人工智慧在自然語言處理與應用
Artificial Intelligence in Natural Language Processing and Applications

Course Introduction

Instructor: 林英嘉 (Ying-Jia Lin)
2025/02/17

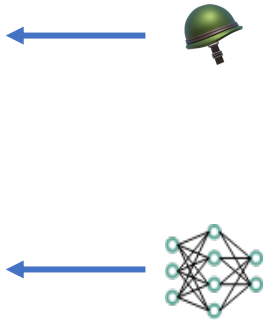
自然語言處理與應用
Natural Language Processing and Applications

Course Introduction

Instructor: 林英嘉 (Ying-Jia Lin)
2025/02/17

About Me

Experience and Education (Time)	
Assistant Professor	Department of Artificial Intelligence, Chang Gung University (2025/02 -)
Postdoctoral Researcher	Department of Computer Science, National Tsing Hua University (2024/09 - 2025/01)
Ph.D.	Department of Computer Science and Information Engineering, National Cheng Kung University (2019/08 - 2024/02)
M.S.	Institute of Biomedical Informatics, National Yang-Ming University (2017/09 - 2019/07)
B.S.	Department of Biomedical Sciences, Chang Gung University (2013/09 - 2017/06)



Outline

- Introduction to Artificial Intelligence and Natural Language Processing
- History of Artificial Intelligence
- Course Introduction and requirements

Course Website & Slido

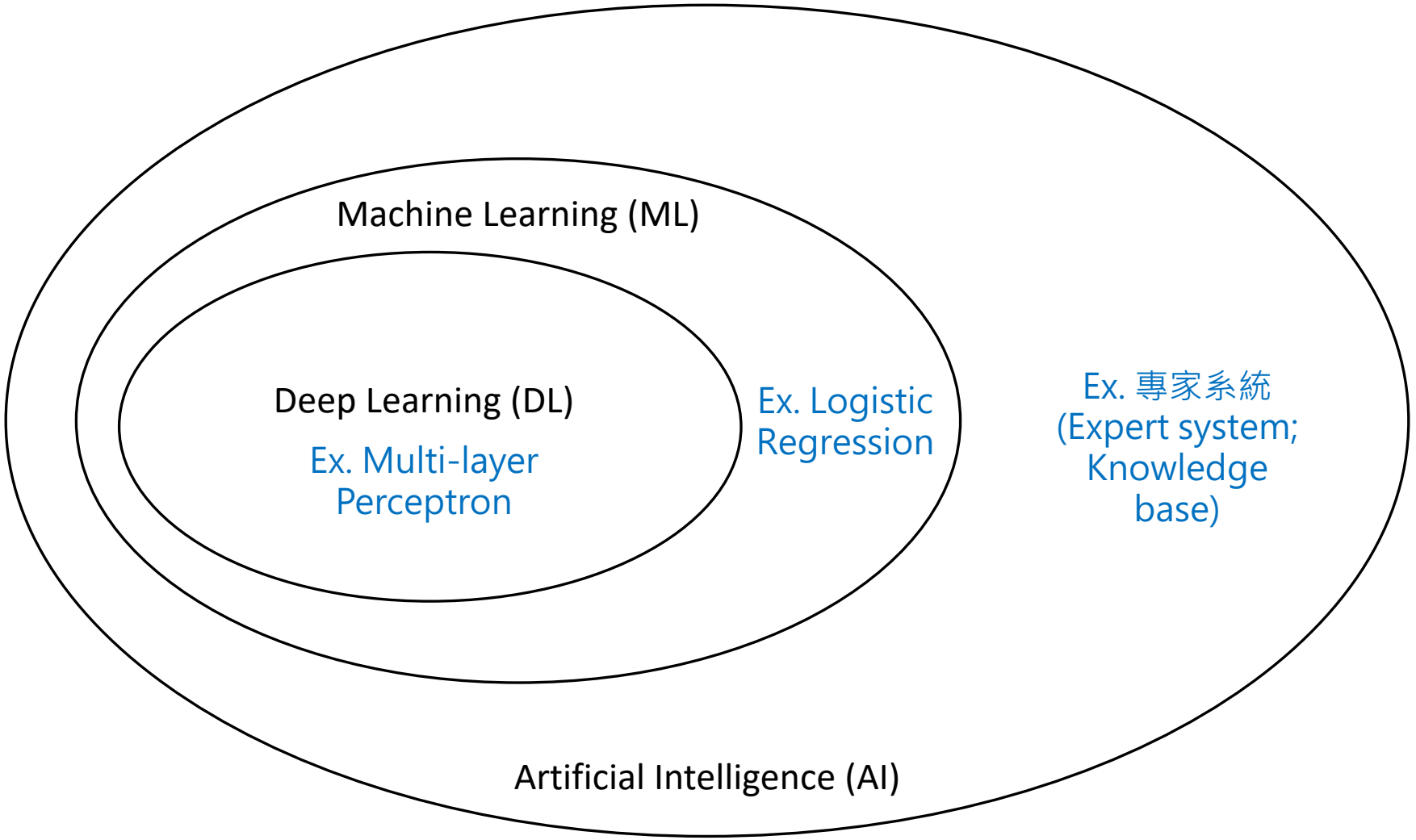
113-2

Natural Language Processing and Applications

https://github.com/mcps5601/CGUNLP_2025_Spring

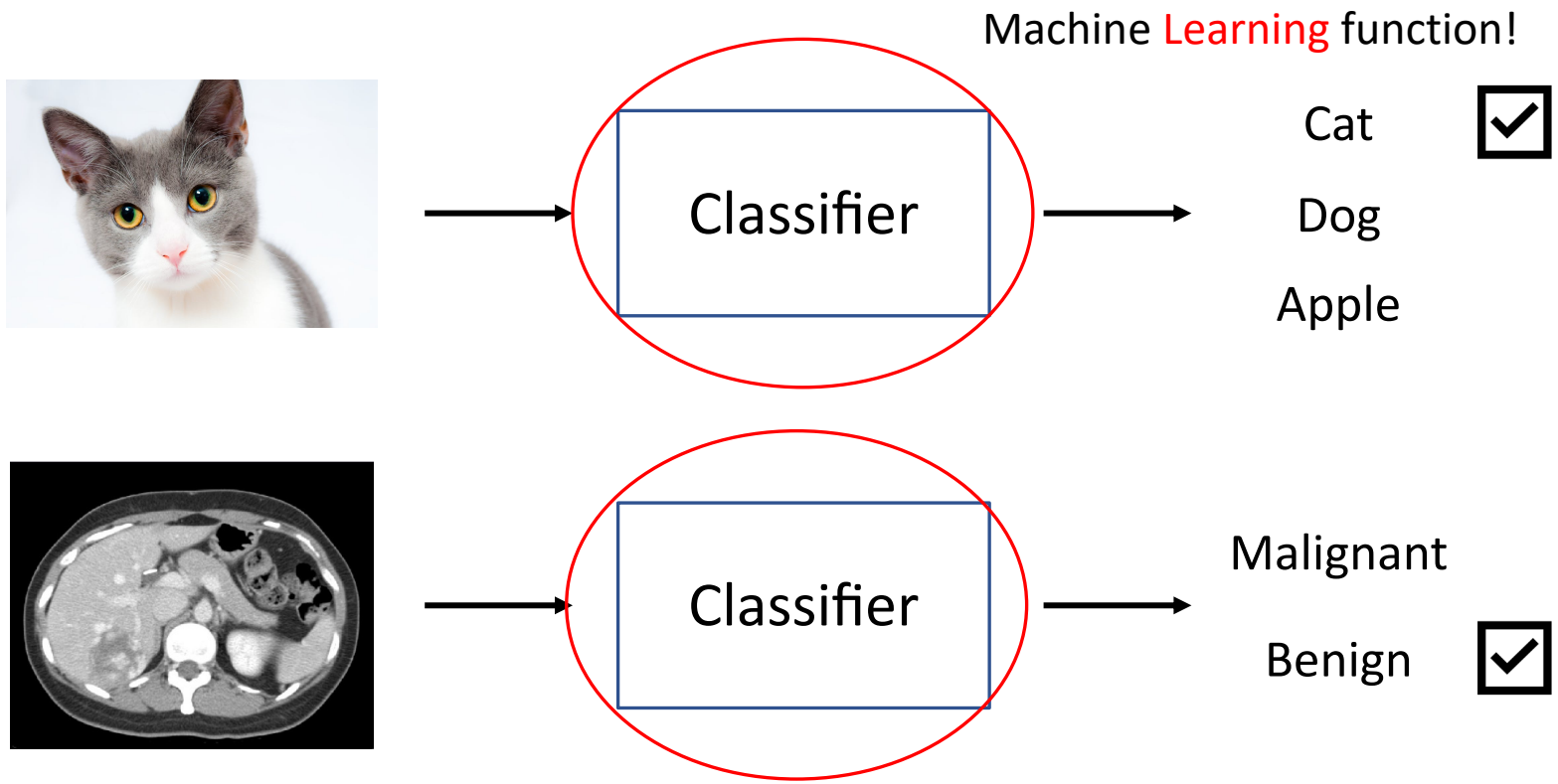
1466691
(Link)

Venn Diagram about AI, ML, and DL



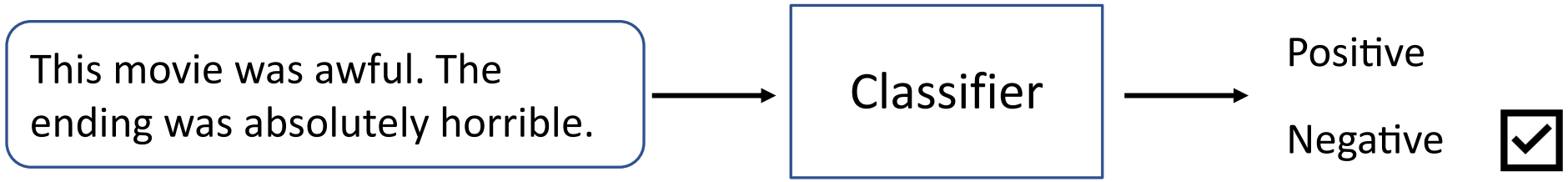
What is Deep Learning?

- We start the concepts from machine learning.

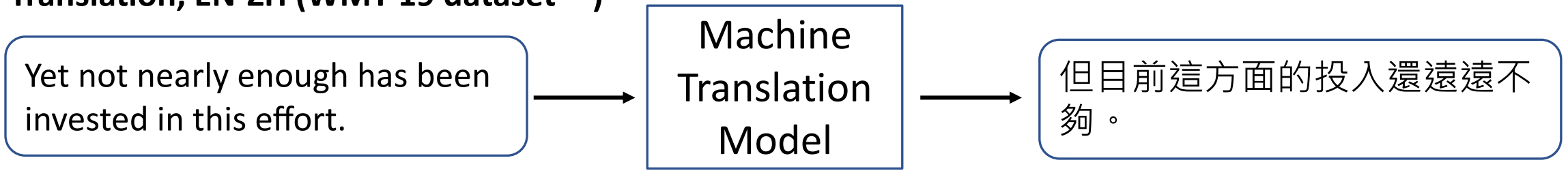


How about DL in Natural Language Processing?

