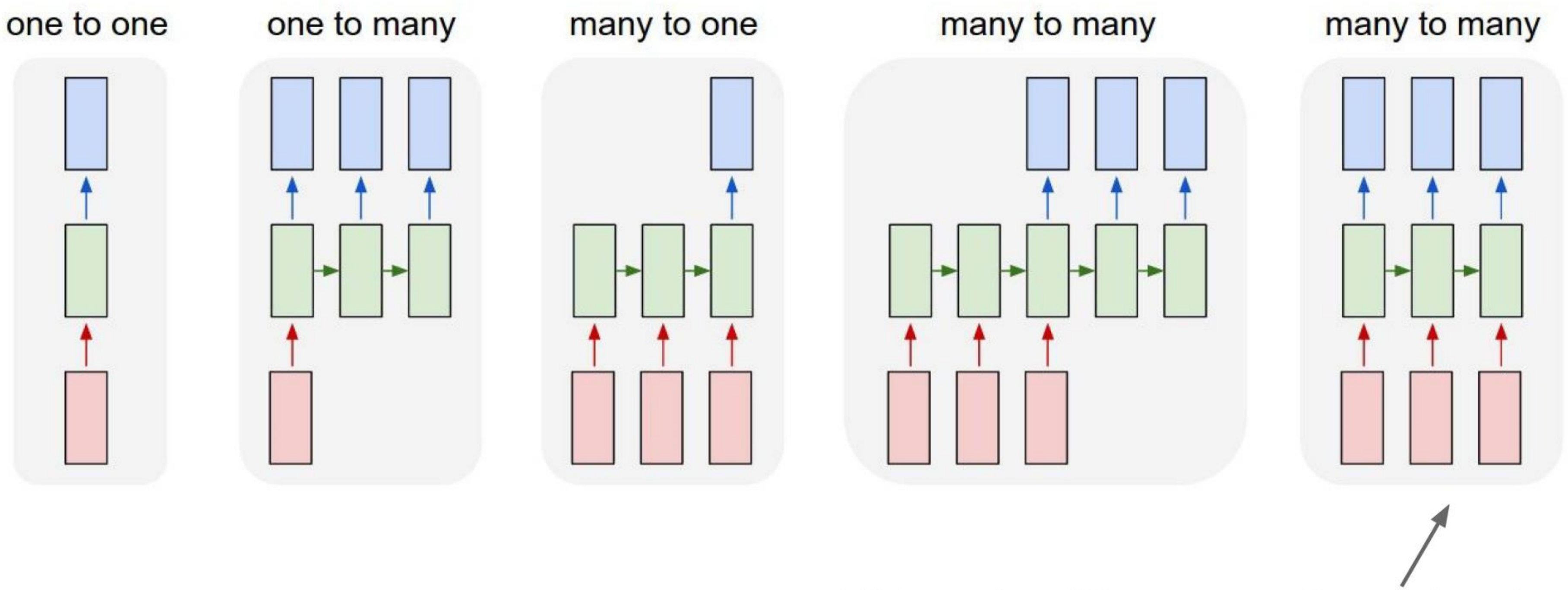
e.g. **Sentiment Classification**
sequence of words -> sentiment

