CSCE 638 Natural Language Processing Foundation and Techniques

Lecture 14: Alignment, Text Similarity, Retrieval-Augmented Generation

Kuan-Hao Huang

Spring 2025

(Some slides adapted from Graham Neubig and ACL 2023 Tutorial: Retrieval-based Language Models and Applications)

Course Project – Proposal

- Due: 3/3 11:59pm
- Page limit: 2 pages (excluding references)
- Format: ACL style
- The proposal should include
  - Introduction to the topic you choose
  - Related literature
  - Novelty and challenges
  - The dataset, models, and approaches you plan to use
  - Evaluation plan

Course Project: Project Highlight

- Put your slides here
  - https://docs.google.com/presentation/d/1FbPJxciLrXIiH3srVR3bSENRfyBM8p86M4DQtoLuBmo/edit?usp=sharing
- Date: 3/5 in person
- Each team has 3 minutes to introduce the project
  - Introduction to the topic you choose
  - Short related literature overview
  - Novelty and challenges
  - The dataset, models, and approaches you plan to use
  - Evaluation plan

Presentation Order

1. Team 10
2. Team 23
3. Team 6
4. Team 9
5. Team 2
6. Team 22
7. Team 5
8. Team 15
9. Team 4
10. Team 13
11. Team 1
12. Team 8
13. Team 11
14. Team 25
15. Team 12
16. Team 3
17. Team 18
18. Team 21
19. Team 17
20. Team 24
21. Team 20
22. Team 26
23. Team 16
24. Team 14
25. Team 7
26. Team 19
27. Team 27

Assignment 2

- https://khhuang.me/CSCE638-S25/assignments/assignment2_0224.pdf
- Due: 3/17 11:59pm
- Summit a .zip file to Canvas
  - submission.pdf for the writing section
  - submission[x].py and submission[x].ipynb for the coding section
- For questions
  - Discuss on Canvas
  - Send an email to csce638-ta-25s@list.tamu.edu, don’t need to CC TA or me

Quiz 2

- Date: 3/17
  - 15 minutes before the end of the lecture
  - 5 questions focusing on high-level concepts

W5	2/10	L8	Transformers [slides]
	2/12	L9	Contextualized Representations, Pre-Training [slides]
W6	2/17	L10	Pre-Training, Model Distillation [slides]
	2/19	L11	Parameter-Efficient Fine-Tuning, Large Language Models [slides]
W7	2/24	L12	Large Language Models, Instruction Tuning [slides]
	2/26	L13	Human Preference Alignment [slides]
W8	3/3	L14	Alignment, Text Similarity, Retrieval-Augmented Generation

Assignment 1

- Average: 97.40
- Median: 98
- Standard deviation: 4.30
- (before applying late penalty)

TA

Rahul Baid

Email: rahulbaid@tamu.edu

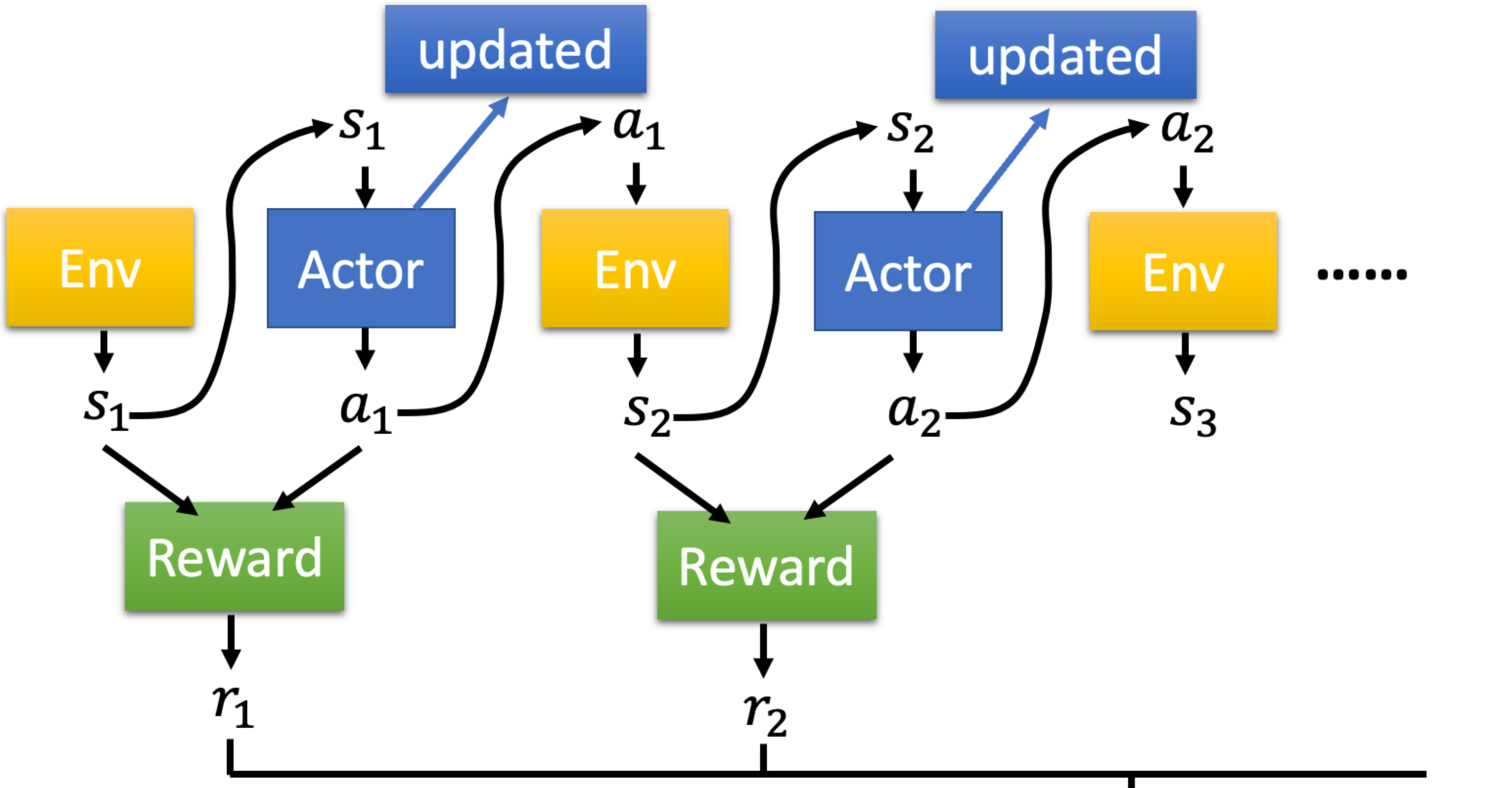
Office Hour: Wed. 12pm – 1pm

Office: PETR 359

Lecture Plan

- Human Preference Optimization
  - Simple Preference Optimization
  - Group Relative Policy Optimization
- Text Similarity
  - Sentence-BERT
  - SimCSE, DIffCSE, DPR
- Retrieval-Augmented Generation

Recap: RLHF/PPO



Expected Reward

$$\bar{R}_\theta = \sum_\tau R(\tau) p_\theta(\tau) = E_{\tau \sim p_\theta(\tau)}[R(\tau)] \qquad R(\tau) = \sum_{t=1}^{T} r_t$$

Recap: RLHF/PPO

- We have the following:
  - A pretrained (possibly instruction-finetuned) LM $p^{PT}(y \mid x)$
  - A reward model $RM_\phi(x, y)$ that produces scalar rewards for LM outputs, trained on a dataset of human comparisons
- Now to do RLHF:
  - Copy the model $p_\theta^{RL}(y \mid x)$, with parameters θ we would like to optimize
  - We want to optimize:

$$\mathbb{E}_{\hat{y} \sim p_\theta^{RL}(\hat{y} \mid x)} \left[RM_\phi(x, \hat{y}) \; - \beta \log \left(\frac{p_\theta^{RL}(\hat{y} \mid x)}{p^{PT}(\hat{y} \mid x)} \right) \right]$$

Recap: RLHF/PPO

An earthquake hit San Francisco. There was minor property damage, but no injuries.	>	A 4.2 magnitude earthquake hit San Francisco, resulting in massive damage.	>	The Bay Area has good weather but is prone to earthquakes and wildfires.
s_1		s_3		s_2

Bradley-Terry [1952] paired comparison model

$$J_{RM}(\phi) = -\mathbb{E}_{(s^w, s^l)\sim D}\left[\log \sigma(RM_\phi(s^w) - RM_\phi(s^l))\right]$$

"winning" sample

"losing" sample

s^w should score higher than s^l

Recap: Direct Preference Optimization (DPO)

RLHF Objective

(get **high reward**, stay **close** to reference model)

$$\max_{\pi} \mathbb{E}_{x\sim\mathcal{D},y\sim\pi(y|x)}\left[r(x,y)\right] - \beta\mathbb{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathrm{ref}}(\cdot \mid x))$$

Maximize reward

Keep similar behavior

Closed-form Optimal Policy

(write **optimal policy** as function of **reward function**; from prior work)

$$\pi^*(y \mid x) = \frac{1}{Z(x)}\pi_{\mathrm{ref}}(y \mid x)\exp\left(\frac{1}{\beta}r(x,y)\right)$$

with $Z(x) = \sum_y \pi_{\mathrm{ref}}(y \mid x)\exp\left(\frac{1}{\beta}r(x,y)\right)$

Note **intractable sum** over possible responses; can't immediately use this

Rearrange

(write **any reward function** as function of **optimal policy**)

Ratio is **positive** if policy likes response more than reference model, **negative** if policy likes response less than ref. model

$$r(x,y) = \underbrace{\beta\log\frac{\pi^*(y \mid x)}{\pi_{\mathrm{ref}}(y \mid x)} + \beta\log Z(x)}_{\text{some parameterization of a reward function}}$$

Direct Preference Optimization (DPO)

Derived from the Bradley-Terry model of human preferences:

A loss function on <u>reward functions</u>

$$\mathcal{L}_R(r, \mathcal{D}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma(r(x, y_w) - r(x, y_l)) \right]$$

+

A transformation between <u>reward functions</u> and <u>policies</u>

$$r_{\pi_\theta}(x, y) = \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x)$$

=

A loss function on <u>policies</u>

Reward of **preferred** response

Reward of **dispreferred** response

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)} \right) \right]$$

Simple Preference Optimization (SimPO)

$$\mathcal{L}_{\mathbf{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}\left[\log \sigma\left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\right)\right]$$

$$\mathcal{L}_{\mathbf{SimPO}}(\pi_\theta) = -\mathbb{E}\left[\log \sigma\left(\frac{\beta}{|y_w|} \log \pi_\theta(y_w \mid x) - \frac{\beta}{|y_l|} \log \pi_\theta(y_l \mid x) - \gamma\right)\right]$$

SimPO: Simple Preference Optimization with a Reference-Free Reward, 2024

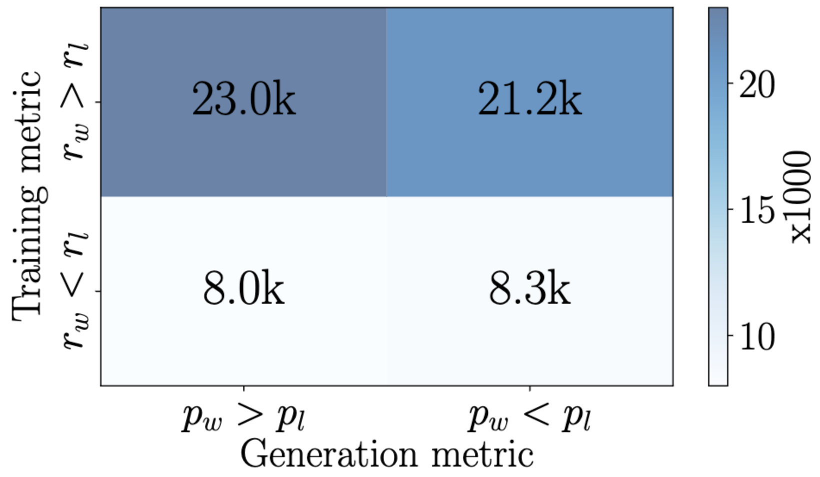
Look Back at DPO

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\boxed{\pi_{\text{ref}}(y_w \mid x)}} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\boxed{\pi_{\text{ref}}(y_l \mid x)}} \right) \right]$$

