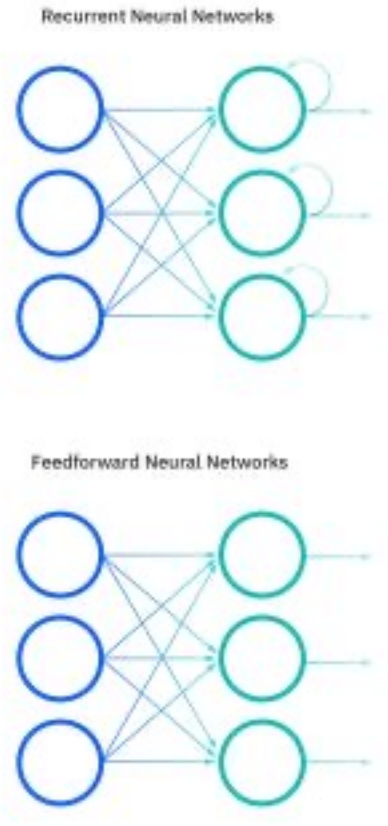
NLP研究热点变迁

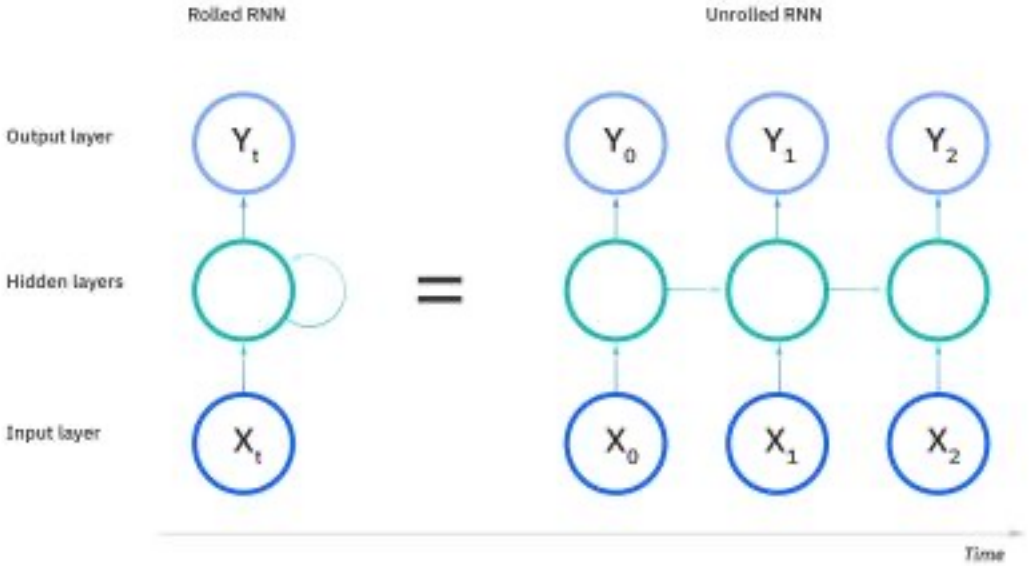
第一课

Scott

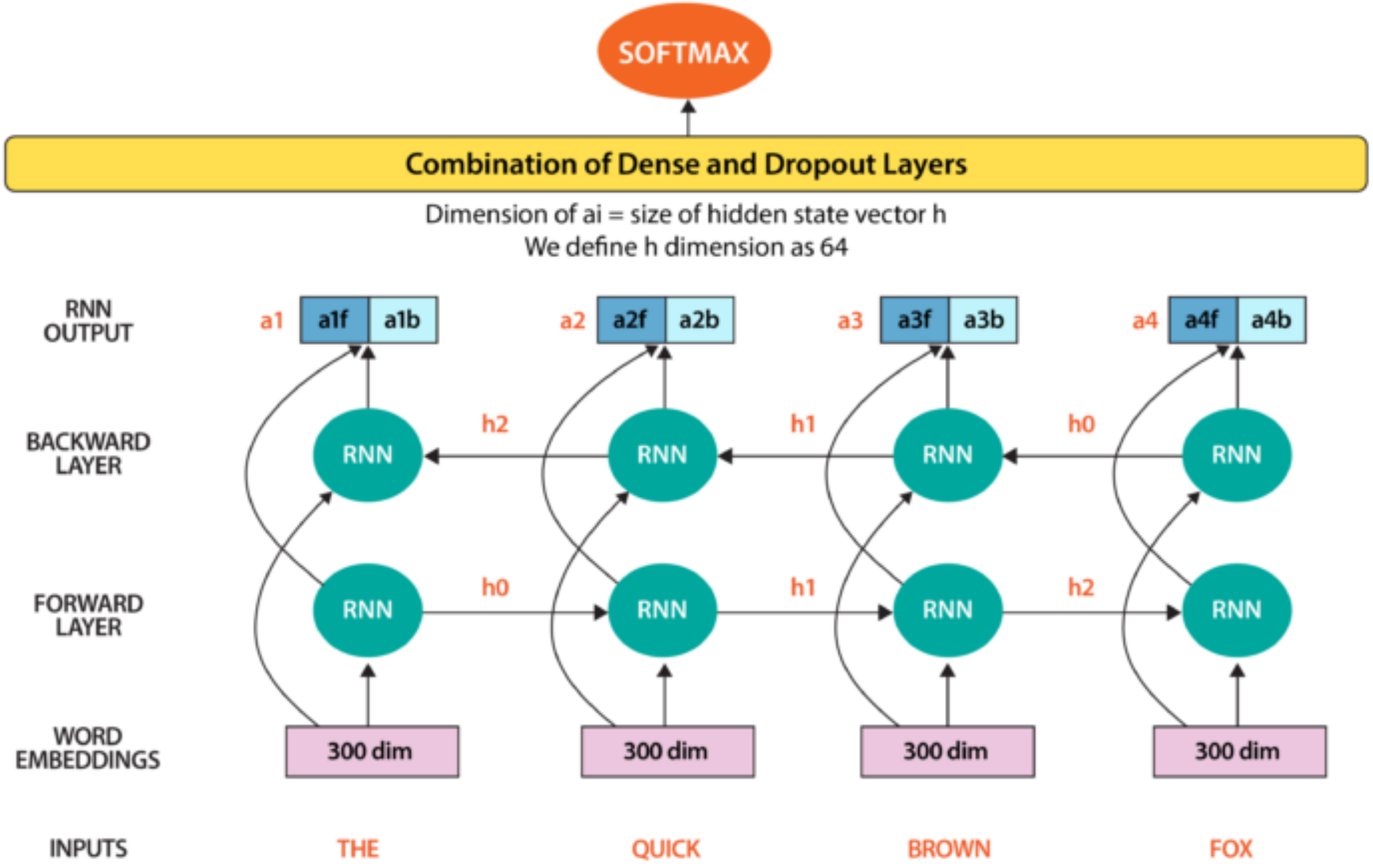
Pre-transformer

- RNN
- LSTM
- GRU



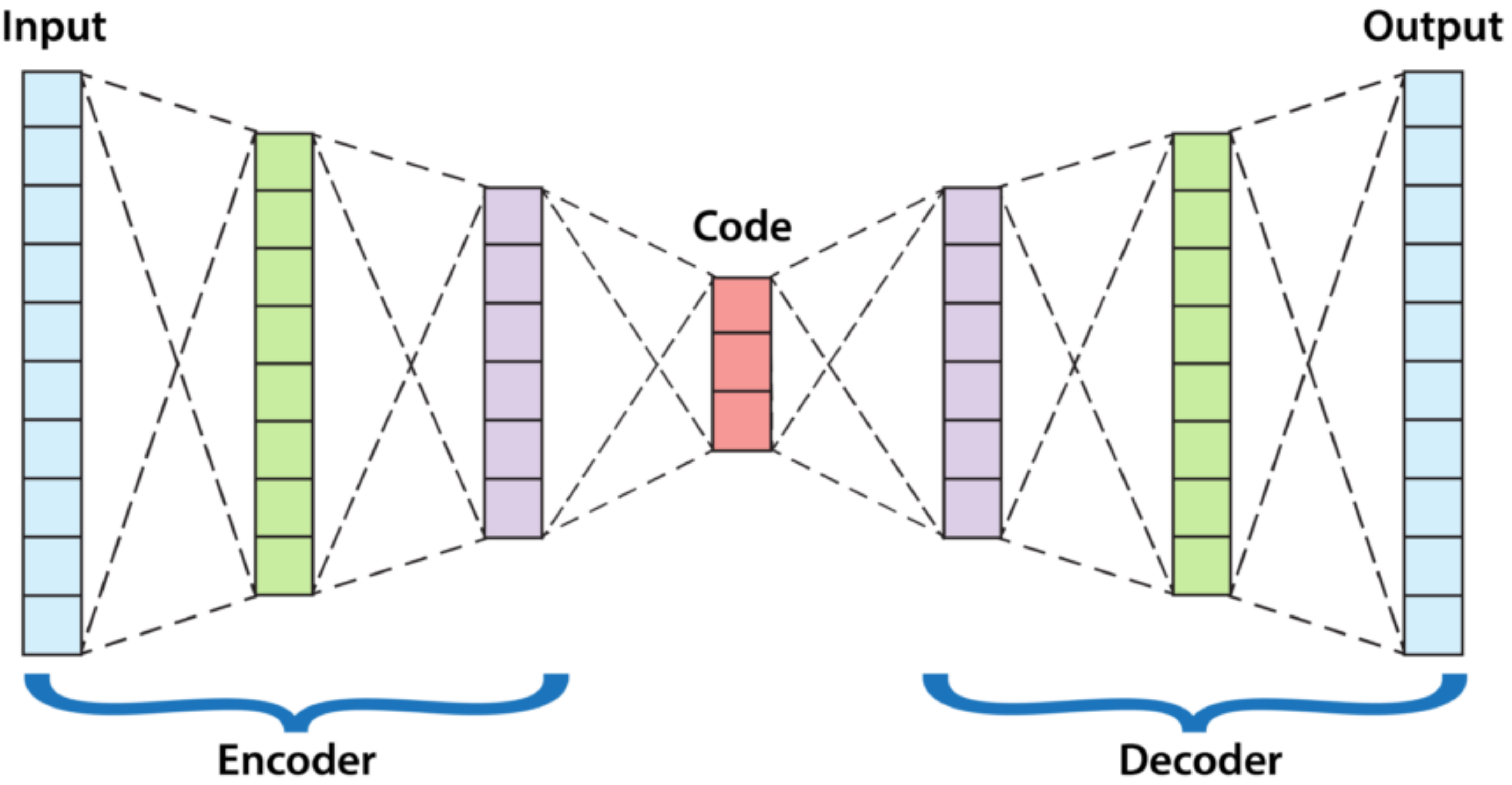


RECURRENT NEURAL NETWORK