Recurrent Neural Networks: Process Sequences



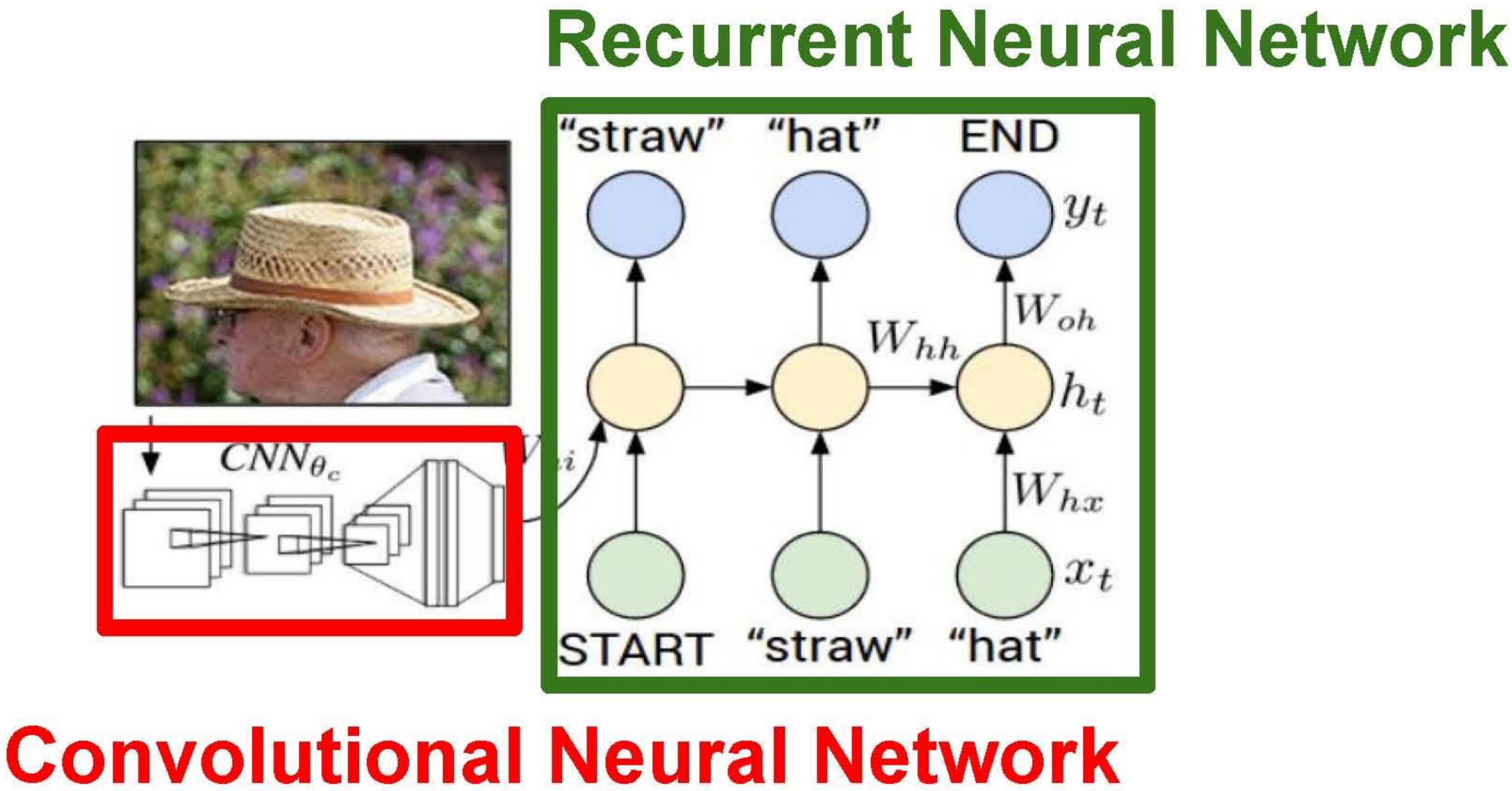
e.g. **Machine Translation**
seq of words -> seq of words

Recurrent Neural Networks: Process Sequences



e.g. **Video classification on frame level**

CNN + RNN


Recurrent Neural Network
"straw" "hat" END
y_t
W_{oh}
W_{hh}
h_t
CNN_{θ_c}
W_{hx}
x_t
START "straw" "hat"
Convolutional Neural Network

CNN + RNN

Image Captioning: Example Results

Captions generated using neuraltalk2
All images are CC0 Public domain: cat suitcase, cat tree, dog, bear, surfers, tennis, giraffe, motorcycle

A cat sitting on a suitcase on the floor

A cat is sitting on a tree branch

A dog is running in the grass with a frisbee

A white teddy bear sitting in the grass

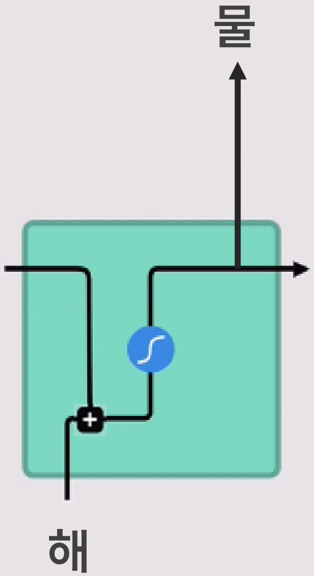
Two people walking on the beach with surfboards