Reward of **preferred** response   Reward of **dispreferred** response

How does reference model affect the behavior?

$$r(x, y_w) > r(x, y_l) \Rightarrow p_\theta(y_w|x) > p_\theta(y_l|x)?$$



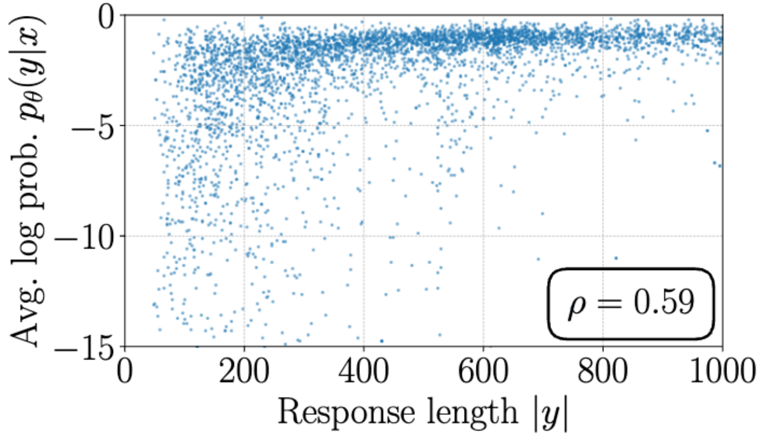
Solution: Reference-Free Reward

$$\mathcal{L}_{\mathrm{DPO}}(\pi_\theta;\pi_{\mathrm{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w\mid x)}{\cancel{\pi_{\mathrm{ref}}(y_w\mid x)}} - \beta\log\frac{\pi_\theta(y_l\mid x)}{\cancel{\pi_{\mathrm{ref}}(y_l\mid x)}}\right)\right]$$

Reward of **preferred** response   Reward of **dispreferred** response

$$r(x,y) = \sum_{i=1}^{|y|} \log \pi_\theta(y_i|x, y_{<i})$$

Length bias!
The model tends to generate longer sequence to maximize reward



(a) Length correlation (DPO).

Solution: Reference-Free Reward

$$\mathcal{L}_{\mathrm{DPO}}(\pi_\theta; \pi_{\mathrm{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\cancel{\pi_{\mathrm{ref}}(y_w \mid x)}} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\cancel{\pi_{\mathrm{ref}}(y_l \mid x)}} \right) \right]$$

Reward of **preferred** response   Reward of **dispreferred** response

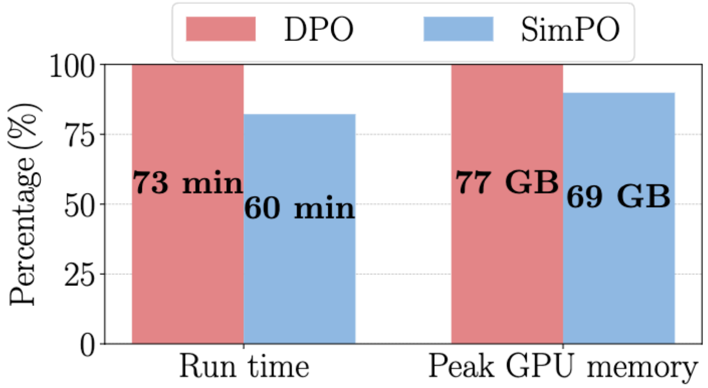
$$r_{\mathrm{SimPO}}(x, y) = \frac{\beta}{|y|} \log \pi_\theta(y \mid x) = \frac{\beta}{|y|} \sum_{i=1}^{|y|} \log \pi_\theta(y_i \mid x, y_{<i})$$

Reward margin

$$\mathcal{L}_{\mathrm{SimPO}}(\pi_\theta) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \left(\frac{\beta}{|y_w|} \log \pi_\theta(y_w|x) - \frac{\beta}{|y_l|} \log \pi_\theta(y_l|x) - \boxed{\gamma} \right) \right]$$

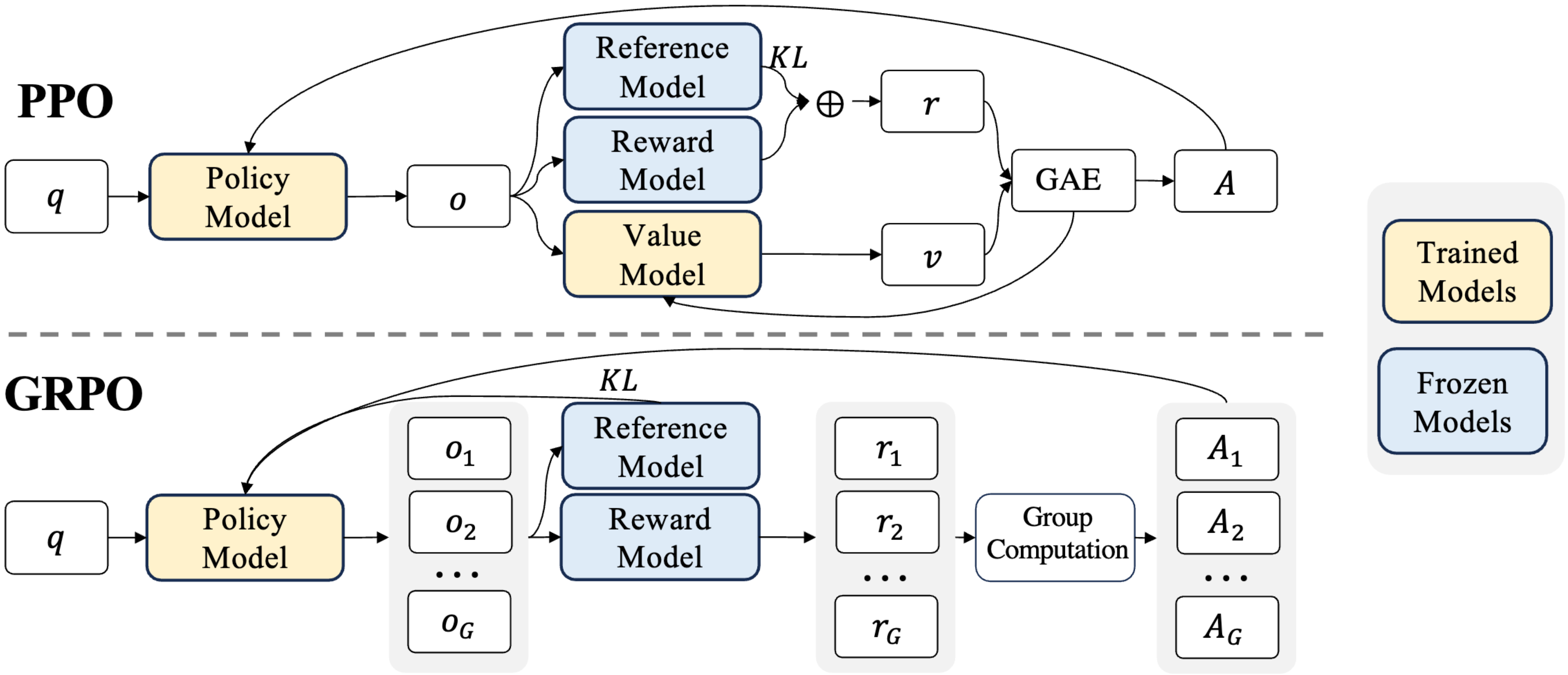
SimPO Performance

Method	Mistral-Base (7B)					Mistral-Instruct (7B)				
	AlpacaEval 2		Arena-Hard	MT-Bench		AlpacaEval 2		Arena-Hard	MT-Bench	
	LC (%)	WR (%)	WR (%)	GPT-4 Turbo	GPT-4	LC (%)	WR (%)	WR (%)	GPT-4 Turbo	GPT-4
SFT	8.4	6.2	1.3	4.8	6.3	17.1	14.7	12.6	6.2	7.5
RRHF [91]	11.6	10.2	5.8	5.4	6.7	25.3	24.8	18.1	6.5	7.6
SLiC-HF [96]	10.9	8.9	7.3	5.8	**7.4**	24.1	24.6	18.9	6.5	**7.8**
DPO [66]	15.1	12.5	10.4	5.9	7.3	26.8	24.9	16.3	6.3	7.6
IPO [6]	11.8	9.4	7.5	5.5	7.2	20.3	20.3	16.2	6.4	**7.8**
CPO [88]	9.8	8.9	6.9	5.4	6.8	23.8	28.8	**22.6**	6.3	7.5
KTO [29]	13.1	9.1	5.6	5.4	7.0	24.5	23.6	17.9	6.4	7.7
ORPO [42]	14.7	12.2	7.0	5.8	7.3	24.5	24.9	20.8	6.4	7.7
R-DPO [64]	17.4	12.8	8.0	5.9	**7.4**	27.3	24.5	16.1	6.2	7.5
SimPO	**21.5**	**20.8**	**16.6**	**6.0**	7.3	**32.1**	**34.8**	21.0	**6.6**	7.6
Method	**Llama-3-Base (8B)**					**Llama-3-Instruct (8B)**				
	AlpacaEval 2		Arena-Hard	MT-Bench		AlpacaEval 2		Arena-Hard	MT-Bench	
	LC (%)	WR (%)	WR (%)	GPT-4 Turbo	GPT-4	LC (%)	WR (%)	WR (%)	GPT-4 Turbo	GPT-4
SFT	6.2	4.6	3.3	5.2	6.6	26.0	25.3	22.3	6.9	8.1
RRHF [91]	12.1	10.1	6.3	5.8	7.0	31.3	28.4	26.5	6.7	7.9
SLiC-HF [96]	12.3	13.7	6.0	6.3	7.6	26.9	27.5	26.2	6.8	8.1
DPO [66]	18.2	15.5	15.9	6.5	7.7	40.3	37.9	32.6	**7.0**	8.0
IPO [6]	14.4	14.2	17.8	6.5	7.4	35.6	35.6	30.5	**7.0**	**8.3**
CPO [88]	10.8	8.1	5.8	6.0	7.4	28.9	32.2	28.8	**7.0**	8.0
KTO [29]	14.2	12.4	12.5	6.3	**7.8**	33.1	31.8	26.4	6.9	8.2
ORPO [42]	12.2	10.6	10.8	6.1	7.6	28.5	27.4	25.8	6.8	8.0
R-DPO [64]	17.6	14.4	17.2	**6.6**	7.5	41.1	37.8	33.1	**7.0**	8.0
SimPO	**22.0**	**20.3**	**23.4**	**6.6**	7.7	**44.7**	**40.5**	**33.8**	**7.0**	8.0



(c) Efficiency of DPO vs. SimPO.