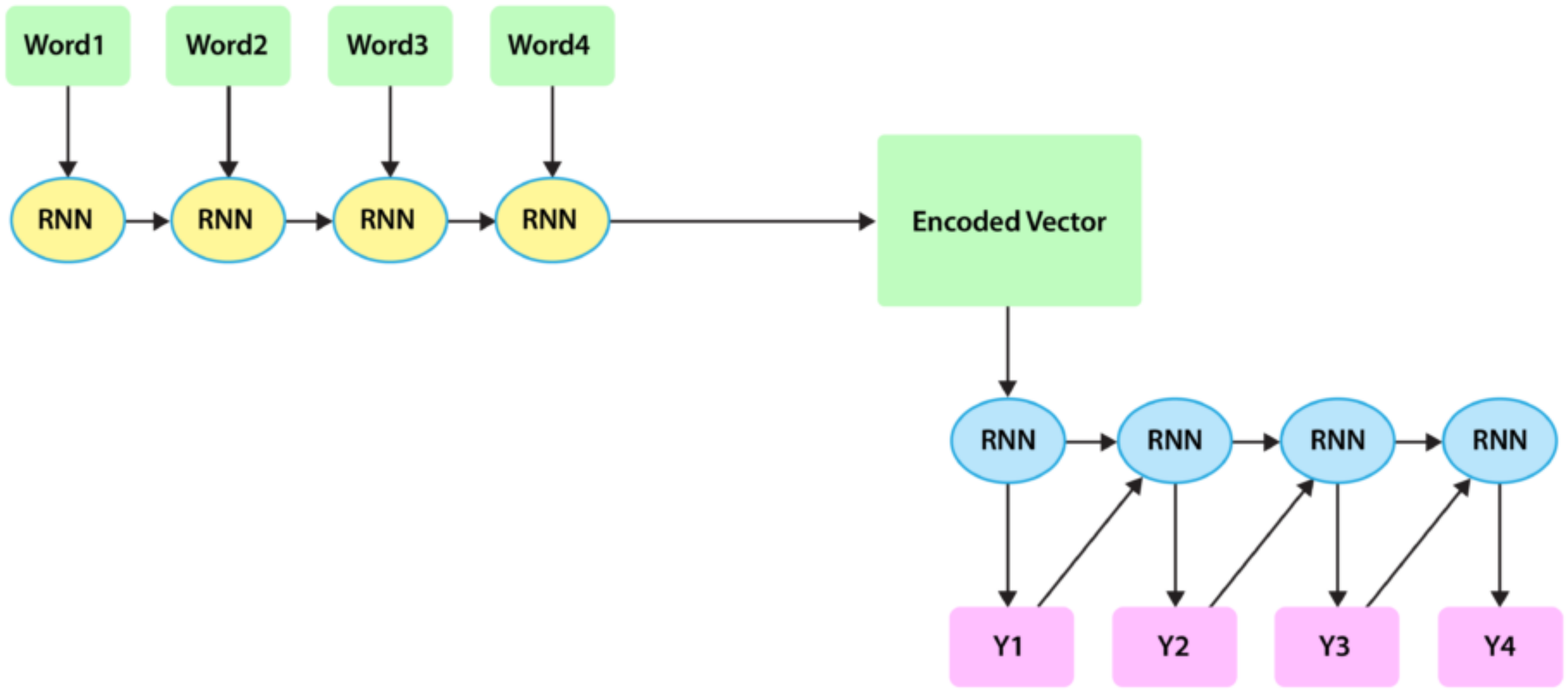
A bidirectional recurrent neural network processes the input both forward and backward to improve the representations it produces.

AUTO-ENCODER



An autoencoder uses an encoder to compress an input into a representation