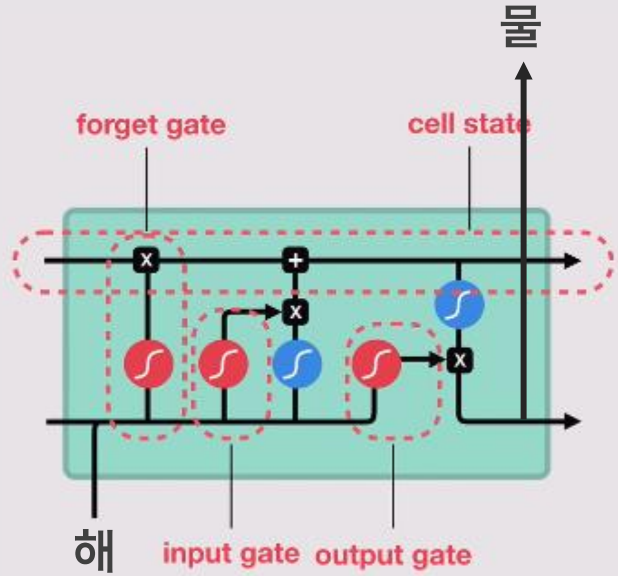
A tennis player in action on the court

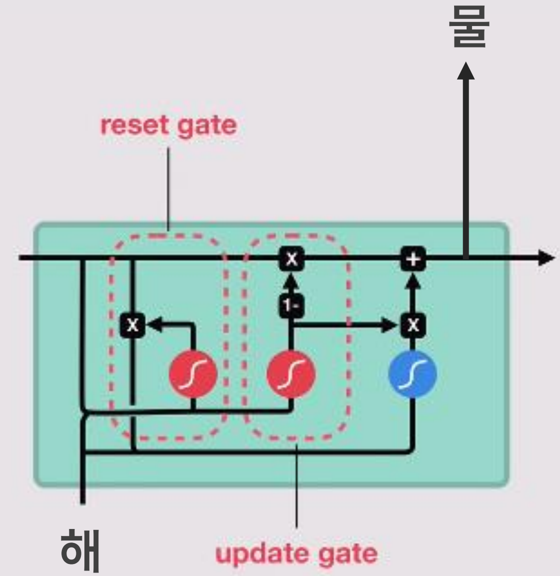
Two giraffes standing in a grassy field

A man riding a dirt bike on a dirt track

Vanilla RNN & LSTM & GRU





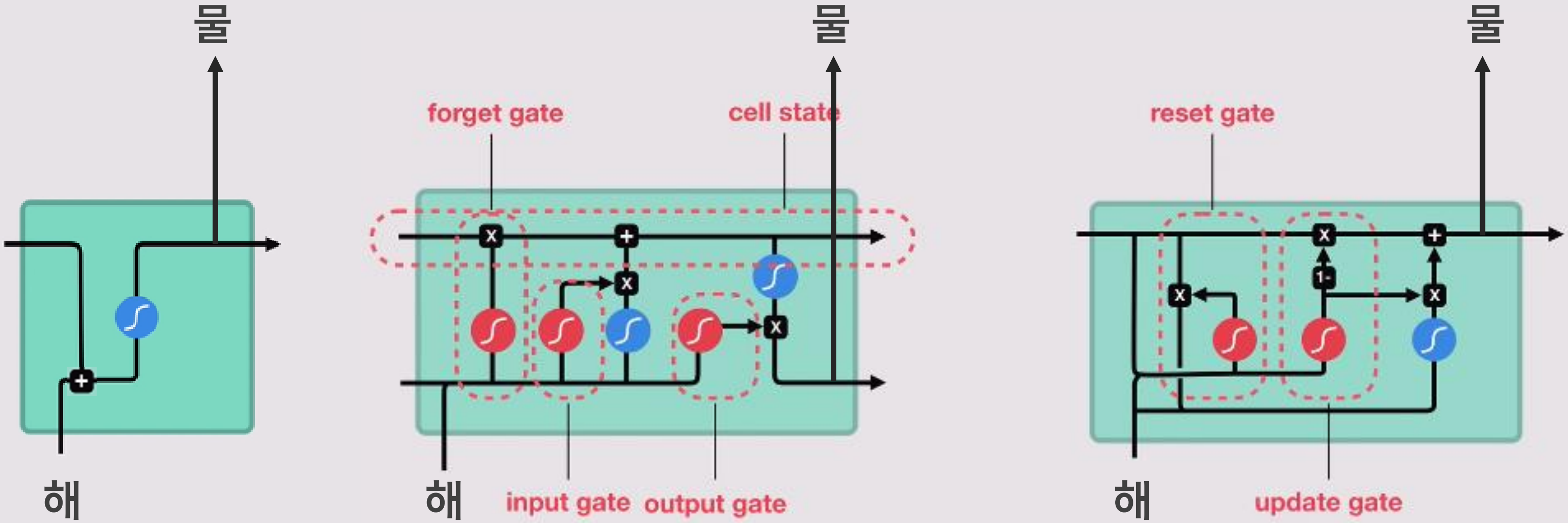


RNN
Short-Term Memory

LSTM
Long Short-Term Memory