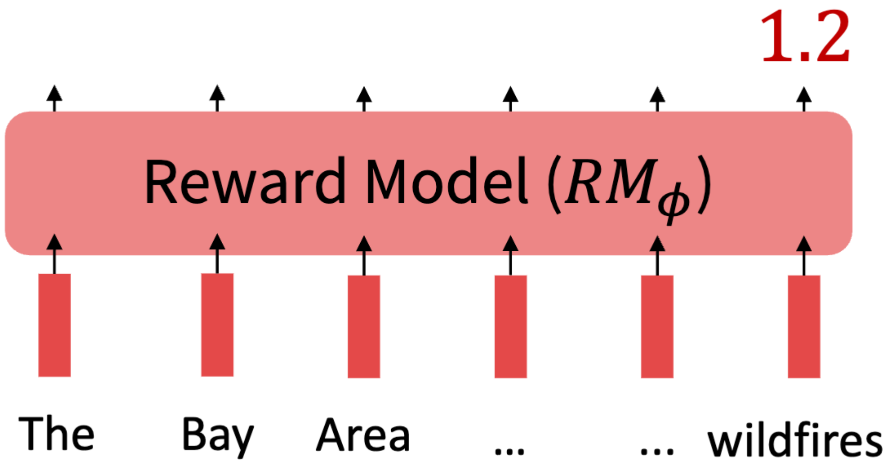
Group Relative Policy Optimization (GRPO)



Deepseek uses it!

Recap: Reward Model in PPO

- Train a reward model (RM) from an annotated dataset



$$\mathbb{E}_{\hat{s} \sim p_\theta(s)}[R(\hat{s})]$$

Group Relative Policy Optimization (GRPO)

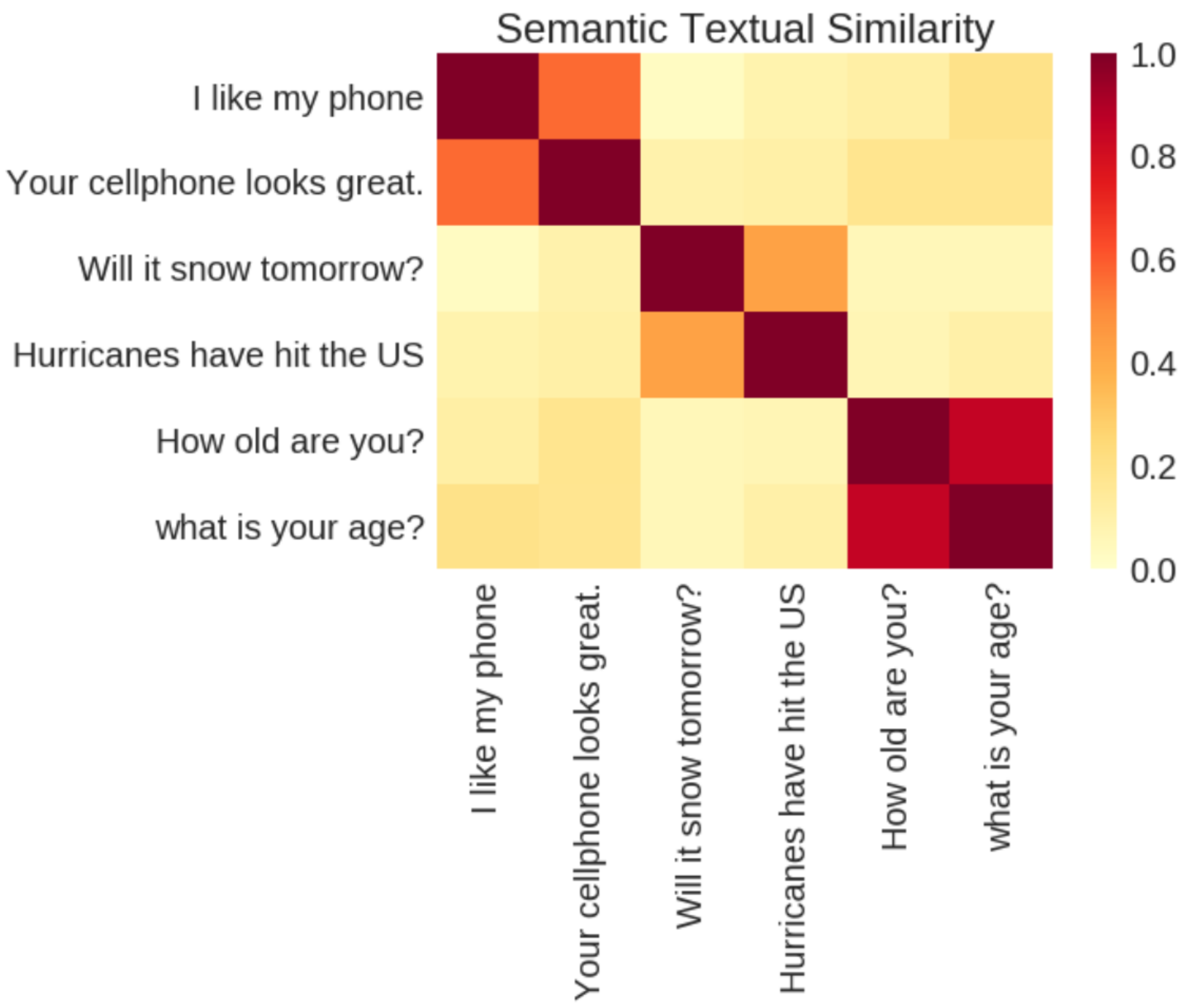
- Consider group relative reward
  - Given x, sample multiple output $y_1, y_2, \dots, y_G$
  - Use reward model to get reward $r_1, r_2, \dots, r_G$

$$A_i = \frac{r_i - mean(r_1, r_2, \dots, r_G)}{std(r_1, r_2, \dots, r_G)}$$

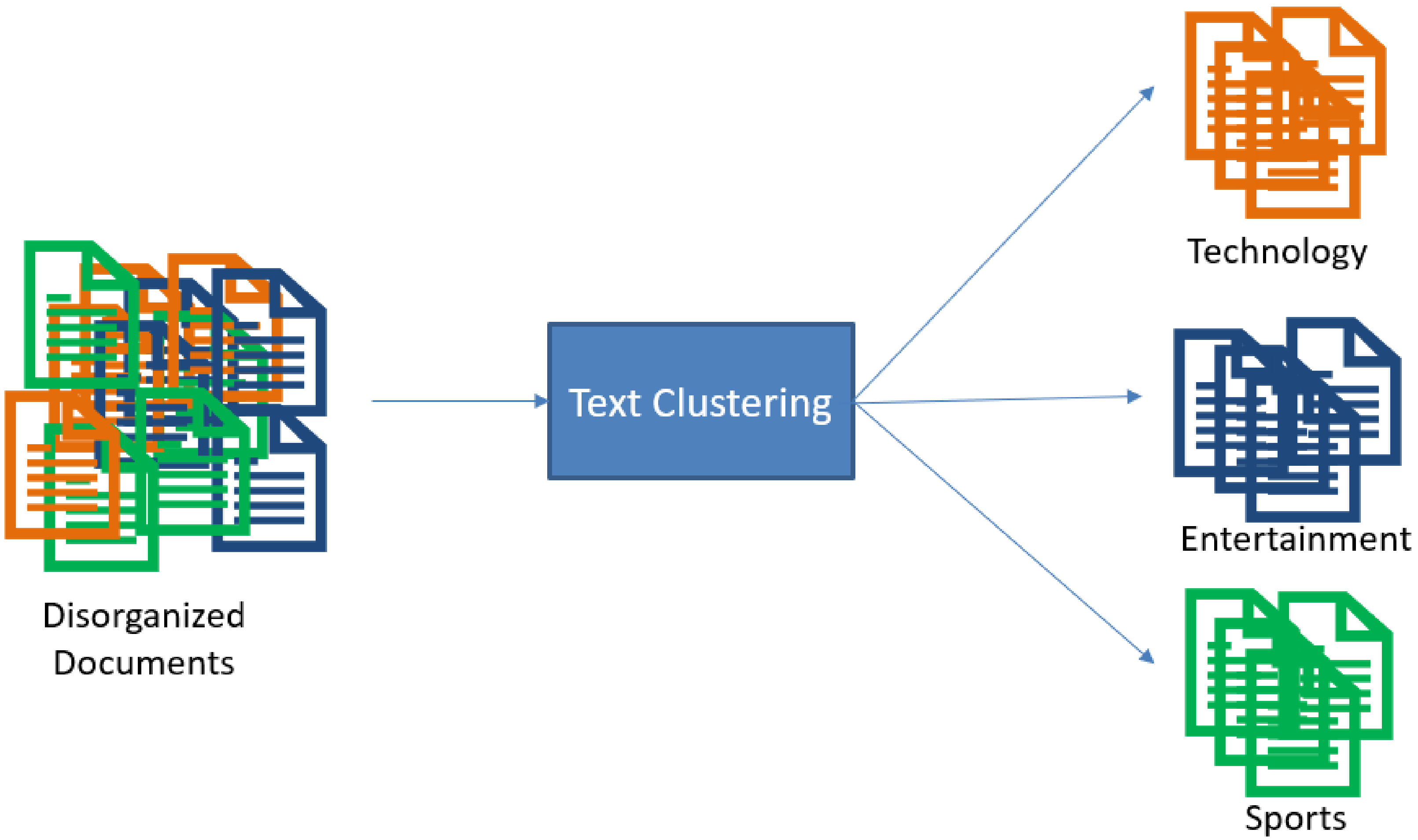
Lecture Plan

- Human Preference Optimization
  - Simple Preference Optimization
  - Group Relative Policy Optimization
- Text Similarity
  - Sentence-BERT
  - SimCSE, DIffCSE, DPR
- Retrieval-Augmented Generation