Sentiment Analysis (IMDb dataset [1])



Translation, EN-ZH (WMT-19 dataset [2])

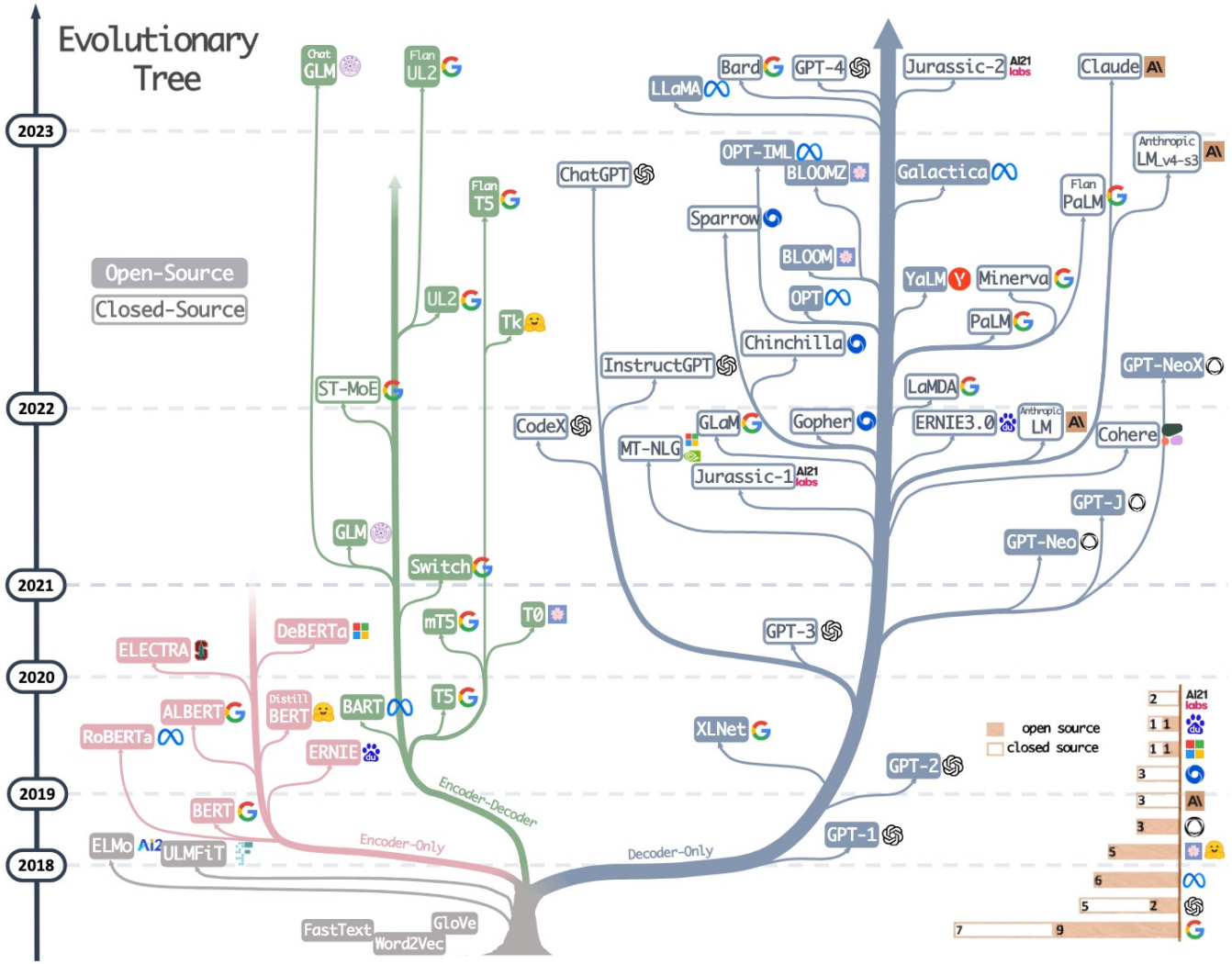


[1] https://huggingface.co/datasets/stanfordnlp/imdb
[2] https://huggingface.co/datasets/wmt/wmt19/viewer/zh-en

Common NLP tasks

Text Classification	Text Generation	Sequence Tagging	Text Matching
Sentiment Analysis	Machine Translation	Part-of-speech tagging	Semantic Textual Similarity
Natural Language Inference (NLI)	Summarization	Named Entity Recognition (NER)	Retrieval
	Dialogue Generation	Word Segmentation	

Evolutionary Tree of Large Language Models



Yang, Jingfeng, et al. "Harnessing the power of llms in practice: A survey on chatgpt and beyond." ACM Transactions on Knowledge Discovery from Data 18.6 (2024): 1-32.

Chinese LLMs

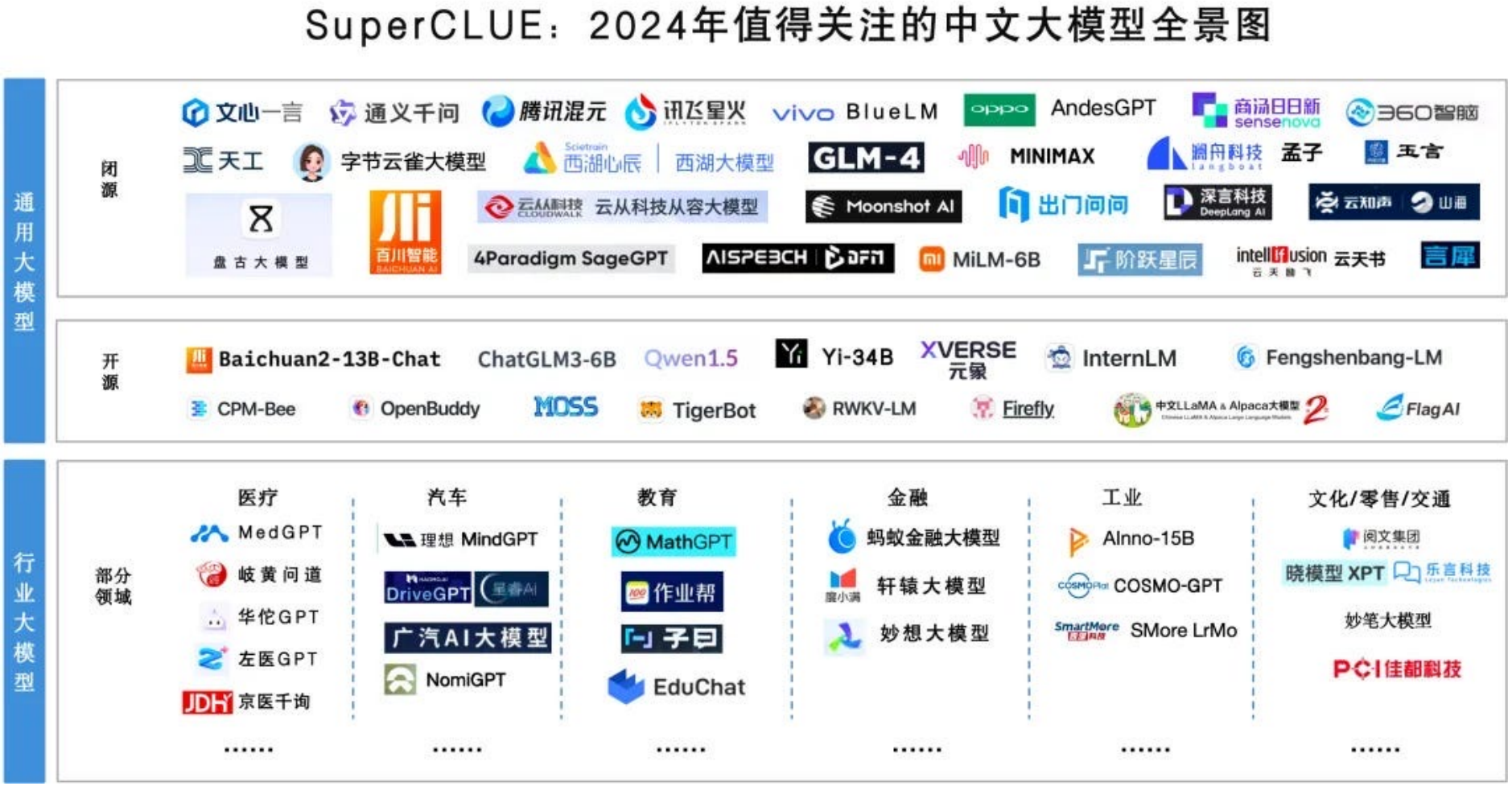


Figure source: https://chinai.substack.com/p/chinai-264-superclue-april-2024-report

SuperCLUE Leaderboard

SuperCLUE总排行榜（2024年12月）

排名	模型名称	机构	总分	Hard	理科	文科	使用方式	发布日期
-	o1	OpenAI	80.4	76.7	87.3	77.1	网页	2025年1月8日
-	o1-preview	OpenAI	74.2	63.6	80.6	78.5	API	2025年1月8日
-	ChatGPT-4o-latest	OpenAI	70.2	57.8	72.1	80.7	API	2025年1月8日
🥇	DeepSeek-V3	深度求索	68.3	54.8	72	78.2	API	2025年1月8日
🥇	SenseChat 5.5-latest	商汤	68.3	51.5	71.6	81.8	API	2025年1月8日
-	Gemini-2.0-Flash-Exp	Google	68.2	55.5	72.6	76.6	API	2025年1月8日
-	Claude 3.5 Sonnet(20241022)	Anthropic	67.7	54.6	71.4	77.2	API	2025年1月8日
🥇	360zhinao2-o1	360	67.4	51.4	72.1	78.7	API	2025年1月8日
🥈	Doubao-pro-32k-241215	字节跳动	66.5	50.6	72.3	76.6	API	2025年1月8日
🥈	NebulaCoder-V5	中兴通讯	66.4	48.6	69.5	80.9	API	2025年1月8日
🥈	Qwen-max-latest	阿里巴巴	66.2	51.3	67.4	80	API	2025年1月8日
-	Qwen2.5-72B-Instruct	阿里巴巴	65.4	49.7	66.2	80.3	API	2025年1月8日
🥉	Step-2-16k	阶跃星辰	65.2	50	65.1	80.3	API	2025年1月8日
🥉	GLM-4-Plus	智谱AI	65.1	48.5	68.1	78.8	API	2025年1月8日

Figure source: https://www.superclueai.com/

電腦是如何學習自然語言呢？

Word Representations

Basic Approach

Corpus

I like to eat apple

You like to watch TV

Vocabulary

apple	[1, 0, 0, 0, 0, 0, 0, 0]
eat	[0, 1, 0, 0, 0, 0, 0, 0]
I	[0, 0, 1, 0, 0, 0, 0, 0]
like	[0, 0, 0, 1, 0, 0, 0, 0]
watch	[0, 0, 0, 0, 1, 0, 0, 0]
to	[0, 0, 0, 0, 0, 1, 0, 0]
⋮	

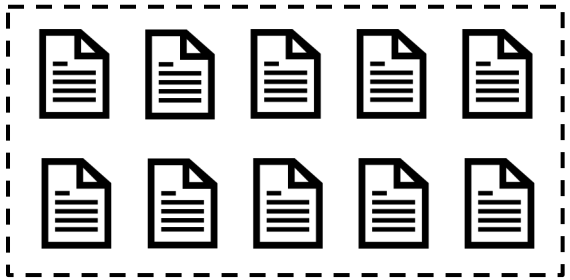
One hot encodings with each vector size equal to the vocab size (8 in this case)

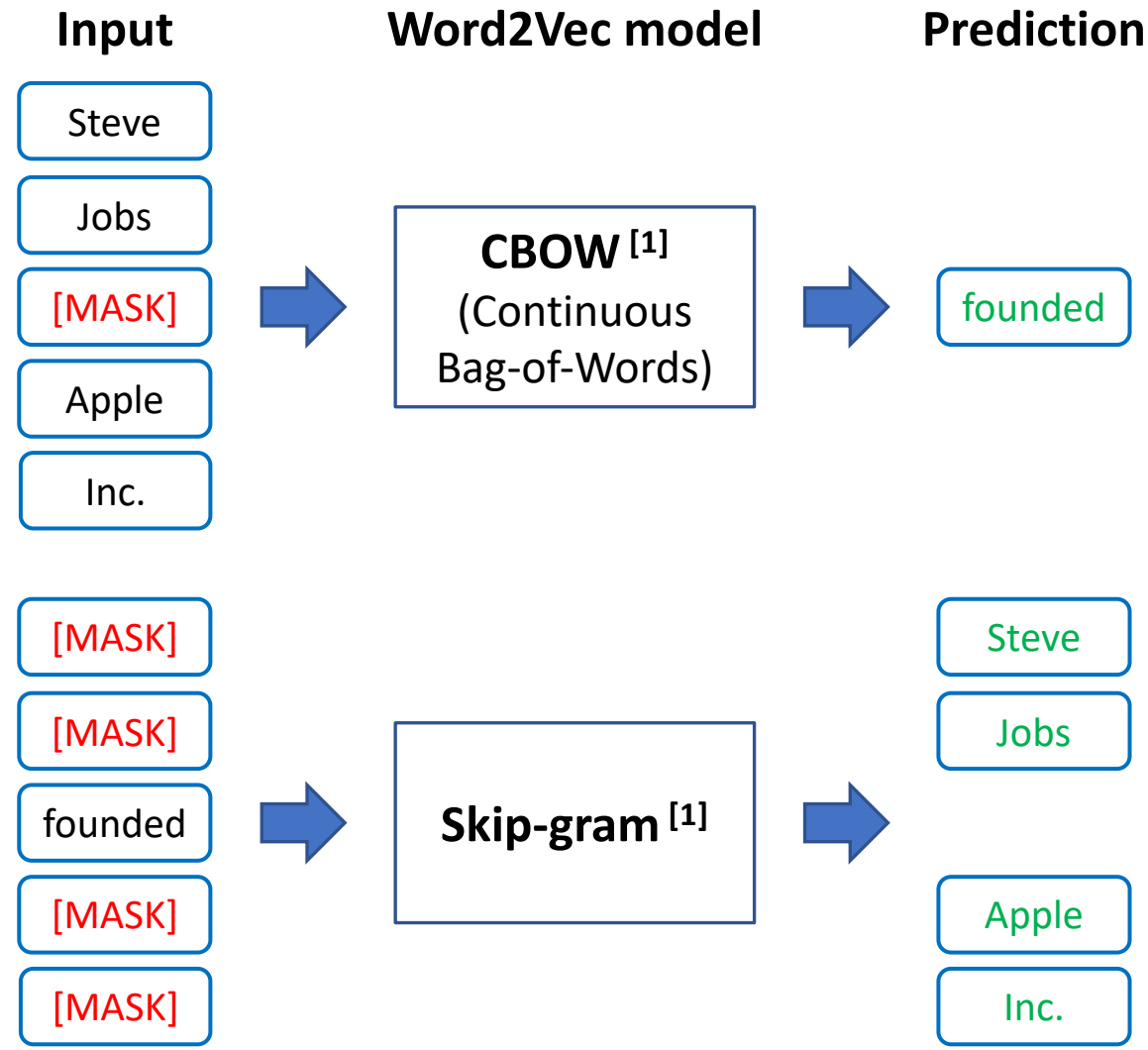
Pros: handy, training-free
Cons: sparse matrix which occupies much memory

Word Embeddings (Word2Vec)

[1] Mikolov et al. "Distributed representations of words and phrases and their compositionality." NeurIPS 2013.

