How AI solve the social issue of the health care business

NLP In the Medical Field

31 January, 2021

#sg_ai_world

Hideaki Hamano

Personal Background

Public Background :

- 15+ years USA/UK/EU/Tokyo Financial Industry (Tokyo/London) :
- Exotic Derivatives Trading Model Development and Analytics
- Managing Director / Researcher in the health care business (Tokyo)

Winning Award:

- Kaggle : Top 8% **Bronze medal** (Zillow Properties Price Expectation)
- Tokyo University : Finalist research group at the 2018 Spring Laboratory Competition (Dr. Matuso Laboratory 2017 DeepLearning JP)
- Tokyo University : No1 Natural Learning Process(Translation) Competition (Dr. Matuso Laboratory 2017 DeepLearning JP)

Today's Topic

01

NLP

Technical Briefing

02

AI impacts Health care Business

03

Conclusion

1.Linguistics model

Natural Language Processing

What is NLP(Natural Language Processing)?

Natural language processing (**NLP**) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. The result is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

(From wikipedia)

Development Story

Low Level Approach
Mechanical method

High Level Approach
Humanical method

Until 1990

Machine learning approach

Hybrid approach using the philosophy and the linear algebra etc.

From 2000s

Deep Leaning Model to Recurrent Neural Network and LTSM (Long Term Short Memory, 1997)

Beyond 2017

Transformer model successfully gained the superb result to analyze the linguistic message. (Translation)

https://arxiv.org/abs/1706.03762

Beyond 2018

B E R T mdoel

BioBert

SciBert

Clinical Bert

AREA

Classification

Classification

Automatic Linguistic Translation

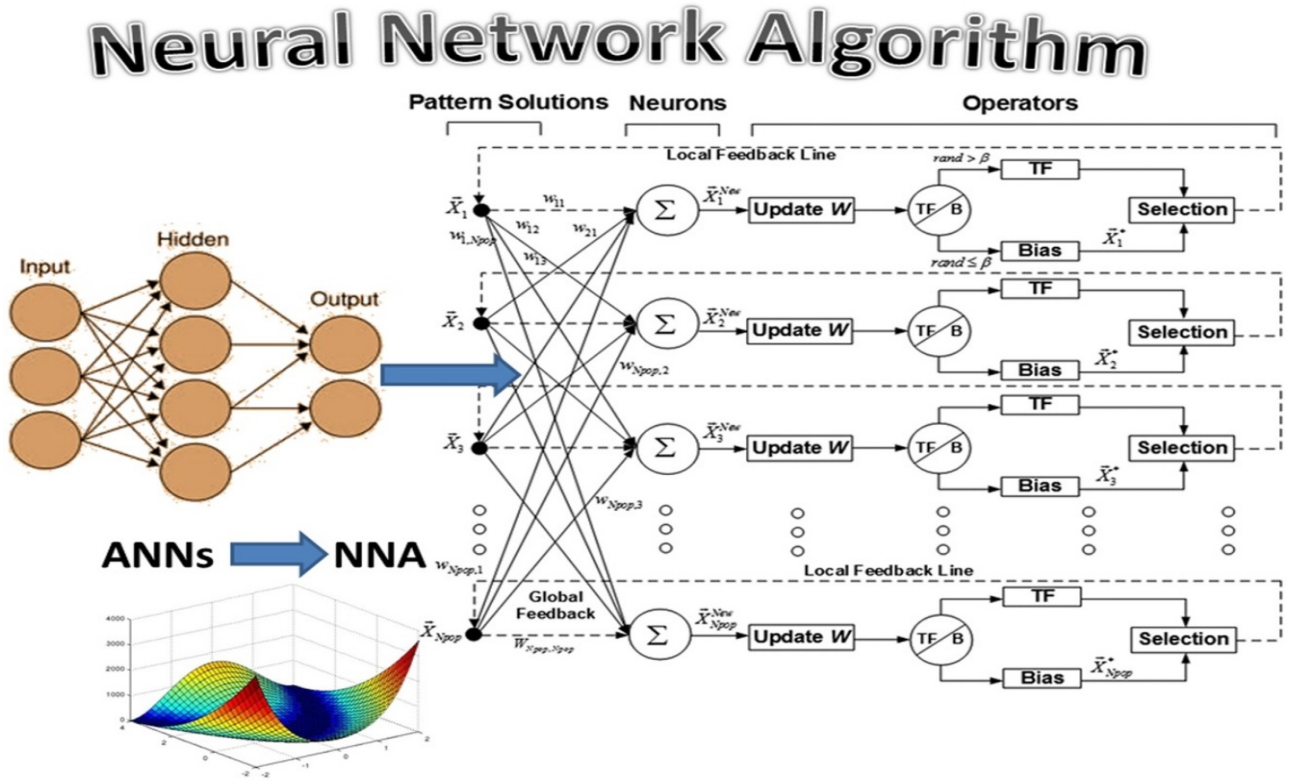
Data Searching

Emotional Analysis

Building the structured sentence

NLP Technical Briefing 1 a

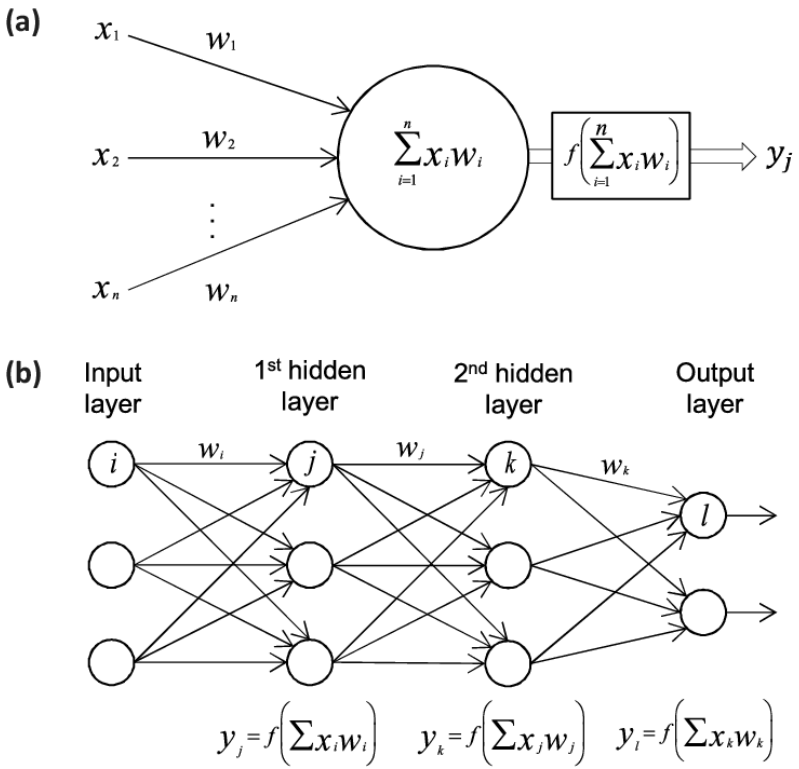
Neural Network

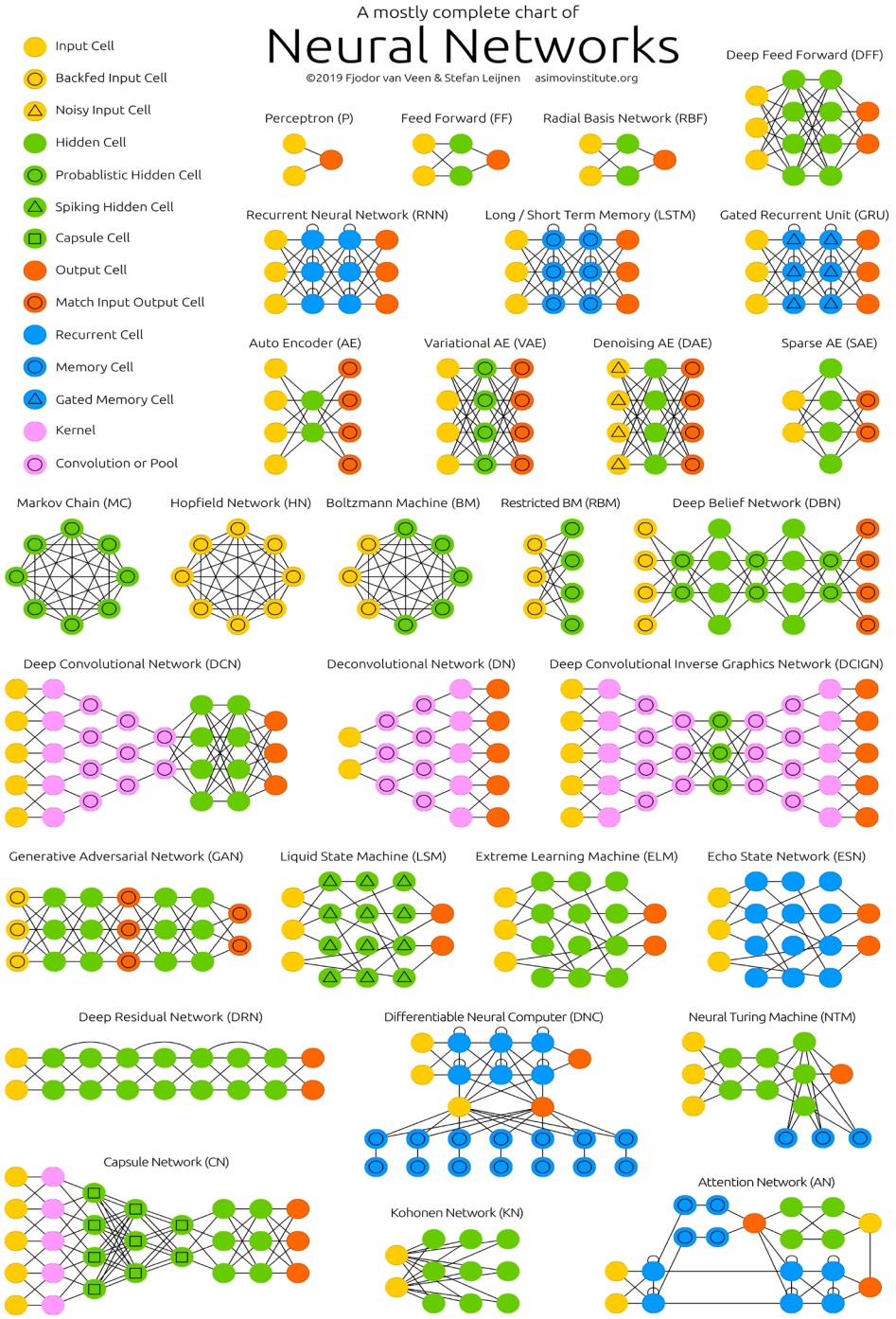


A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria.