Text Similarity

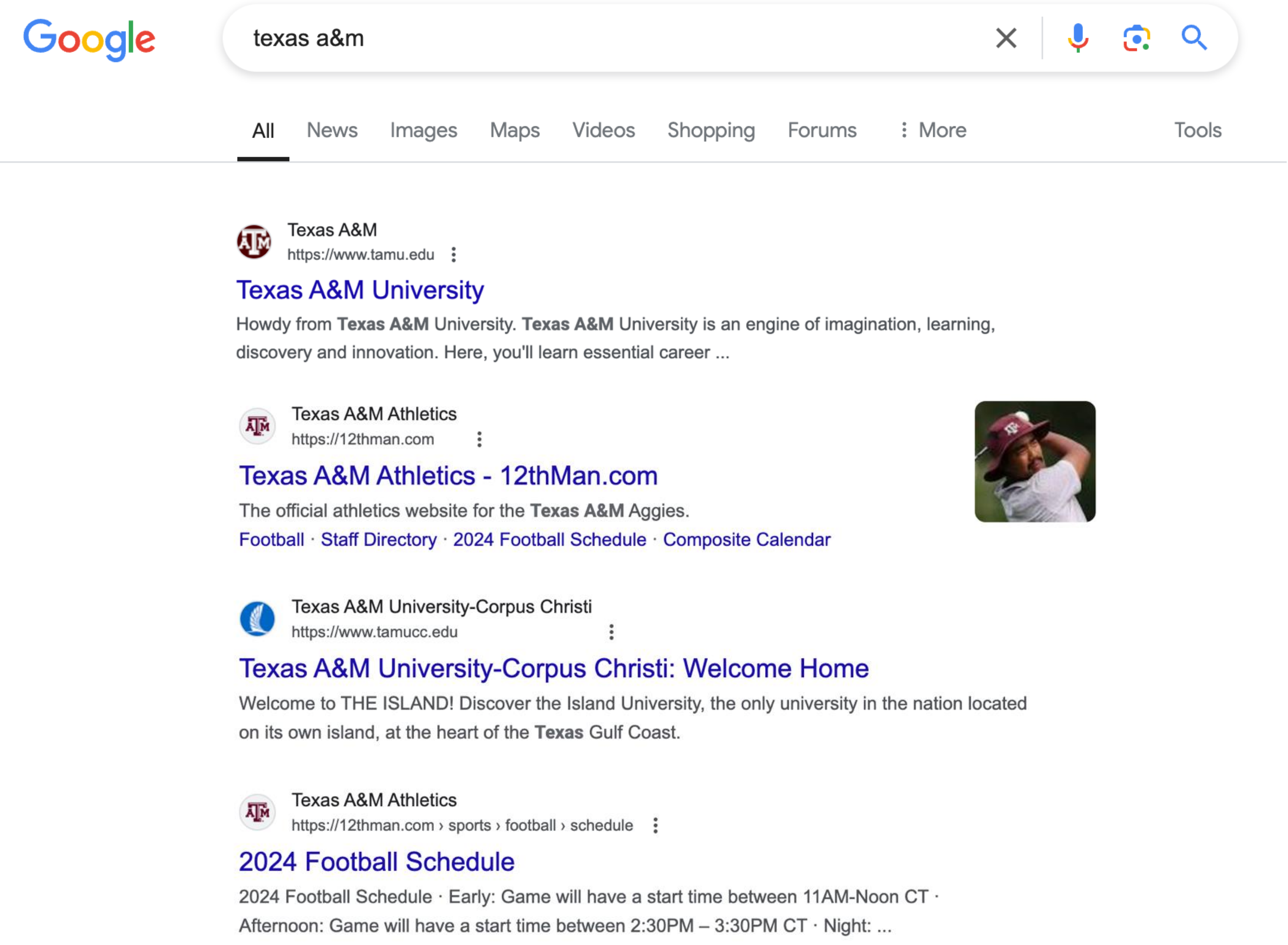


Document Clustering

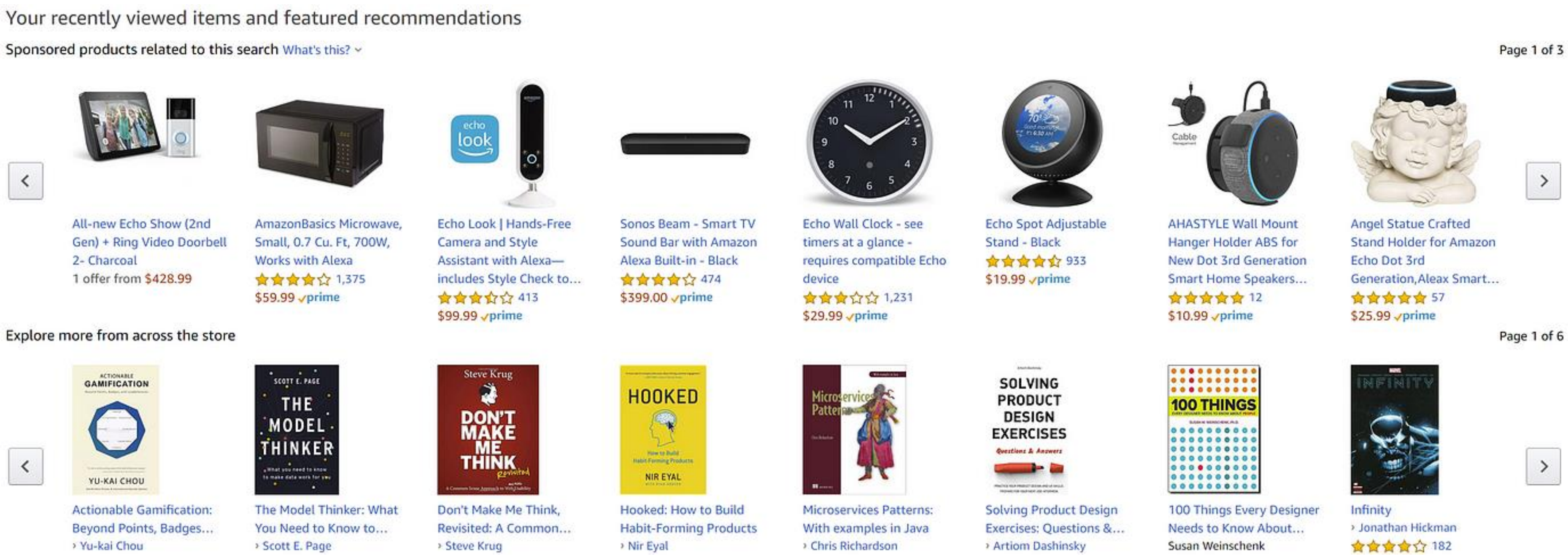


https://medium.com/@danielafrimi/text-clustering-using-nlp-techniques-c2e6b08b6e95

Information Retrieval



Recommendation Systems



Semantic Quality Control

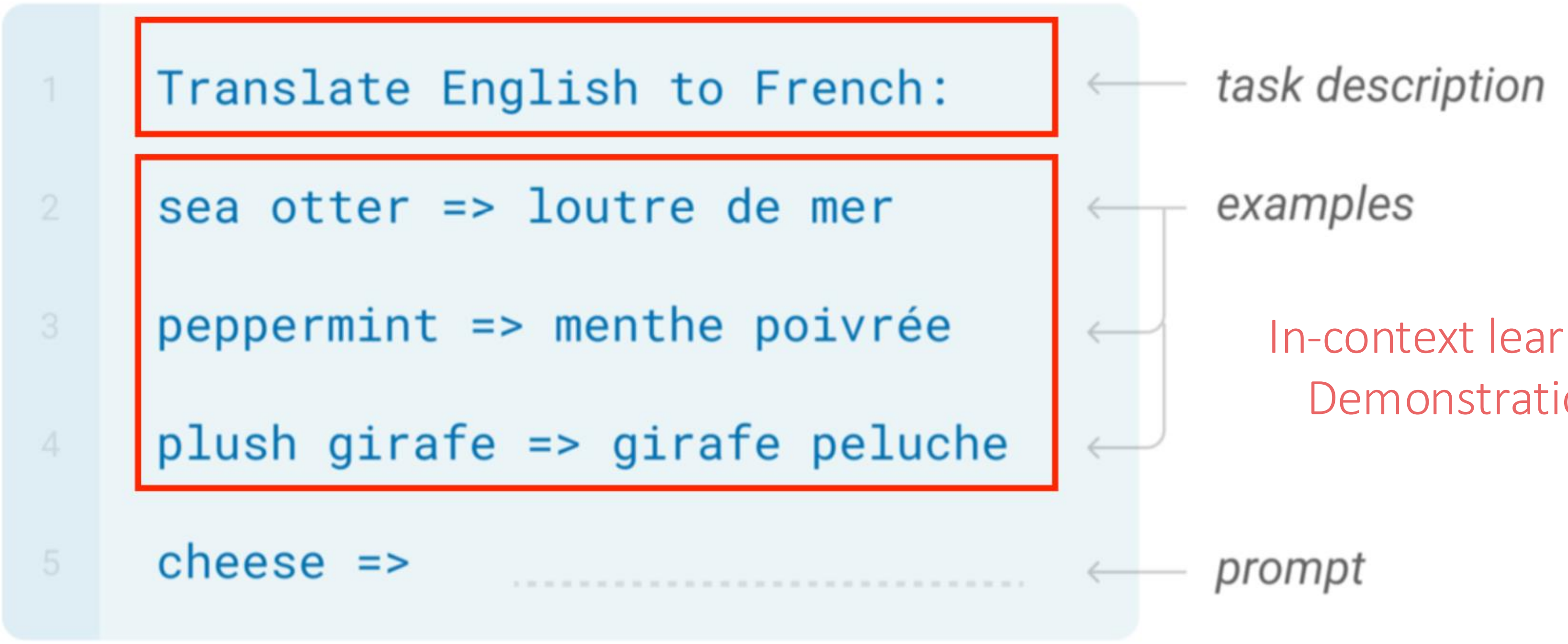
- Paraphrase generation

We will go hiking if tomorrow is a sunny day.

If it is sunny tomorrow, we will go hiking.

- Style transfer
- Plagiarism detection

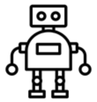
In-Context Example Selection



In-context learning examples
Demonstration examples

Language Models are Few-Shot Learners, 2020

Semantic Textual Similarity Benchmark

			
A soccer player is kicking the soccer ball into the goal from a long way down the field.	*A soccer player kicks the ball into the goal.*	3.25	3.94
Earlier this month, RIM had said it expected to report second-quarter earnings of between 7 cents and 11 cents a share.	*Excluding legal fees and other charges it expected a loss of between 1 and 4 cents a share.*	1.2	0.5
…	…	…	…
David Beckham Announces Retirement From Soccer.	*David Beckham retires from football.*	4.4	3.8

Pearson's Correlation Coefficient

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

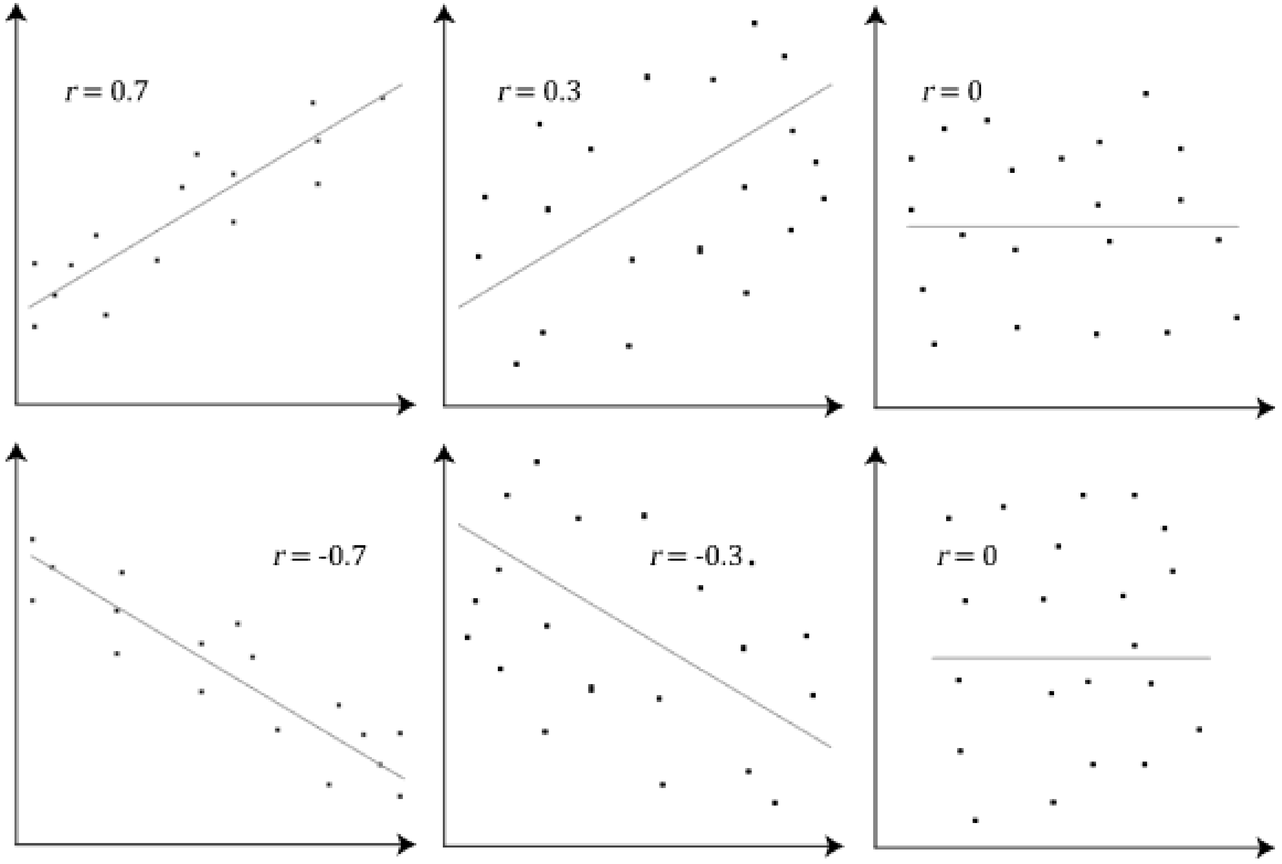
r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