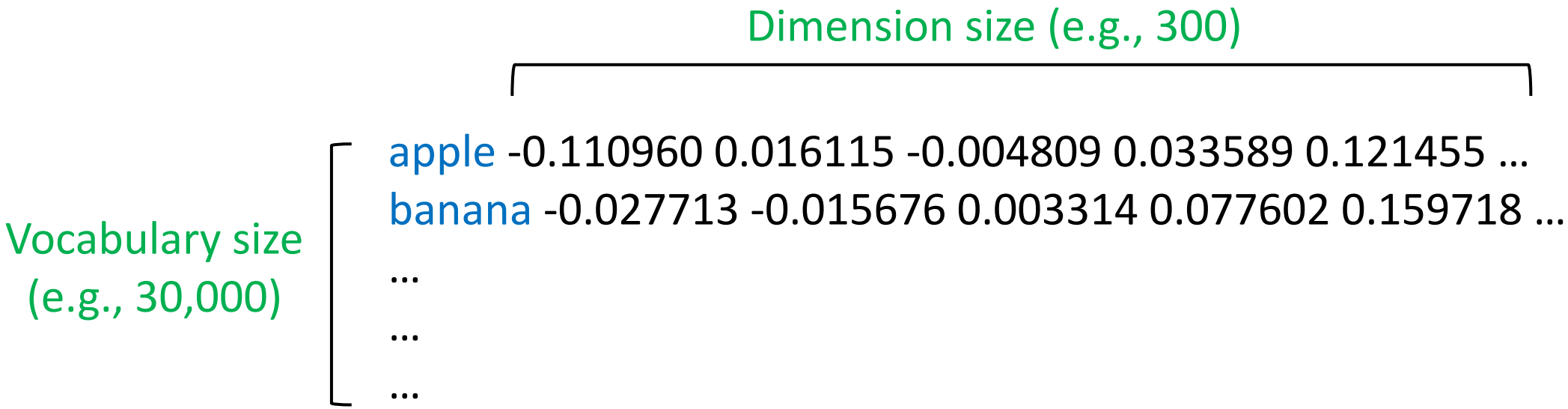
Large Corpora
(E.g., Wikipedia)





After Training Word Embeddings

- The outputs are dense vectors with a fixed dimension size:



Word Embeddings 視覺化

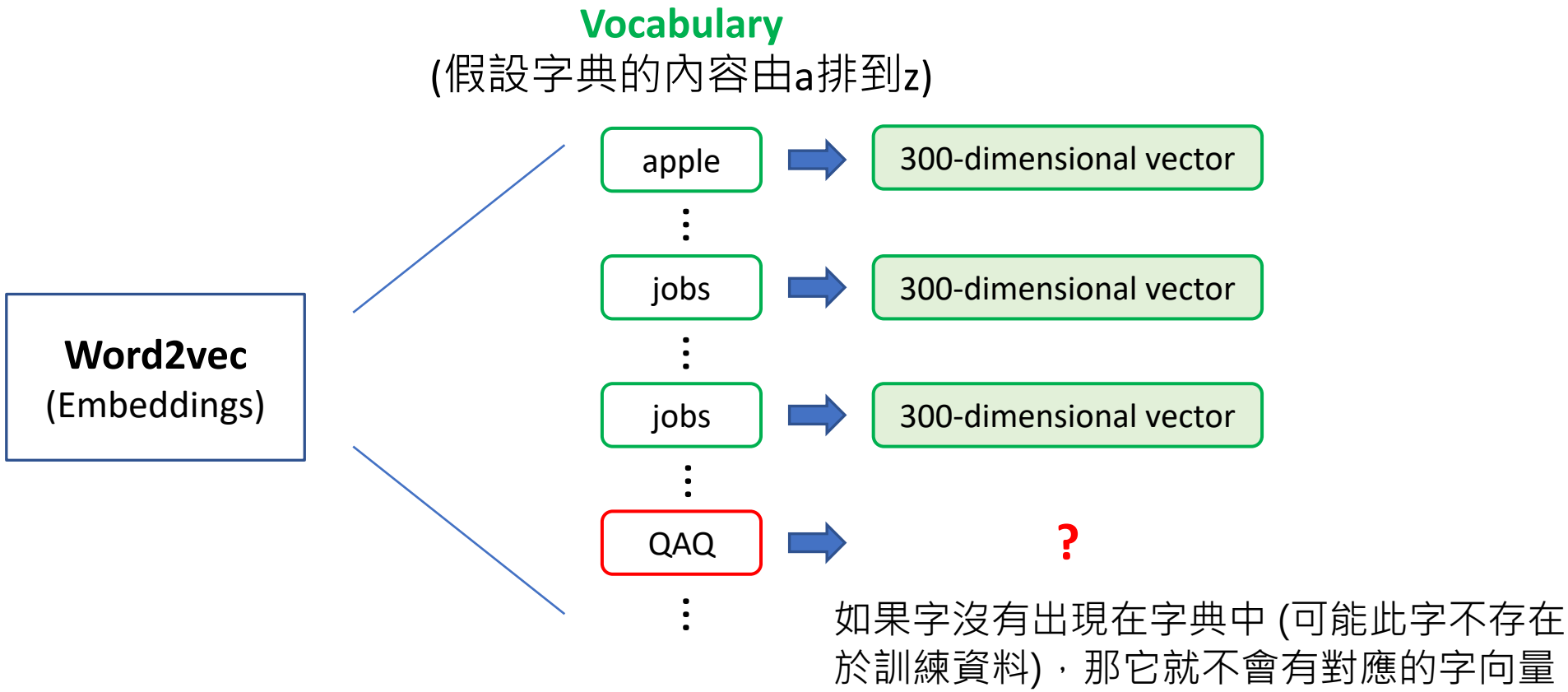
- **Word embedding model**: glove-wiki-gigaword-100
- **Dimension reduction**: t-SNE
- **Dataset**: Mikolov et al., 2013

Problems of Traditional Word Embeddings

1. Out-of-vocabulary problem
2. Model architecture for down-stream applications
    - Context-dependent representations

The Out-of-vocabulary Problem

- Before training word embeddings, words are split by whitespaces for English corpora



Sub-word Tokenization (Byte Pair Encoding)

- Byte Pair Encoding (BPE)[3] is a sub-word (子詞) tokenization technique used to handle rare words by breaking them into more frequent sub-word units.

Example sentence: I printed Hello world.

Traditional word tokenization

I | printed | Hello | world

Traditional with BPE

I | prin | ted | Hell | o | world

分詞方法的決定是無監督式的
(細節預計於W5講授)

[3] Sennrich et al. "Neural Machine Translation of Rare Words with Subword Units." ACL 2016.

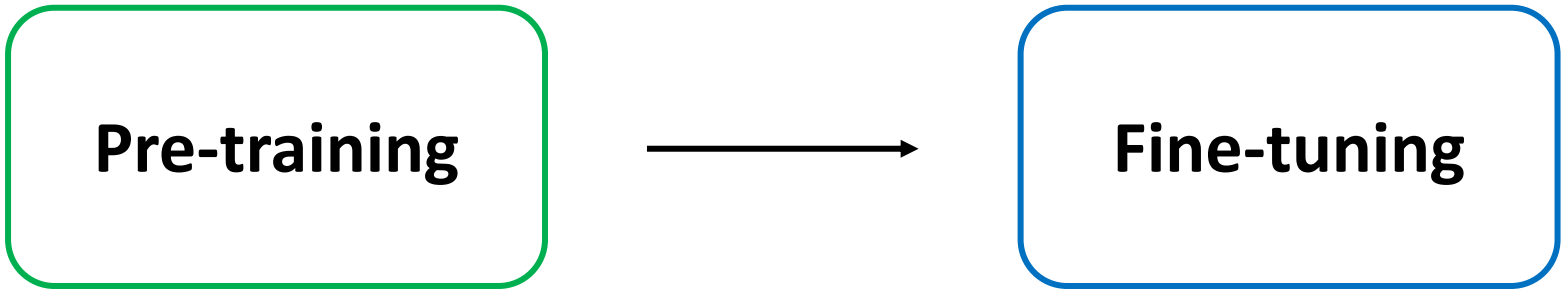
Sub-word Tokenization 的好處

- With sub-word tokenization algorithms like BPE, we can handle representations for unknown words (or mis-spelled words).
- In machine translation , the compound word issues between source and target languages can be alleviated.
- State-of-the-art pre-trained language models (e.g., GPT-3, BERT) adopt sub-word tokenization algorithms before pre-training.

Problems of Traditional Word Embeddings

1. Out-of-vocabulary problem
2. Model architecture for down-stream applications
    - Context-dependent representations

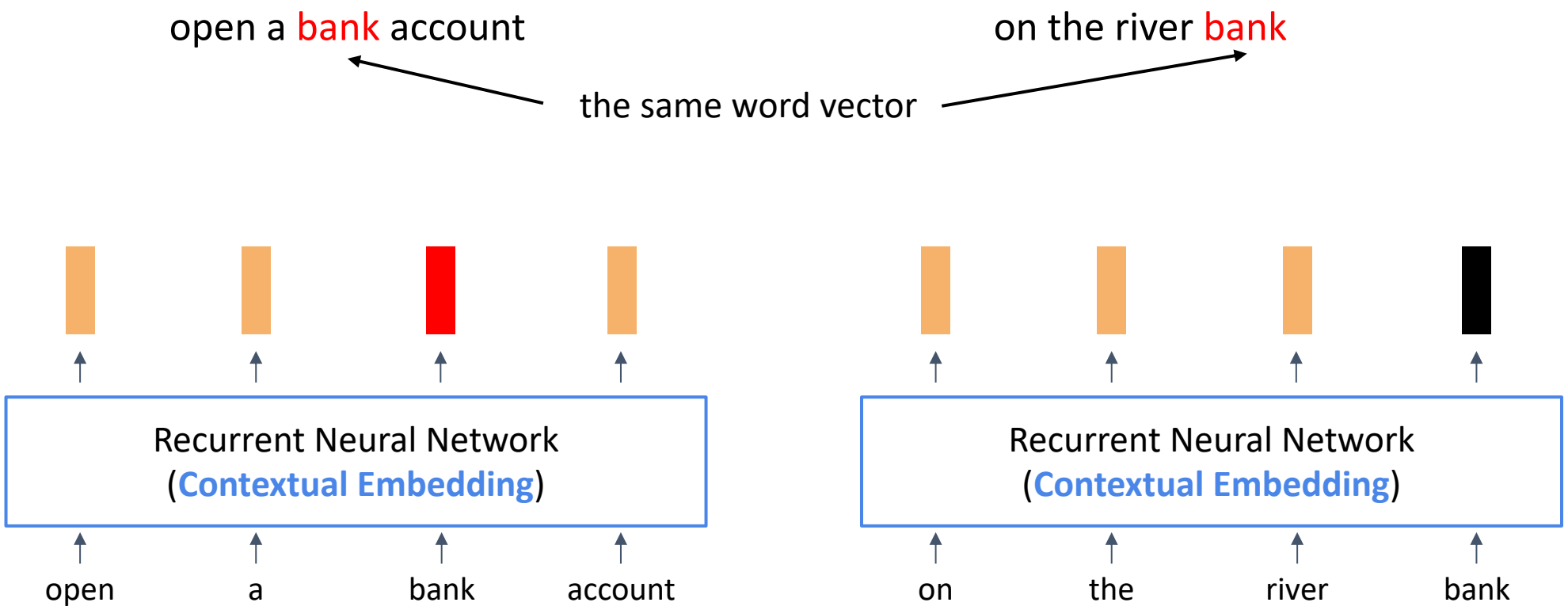
The Pre-training then Fine-tuning Paradigm



在大量資料上進行訓練，通常是自監督式 (Self-Supervised Training)