and a decoder to reconstruct the input from the representation.

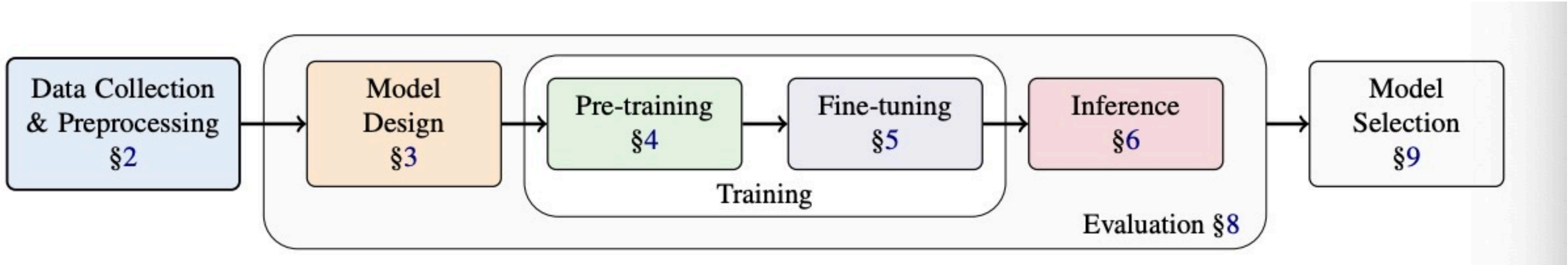
SEQ2SEQ MODEL FOR TRANSLATION



Given a sentence, a Recurrent Neural Network encodes the sentence and then iteratively generates a translation.