(Figure : MathWorks Wikipedia)

NLP Technical Briefing 1 b
Neural Network





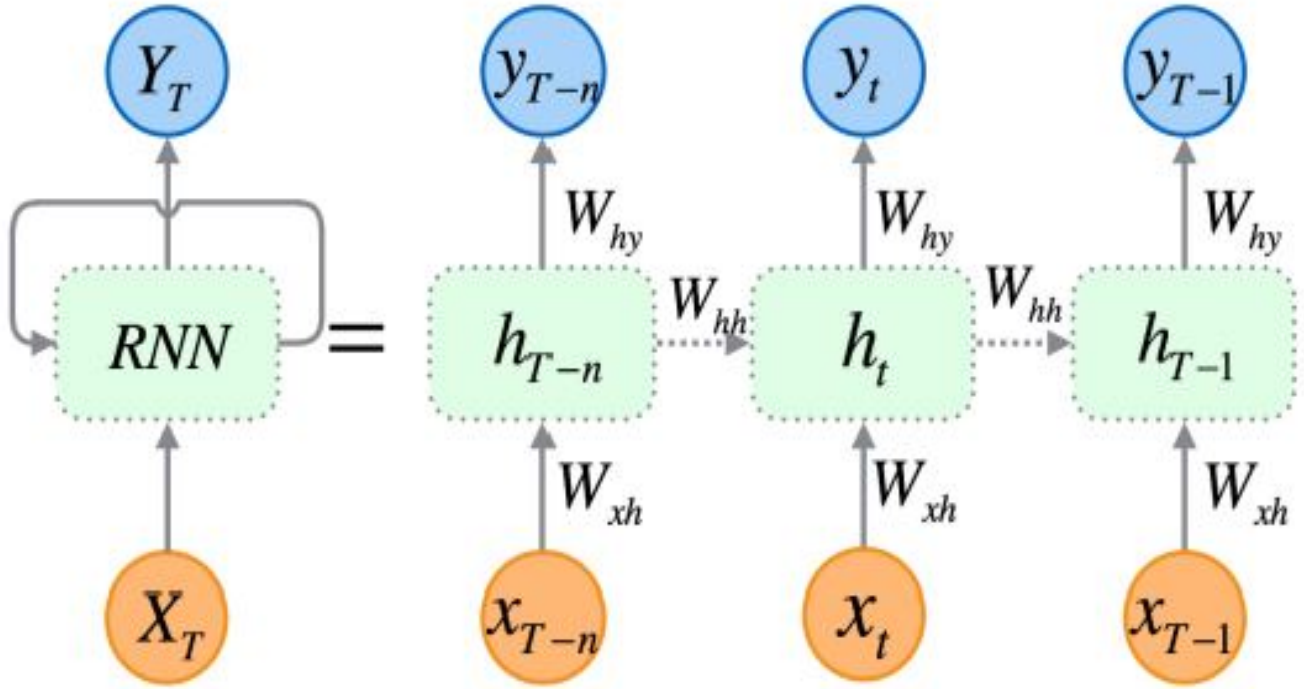
Van Veen, F. & Leijnen, S. (2019). The Neural Network Zoo. Retrieved from https://www.asimovinstitute.org/neural-network-zoo

Sequence data

$$x_0, x_1, x_2, x_3 \ \ldots\ x_n$$

Think about climate ?

NLP Technical Briefing 1 b
RNN(Recurrent Neural Network)

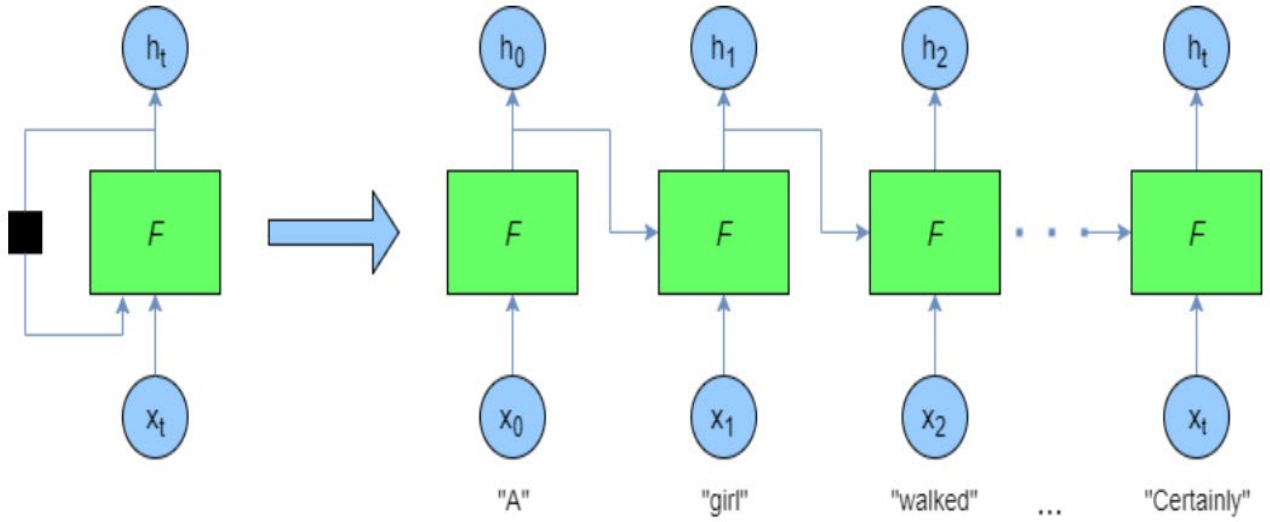


1 Standard RNN architecture and an unfolded structure with T time

Recurrent Neural Network(RNN) accepts the inputs of some state information data.

(Figure : MathWorks Wikipedia)

NLP Technical Briefing 1 c
RNN(Recurrent Neural Network)



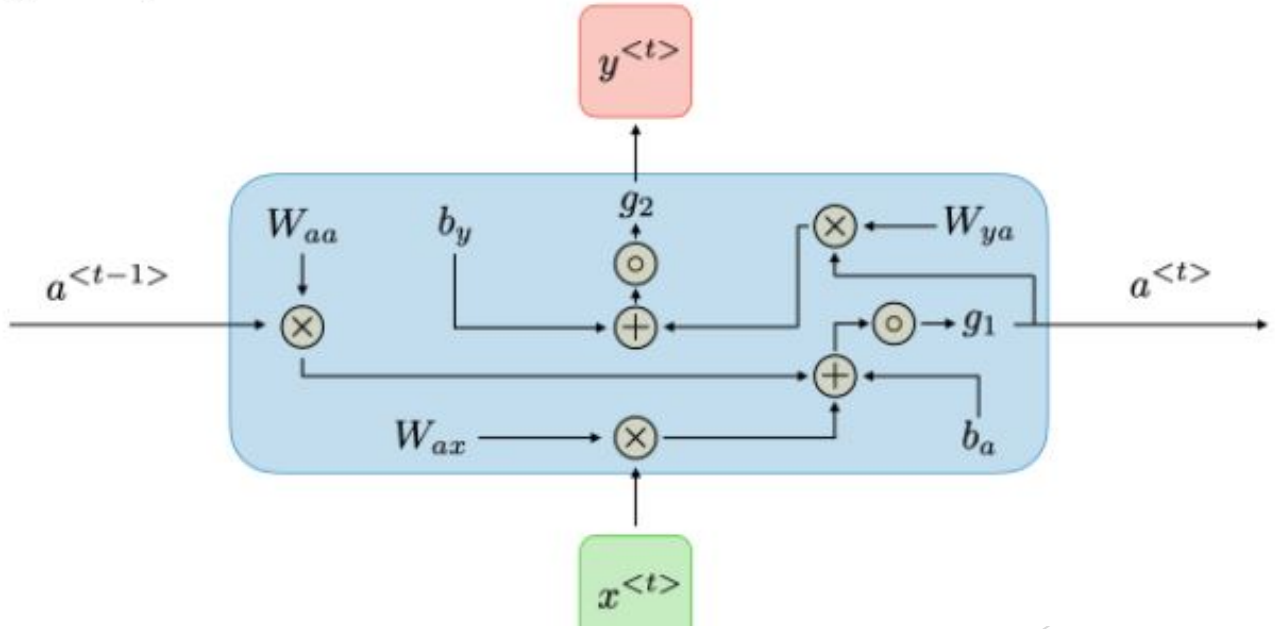
A Sequence Model : Predict Next World

NLP Technical Briefing 1 d
RNN(Recurrent Neural Network)

For each timestep t, the activation $a^{<t>}$ and the output $y^{<t>}$ are expressed as follows:

$$\boxed{a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)} \quad \text{and} \quad \boxed{y^{<t>} = g_2(W_{ya}a^{<t>} + b_y)}$$

where $W_{ax}, W_{aa}, W_{ya}, b_a, b_y$ are coefficients that are shared temporally and g_1, g_2 activation functions.



RNN Pro and Con

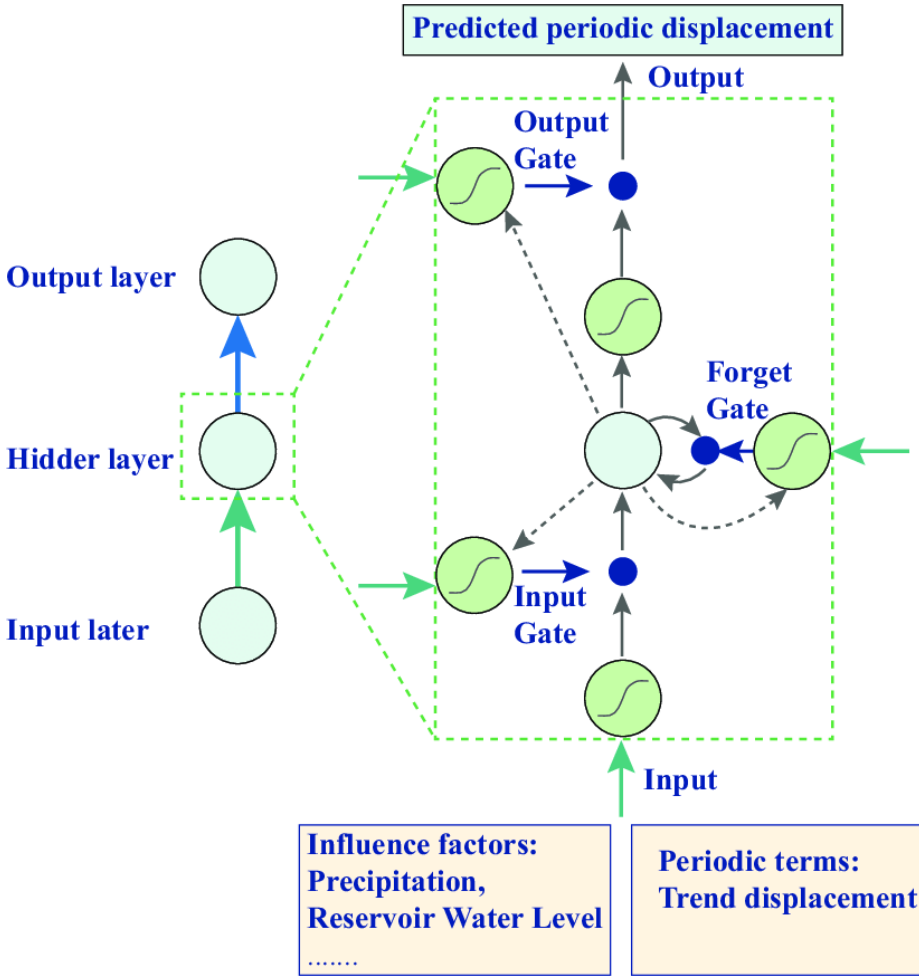
Pro

- **Handle the essence of sentence**
- **Detect the linguistic pattern.**

Con

- **Hard to maintain the Long sentence.**
- **Vanishing and Exploding Gradients**

NLP Technical Briefing 2 a
LSTM(Long Short Term Memory)