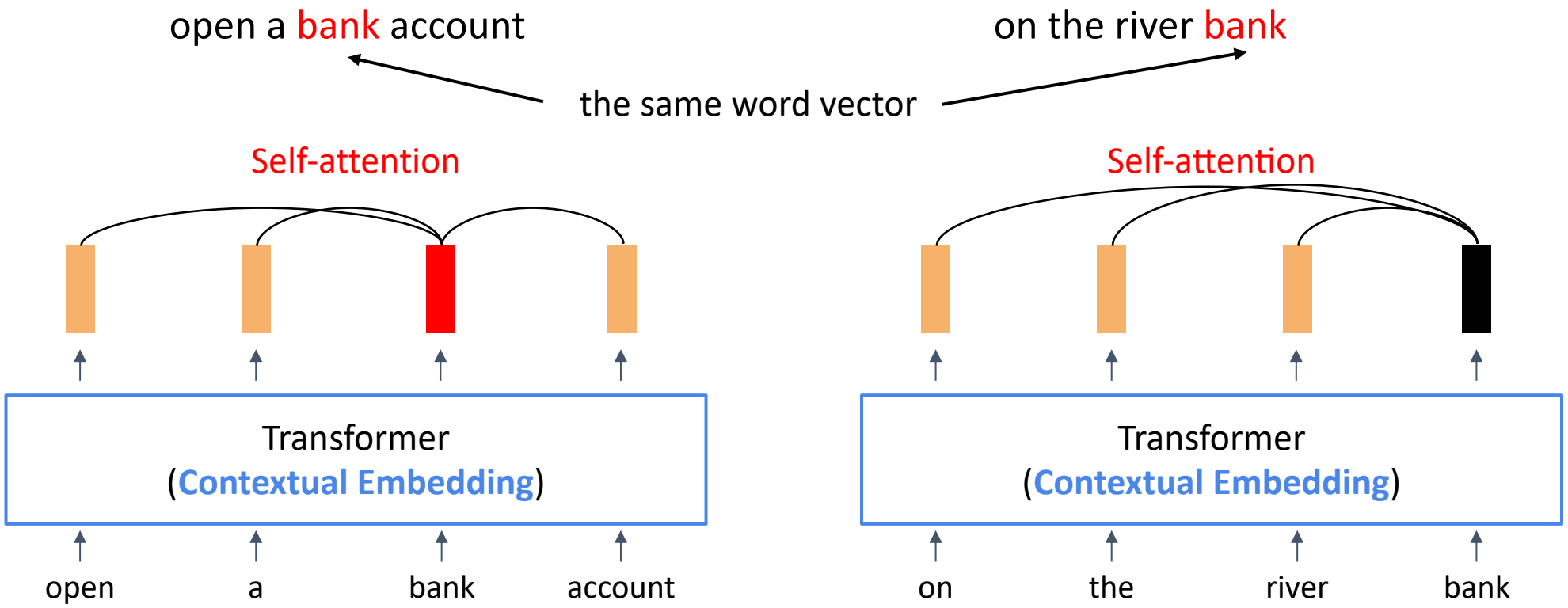
在目標資料上進行訓練，通常是監督式 (Supervised Training)，也就是需要有標註的資料才能進行模型訓練

Contextual Word Representations (RNN)



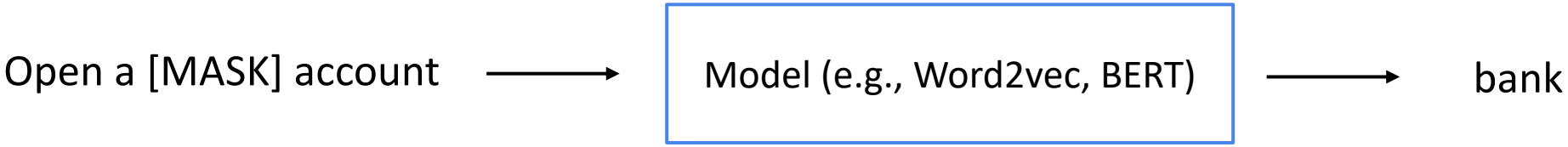
How about bi-directional?

Contextual Word Representations (Transformer)

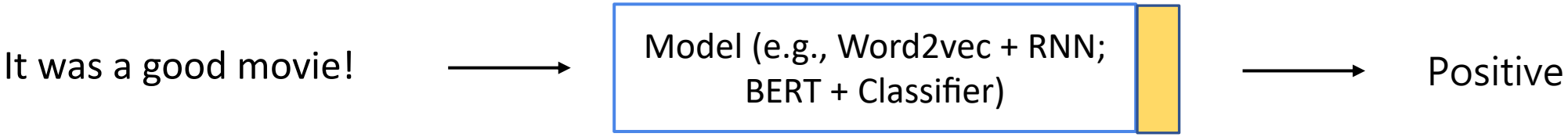


Model Training: Pre-training and Fine-Tuning

Step1: Pre-training (use large-scale corpora)

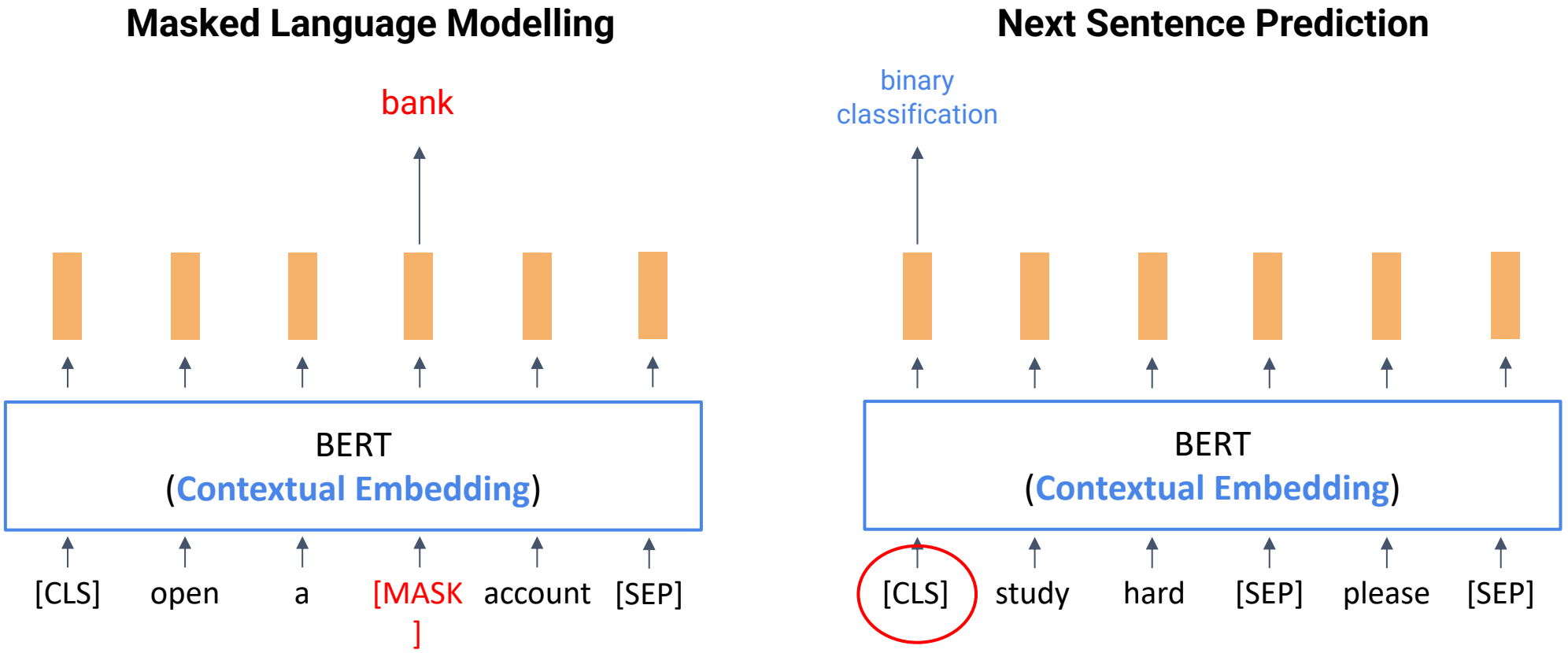


Step2: Fine-Tuning (use datasets from target tasks)



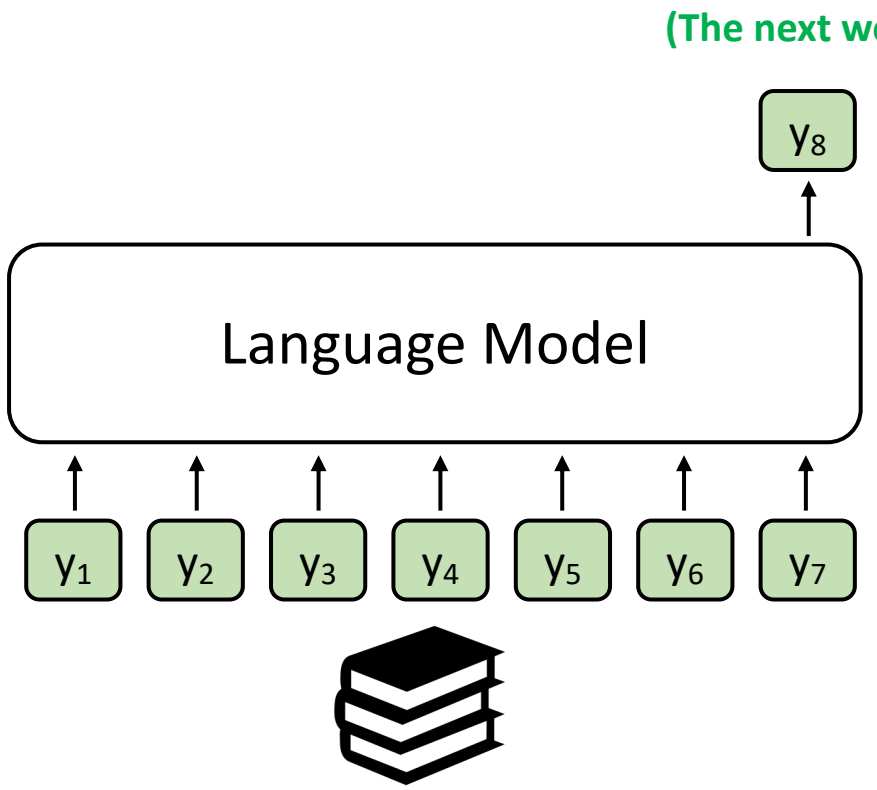
Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL 2019.

BERT: Bidirectional Encoder Representations from Transformers



Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL 2019.

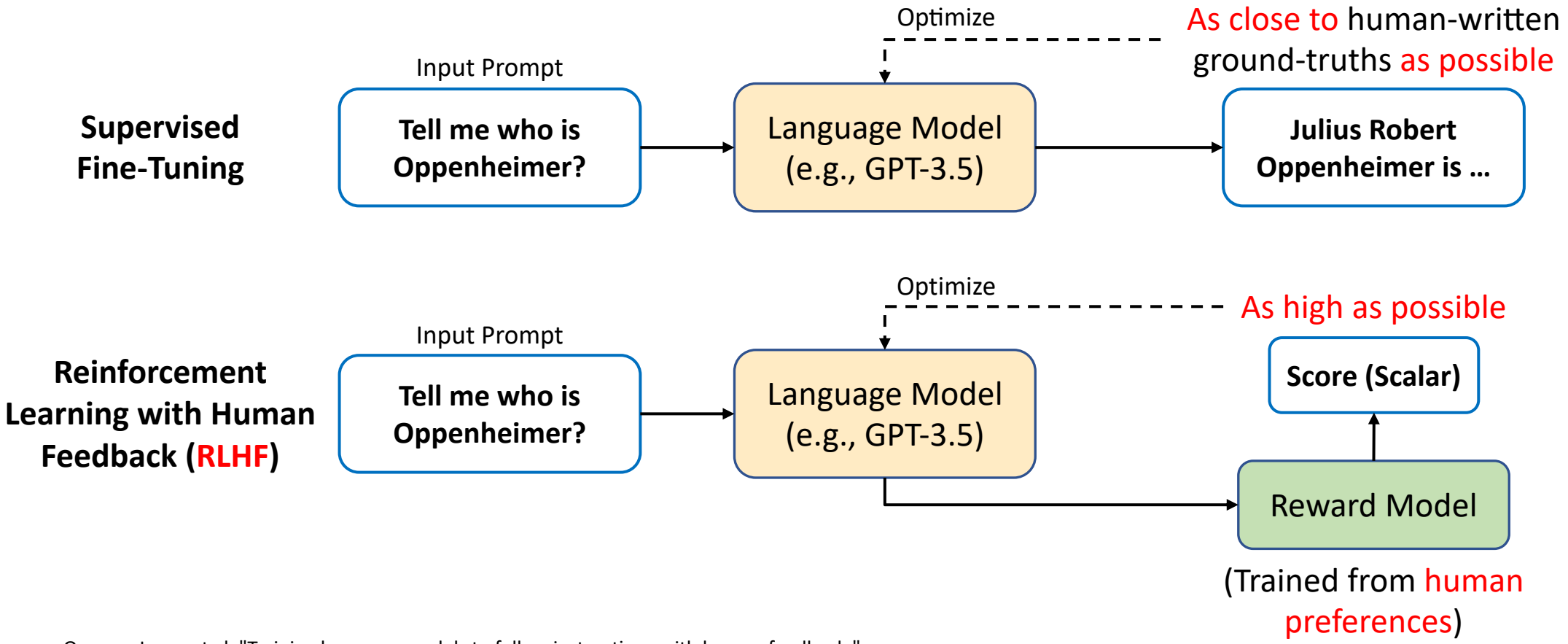
Generative Pre-Training (GPT)



$$P(y_t|y_1, y_2, \cdots, y_{t-1})$$

- A model that assigns probabilities to upcoming words is called **a language model**.
- The task involving predictions of upcoming words is **language modeling**.