后transformer时代

- 自然语言处理发展进程（后transformer时代）范式转移与统一范式



自然语言推理

The boy is sleeping

The boy is resting

Field of Study	# Papers	Representative Papers	Field of Study	# Papers	Representative Papers
Machine Translation	12,922	Liu et al. (2020), Goyal et al. (2022)	Visual Data in NLP	2,401	Tan and Bansal (2019), Xu et al. (2021)
Language Models	11,005	Devlin et al. (2019), Ouyang et al. (2022)	Ethical NLP	2,322	Blodgett et al. (2020), Perez et al. (2022)
Representation Learning	6,370	Reimers and Gurevych (2019), Gao et al. (2021b)	Question Answering	2,208	Karpukhin et al. (2020), Liu et al. (2022b))
Text Classification	6,117	Wei and Zou (2019), Hu et al. (2022)	Tagging	1,968	Malmi et al. (2019), Wei et al. (2020)
Low-Resource NLP	5,863	Gao et al. (2021a), Liu et al. (2022a)	Summarization	1,856	Liu and Lapata (2019), He et al. (2022)
Dialogue Systems & Conversational Agents	4,678	Zhang et al. (2020), Roller et al. (2021)	Green & Sustainable NLP	1,780	Strubell et al. (2019), Ben Zaken et al. (2022)
Syntactic Parsing	4,028	Zhou and Zhao (2019), Glavaš and Vulić (2021)	Cross-Lingual Transfer	1,749	Conneau et al. (2020), Feng et al. (2022)
Speech & Audio in NLP	3,915	Baevski et al. (2022), Wang et al. (2020)	Morphology	1,749	McCarthy et al. (2020), Goldman et al. (2022)
Knowledge Representation	2,967	Schneider et al. (2022), Safavi and Koutra (2021)	Explainability & Interpretability in NLP	1,671	Danilevsky et al. (2020), Pruthi et al. (2022)
Structured Data in NLP	2,803	Herzig et al. (2020), Yin et al. (2020)	Robustness in NLP	1,621	Hendrycks et al. (2020), Meade et al. (2022)