Input gate

- Gate to adjust the inuput data in the previous time

Output gate

- Gete to adjust the output data in the previous time

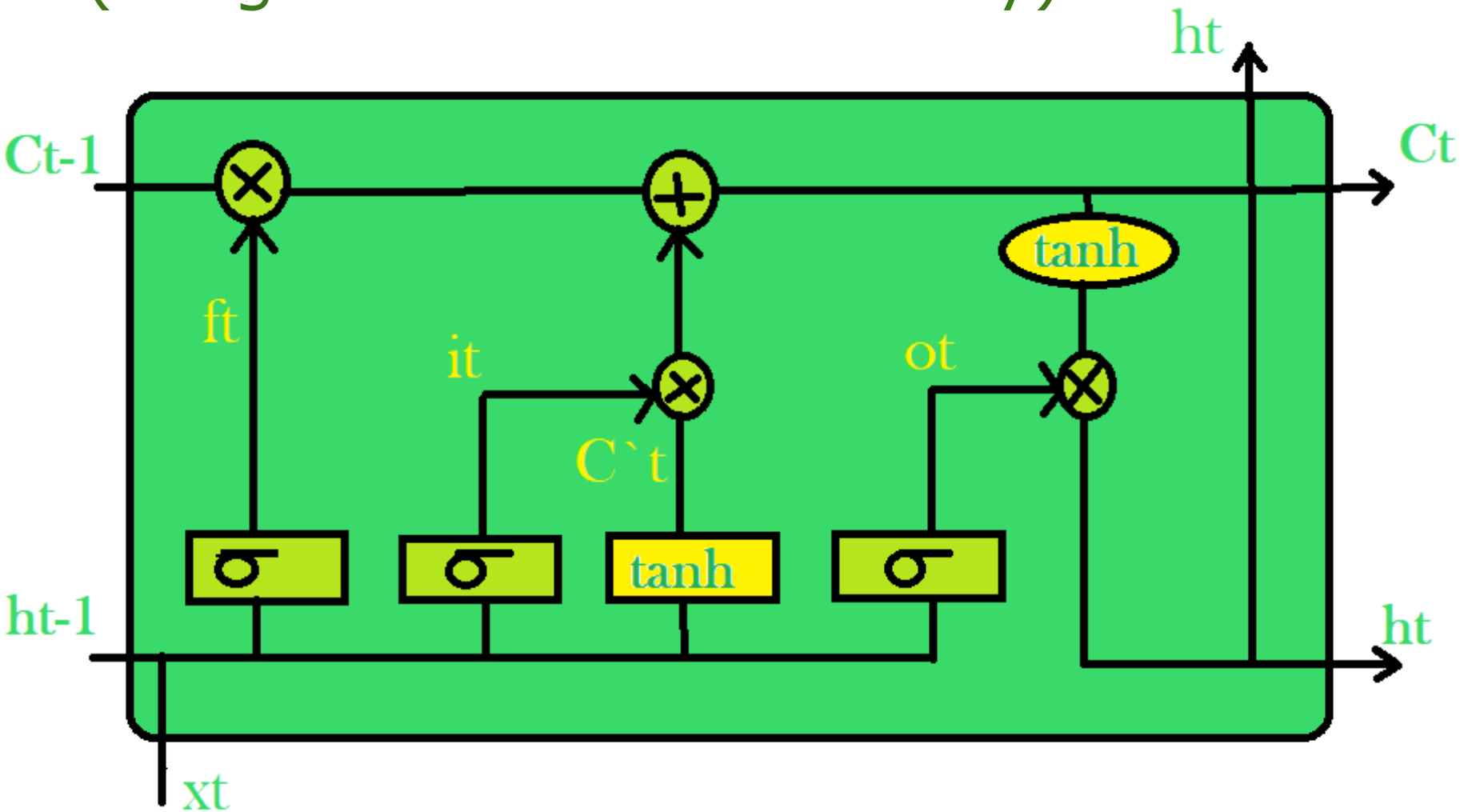
Foregotten gate

- Gate to adjust the past data

LSTM, Hochreiter & Schmidnuber, 1997
http://www.bioinf.jku.at/publications/older/2604.pdf

NLP Technical Briefing 2 b

LSTM(Long Short Term Memory)



https://www.geeksforgeeks.org/understanding-of-lstm-networks/

LSTM Pro and Con

Pro

- Long Sentence
- Effective Data Flow

Con

- Much Memory (Gate)
- Time Consuming (Training)
- Complex/Difficult train

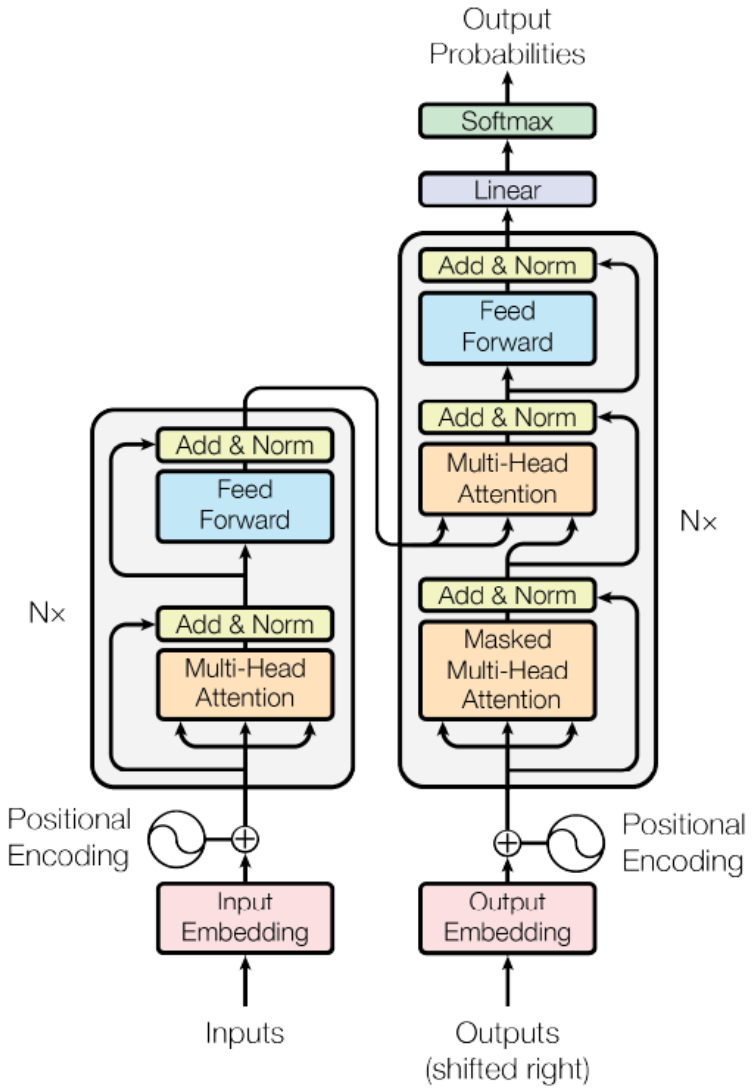
NLP Technical Briefing 3a
Transformer

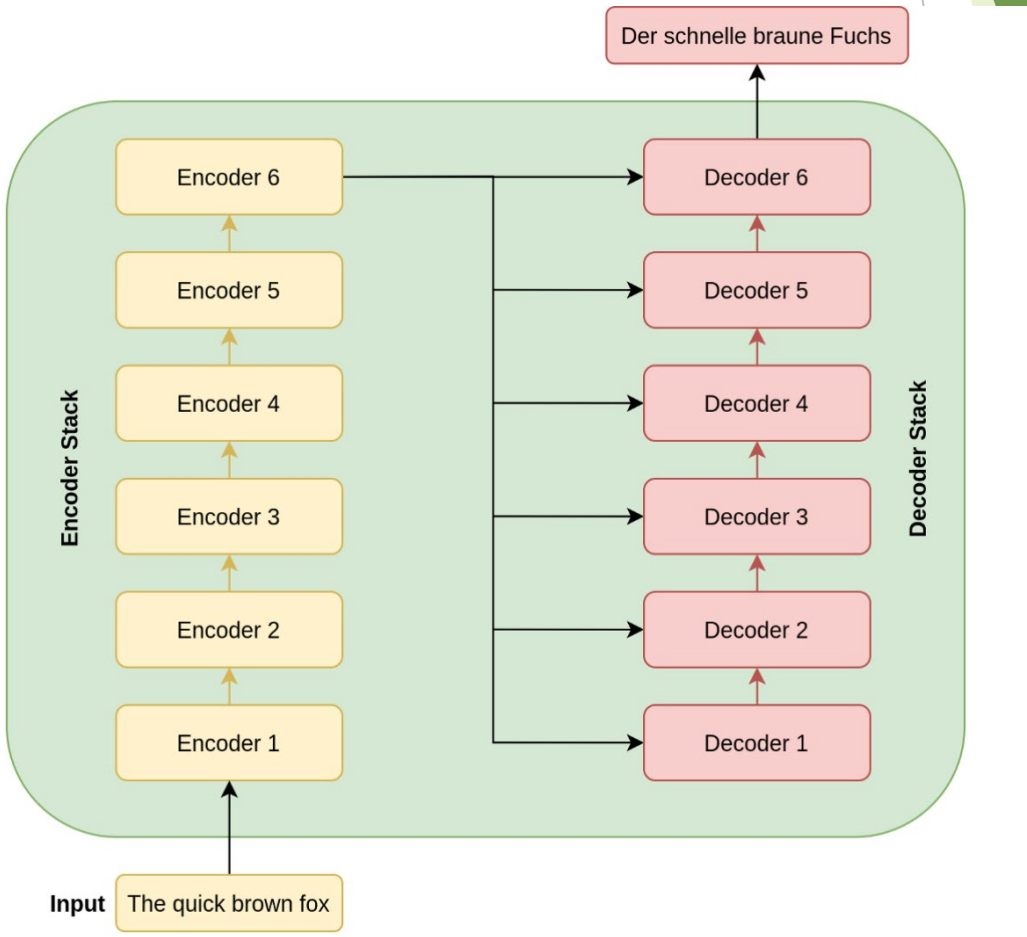
Ashishi Vaswani, Attention Is All You Need

Adobe Acrobat Document

Features

Encoder and Decoder、but never applied CNN(Convolutional Neural Network).