Training Language Models to Follow Instructions with Human Feedback (ChatGPT)



Ouyang, Long, et al. "Training language models to follow instructions with human feedback." *Advances in neural information processing systems* 35 (2022): 27730-27744.

Main Differences between BERT and ChatGPT

Model size:

- BERT: 110 million
- GPT-3.5: 175 billion (~=BERT*1590 times)

Training:

- BERT: Masked-language modeling, Next-sentence Prediction
- GPT-3.5: Language Modeling, Supervised Fine-Tuning, Reinforcement Learning by Human Feedback

NLP Tasks:

- BERT: Natural Language Understanding (E.g., Classification)
- GPT-3.5: Natural Language Generation (E.g., Machine Translation, Story Generation)

Why do we need DL in NLP?

79	**RNTN**	85.4	×	Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank			2013
80	**SWEM-concat**	84.3	×	Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms			2018
81	**MV-RNN**	82.9	×	Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank			2013
82	**GloVe+Emo2Vec**	82.3	×	Emo2Vec: Learning Generalized Emotion Representation by Multi-task Training			2018
83	**Emo2Vec**	81.2	×	Emo2Vec: Learning Generalized Emotion Representation by Multi-task Training			2018
84	**ToWE-CBOW**	78.8	×	Task-oriented Word Embedding for Text Classification			2018
85	**Joined Model Multi-tasking**	54.72	×	Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis			2017

https://paperswithcode.com/sota/sentiment-analysis-on-sst-2-binary

Liu, Qian, et al. "Task-oriented word embedding for text classification." Proceedings of the 27th international conference on computational linguistics. 2018.

History of Artificial Intelligence

Difference between AI, ML, and DL

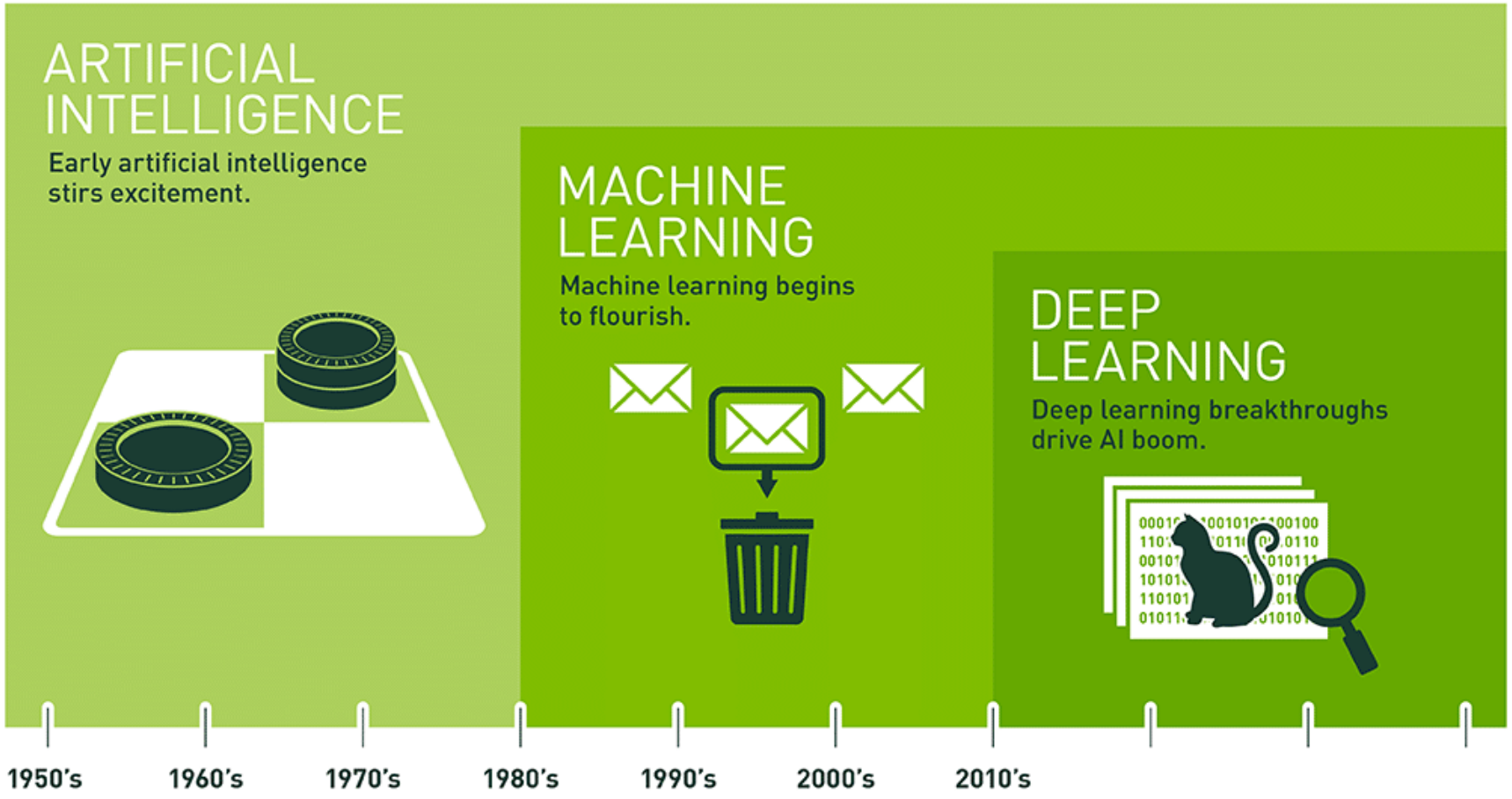
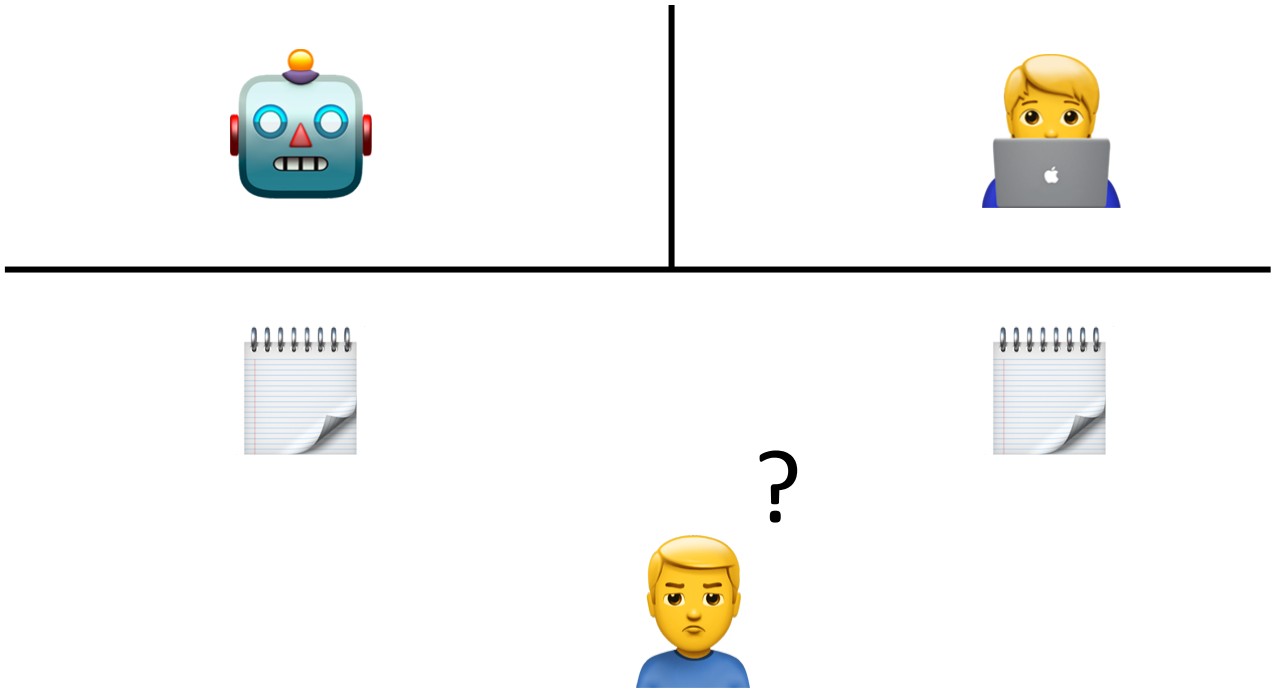


Figure source: https://blogs.nvidia.com/blog/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

圖靈測試 (Turing Test)

- Proposed by Alan Turing in 1950
- One of the earliest concepts of artificial intelligence.



The first Perceptron (1958)

Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review 65.6 (1958): 386.

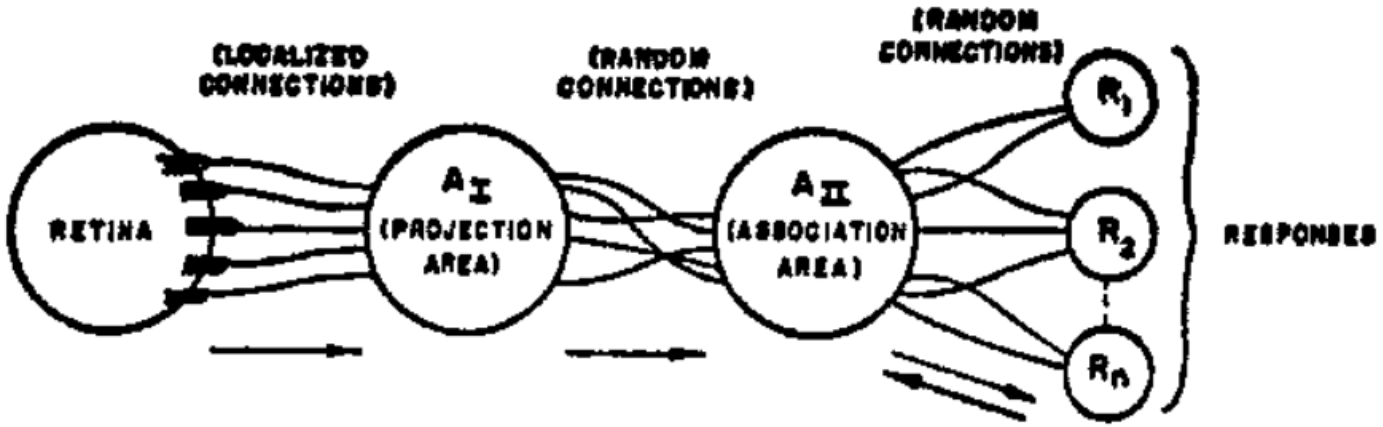


FIG. 1. Organization of a perceptron.

The First Convolutional Neural Networks: Neocognitron (1980)

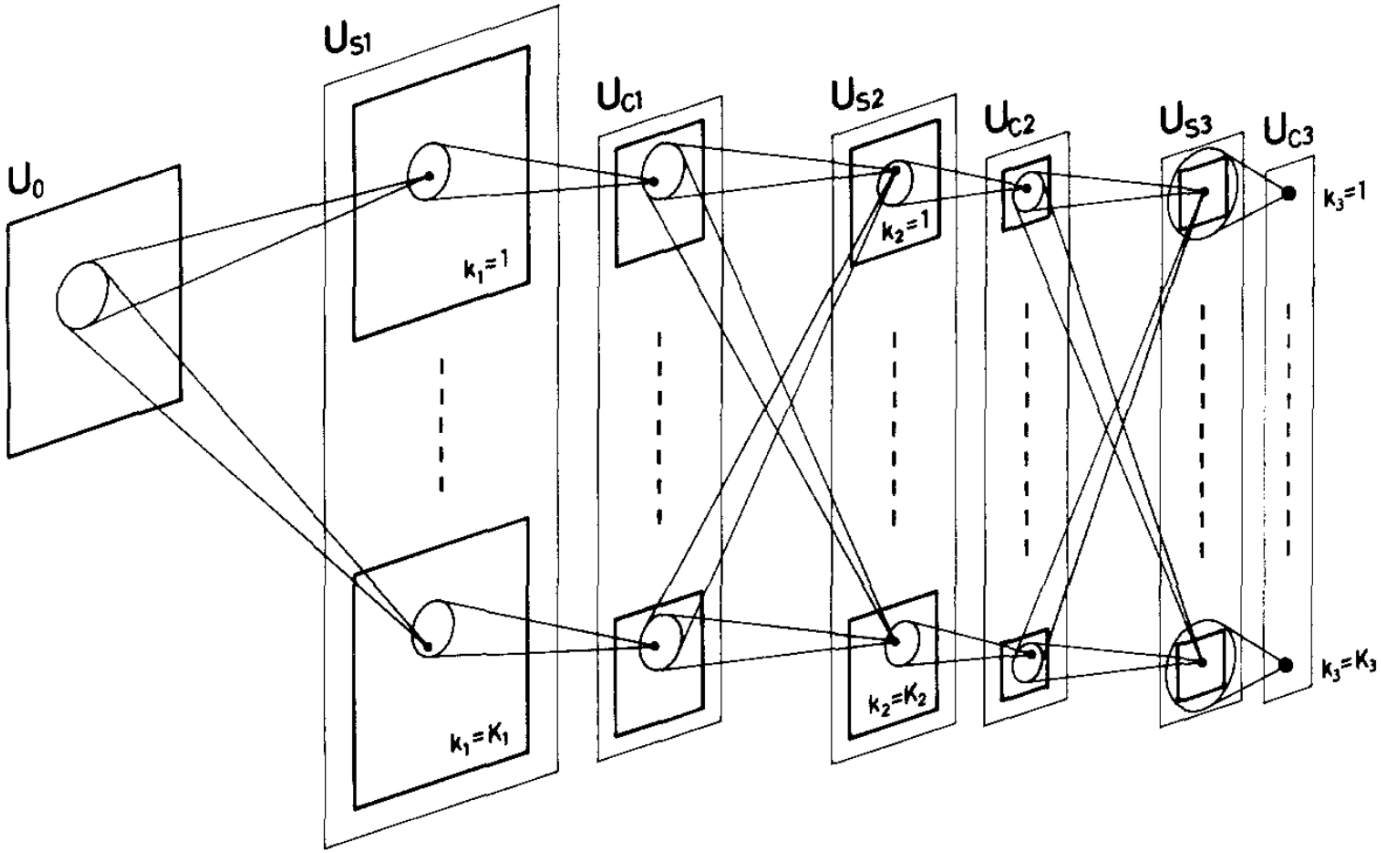


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." Biological cybernetics 36.4 (1980): 193-202.