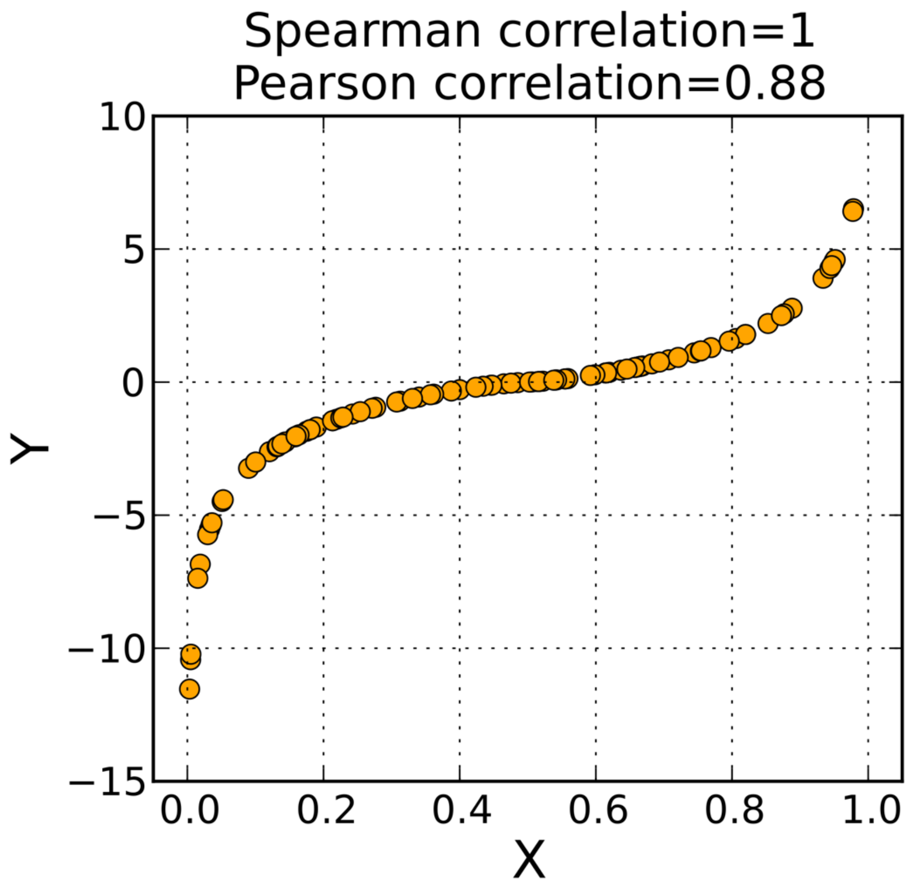
$\bar{y}$ = mean of the values of the y-variable



https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

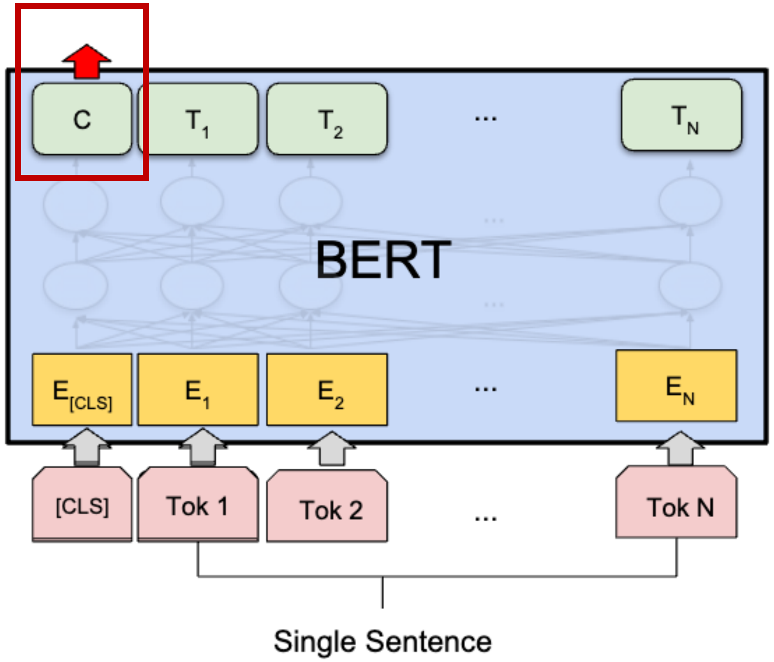
Spearman’s Correlation Coefficient

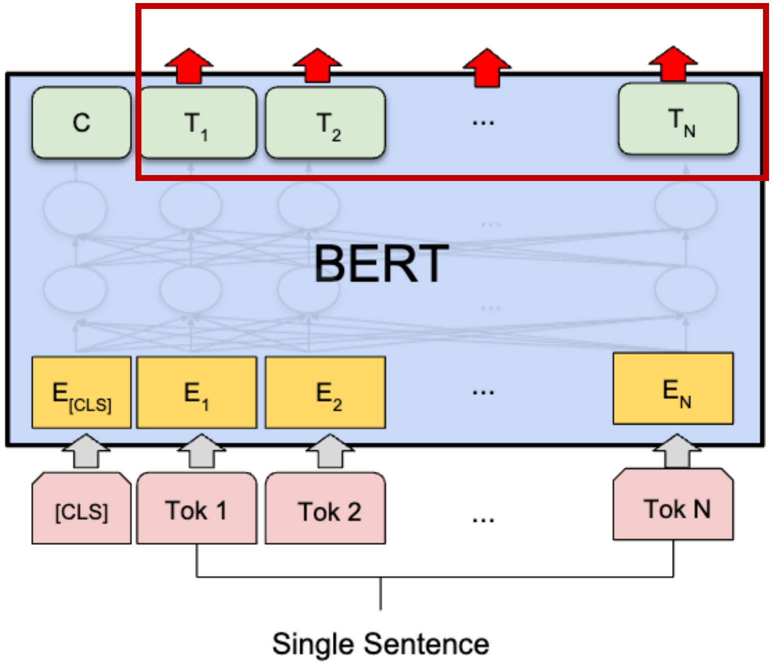
- Pearson’s correlation coefficient on rank
- Score
  - Human: [1.2, 3.4, 2.5, 0.7, 4.0]
  - Machine: [0.5, 3.3, 1.0, 1.2, 3.4]
- Rank
  - Human: [4, 2, 3, 5, 1]
  - Machine: [5, 2, 4, 3, 1]
- Assesses monotonic relationships
  - whether linear or not



https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

A Simple Approach: Text Encoder + Cosine Similarity





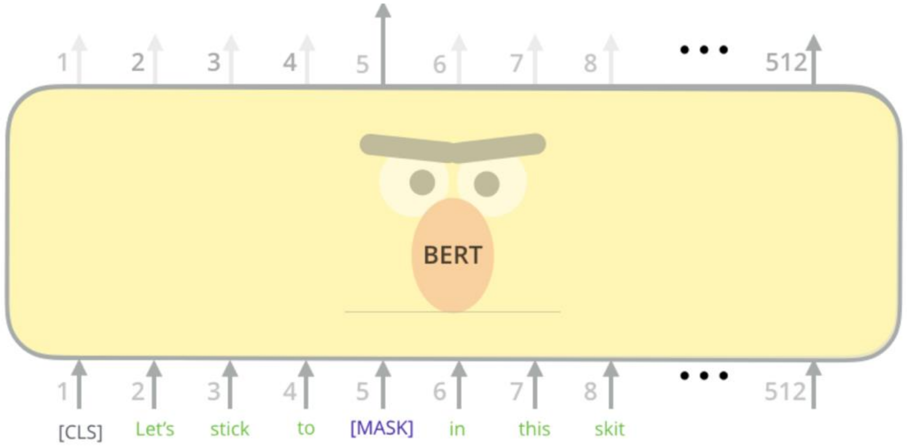
$$E_1 = \text{Encoder}(S_1)$$

$$E_2 = \text{Encoder}(S_2)$$

$$\text{Similarity}(S_1, S_2) = \frac{E_1 \cdot E_2}{\|E_1\|\|E_2\|}$$

Unfortunately, the performance is bad (why?)

A Simple Approach: Text Encoder + Cosine Similarity



Pre-trained BERT embeddings are more about lexical information

If it is sunny tomorrow, we will go hiking.

Let's go to hike once tomorrow is sunny.

We will go hiking if tomorrow is a sunny day.

Good classification performance ≠ Good similarity

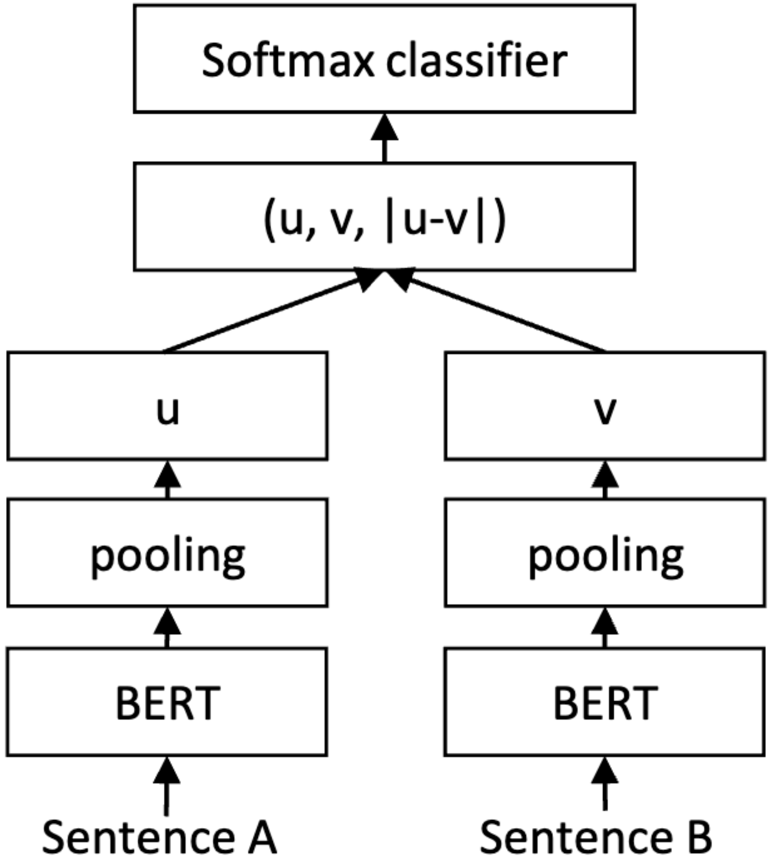
Sentence-BERT

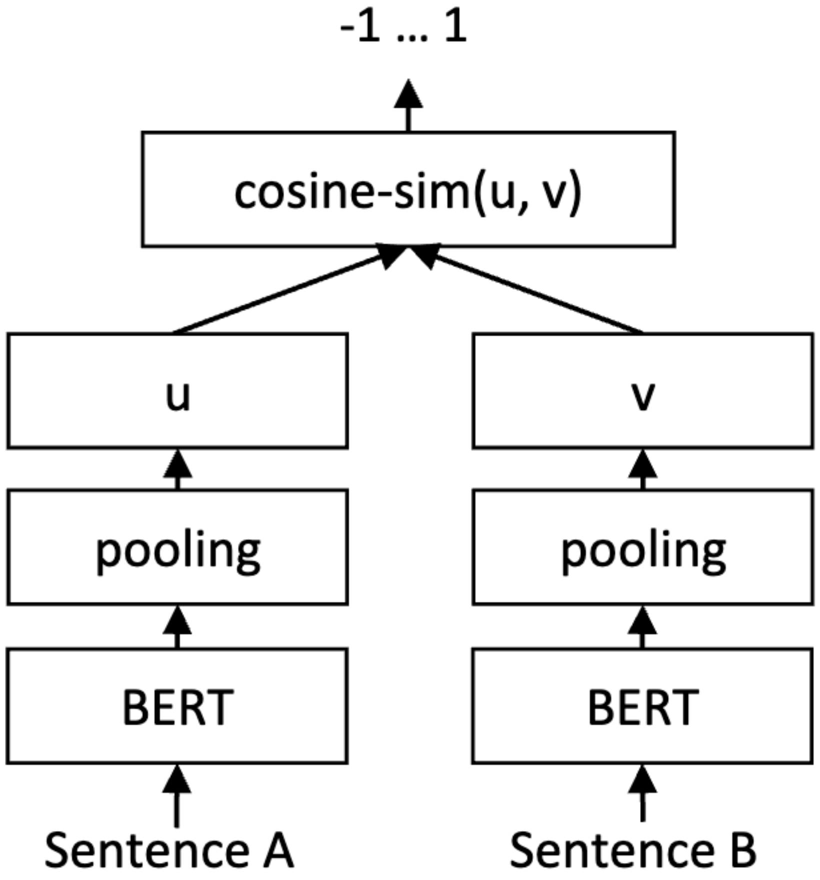
- Consider SNLI dataset
  - Stanford Natural Language Inference