Jay Alammar, The Illustrated Transformer

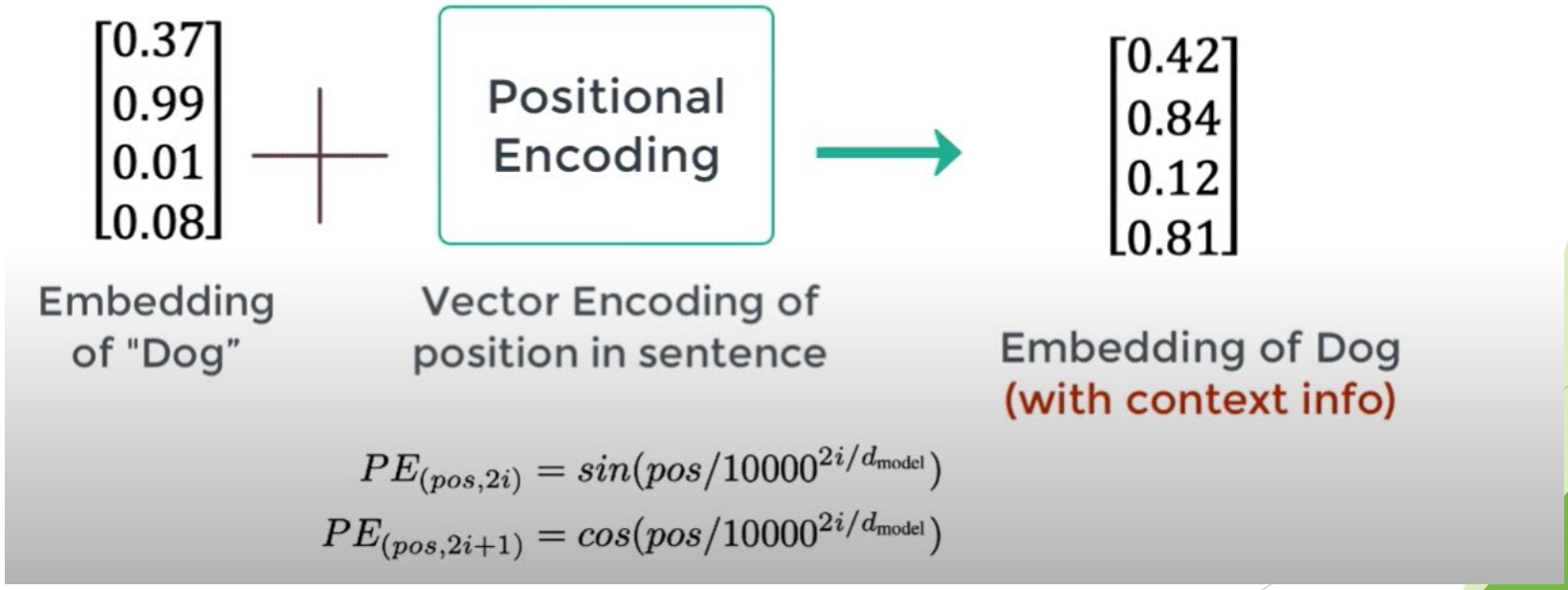
Ashishi Vaswani, Attention Is All You Need

NLP Technical Briefing 3b
Transformer

AJ's dog is a cutie → Position 2

AJ looks like a dog → Position 5

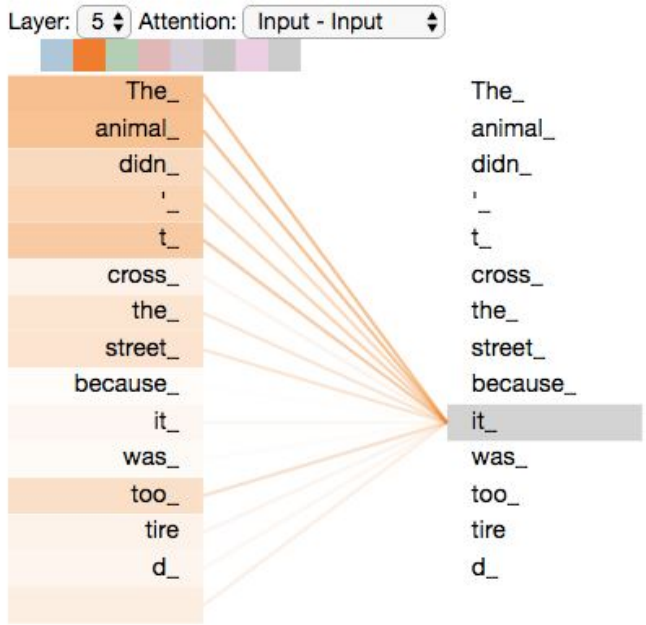
Positional Encoder :vector that gives context based on position of word in sentence

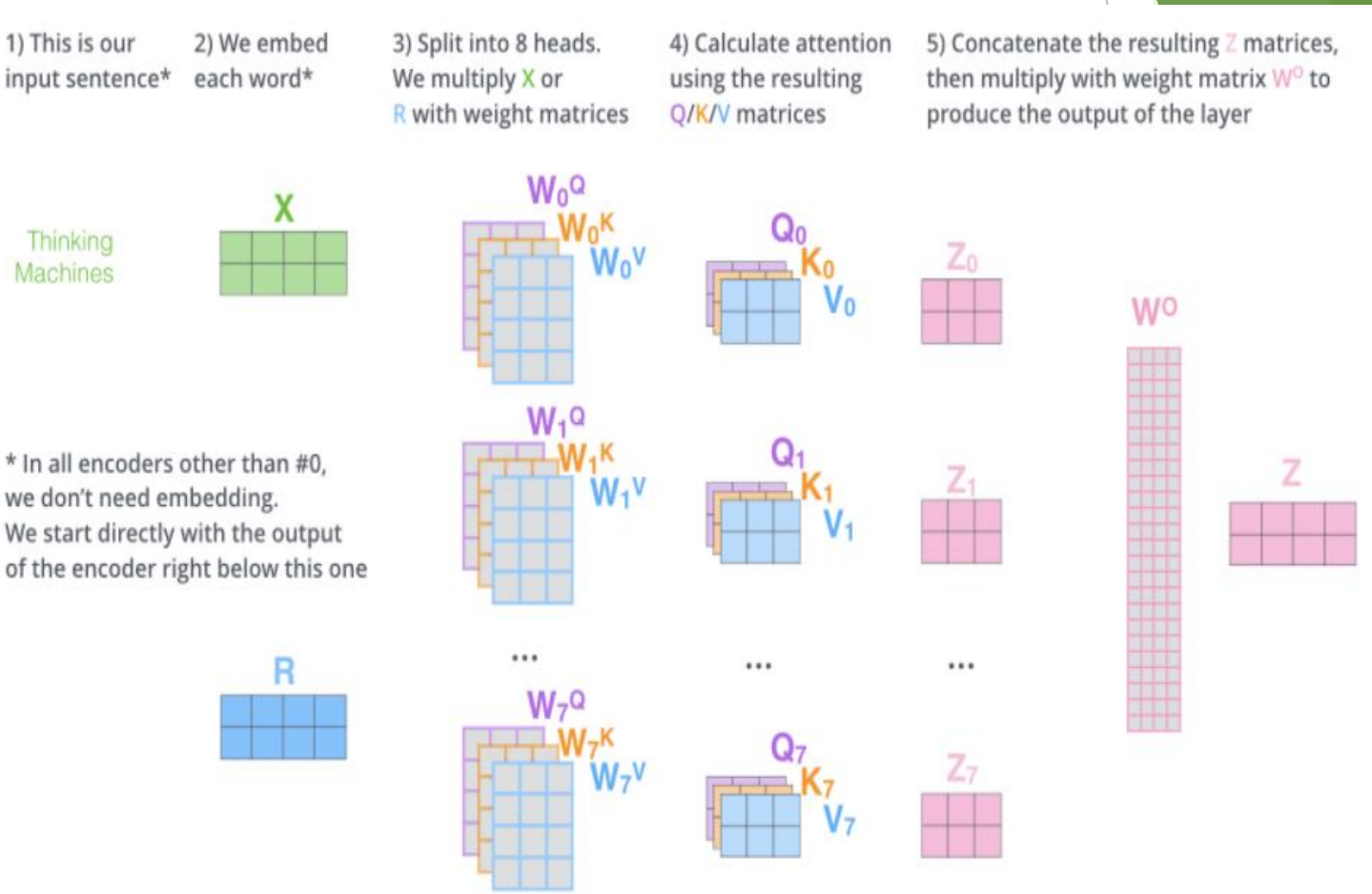


Transformer Neural Networks - EXPLAINED! (Attention is all you need)

https://www.youtube.com/watch?v=TQQlZhbC5ps&ab_channel=CodeEmporium

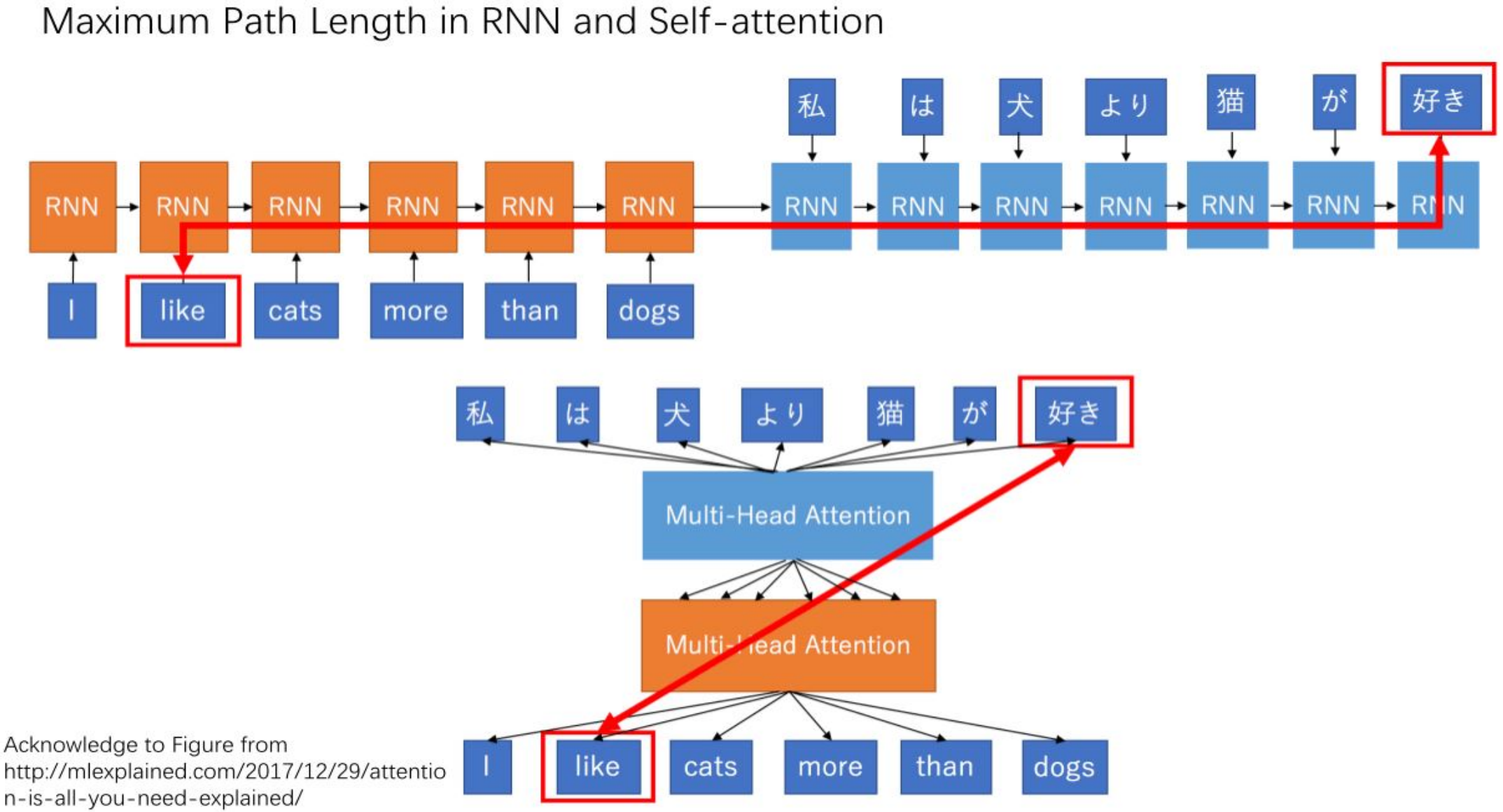
NLP Technical Briefing 3c
Transformer – Self Attention





Jay Alammar, The Illustrated Transformer

NLP Technical Briefing 3d
Transformer – Paradigm Shift



Transformer Pro and Con

Pro

- **Short Transaction time**
- **Access any position at words**
- **Long Sentence**