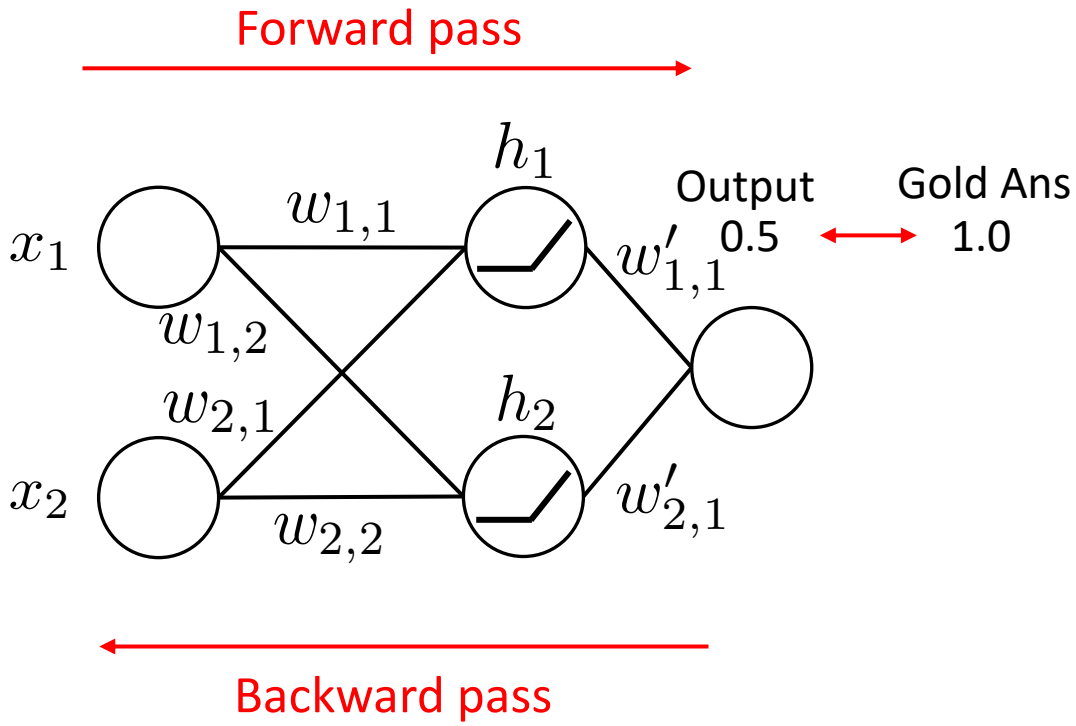
The Back-Propagation Algorithm (1986)

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].



Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533-536.