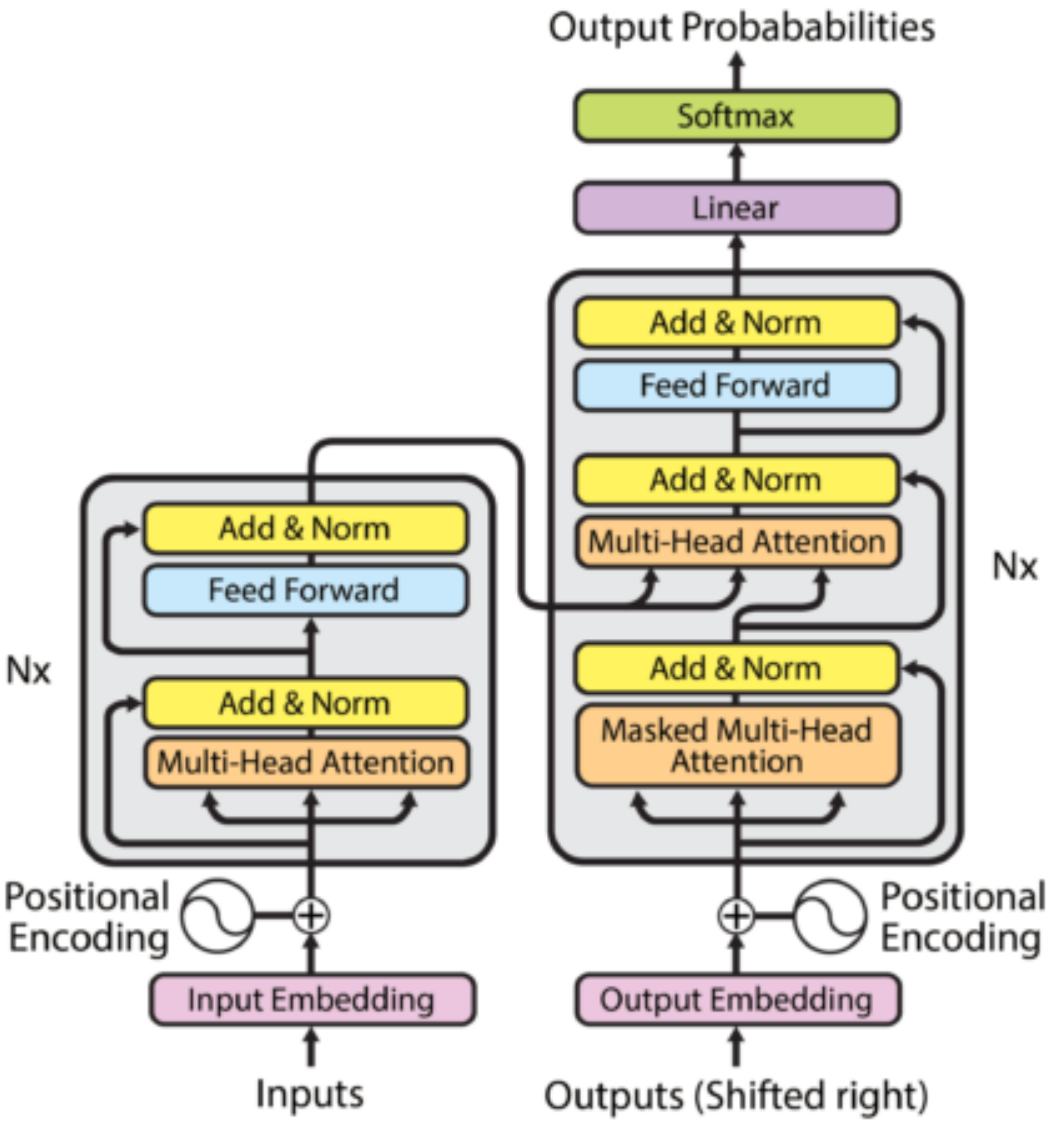
Transformers

- Self-attention

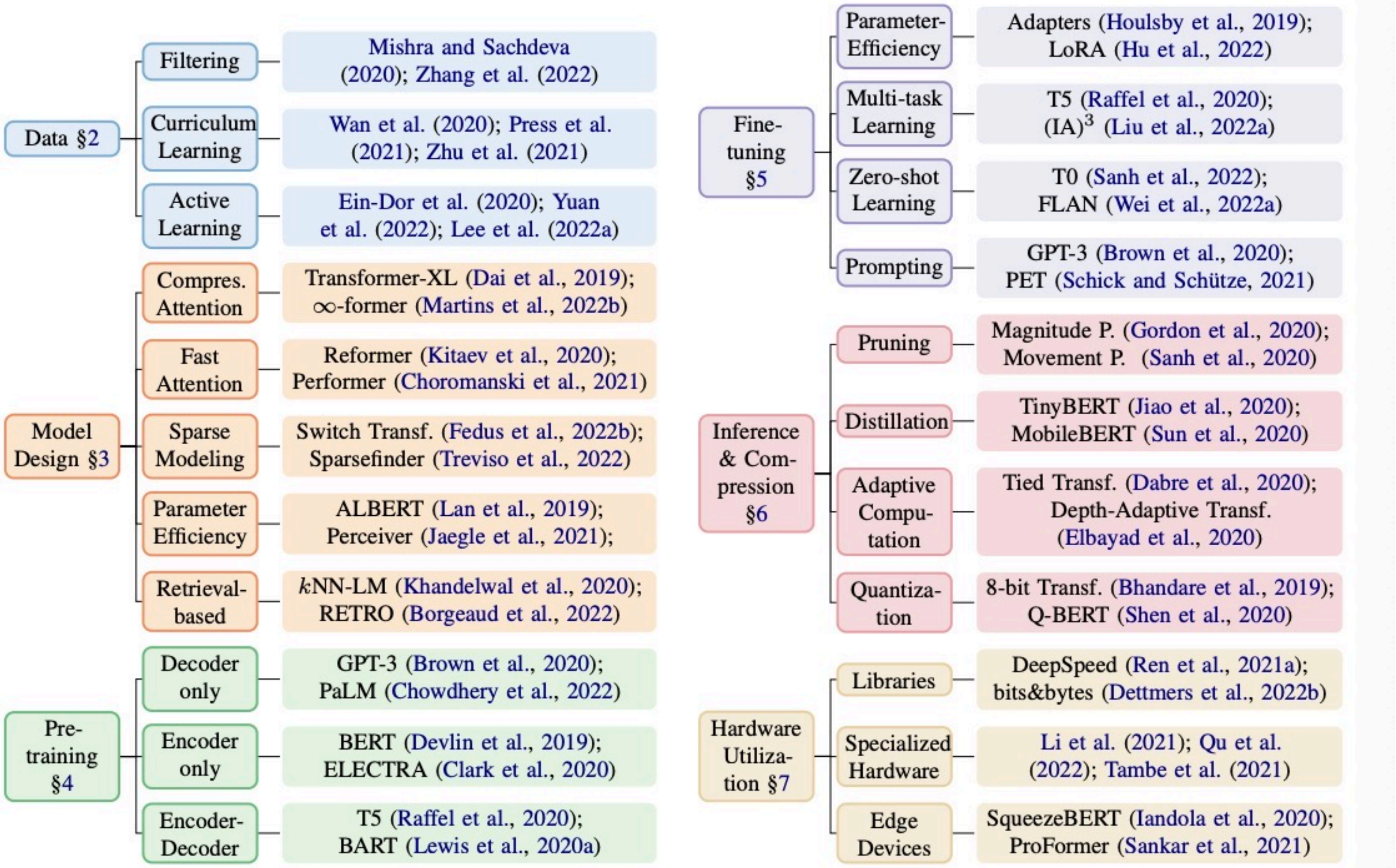
Attention is all you need

https://github.com/huggingface/transformers

TRANSFORMER



The encoder-decoder transformer used for translation. Encoder on the left, decoder on the right. Note that the decoder takes in its previously generated words during generation.