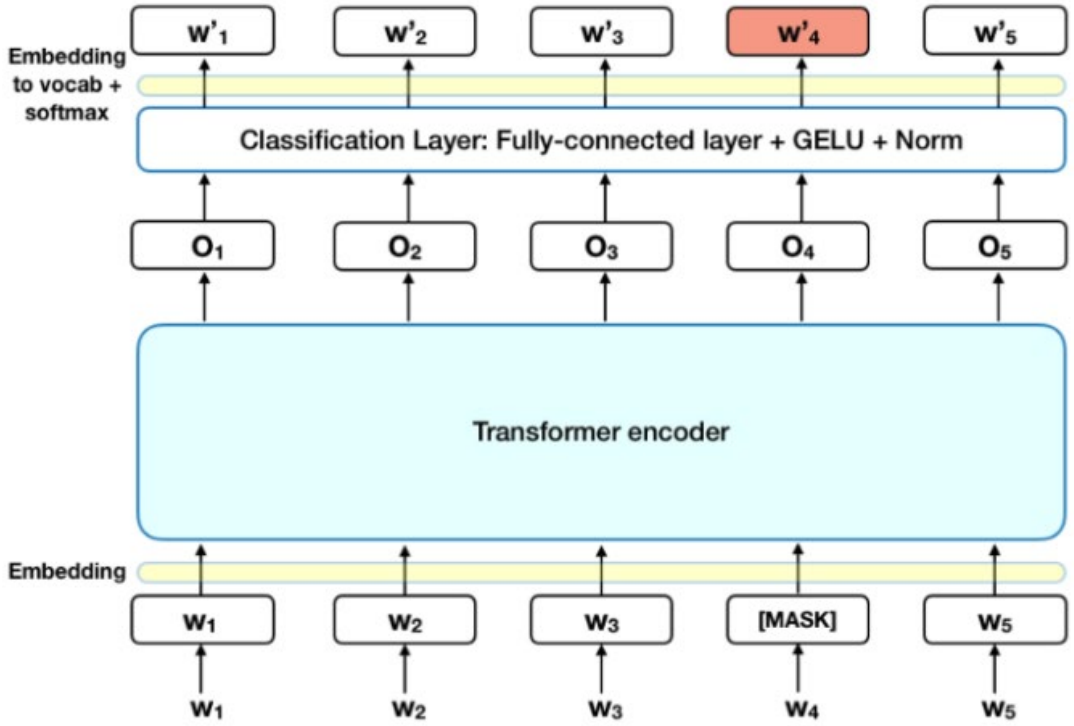
Con

- **Long Sequence needs Long Runtime**
- **No Support regular langueages**
- **No support basic recursion**

Michaek Hahn, Theretical Limitations of Self-Attention in Neural Sequence Models, 2019

NLP Technical Briefing 4a
Bert (Bidirectional Encoder Represetations from Transformers)

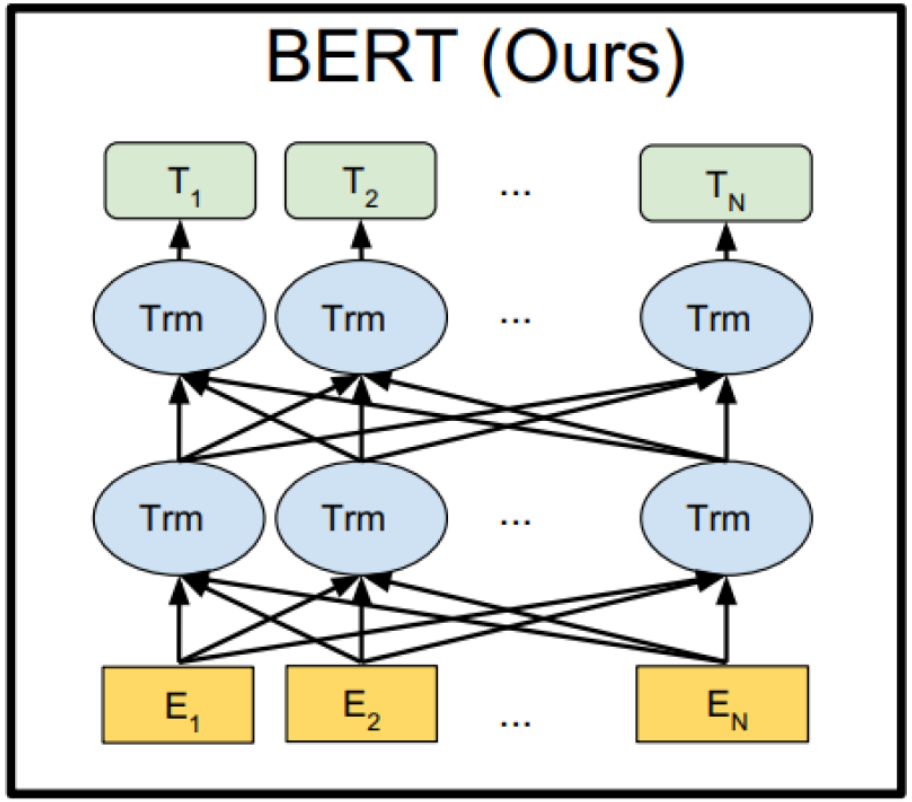
Feature	
	bi-directional model
	Masked Language Mode (missing word)
	Next sentence Prediction
	Pretrained model
	Fine Tuning



Yacob Devlin、BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

NLP Technical Briefing 4b
Bert

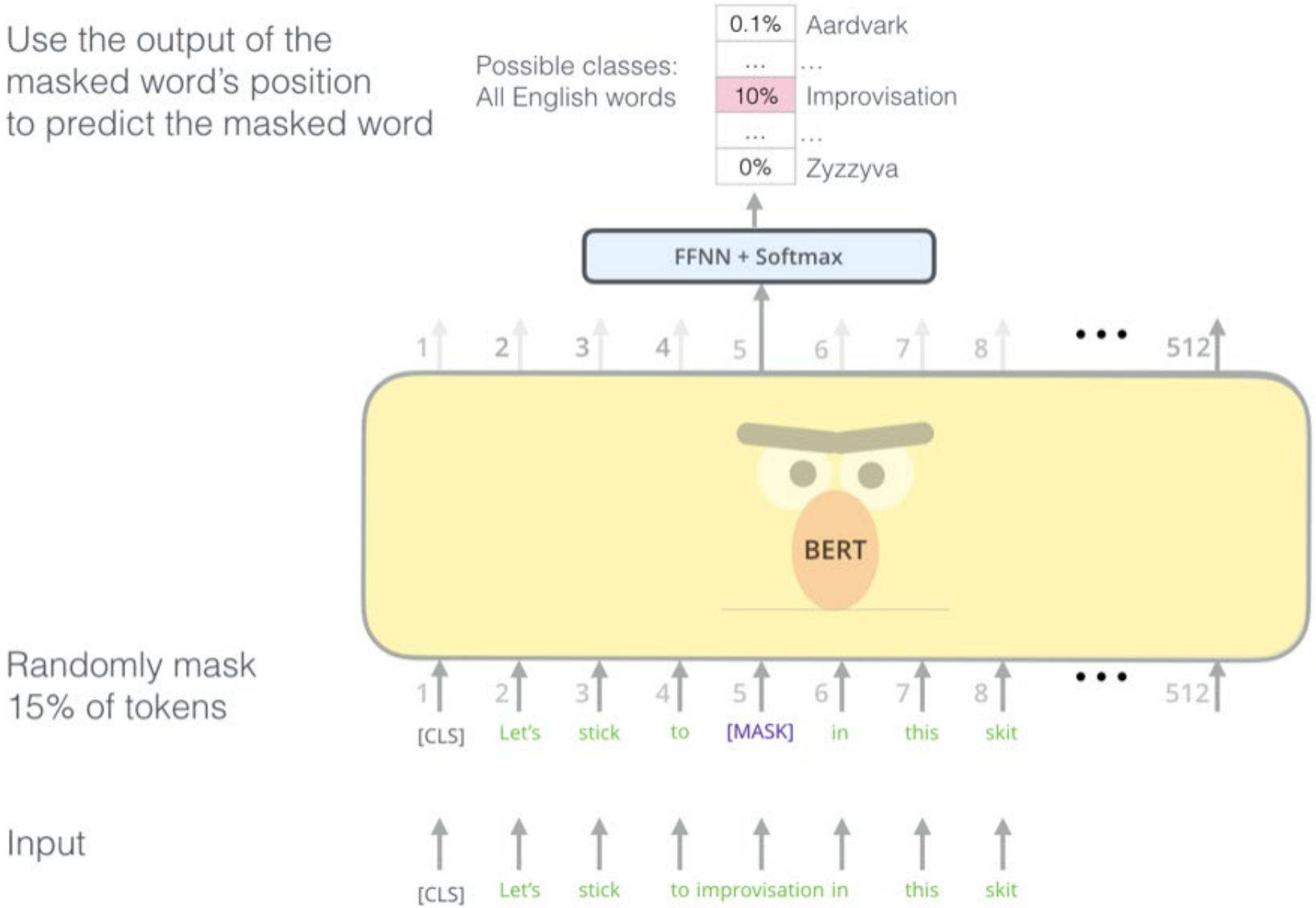
BiDirectional Model



Yacob Devlin、BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

NLP Technical Briefing 4c
Bert
Masked LM

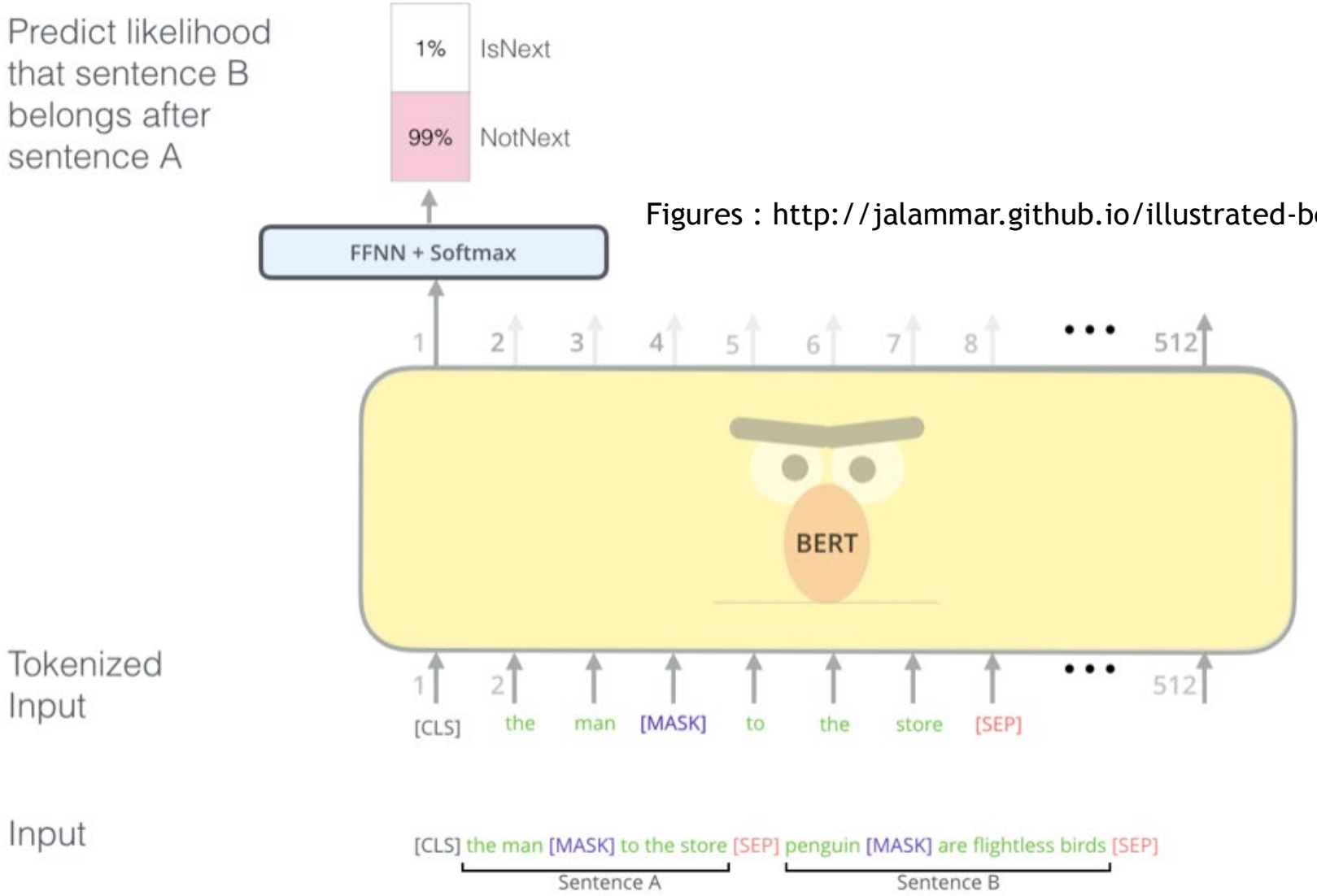
Figures : http://jalammar.github.io/illustrated-bert/