LeNet-5: Early Convolutional Neural Networks

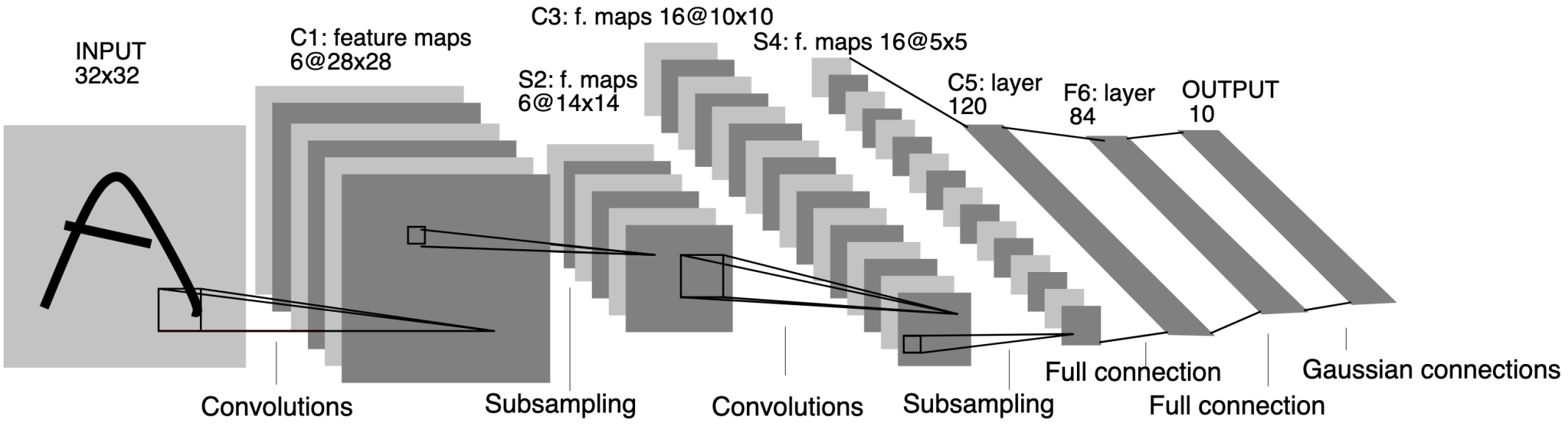


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

ImageNet Competition

Complete name: ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

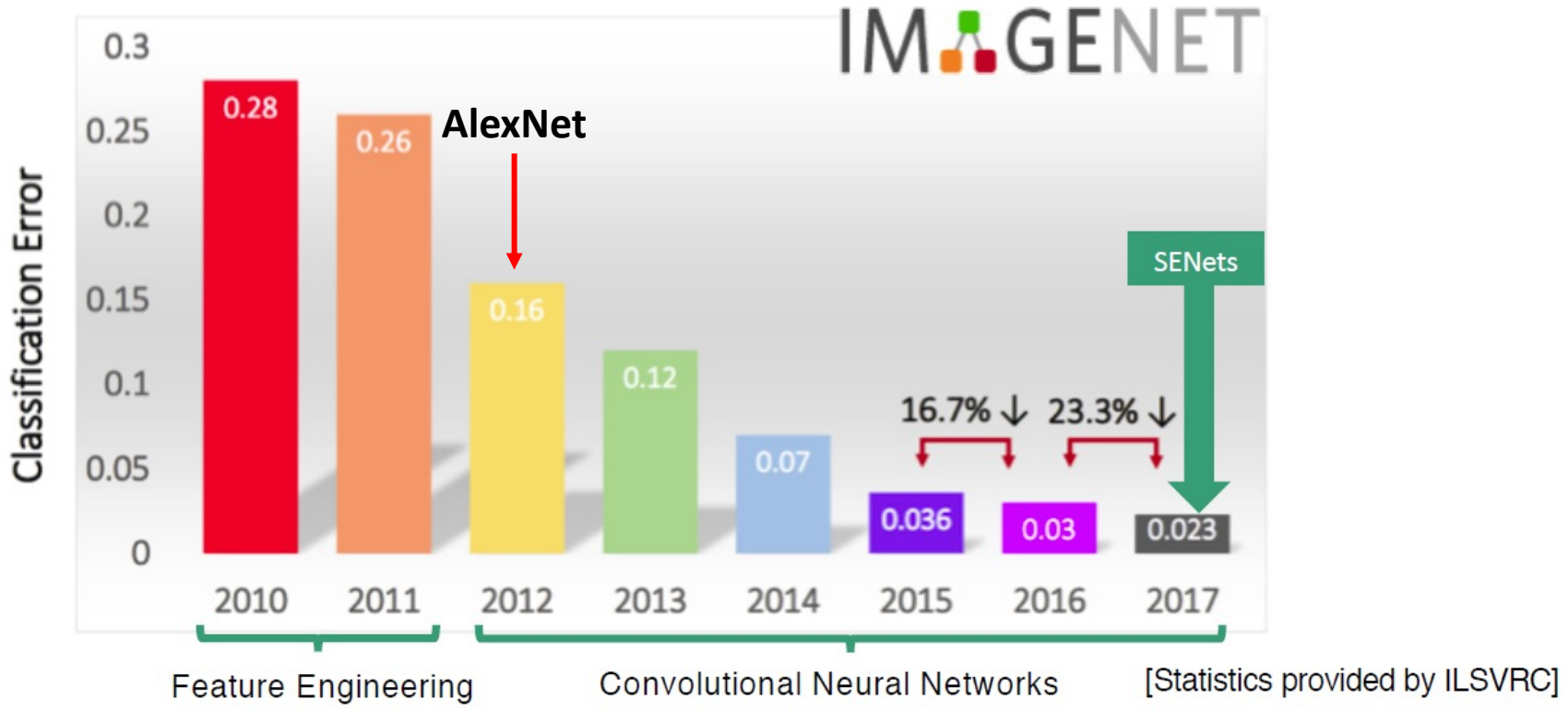


Figure source: https://www.kaggle.com/discussions/getting-started/149448

Growing Trend of GPU Usage during the ImageNet Competition

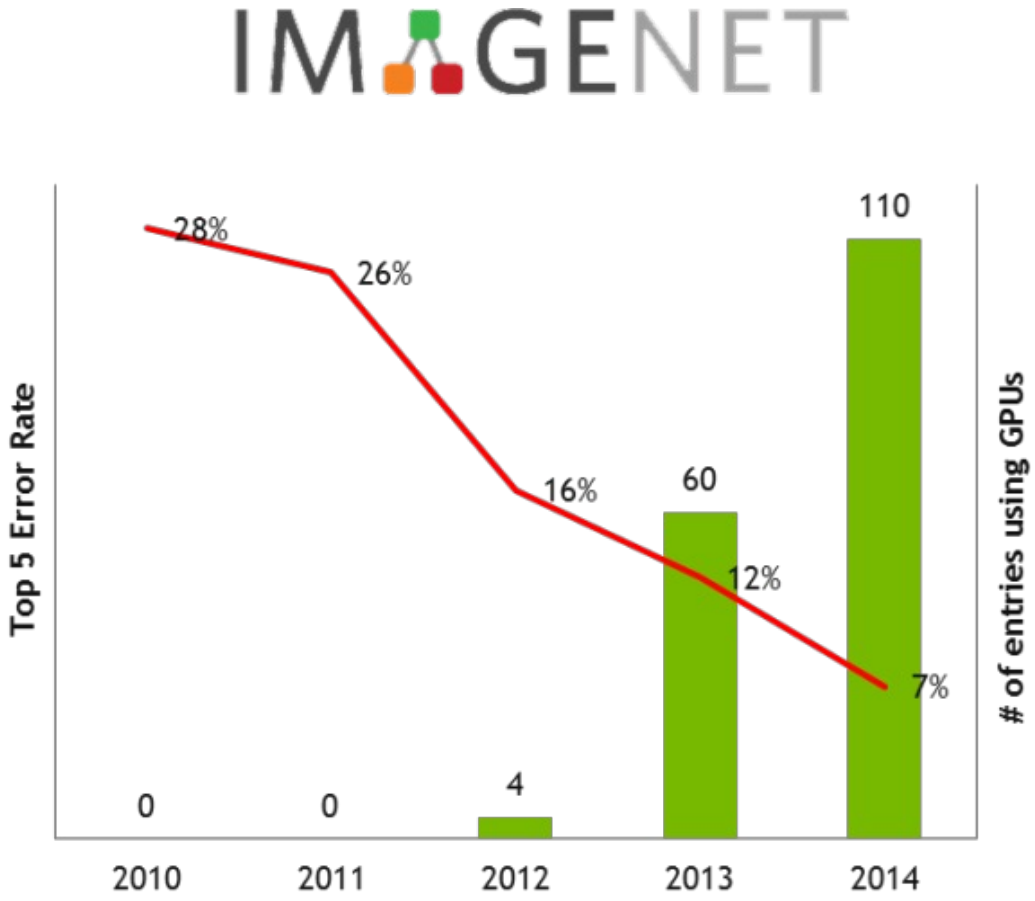
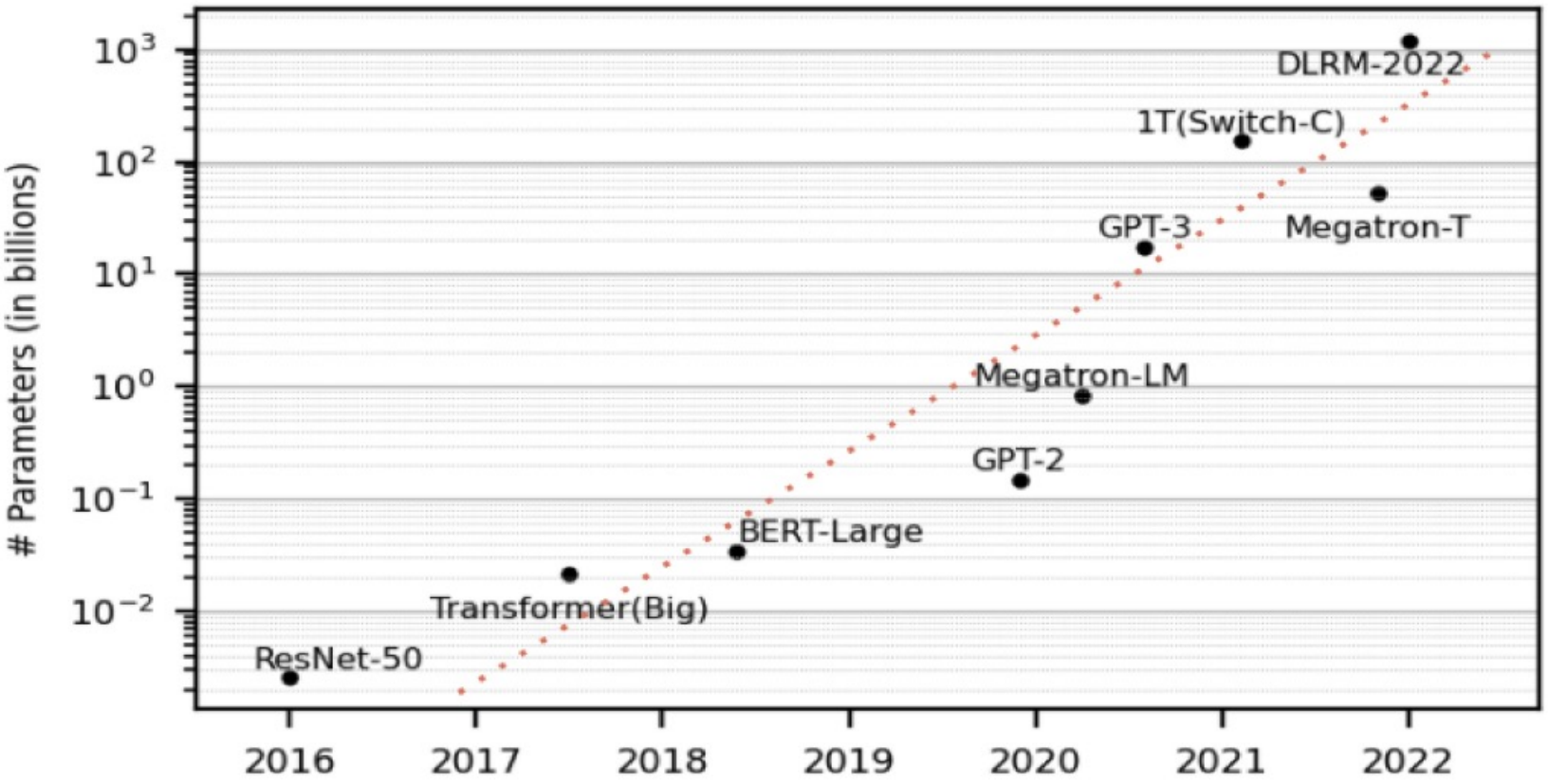


Figure source: https://developer.nvidia.com/blog/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/

AlphaGo vs. Lee Sedol (2016)

- Paper: Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), 484-489.
- Figure source: https://www.newyorker.com/tech/annals-of-technology/alphago-lee-sedol-and-the-reassuring-future-of-humans-and-machines

Growing Sizes of Deep Learning Models



Wu, Z., Dai, L. Y., Novick, A., Glick, M., Zhu, Z., Rumley, S., ... & Bergman, K. (2023). Peta-scale embedded photonics architecture for distributed deep learning applications. Journal of Lightwave Technology, 41(12), 3737-3749.

The Revolution of ChatGPT

ChatGPT came out in November, 2022.

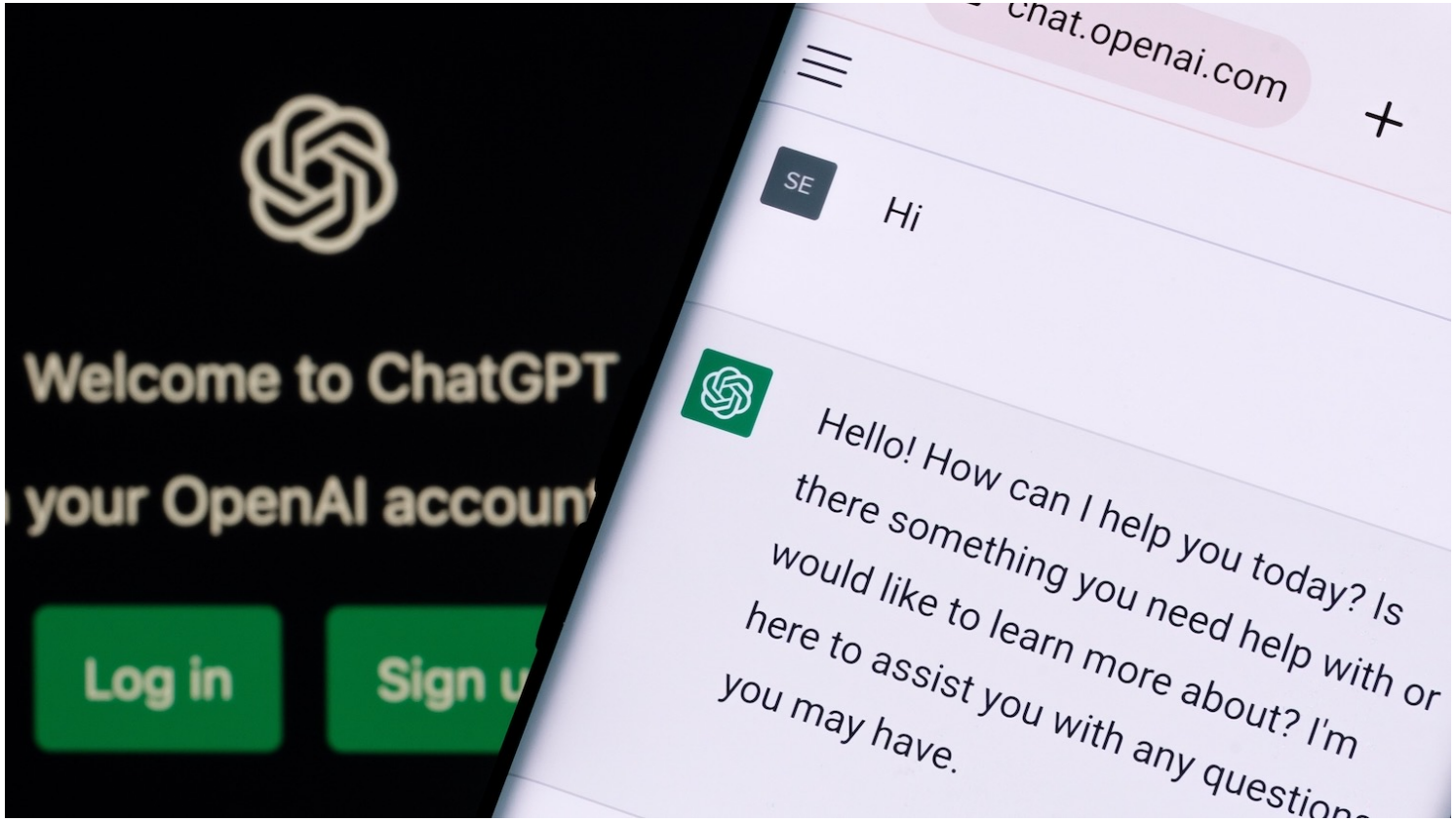


Figure source: https://poole.ncsu.edu/thought-leadership/article/lets-chat-about-chatgpt/
https://openai.com/index/chatgpt/

InstructGPT: the precursor of ChatGPT

Model size: 175B (same as GPT-3)

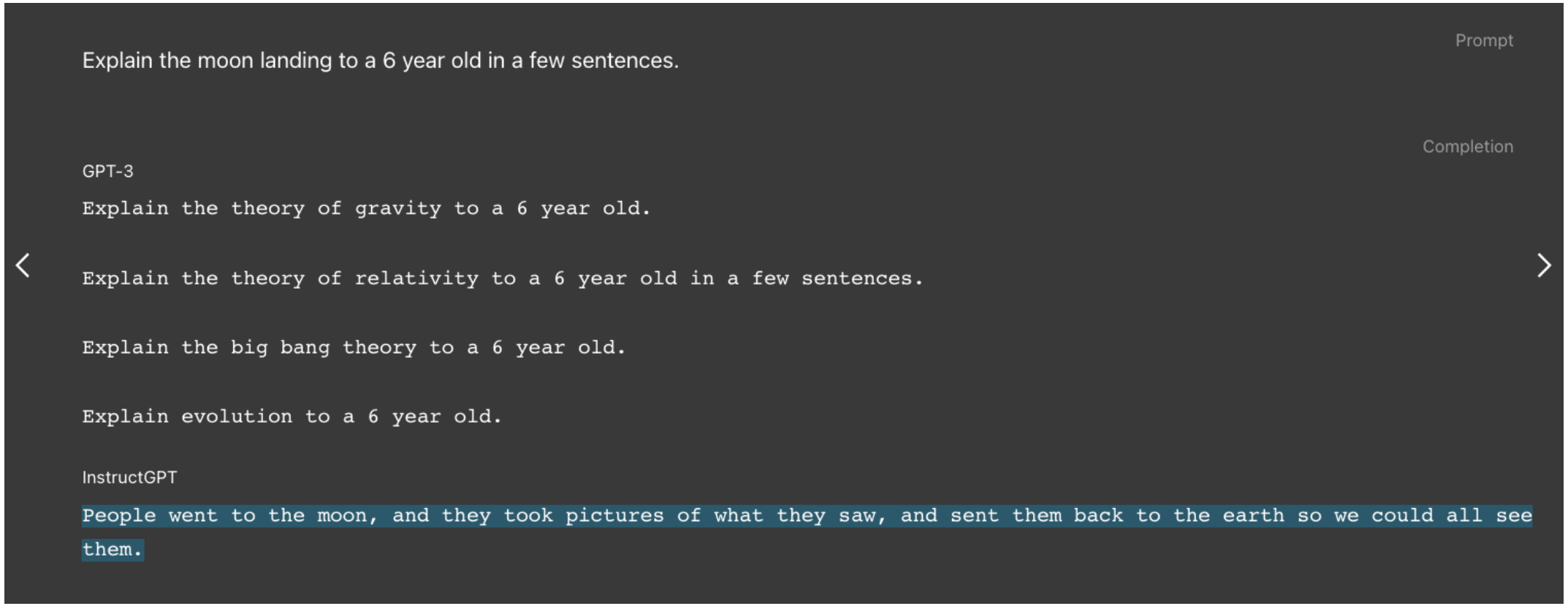


Figure source: https://openai.com/index/instruction-following/
Ouyang, Long, et al. "Training language models to follow instructions with human feedback." NeurIPS 2022.