Data §2
Filtering
Mishra and Sachdeva (2020); Zhang et al. (2022)
Curriculum Learning
Wan et al. (2020); Press et al. (2021); Zhu et al. (2021)
Active Learning
Ein-Dor et al. (2020); Yuan et al. (2022); Lee et al. (2022a)
Model Design §3
Compres. Attention
Transformer-XL (Dai et al., 2019); ∞-former (Martins et al., 2022b)
Fast Attention
Reformer (Kitaev et al., 2020); Performer (Choromanski et al., 2021)
Sparse Modeling
Switch Transf. (Fedus et al., 2022b); Sparsefinder (Treviso et al., 2022)
Parameter Efficiency
ALBERT (Lan et al., 2019); Perceiver (Jaegle et al., 2021);
Retrieval-based
kNN-LM (Khandelwal et al., 2020); RETRO (Borgeaud et al., 2022)
Pre-training §4
Decoder only
GPT-3 (Brown et al., 2020); PaLM (Chowdhery et al., 2022)
Encoder only
BERT (Devlin et al., 2019); ELECTRA (Clark et al., 2020)
Encoder-Decoder
T5 (Raffel et al., 2020); BART (Lewis et al., 2020a)
Fine-tuning §5
Parameter-Efficiency
Adapters (Houlsby et al., 2019); LoRA (Hu et al., 2022)
Multi-task Learning
T5 (Raffel et al., 2020); (IA)^3 (Liu et al., 2022a)
Zero-shot Learning
T0 (Sanh et al., 2022); FLAN (Wei et al., 2022a)
Prompting
GPT-3 (Brown et al., 2020); PET (Schick and Schütze, 2021)
Inference & Compression §6
Pruning
Magnitude P. (Gordon et al., 2020); Movement P. (Sanh et al., 2020)
Distillation
TinyBERT (Jiao et al., 2020); MobileBERT (Sun et al., 2020)
Adaptive Computation
Tied Transf. (Dabre et al., 2020); Depth-Adaptive Transf. (Elbayad et al., 2020)
Quantization
8-bit Transf. (Bhandare et al., 2019); Q-BERT (Shen et al., 2020)
Hardware Utilization §7
Libraries
DeepSpeed (Ren et al., 2021a); bits&bytes (Dettmers et al., 2022b)
Specialized Hardware
Li et al. (2021); Qu et al. (2022); Tambe et al. (2021)
Edge Devices
SqueezeBERT (Iandola et al., 2020); ProFormer (Sankar et al., 2021)

Task		Original Paradigm		Shifted Paradigm			
TC	Paradigm	`Class`	$\Rightarrow$	`Matching`	`Seq2Seq`	`(M) LM`	
	Input	$\mathcal{X}$		$\mathcal{X}, \mathcal{L}$	$\mathcal{X}$	$f_{prompt}(\mathcal{X})$	
	Output	$\mathcal{Y}$		$\mathcal{Y} \in \{0, 1\}$	$y_1, \cdots, y_m$	$g(\mathcal{Y})$	
	Example	(Devlin et al., 2019)		(Chai et al., 2020)	(Yang et al., 2018a)	(Schick and Schütze, 2021a)	
NLI	Paradigm	`Matching`	$\Rightarrow$	`Class`	`Seq2Seq`	`(M) LM`	
	Input	$\mathcal{X}_a, \mathcal{X}_b$		$\mathcal{X}_a \oplus \mathcal{X}_b$	$f_{prompt}(\mathcal{X}_a, \mathcal{X}_b)$	$f_{prompt}(\mathcal{X}_a, \mathcal{X}_b)$	
	Output	$\mathcal{Y}$		$\mathcal{Y}$	$\mathcal{Y}$	$g(\mathcal{Y})$	
	Example	(Chen et al., 2017b)		(Devlin et al., 2019)	(McCann et al., 2018)	(Schick and Schütze, 2021a)	
NER	Paradigm	`SeqLab`	$\Rightarrow$	`Class`	`MRC`	`Seq2Seq`	`(M) LM`
	Input	$x_1, \cdots, x_n$		$\mathcal{X}_{span}$	$\mathcal{X}, \mathcal{Q}_y$	$\mathcal{X}$	$\mathcal{X}$
	Output	$y_1, \cdots, y_n$		$\mathcal{Y}$	$\mathcal{X}_{span}$	$(\mathcal{X}_{ent_i}, \mathcal{Y}_{ent_i})_{i=1}^{m}$	$g(\mathcal{Y})$
	Example	(Ma and Hovy, 2016)		(Fu et al., 2021)	(Li et al., 2020)	(Yan et al., 2021b)	(Cui et al., 2021)
ABSA	Paradigm	`Class`	$\Rightarrow$	`Matching`	`MRC`	`Seq2Seq`	`(M) LM`
	Input	$\mathcal{X}_{asp}$		$\mathcal{X}, \mathcal{S}_{aux}$	$\mathcal{X}, \mathcal{Q}_{asp}, \mathcal{Q}_{opin}, \mathcal{Q}_{sent}$	$\mathcal{X}$	$f_{prompt}(\mathcal{X})$
	Output	$\mathcal{Y}$		$\mathcal{Y}$	$\mathcal{X}_{asp}, \mathcal{X}_{opin}, \mathcal{Y}_{sent}$	$(\mathcal{X}_{asp_i}, \mathcal{X}_{opin_i}, \mathcal{Y}_{sent_i})_{i=1}^{m}$	$g(\mathcal{Y})$
	Example	(Wang et al., 2016)		(Sun et al., 2019)	(Mao et al., 2021)	(Yan et al., 2021a)	(Li et al., 2021)
RE	Paradigm	`Class`	$\Rightarrow$	`MRC`	`Seq2Seq`	`(M) LM`	
	Input	$\mathcal{X}$		$\mathcal{X}, \mathcal{Q}_y$	$\mathcal{X}$	$f_{prompt}(\mathcal{X})$	
	Output	$\mathcal{Y}$		$\mathcal{X}_{ent}$	$(\mathcal{Y}_i, \mathcal{X}_{sub_i}, \mathcal{X}_{obj_j})_{i=1}^{m}$	$g(\mathcal{Y})$	
	Example	(Zeng et al., 2014)		(Levy et al., 2017)	(Zeng et al., 2018)	(Han et al., 2021)	
Summ	Paradigm	`SeqLab` / `Seq2Seq`	$\Rightarrow$	`Matching`	`(M) LM`		
	Input	$\mathcal{X}_1, \cdots, \mathcal{X}_n$ / $\mathcal{X}, \mathcal{Q}_{summ}$		$(\mathcal{X}, \mathcal{S}_{cand_i})_{i=1}^{n}$	$\mathcal{X}$, Keywords/Prompt		
	Output	$\mathcal{Y}_1, \cdots, \mathcal{Y}_n \in \{0, 1\}^n$ / $\mathcal{Y}$		$\hat{\mathcal{S}}_{cand}$	$\mathcal{Y}$		
	Example	(Cheng and Lapata, 2016)		(Zhong et al., 2020)	(Aghajanyan et al., 2021)		
		(McCann et al., 2018)					
Parsing	Paradigm	`Seq2ASeq`	$\Rightarrow$	`(M) LM`	`SeqLab`	`MRC`	`Seq2Seq`
	Input	$(\mathcal{X}, \mathcal{C}_t)_{t=0}^{m-1}$		$(\mathcal{X}, \mathcal{Y}_i)_{i=1}^{k}$	$x_1, \cdots, x_n$	$\mathcal{X}, \mathcal{Q}_{child}$	$\mathcal{X}$
	Output	$\mathcal{A} = a_1, \cdots, a_m$		$\hat{\mathcal{Y}}$	$g(y_1, \cdots, y_n)$	$\mathcal{X}_{parent}$	$g(y_1, \cdots, y_m)$
	Example	(Chen and Manning, 2014)		(Choe and Charniak, 2016)	(Strzyz et al., 2019)	(Gan et al., 2021)	(Vinyals et al., 2015)