BERT's clever language modeling task masks 15% of words in the input and asks the model to predict the missing word.

Yacob Devlin、BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

NLP Technical Briefing 4d Bert NextSentece



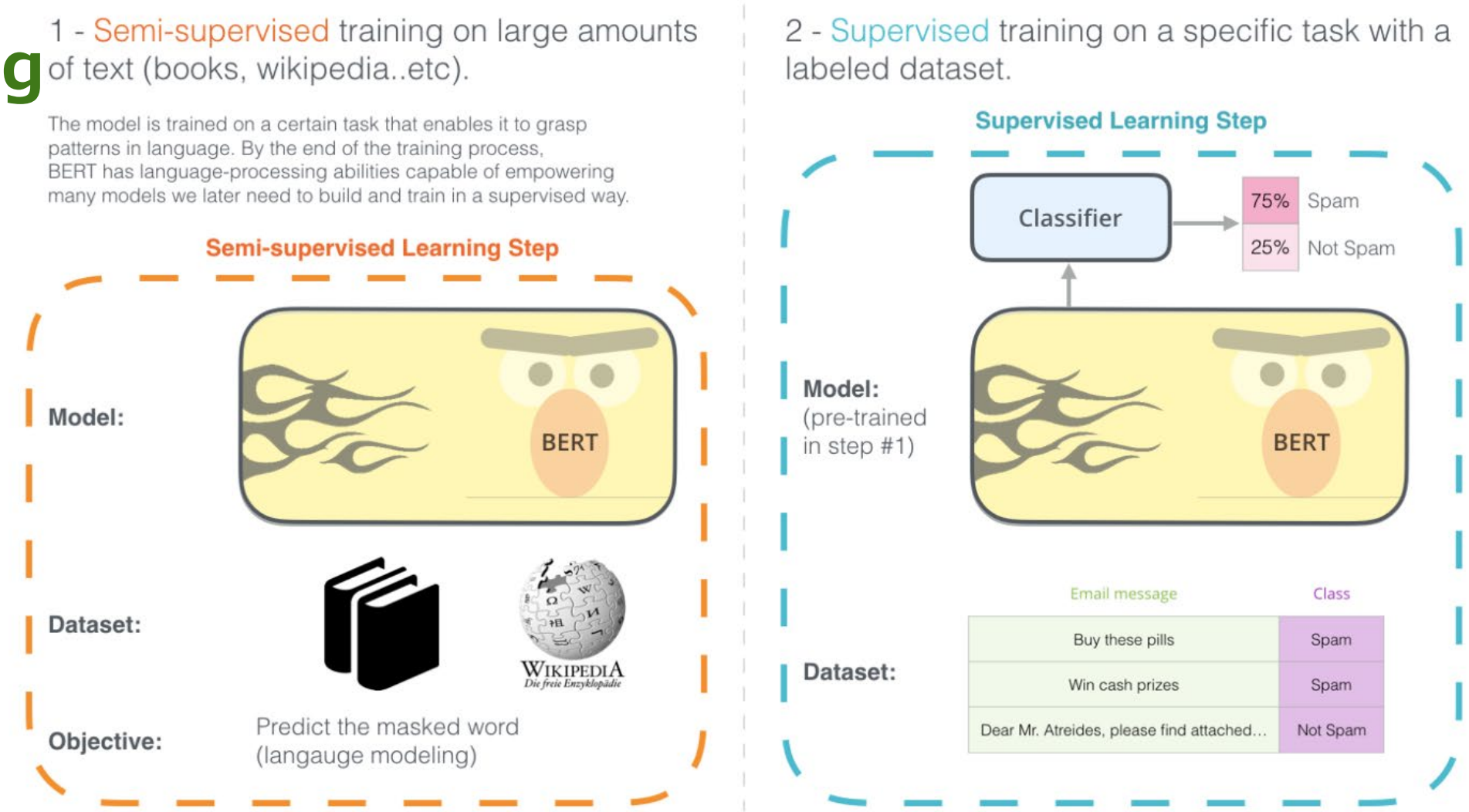
Figures : http://jalammar.github.io/illustrated-bert/

The second task BERT is pre-trained on is a two-sentence classification task. The tokenization is oversimplified in this graphic as BERT actually uses WordPieces as tokens rather than words --- so some words are broken down into smaller chunks.

Yacob Devlin、BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

NLP Technical Briefing 4e
Bert
Fine-Tuning

Figures : http://jalammar.github.io/illustrated-bert/



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

Yacob Devlin、BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Bert Pro and Con

Pro

- **Short Transaction time**
- **Understand context**
- **Generative model**
- **Good fit unstructured context**

Con

- **Big Model size**
- **Need Several models**

https://github.com/thunlp/PLMpapers/blob/master/PLMfamily.pptx

AFTER BERT

ELMo

ULMFiT

Multi-lingual

MultiFiT

Transformer

BERT

GPT

Bidirectional LM

Larger model
More data

GPT-2

Defense

Grover

Cross-lingual

XLM
UDify

Multi-task

MT-DNN

Knowledge distillation

$MT\text{-}DNN_{KD}$

+ Generation

MASS
UniLM

Span prediction
Remove NSP

SpanBERT

Longer time
Remove NSP
More data

RoBERTa

Permutation LM
Transformer-XL
More data

XLNet

+Knowledge Graph

ERNIE
(Tsinghua)

Neural entity linker

KnowBert

Cross-modal

VideoBERT
CBT
ViLBERT
VisualBERT
B2T2
Unicoder-VL
LXMERT
VL-BERT
UNITER

Whole Word Masking

ERNIE (Baidu)
BERT-wwm

By Xiaozhi Wang & Zhengyan Zhang @THUNLP

2.
AI(NLP) impacts Healthcare business

2 Big Trends Natural Language Processing

structured and unstructured

- Sentence, Summary
- Auto Translation

Information Extraction

- Information Search (Article)

Medical
Pharmaceutical
Health Care

Traditional approaches (machine learning) Encountered Many Issues in the health care business (EHR)

- Integration of different Data Format (different field format)
- Doctor(nurse) Free Text – Thousand number variables harm the prediction model, in a result to make false-positive.

NLP Research Motivation in the health care business

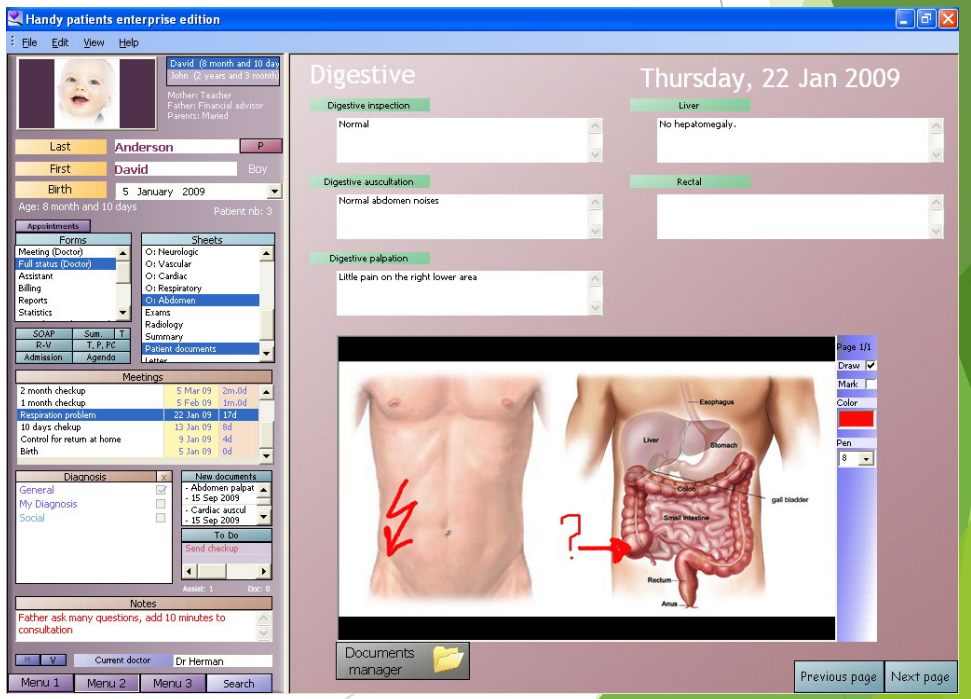
- Clinical Notes Analysis (EHR = Electronic Health Record) :
- Robust Professional Document=200,000 years DR wisdom
- 100 mio patient Data in EHR
- Leverage with Deep Learning : Uncover unstructured and complex data and predict the patient future status.
- Save time to find out the complex diagnosis result.