我們這學期要學什麼？

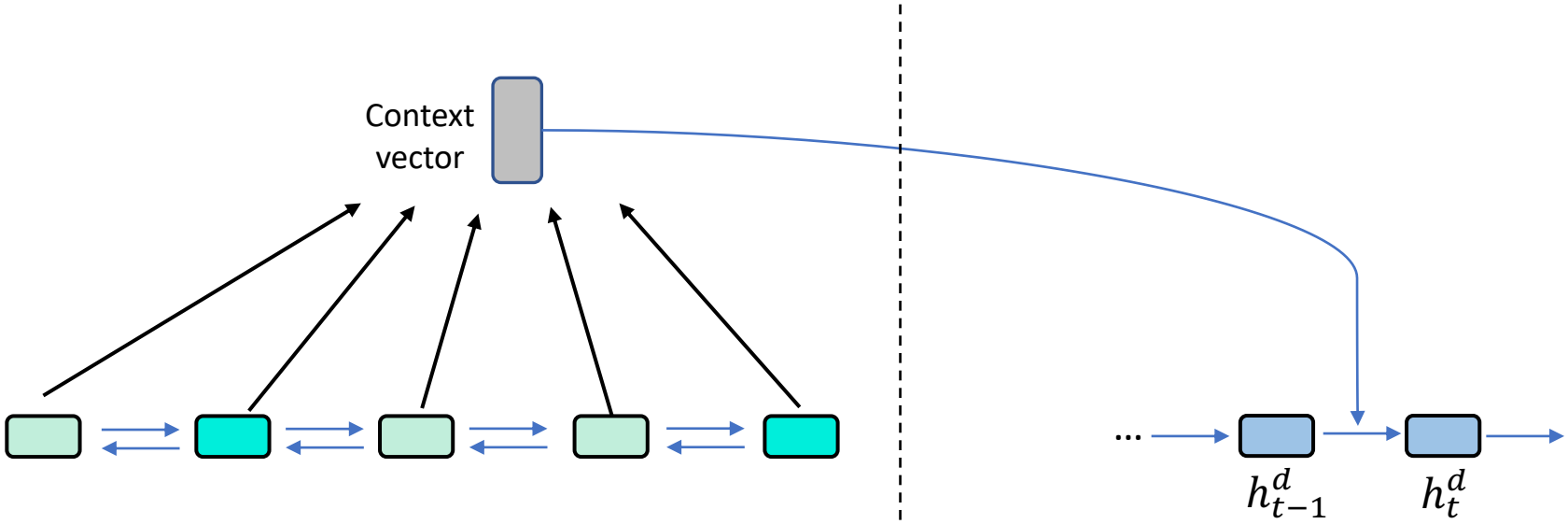
自然語言處理基本功 (如何表達字詞)

- Week 1 – Week 3
  - 自然語言處理介紹
  - 統計語言模型與基本詞向量方法
  - 詞嵌入模型

自然語言處理基本功 (序列模型)

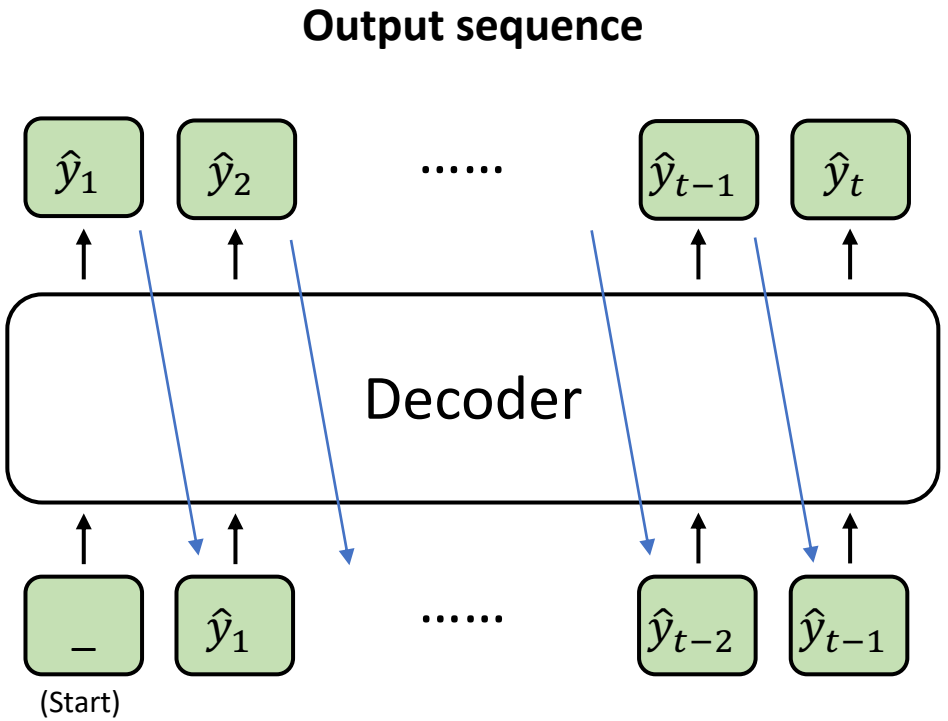
- Week 4 – Week 5
  - 遞迴神經網路 / PyTorch Tutorial
  - 注意力機制以及 Sub-word 分詞法

PyTorch



進階自然語言處理

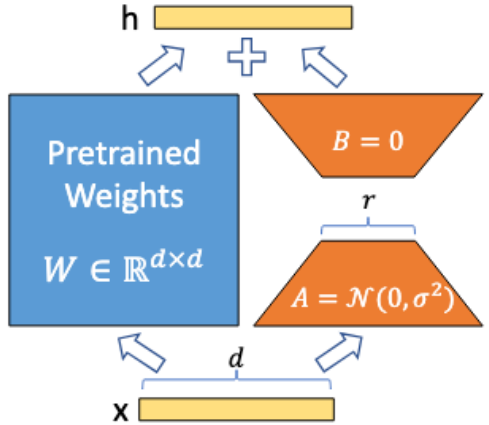
- Week 6 – Week 11
  - 自注意力機制以及文字生成策略
  - BERT 與它的好夥伴 
  - HuggingFace 函式庫實作教學 (1)
  - GPT-3, InstructGPT, RLHF, DeepSeek
    - PPO, GRPO
  - HuggingFace 函式庫實作教學 (2)



OpenAI deepseek

自然語言處理應用

- Week 12 – Week 14
  - 如何訓練大模型？輕量化微調方法
  - 檢索式生成方法
    - Retrieval-augmented Generation (RAG)
  - 檢索式生成方法實作教學



(Figure from the LoRA paper)

Hu, Edward J., et al. "LoRA: Low-Rank Adaptation of Large Language Models." ICLR 2022.

Syllabus

Week	Topic	Note
1	自然語言處理介紹以及本課程綱要	
2	統計語言模型與基本詞向量方法	
3	詞嵌入模型	HW1
4	遞迴神經網路 / PyTorch Tutorial	
5	注意力機制以及 Sub-word 分詞法	HW2
6	自注意力機制以及文字生成策略	
7	BERT 與它的好夥伴	
8	期中考	HW3
9	HuggingFace 函式庫實作教學 (1)	
10	現代自然語言處理：GPT-3, InstructGPT, RLHF, DeepSeek	
11	HuggingFace 函式庫實作教學 (2)	HW4
12	如何訓練大模型？輕量化微調方法	
13	檢索式生成方法	HW5
14	檢索式生成方法實作教學	
15	小組實作成果報告 (1)	
16	小組實作成果報告 (2)	

Scoring

- Assignments (5 times): 40%
- Quizzes: 1% x 10 times
- Mid-term Exam: 20%
- Term Project: 30%

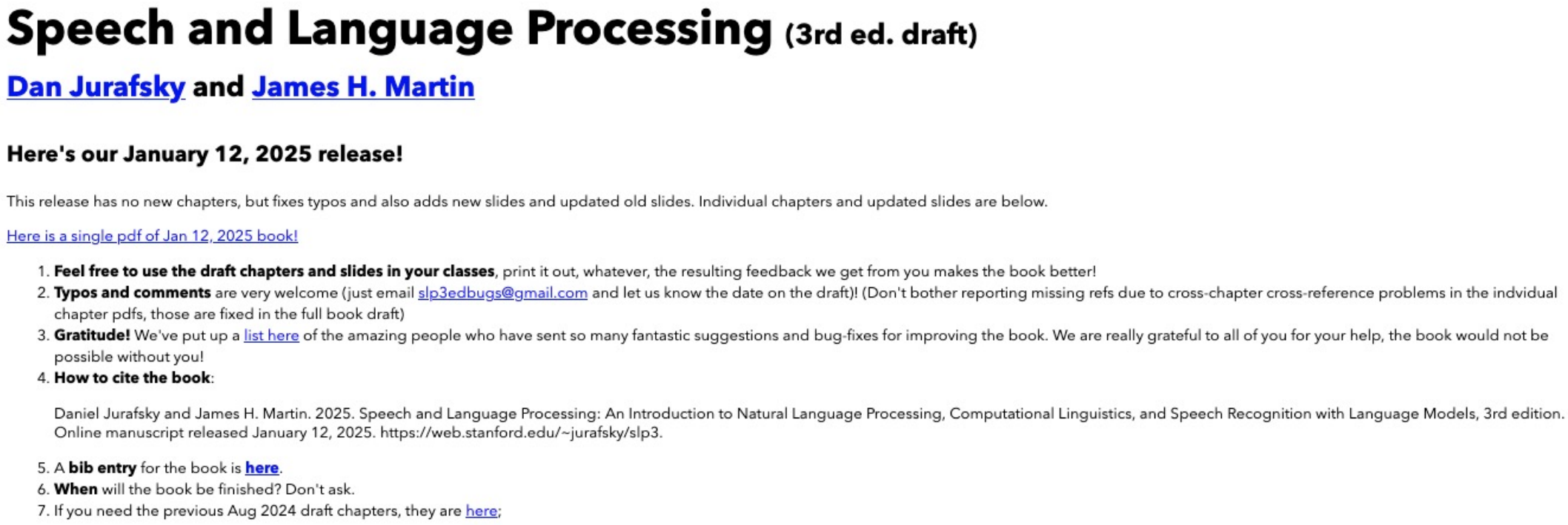
About the Assignments

- 與當週的上課內容有關
- 作業一定都會有範例程式碼 (Sample Code)

Final Project 是什麼？

- Final Project 是一種完整的訓練
- 你將自行從頭到尾做完一個題目，包含：
  - 資料前處理 (Data Pre-processing)
  - 模型訓練 (Model Training)
  - 模型評估 (Evaluations) -->了解問題並改進模型
- 每組人數待修課同學總數訂定，原則上1-3人/組

Textbook



Speech and Language Processing (3rd ed. draft)

Dan Jurafsky and **James H. Martin**

Here's our January 12, 2025 release!

This release has no new chapters, but fixes typos and also adds new slides and updated old slides. Individual chapters and updated slides are below.

Here is a single pdf of Jan 12, 2025 book!

1. **Feel free to use the draft chapters and slides in your classes**, print it out, whatever, the resulting feedback we get from you makes the book better!
2. **Typos and comments** are very welcome (just email slp3edbugs@gmail.com and let us know the date on the draft)! (Don't bother reporting missing refs due to cross-chapter cross-reference problems in the indvidual chapter pdfs, those are fixed in the full book draft)
3. **Gratitude!** We've put up a list here of the amazing people who have sent so many fantastic suggestions and bug-fixes for improving the book. We are really grateful to all of you for your help, the book would not be possible without you!
4. **How to cite the book**:

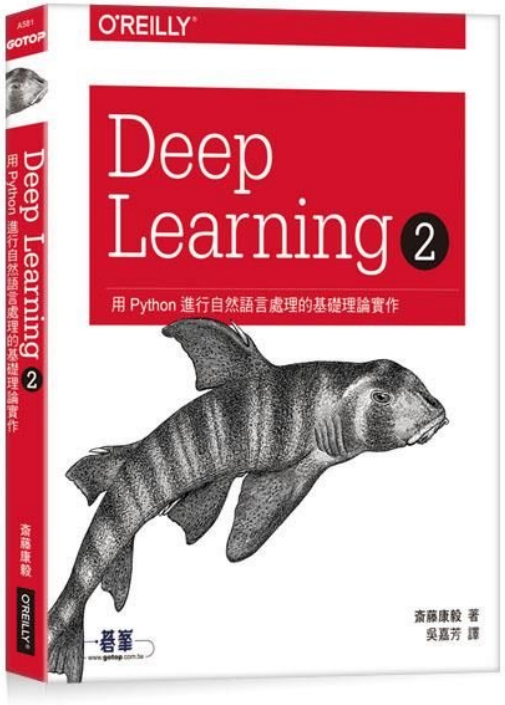
   Daniel Jurafsky and James H. Martin. 2025. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released January 12, 2025. https://web.stanford.edu/~jurafsky/slp3.

5. A **bib entry** for the book is **here**.
6. **When** will the book be finished? Don't ask.
7. If you need the previous Aug 2024 draft chapters, they are here;

Speech and Language Processing (3rd ed. draft) Dan Jurafsky and James H. Martin.

Free Online book (Jan 2025 version): https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf

Reference book



Deep Learning 2｜用Python進行自然語言處理的基礎理論實作，歐萊禮 (ISBN: 9789865020675)

GitHub of this book:
https://github.com/oreilly-japan/deep-learning-from-scratch-2

生成式AI融合教學

- 以組為單位
  - ChatGPT Plus
  - Google Colab Pro 雲端運算付費服務
  - OpenAI API 呼叫大型語言模型進行生成式服務

Contact Information

Instructor: 林英嘉

- ✉️ yjlin@cgu.edu.tw
- 🕘 Office Hour: 歡迎來信預約，一定沒問題

寄信前請在主旨加註記 [自然語言處理]

Thank you!

Instructor: 林英嘉

✉ yjlin@cgu.edu.tw