A boy is jumping on skateboard in the middle of a red bridge.	*The boy skates down the sidewalk.*	Contradiction
A boy is jumping on skateboard in the middle of a red bridge.	*The boy is wearing safety equipment.*	Neutral
A boy is jumping on skateboard in the middle of a red bridge.	*The boy does a skateboarding trick.*	Entailment

Sentence-BERT

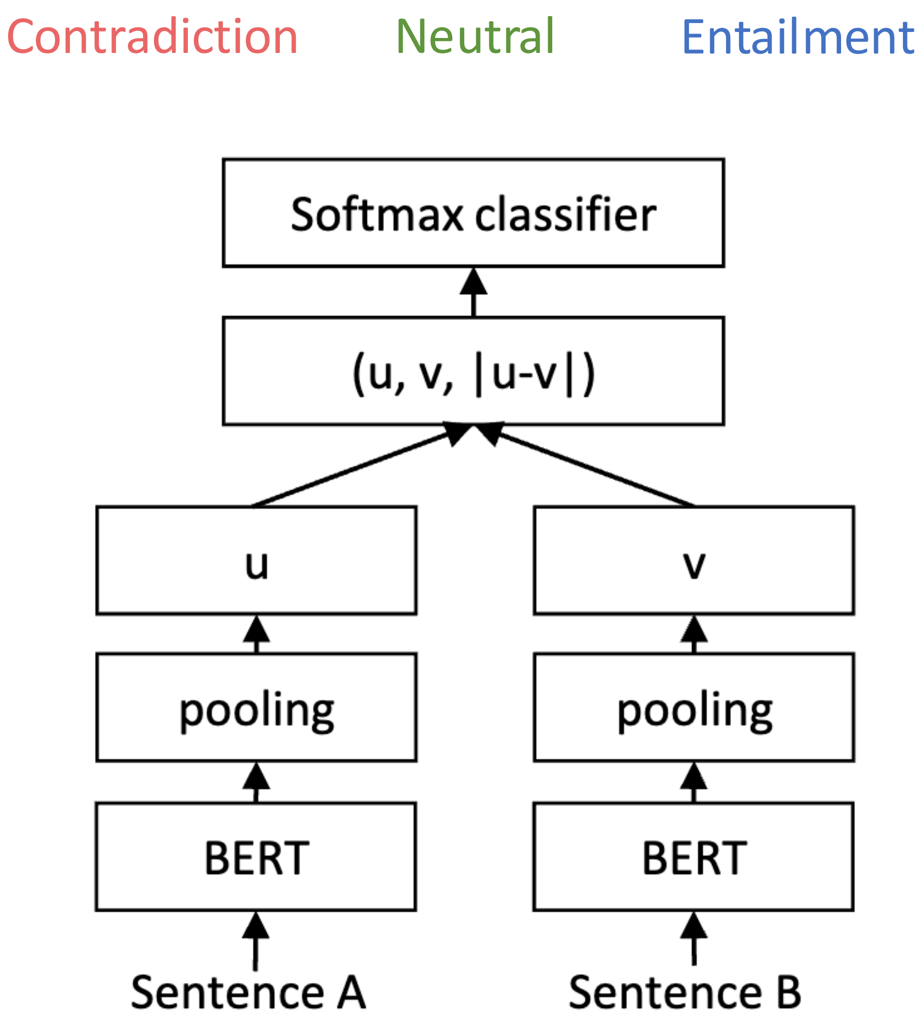
Contradiction Neutral Entailment





Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Sentence-BERT



Cross Entropy Loss

$$o = \mathrm{softmax}(W_t(u, v, |u - v|))$$

Triplet Loss

$$max(||s_a - s_p|| - ||s_a - s_n|| + \epsilon, 0)$$

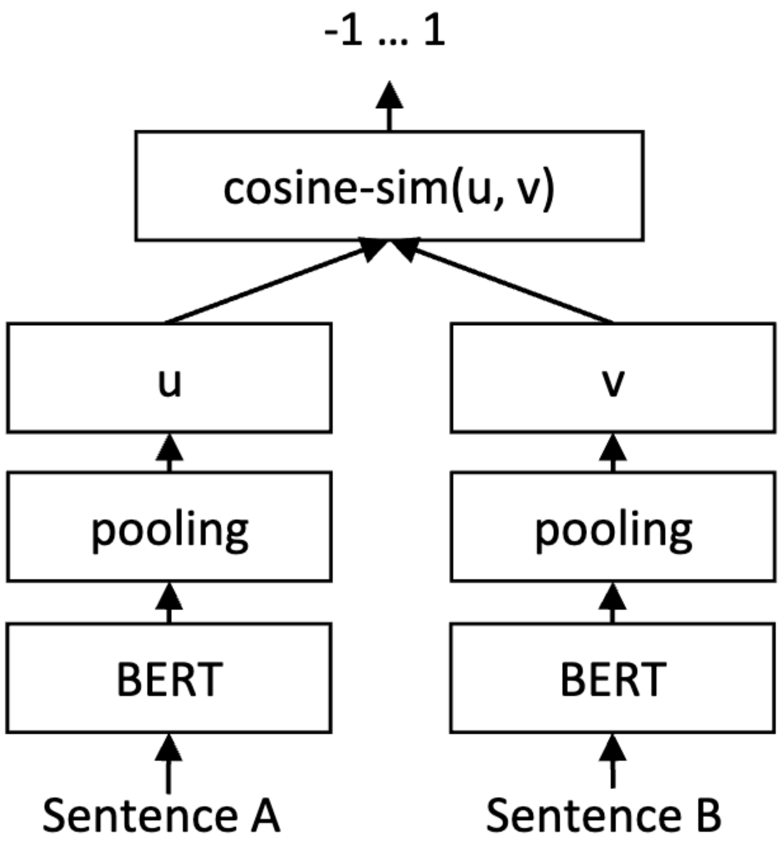
Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Sentence-BERT: Performance

Model	STS12	STS13	STS14	STS15	STS16	STSb	SICK-R	Avg.
Avg. GloVe embeddings	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
Avg. BERT embeddings	38.78	57.98	57.98	63.15	61.06	46.35	58.40	54.81
BERT CLS-vector	20.16	30.01	20.09	36.88	38.08	16.50	42.63	29.19
InferSent - Glove	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder	64.49	67.80	64.61	76.83	73.18	74.92	**76.69**	71.22
SBERT-NLI-base	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
SBERT-NLI-large	72.27	**78.46**	**74.90**	80.99	76.25	**79.23**	73.75	76.55
SRoBERTa-NLI-base	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
SRoBERTa-NLI-large	**74.53**	77.00	73.18	**81.85**	**76.82**	79.10	74.29	**76.68**

SimCSE

- Simple Contrastive Learning of Sentence Embeddings



Contrastive Loss

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau}}$$

Contrastive Learning

Sentence 1A | *Sentence 1B*

Sentence 1A	*Sentence 1B*
Sentence 2A	*Sentence 2B*
Sentence 3A	*Sentence 3B*
Sentence 4A	*Sentence 4B*
Sentence 5A	*Sentence 5B*

Contrastive Loss

$$\ell_i = -\log \frac{\boxed{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)}}/\tau}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau}}$$

Contrastive Learning

Sentence 1A	*Sentence 1B*
Sentence 2A	*Sentence 2B*
Sentence 3A	*Sentence 3B*
Sentence 4A	*Sentence 4B*
Sentence 5A	*Sentence 5B*

Contrastive Loss

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} \boxed{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)}}/\tau}$$

Contrastive Learning

Sentence 1A	*Sentence 1B*
Sentence 2A	*Sentence 2B*
Sentence 3A	*Sentence 3B*
Sentence 4A	*Sentence 4B*
Sentence 5A	*Sentence 5B*

Contrastive Loss

$$\ell_i = -\log \frac{\boxed{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)}}/\tau}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau}}$$

Contrastive Learning

Sentence 1A	*Sentence 1B*
Sentence 2A	*Sentence 2B*
Sentence 3A	*Sentence 3B*
Sentence 4A	*Sentence 4B*
Sentence 5A	*Sentence 5B*

Contrastive Loss

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} \boxed{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)}}/\tau}$$

Unsupervised Contrastive Learning

Sentence 1

Sentence 1'

Sentence 2

Sentence 3

Sentence 4

Sentence 5

Contrastive Loss

$$\ell_i = -\log \frac{\boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z_i'})/}}\tau}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z_j'})/\tau}}$$

Unsupervised Contrastive Learning

Sentence 1

Sentence 1'

Sentence 2

Sentence 3

Sentence 4

Sentence 5

Contrastive Loss

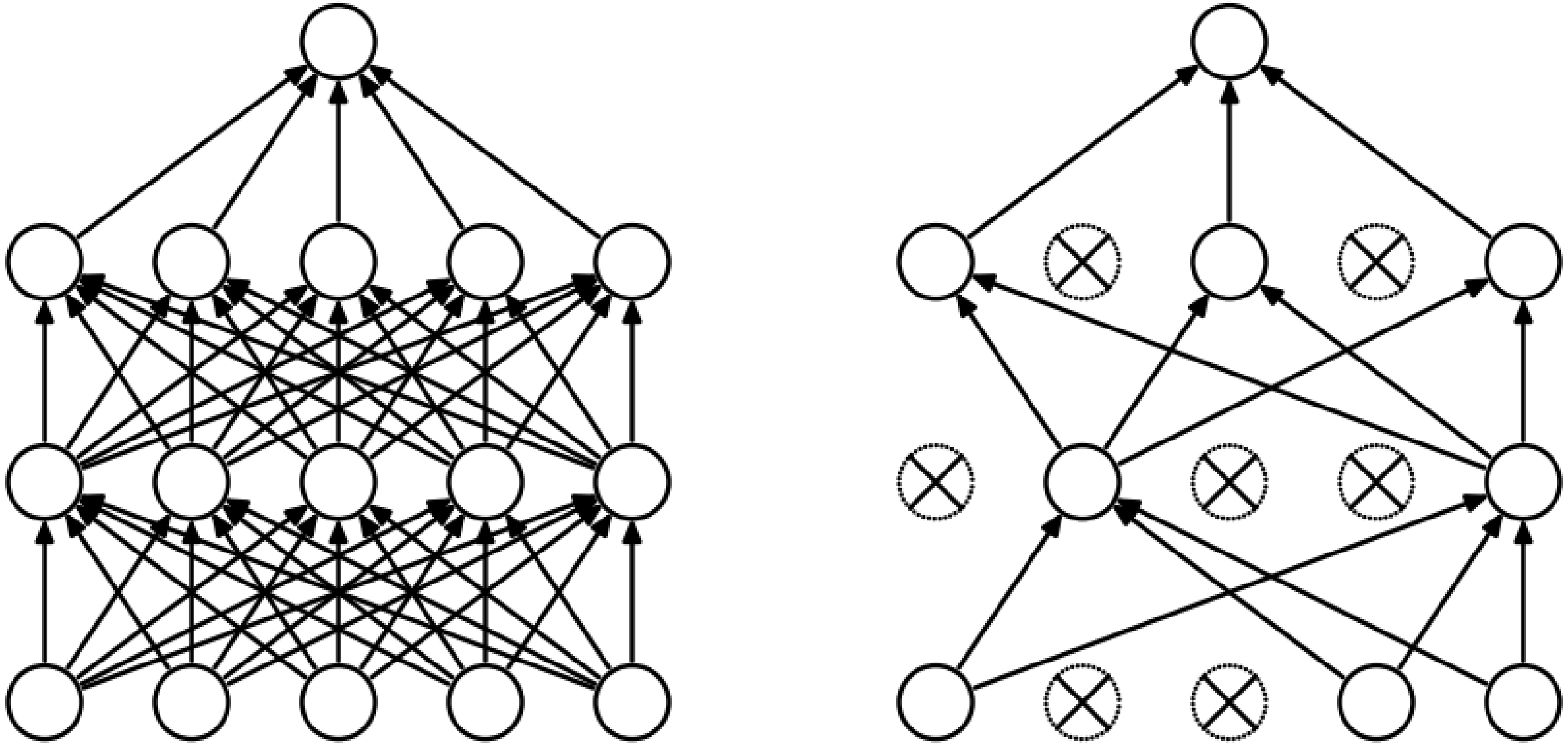
$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z'_i})/\tau}}{\sum_{j=1}^{N} \boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z'_j})}}/\tau}$$