Example of Electronic Health Records (EHR)

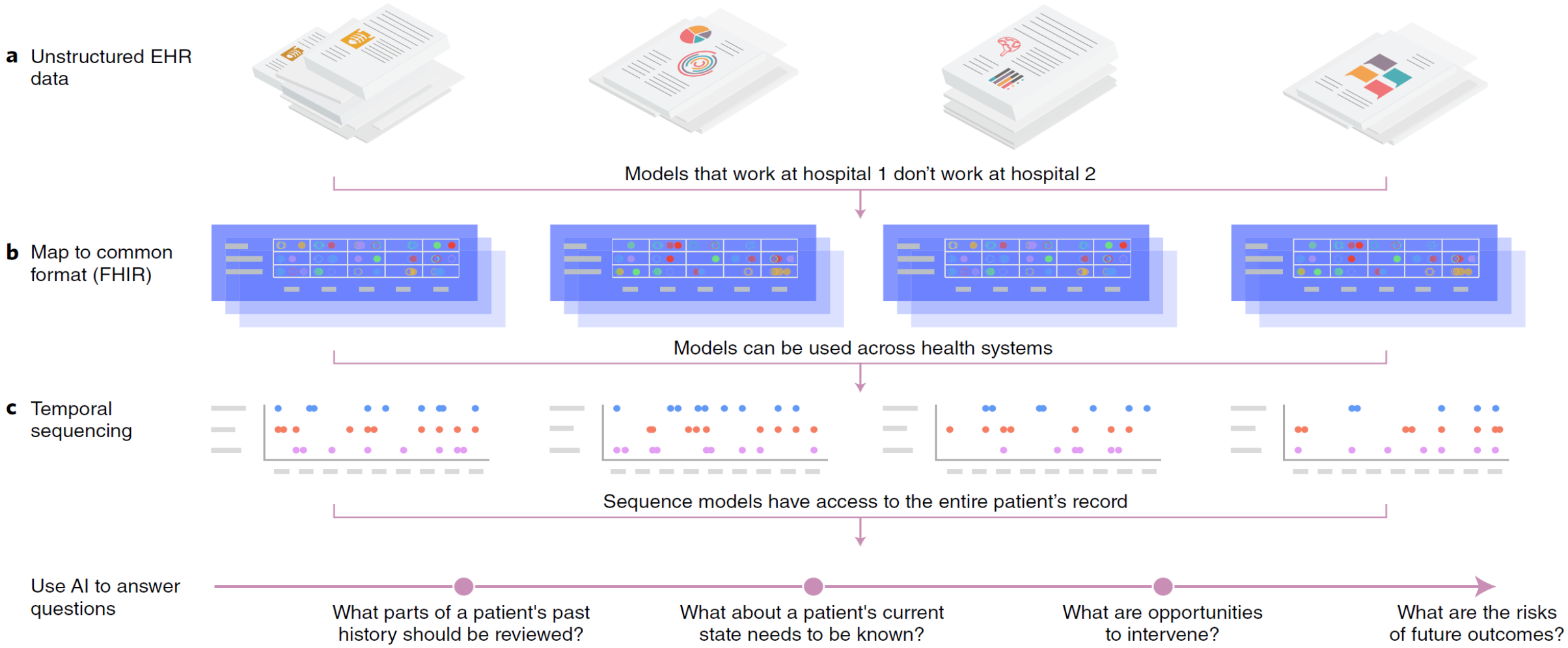
In the era of data digitalization, the data size of storing the medical records is massively increasing. For example, neurological imaging data expanded by a significant factor 25,000, from about 200 gigabytes of data a year worldwide in the late 1980s to 5 petabytes a year in the early 2010s.(nature.com)

EHR in Japan : Hybrid type of images and text message

EHR sample image from Wikipedia

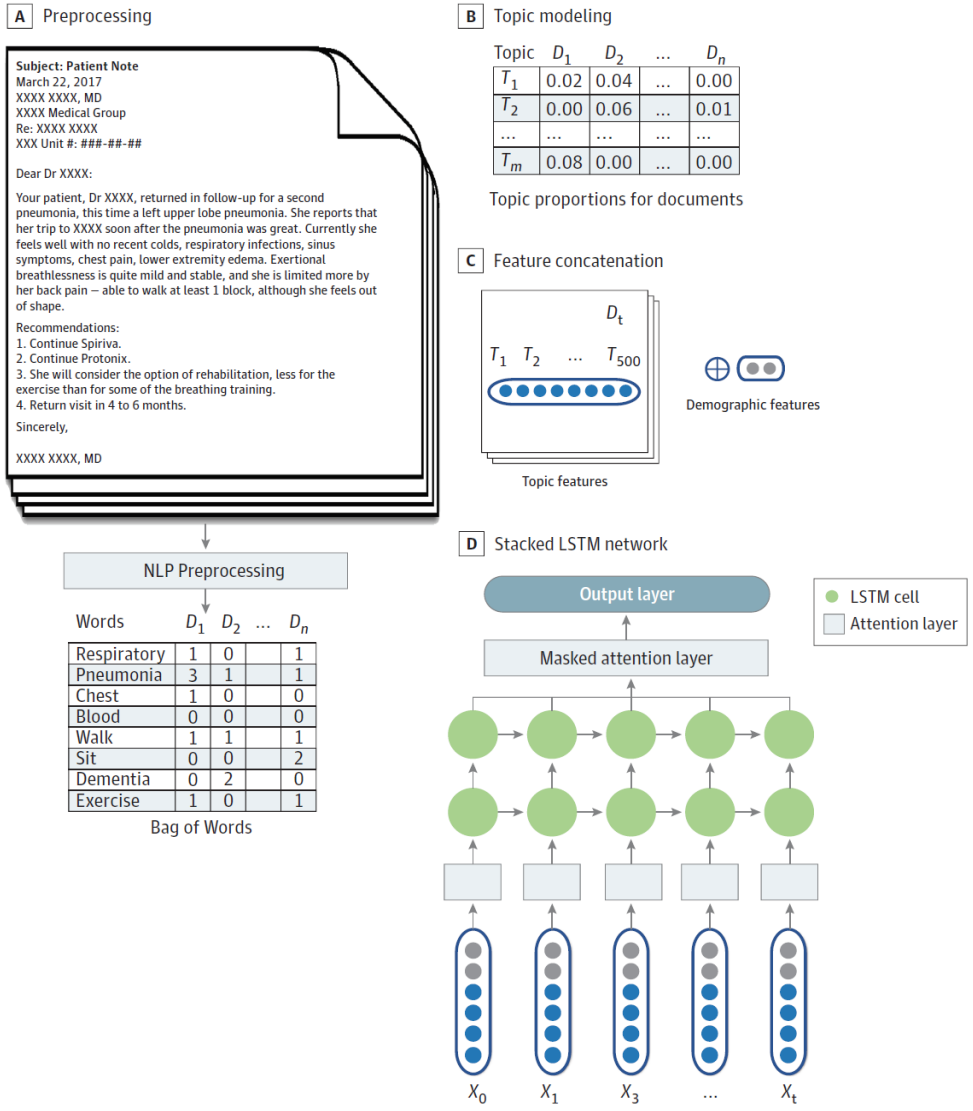


Prediction step from EHR (Traditional approach - ex. Large Scale RNN)



Ex1. : Biomedical NLP (LSTM) for EHR

Figure 1. Overview of the Predictive Modeling Using Longitudinal Clinical Notes and Demographics of Patients With Dementia



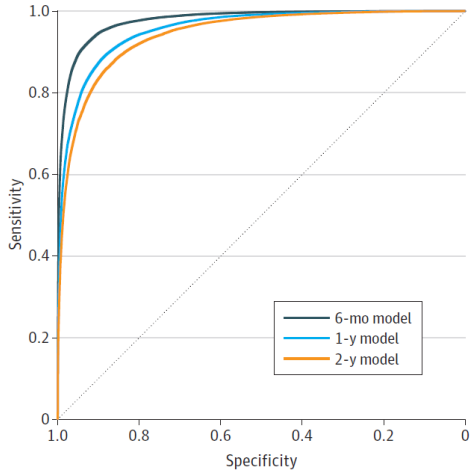
DATA: Electronic Health Record (26921 Alzheimer patients)

- Clinical Visit Notes
- Discharge Summaries
- Doctor Consultation Notes
- (except radiology and pathology reports)

Model: Attention Layer + LSTM

Result: Predicts 0.5, 1 and 2-year mortality prediction

Figure 2. Receiver Operating Characteristic Curves of the Deep Learning Models in Predicting Patient Mortality



JAMA Network Open. 2019;2(7):e196972. doi:10.1001/jamanetworkopen.2019.6972

(Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions)
Liquin, Jul 2019, Wang

Ex2. : Biomedical NLP (BioBert) for Text Mining

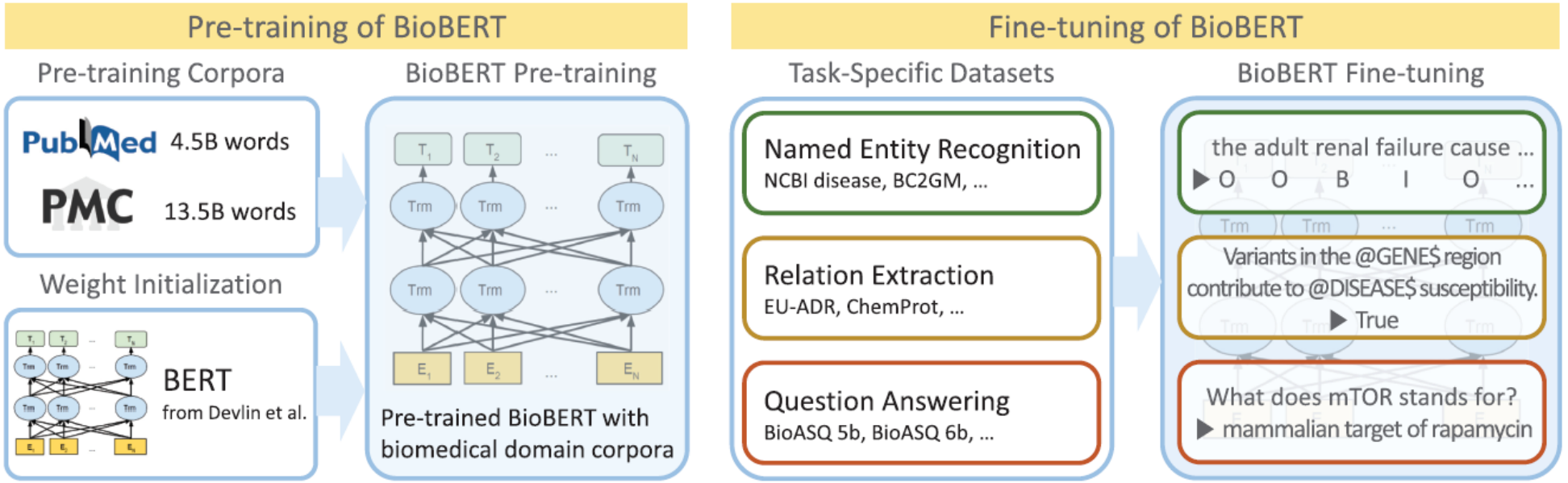
BioBert Model

Biomedical Natural Language Processing Model based on PubMed

Data : Book Corpus + English Wikipedia
+ PubMed Abstraction + PMC Full Text

Model: BioBert (implemented from Bert)

Result: Improve 12.24% Medical Record Review biomedical questioning and answering, Improve Medical Entity



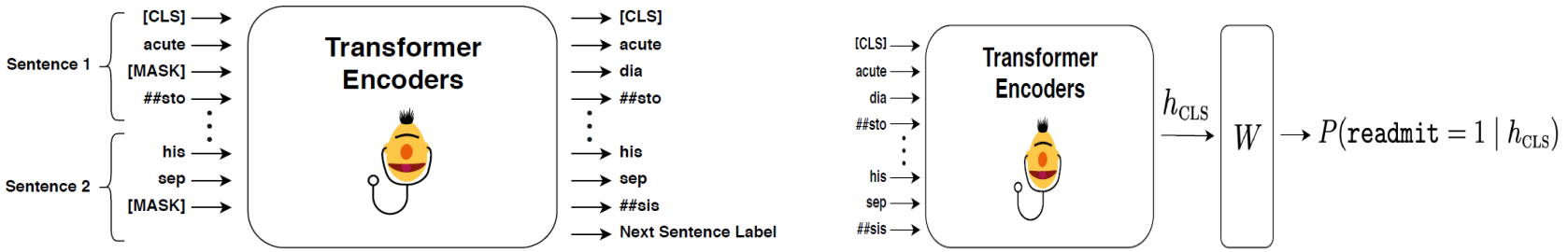
(Yu Gu、BioBERT: a pre-trained biomedical language representation model for biomedical text mining)