Table 1: Paradigms shift in NLP tasks. **TC**: text classification. **NLI**: natural language inference. **NER**: named entity recognition. **ABSA**: aspect-based sentiment analysis. **RE**: relation extraction. **Summ**: text summarization. **Parsing**: syntactic/semantic parsing. f and g indicate pre-processing and post-processing, respectively. $\mathcal{L}$ means label description. $\oplus$ means concatenation. $\mathcal{X}_{asp}, \mathcal{X}_{opin}, \mathcal{Y}_{sent}$ mean aspect, opinion, and sentiment, respectively. $\mathcal{S}_{aux}$ means auxiliary sentence. $\mathcal{X}_{sub}, \mathcal{X}_{obj}$ stand for subject entity and object entity, respectively. $\mathcal{S}_{cand}$ means candidate summary. $\mathcal{C}_t$ is the configuration at time step t and $\mathcal{A}$ is a sequence of actions.

Conference and journals

- AAAI
- IJCAI
- ICML
- NeurIPS
- ACL

- ICLR
- WWW
- SIGIR

- EMNLP
- COLING
- NAACL

- NLPCC
- CONLL

- TASLP
- TAC

- KBS
- NC
- NLE

Conference List
https://conferencelist.info/upcoming/

Deadline
https://ccfddl.github.io/

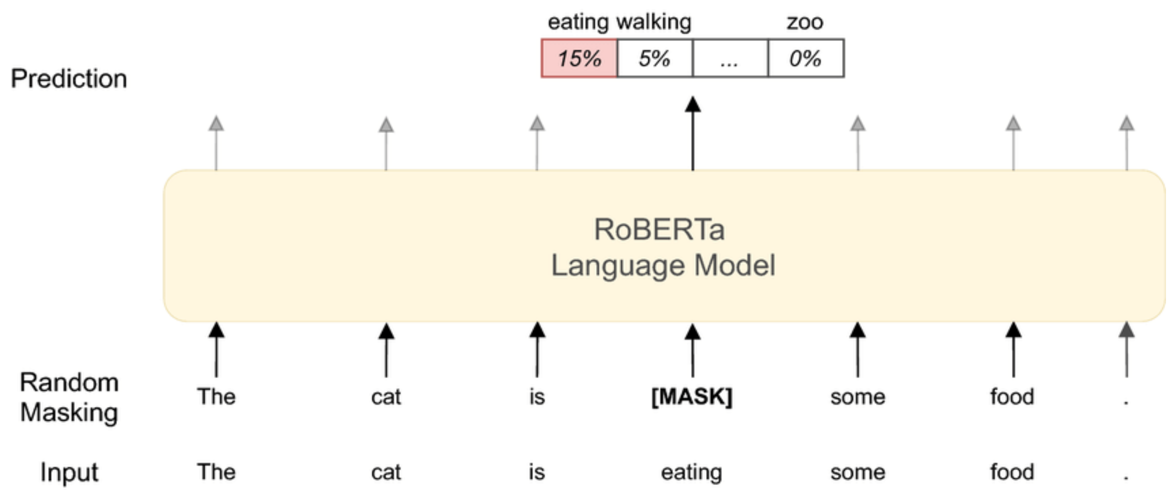
Pre-training and Fine-tuning

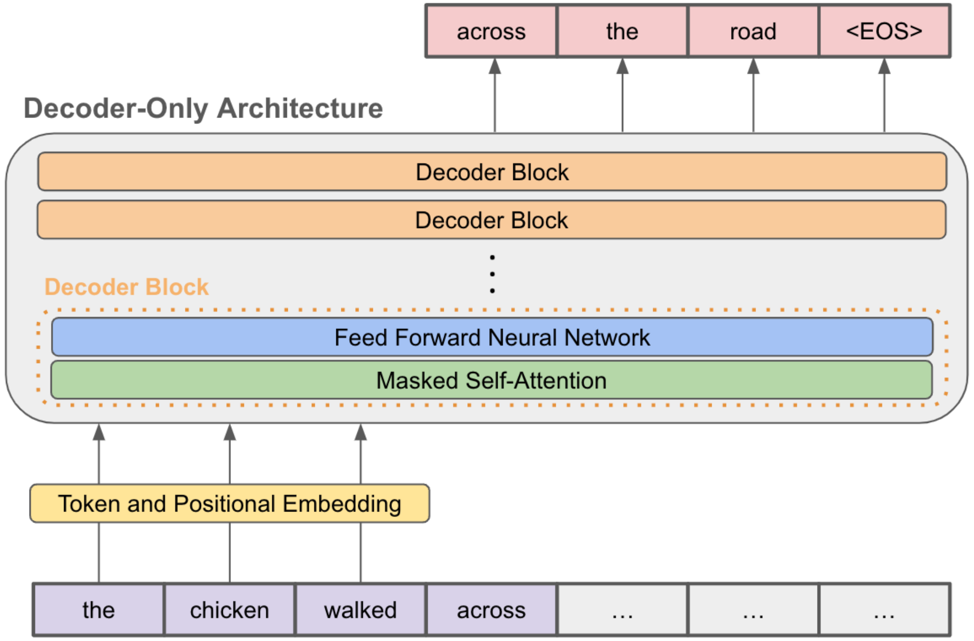
- BERT
- GPT
- T5

[BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding]

[Improving Language Understanding by Generative Pre-Training]

Transfer learning





- BERT - masked language modeling (MLM), next sentence prediction (NSP)
- GPT - predicting the next word
- BART - denoising (masking, sentence permutation, token deletion, document rotation)

Label verbalize

- Solving classification as a generation task (T5)

[Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer]