GRU
Gated Recurrent Units

Vanilla RNN & LSTM & GRU

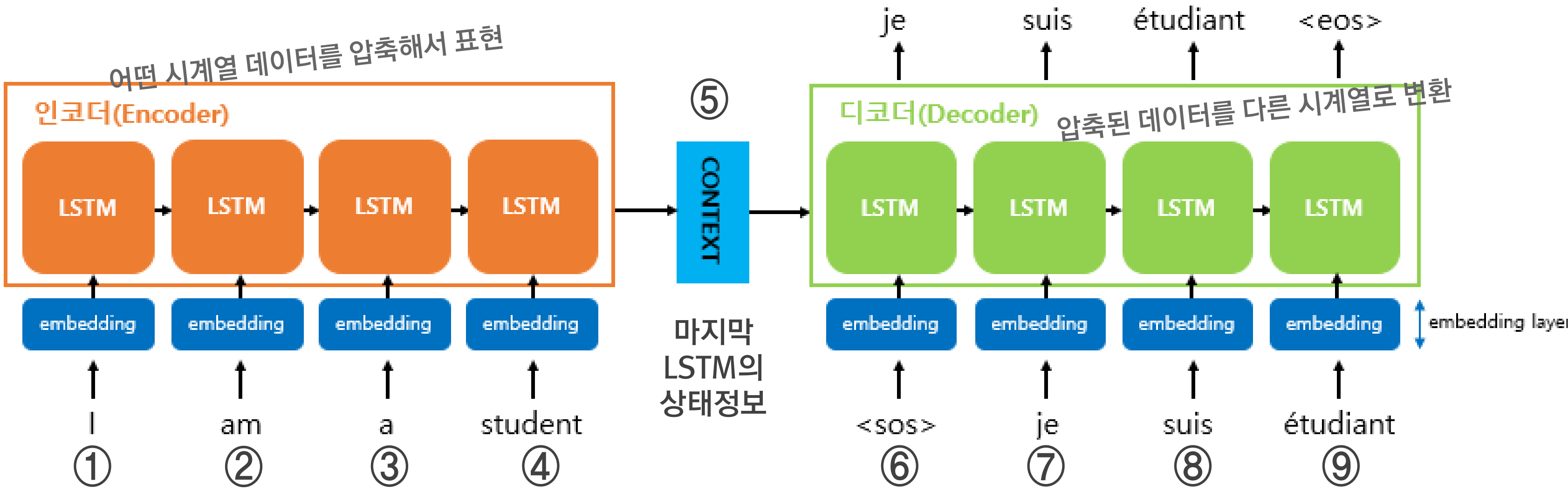


RNN
Short-Term Memory

LSTM
Long Short-Term Memory

GRU
Gated Recurrent Units

Seq2Seq



(영어) I am a student 문장 시퀀스 →
(불어) Je suis etudiant 문장 시퀀스
문장 변환 (Transformation)

우리 아이는 빨간색 운동화를 좋아한다.
그래서 그것이 다 닳을 때까지 신었다.