Ex3. : Biomedical NLP (ClinicalBert) - a

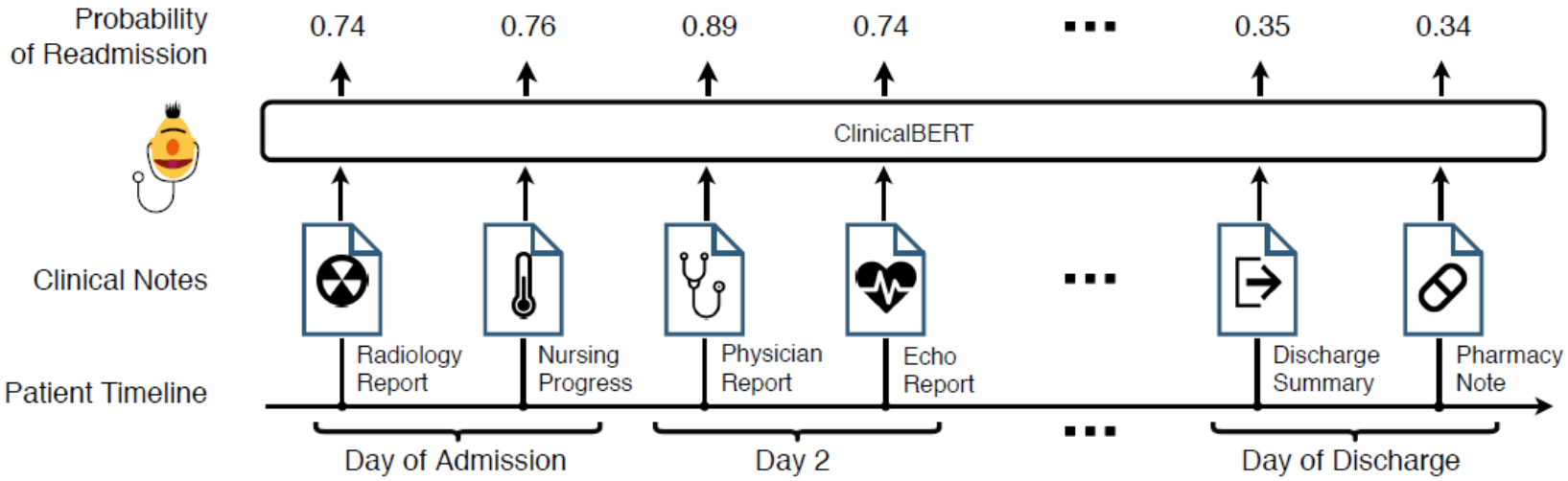
ClinicalBert

Data : Medical Information Mart for Intensive Care III
2,083,180 de-identified notes (admission)

Model: ClinicalBert (implemented from Bert)



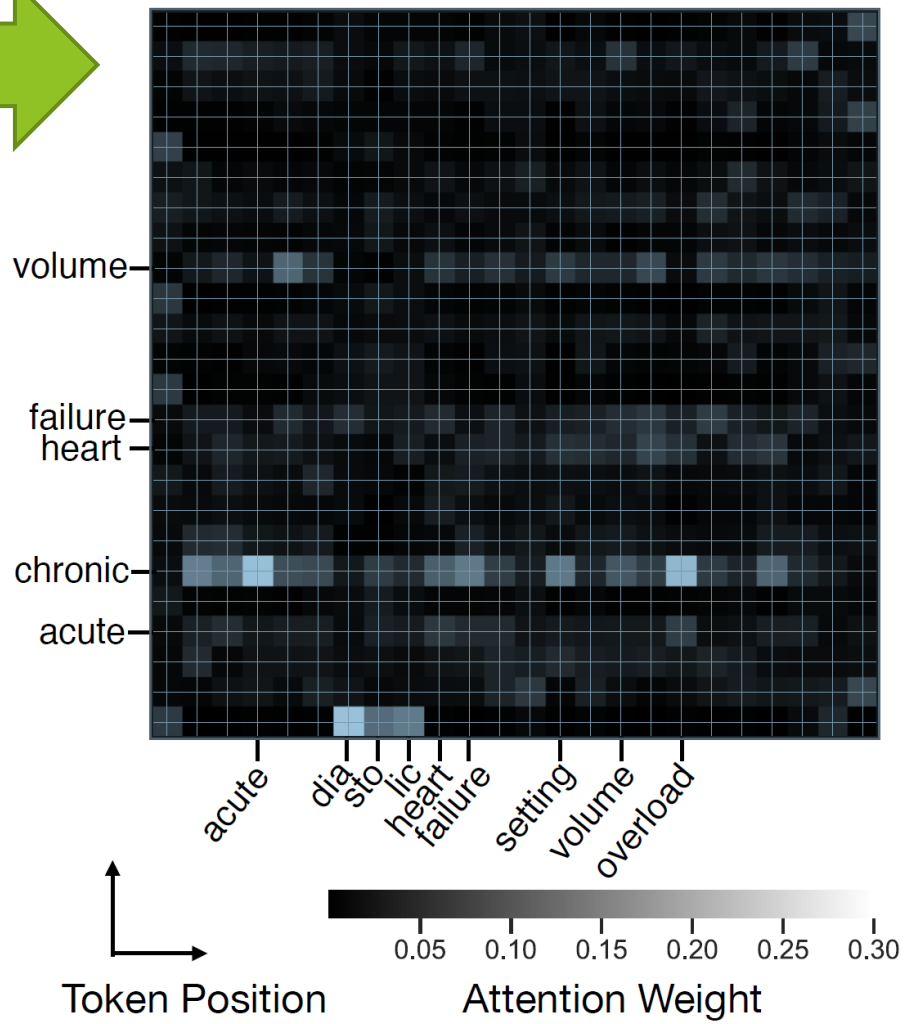
Predicts 30-day readmission data process



Ex3. : Biomedical NLP (ClinicalBert) - b

Original sentence : He has experienced acute chronic diastolic heart failure in the setting of volume overload due to his sepsis.

"chronic" has strong relation inside Attention layer.



Improve the masked language modeling and next sentence prediction

Model	Masked language modeling	Next sentence prediction
ClinicalBERT	**86.80%**	**99.25%**
BERT	56.80%	80.50%

Accurate 30-day readmission prediction

Model	AUROC	AUPRC	RP80
ClinicalBERT	0.714 ± 0.018	0.701 ± 0.021	0.242 ± 0.111
Bag-of-words	0.684 ± 0.025	0.674 ± 0.027	0.217 ± 0.119
BI-LSTM	0.694 ± 0.025	0.686 ± 0.029	0.223 ± 0.103
BERT	0.692 ± 0.019	0.678 ± 0.016	0.172 ± 0.101

3. Conclusion

AI (NPR) = Variable Service

NPR Analysis Expanding

Research from Commerce (GE)

Big Data Analysis (GAFA)

Data is Borderless

NLP/BigData

Express Scripts (USA)

(Pharmacy Benefit Management Org.)

NLP Analysis based on 83mio Patient Data

Data (22peta byte)

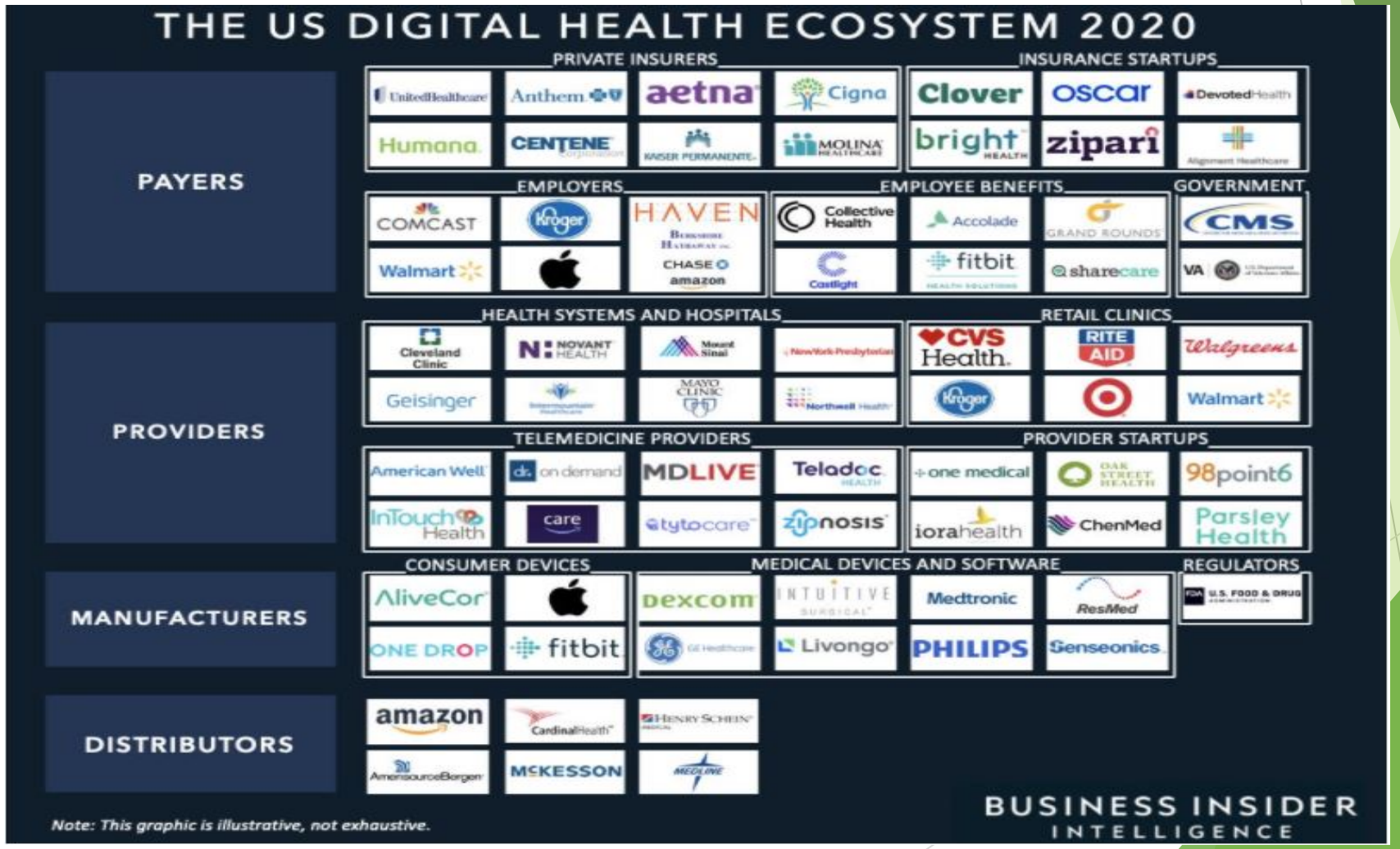
EHR, Voice Message, Public Twitter message, Patient's Twitter message, Email

Result:

Patient Sentiment, Patient Character, Diagnosis, Predict Low Adherence

https://www.express-scripts.com/corporate/articles/using-ai-improve-patient-experience

AI accelerates in the Health Care Business



Business Insider 2020

AI (NLP) Issue

AI Ethics and Privacy

AI Model
vs
socio economics

Future NLP

- Hand writing
- MultiModal
- XLNet (SoTA)