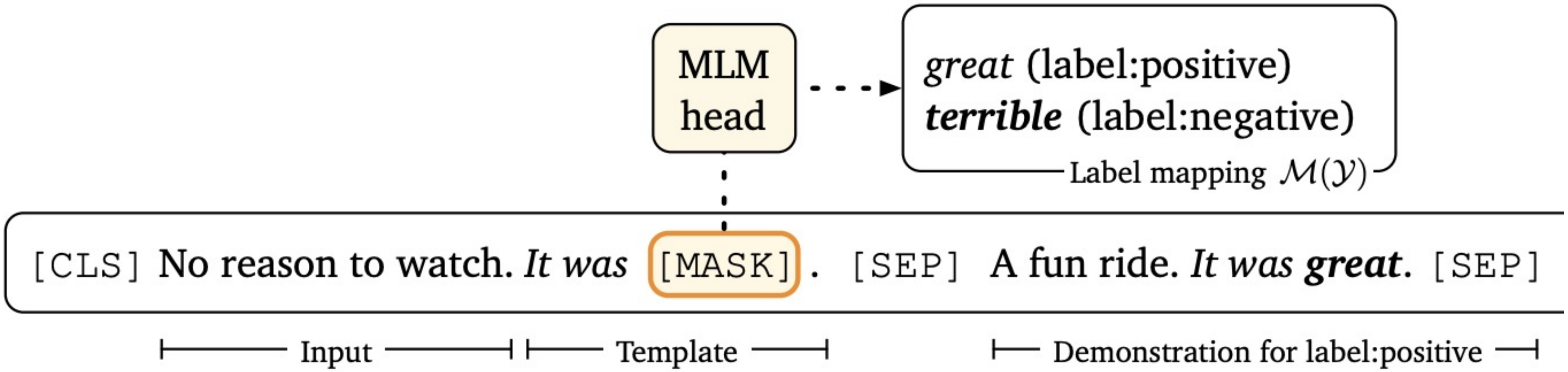
Graph / Contrastive learning

SimCSE: Simple Contrastive Learning of Sentence Embeddings

https://arxiv.org/abs/2104.08821

Prompt-based Fine-tuning

- [Making Pre-trained Language Models Better Few-shot Learners]
- [FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS]



Instruction tuning

- Instruction Induction: From Few Examples to Natural Language Task Descriptions
- Training language models to follow instructions with human feedback

CoT

- **Tree of Thoughts: Deliberate Problem Solving with Large Language Models**
- https://arxiv.org/abs/2305.10601
- **Automatic Chain of Thought Prompting in Large Language Models**
- https://arxiv.org/abs/2210.03493
- **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models**
- https://arxiv.org/abs/2201.11903
- **Self-Consistency Improves Chain of Thought Reasoning in Language Models**
- https://arxiv.org/abs/2203.11171
- **Large Language Models Are Human-Level Prompt Engineers**
- https://arxiv.org/abs/2211.01910
- **Distilling Reasoning Capabilities into Smaller Language Models**
- https://arxiv.org/abs/2212.00193
- **Analysing Mathematical Reasoning Abilities of Neural Models**
- https://arxiv.org/abs/1904.01557
- **Large Language Models are Zero-Shot Reasoners**
- https://arxiv.org/abs/2205.11916
- **Selection-Inference: Exploiting Large Language Models for Interpretable Logical Reasoning**
- https://arxiv.org/abs/2205.09712
- **ART: Automatic multi-step reasoning and tool-use for large language models**
- https://arxiv.org/abs/2303.09014

Baseline
https://github.com/kyegomez/tree-of-thoughts