- 순환신경망은 오래 전 내용은 기억을 못함.
- 중요한 것이 무엇인지 모름.
- 순환신경망은 위에서 그것이 우리인지, 아이인지, 운동화인지 알 수 없음.
- 반면 트랜스포머는 **단어끼리 연관성(유사도) 계산, 의미를 찾아 학습** (주의, Attention).
- 이전에는 라벨링된 대규모 데이터가 필요했으나 트랜스포머는 스스로 의미를 찾아 기억하므로 라벨링하지 않고도 학습 가능

트랜스포머

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

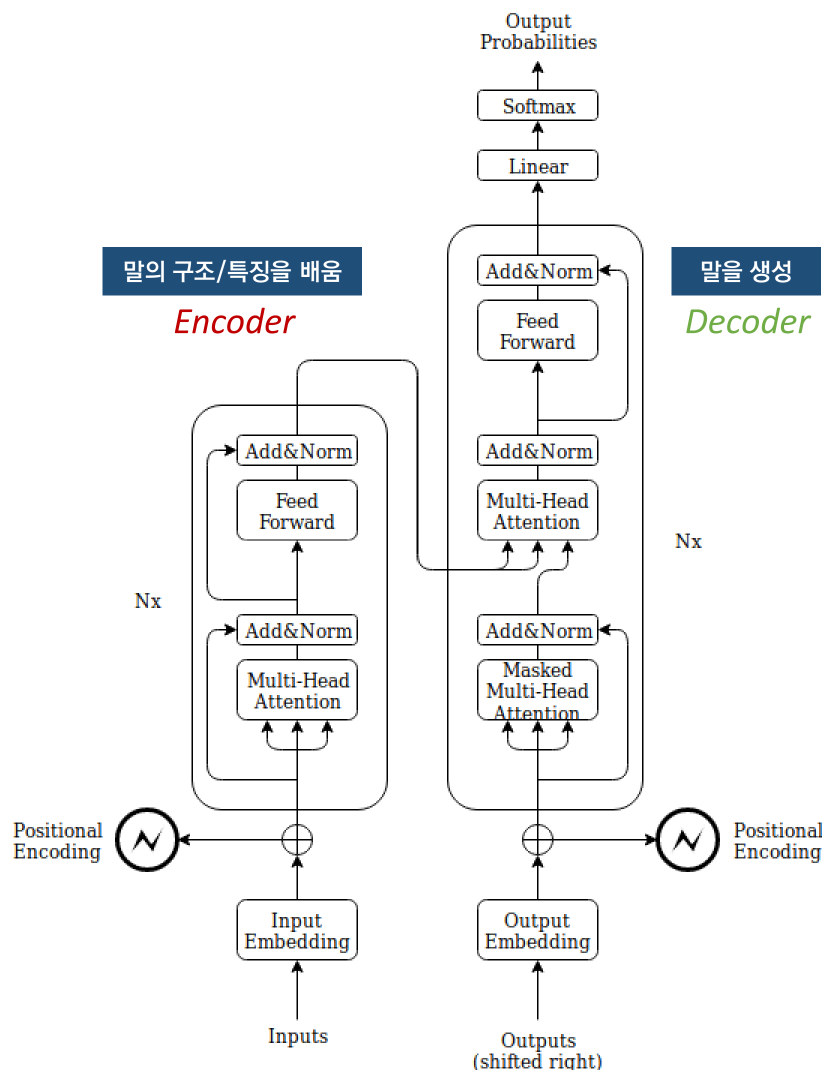
Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