Generate positive example with masking

If it is sunny tomorrow, we will go hiking.

If [mask] is sunny tomorrow, we [mask] go hiking.

Dropout



(a) Standard Neural Net (b) After applying dropout.

Generate positive example with neuron masking

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Unsupervised Contrastive Learning

Sentence 1

Sentence 1’

Sentence 2

Sentence 3

Sentence 4

Sentence 5

Contrastive Loss

$$\ell_i = -\log \frac{\boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z'_i})}}/\tau}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z'_j})/\tau}}$$

Unsupervised Contrastive Learning

Sentence 1

Sentence 1'

Sentence 2

Sentence 3

Sentence 4

Sentence 5

Contrastive Loss

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z'_i})/\tau}}{\sum_{j=1}^{N} \boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z'_j})}}/\tau}$$

Unsupervised Contrastive Learning

Sentence 1

Sentence 2 *Sentence 2'*

Sentence 3

Sentence 4

Sentence 5

Contrastive Loss

$$\ell_i = -\log \frac{\boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z_i'})/\tau}}}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z_j'})/\tau}}$$

Unsupervised Contrastive Learning

Sentence 2

Sentence 1'

Sentence 2

Sentence 3

Sentence 4

Sentence 5

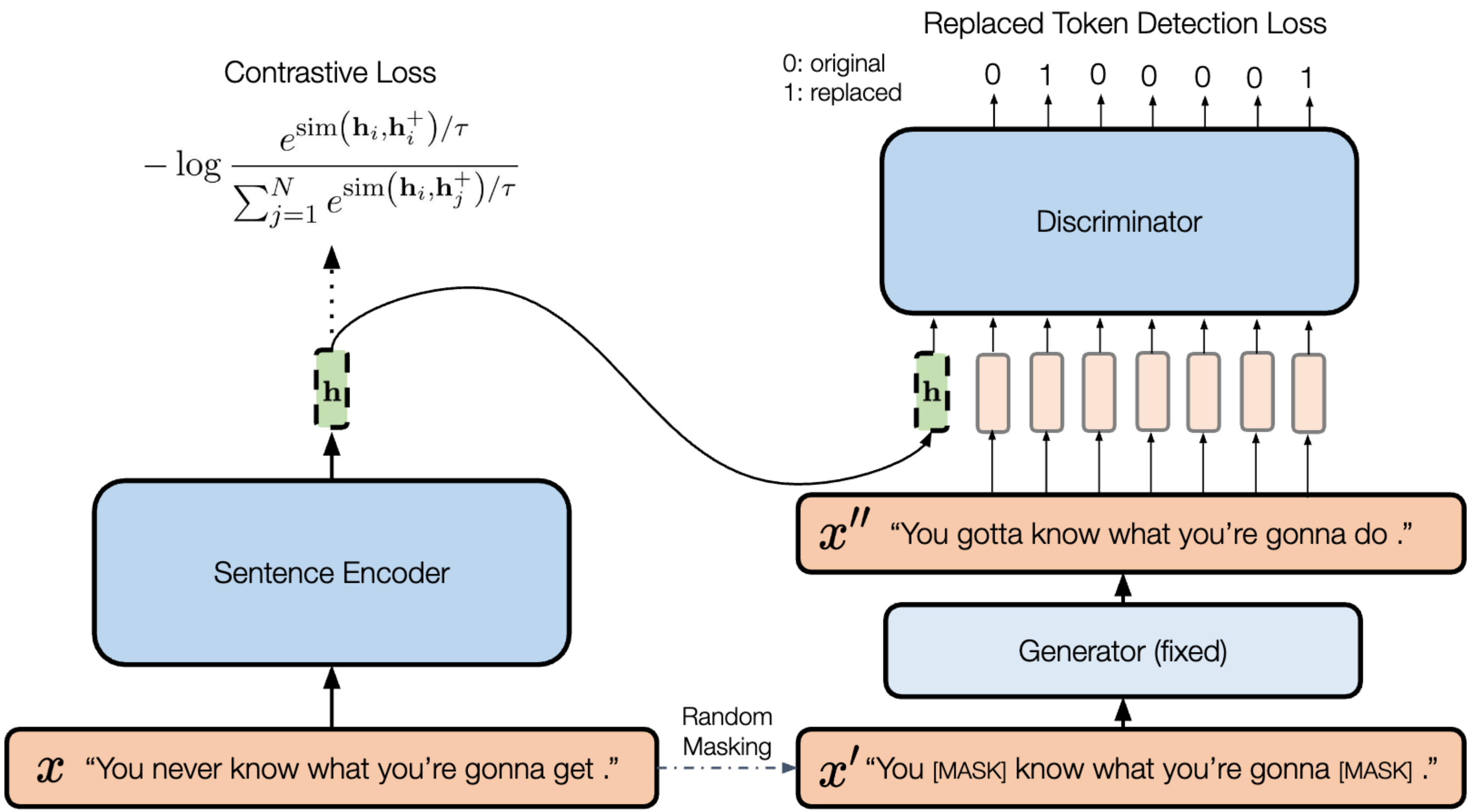
Contrastive Loss

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z'_i})/\tau}}{\sum_{j=1}^{N} \boxed{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z'_j})}}/\tau}$$

SimCSE: Performance

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
Unsupervised models								
GloVe embeddings (avg.)♣	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
$\text{BERT}_{\text{base}}$ (first-last avg.)	39.70	59.38	49.67	66.03	66.19	53.87	62.06	56.70
$\text{BERT}_{\text{base}}$-flow	58.40	67.10	60.85	75.16	71.22	68.66	64.47	66.55
$\text{BERT}_{\text{base}}$-whitening	57.83	66.90	60.90	75.08	71.31	68.24	63.73	66.28
IS-$\text{BERT}_{\text{base}}$♡	56.77	69.24	61.21	75.23	70.16	69.21	64.25	66.58
CT-$\text{BERT}_{\text{base}}$	61.63	76.80	68.47	77.50	76.48	74.31	69.19	72.05
∗ SimCSE-$\text{BERT}_{\text{base}}$	**68.40**	**82.41**	**74.38**	**80.91**	**78.56**	**76.85**	**72.23**	**76.25**
$\text{RoBERTa}_{\text{base}}$ (first-last avg.)	40.88	58.74	49.07	65.63	61.48	58.55	61.63	56.57
$\text{RoBERTa}_{\text{base}}$-whitening	46.99	63.24	57.23	71.36	68.99	61.36	62.91	61.73
DeCLUTR-$\text{RoBERTa}_{\text{base}}$	52.41	75.19	65.52	77.12	78.63	72.41	**68.62**	69.99
∗ SimCSE-$\text{RoBERTa}_{\text{base}}$	**70.16**	**81.77**	**73.24**	**81.36**	**80.65**	**80.22**	68.56	**76.57**
∗ SimCSE-$\text{RoBERTa}_{\text{large}}$	**72.86**	**83.99**	**75.62**	**84.77**	**81.80**	**81.98**	**71.26**	**78.90**
Supervised models								
InferSent-GloVe♣	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder♣	64.49	67.80	64.61	76.83	73.18	74.92	76.69	71.22
$\text{SBERT}_{\text{base}}$♣	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
$\text{SBERT}_{\text{base}}$-flow	69.78	77.27	74.35	82.01	77.46	79.12	76.21	76.60
$\text{SBERT}_{\text{base}}$-whitening	69.65	77.57	74.66	82.27	78.39	79.52	76.91	77.00
CT-$\text{SBERT}_{\text{base}}$	74.84	83.20	78.07	83.84	77.93	81.46	76.42	79.39
∗ SimCSE-$\text{BERT}_{\text{base}}$	**75.30**	**84.67**	**80.19**	**85.40**	**80.82**	**84.25**	**80.39**	**81.57**
$\text{SRoBERTa}_{\text{base}}$♣	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
$\text{SRoBERTa}_{\text{base}}$-whitening	70.46	77.07	74.46	81.64	76.43	79.49	76.65	76.60
∗ SimCSE-$\text{RoBERTa}_{\text{base}}$	**76.53**	**85.21**	**80.95**	**86.03**	**82.57**	**85.83**	**80.50**	**82.52**
∗ SimCSE-$\text{RoBERTa}_{\text{large}}$	**77.46**	**87.27**	**82.36**	**86.66**	**83.93**	**86.70**	**81.95**	**83.76**

DiffCSE



DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

DiffCSE: Performance

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
GloVe embeddings (avg.)$^\clubsuit$	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
$\text{BERT}_{\texttt{base}}$ (first-last avg.)$^\diamondsuit$	39.70	59.38	49.67	66.03	66.19	53.87	62.06	56.70
$\text{BERT}_{\texttt{base}}$-flow$^\diamondsuit$	58.40	67.10	60.85	75.16	71.22	68.66	64.47	66.55
$\text{BERT}_{\texttt{base}}$-whitening$^\diamondsuit$	57.83	66.90	60.90	75.08	71.31	68.24	63.73	66.28
IS-$\text{BERT}_{\texttt{base}}$ $^\heartsuit$	56.77	69.24	61.21	75.23	70.16	69.21	64.25	66.58
CMLM-$\text{BERT}_{\texttt{base}}$ $^\spadesuit$ (1TB data)	58.20	61.07	61.67	73.32	74.88	76.60	64.80	67.22
CT-$\text{BERT}_{\texttt{base}}$ $^\diamondsuit$	61.63	76.80	68.47	77.50	76.48	74.31	69.19	72.05
SG-OPT-$\text{BERT}_{\texttt{base}}$ †	66.84	80.13	71.23	81.56	77.17	77.23	68.16	74.62
SimCSE-$\text{BERT}_{\texttt{base}}$ $^\diamondsuit$	68.40	82.41	74.38	80.91	78.56	76.85	**72.23**	76.25
$*$ SimCSE-$\text{BERT}_{\texttt{base}}$(reproduce)	70.82	82.24	73.25	81.38	77.06	77.24	71.16	76.16
$*$ DiffCSE-$\text{BERT}_{\texttt{base}}$	**72.28**	**84.43**	**76.47**	**83.90**	**80.54**	**80.59**	71.23	**78.49**
$\text{RoBERTa}_{\texttt{base}}$ (first-last avg.)$^\diamondsuit$	40.88	58.74	49.07	65.63	61.48	58.55	61.63	56.57
$\text{RoBERTa}_{\texttt{base}}$-whitening$^\diamondsuit$	46.99	63.24	57.23	71.36	68.99	61.36	62.91	61.73
DeCLUTR-$\text{RoBERTa}_{\texttt{base}}$ $^\diamondsuit$	52.41	75.19	65.52	77.12	78.63	72.41	68.62	69.99
SimCSE-$\text{RoBERTa}_{\texttt{base}}$ $^\diamondsuit$	**70.16**	81.77	73.24	81.36	80.65	80.22	68.56	76.57
$*$ SimCSE-$\text{RoBERTa}_{\texttt{base}}$(reproduce)	68.60	81.36	73.16	81.61	80.76	80.58	68.83	76.41
$*$ DiffCSE-$\text{RoBERTa}_{\texttt{base}}$	70.05	**83.43**	**75.49**	**82.81**	**82.12**	**82.38**	**71.19**	**78.21**