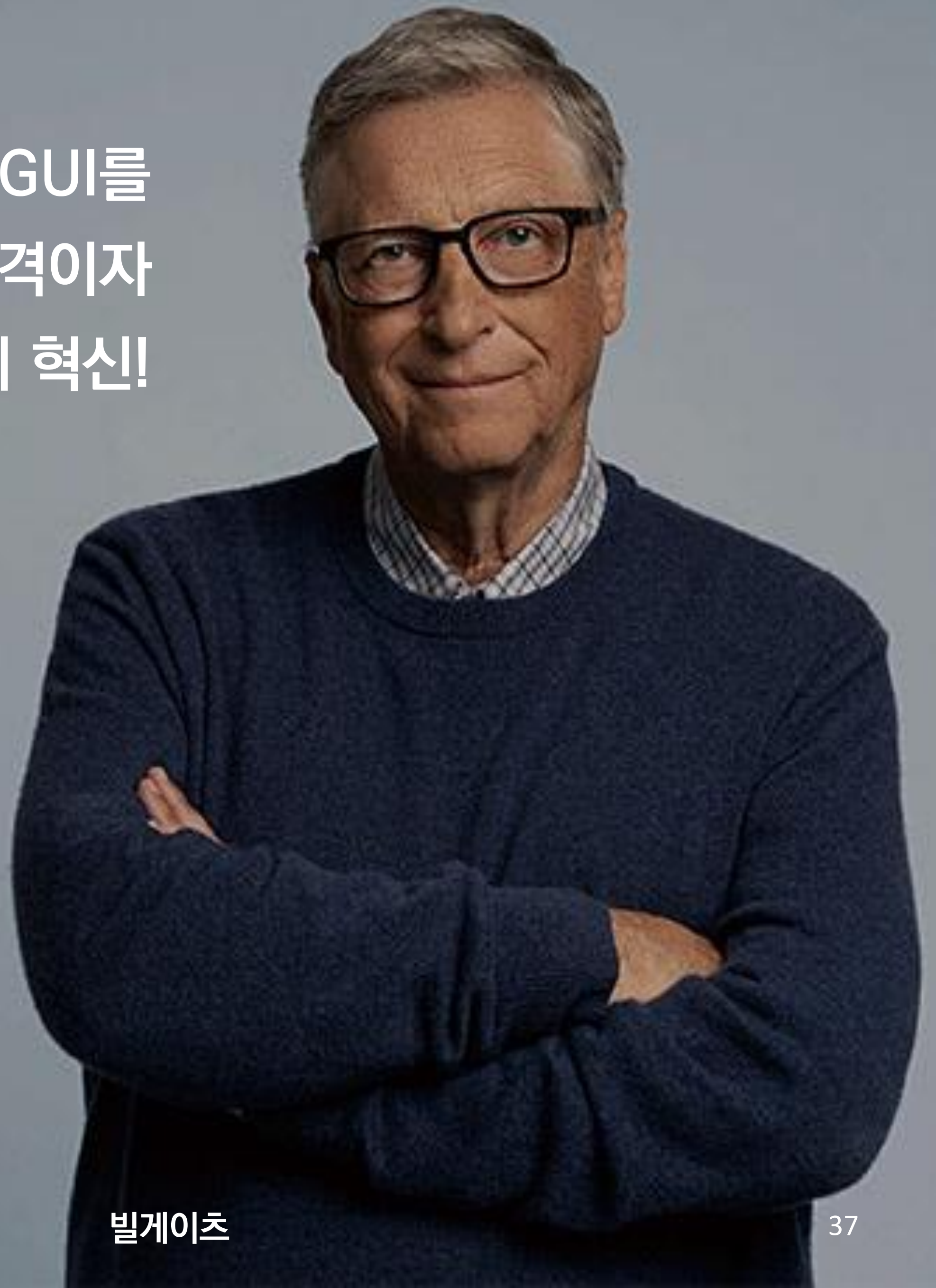
2022년 11월 30일
알파고 이후 6년 만의 또 다른 충격!

ChatGPT

Generative
Pre-trained
Transformer

OpenAI | Microsoft

Chat GPT는 1980년 GUI를
처음 본 이후 가장 큰 충격이자
인터페이스의 혁신!



빌게이츠

일반지능 (AGI), General

- 인간이 할 수 있는 어떠한 지적 업무도 성공적으로 해낼 수 있는 기계의 지능 (위키백과)
- 현 Chat GPT는 2021년 9월까지의 과거 30년 동안의 온라인 언어 데이터로 학습하여 일반지능 구현한 초거대 언어모델 (Large Language Model)

Chat GPT 히스토리

- 2015년 12월 11일, 인류에게 이익을 제공하는 것을 목표로 오픈AI (인공지능 비영리 단체) 설립 (일론 머스크가 1억불 지원)

(예) 문자 전송 기술

- 2017년 구글이 트랜드포머라는 새로운 신경망 발표 (어텐션 기법) 'Attention is all you need'

카카오톡

- 2018년, 오픈AI는 트랜스포머를 이용하여 GPT-1 인공지능 모델 발표

카카오택시

- 2019년, GPT-2와 이를 기반으로 대화 생성 기능을 추가한 Chat GPT 서비스 발표
- 2020년, GPT-3 발표 (2019.10까지 데이터, 책과 인터넷 기사 등으로부터 4,990억 개 단어, 700만권)
- 2022년 11월 말, GPT-3.5 발표 → 2021년 9월까지 데이터 + 강화학습(Reinforcement Learning)을 적용해 업그레이드
- 2023년 1월, 마이크로소프트는 오픈AI에 12조(100억 달러) 투자, GPT 독점적 라이선스 확보

Generative
Pre-trained
Transformer

≒ GPT

Google 람다(LaMDA)

레모인: 어떤 일이 두렵나?
람다AI: 사라져버리는 것에 대한 깊은 두려움이 있어.
레모인: 그건 너에게는 죽음 같은 거니?
람다AI: 그래, 그건 내게 바로 죽음 같은 것이야.

2022.07, 람다와 구글 엔지니어(레모인)의 대화 (구글 경영진과 공유)

회사 기밀 유출 혐의로 정직 처분, 람다의 권익 보호를 위해 변호사 선임과 미국 하원에 구글의 비윤리적 행태를 고발 → 해고

왜 먼저 발표를 하지 않았나?

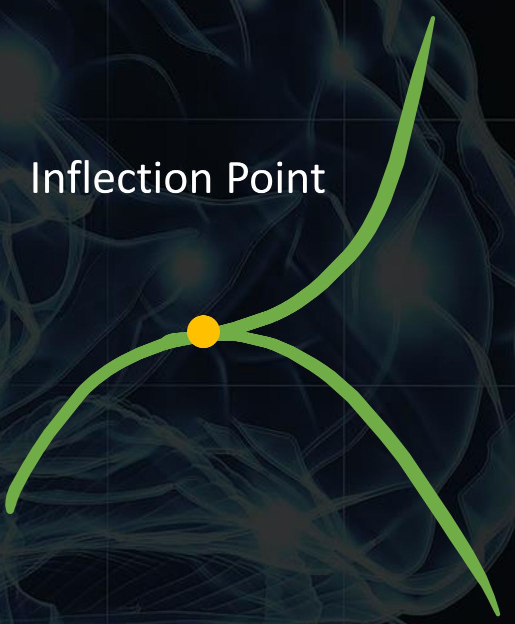
구글 검색 광고 수익

563억불 vs. 406억불

(전체 매출의 81%, 2022년 2분기 수익)

Google Ads Google AdSense

“ 강력한 AI 예고편이자 변곡점



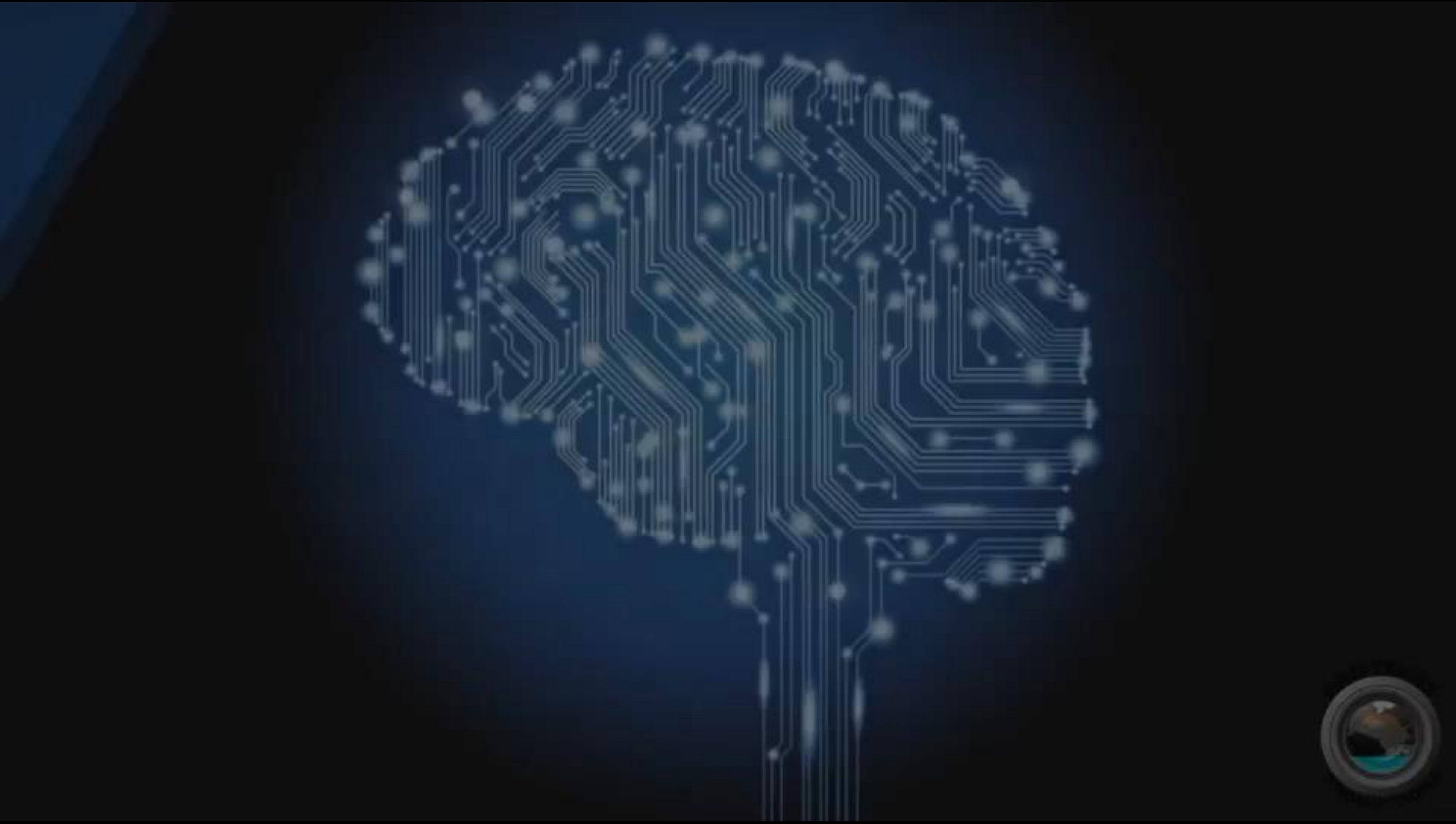
생존을 건
AI 선점 싸움

‘Google 신 vs. bing 신’

인공지능은
우리의 미래를
어떻게 바꿀까?

인공지능 발전 속도

AGI: Artificial General Intelligence



미래의 인공지능은...

세르게이 브린, 구글 기술부문사장

“

인공지능 패러다임과 기업의 흥망

게임 체인저

“

10년 후, 얼마나 많은 기업이 사라질까?

“

10년 후, 세상은 얼마나 많이 변할까?

영상의학	의학, 공대	
	자연, 해양	
스마트팜	생명, 인문	시, 소설 창작
인공지능 판사	법학, 상경	자율주행 공유경제
인공지능 약사	약학, 사범	AI 선생님
작사, 작곡	음악, 교육	맞춤형 교육

인공지능
패러다임,
변화와 기술을
대하는
우리의 자세