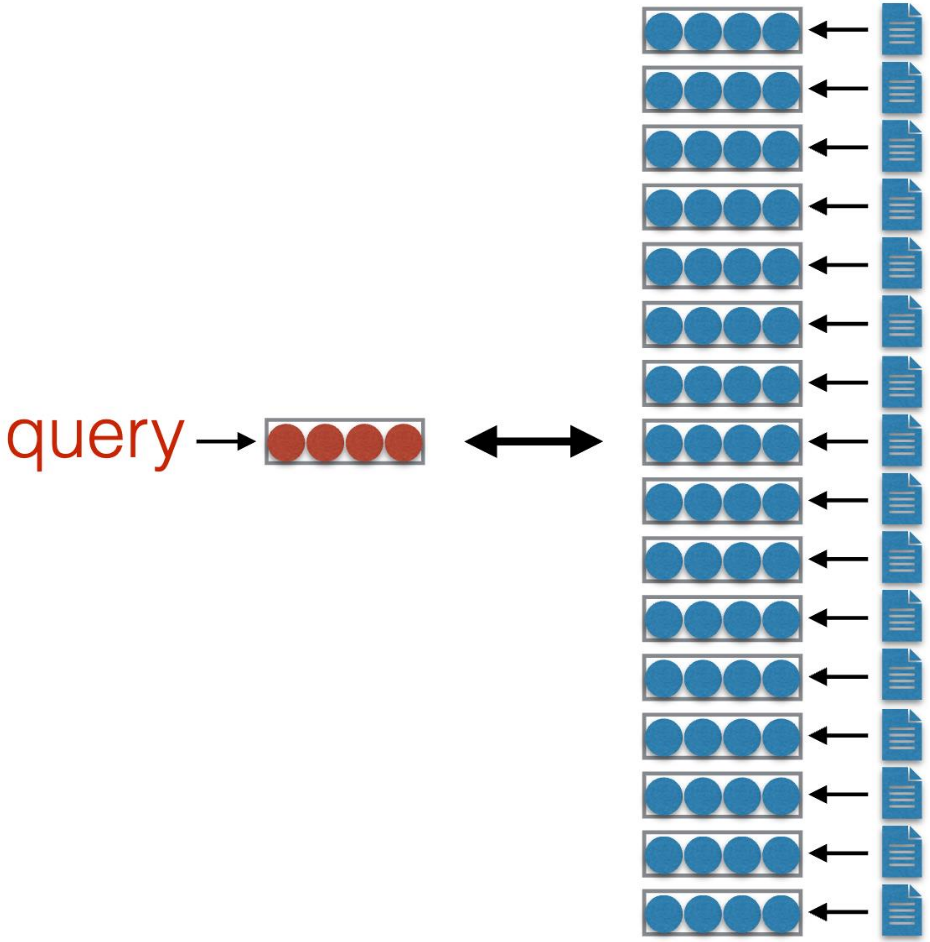
Dense Passage Retrieval

Similarity between query and documents

Similarity between two sentences

We will go hiking if tomorrow is a sunny day.

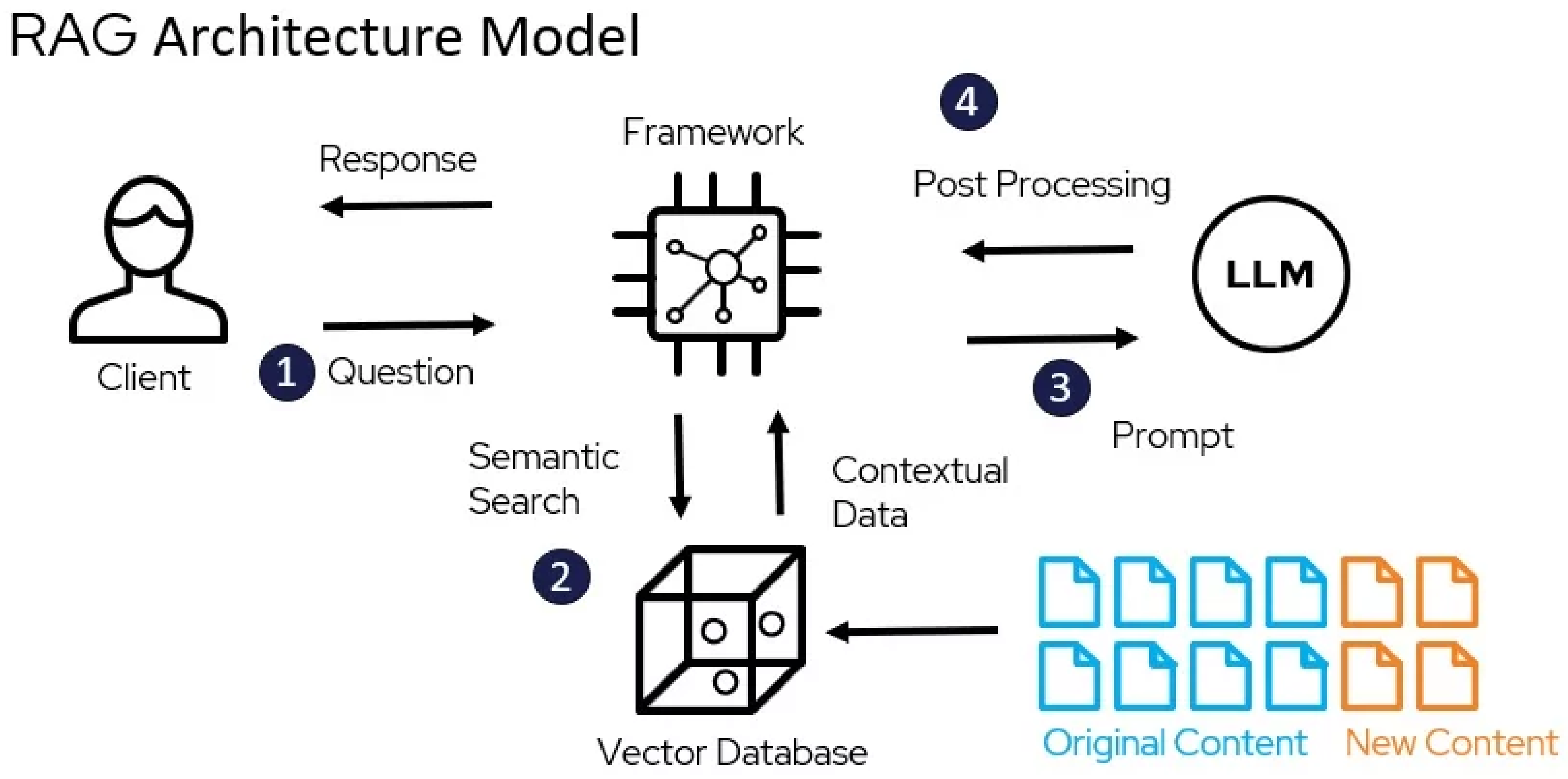
If it is sunny tomorrow, we will go hiking.



Lecture Plan

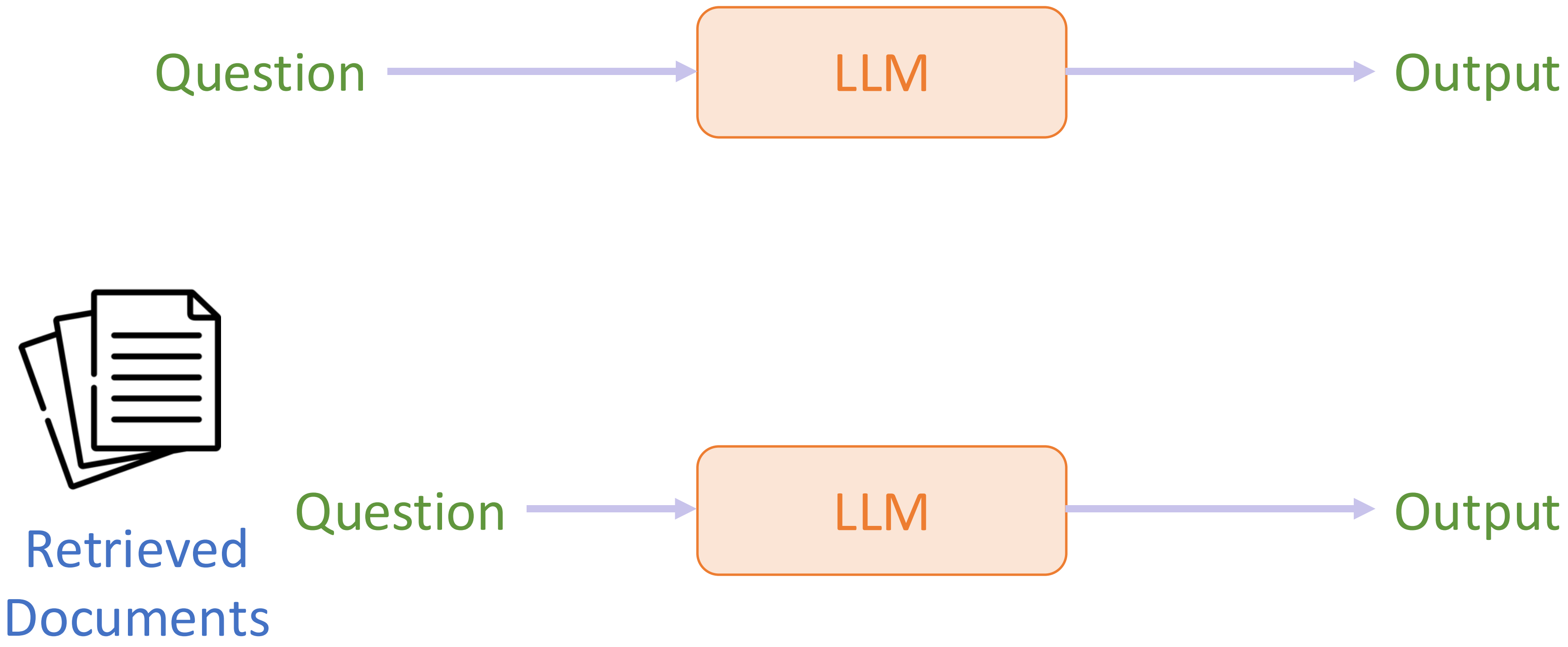
- Human Preference Optimization
  - Simple Preference Optimization
  - Group Relative Policy Optimization
- Text Similarity
  - Sentence-BERT
  - SimCSE, DIffCSE, DPR
- Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG)



https://pureinsights.com/blog/2023/what-is-retrieval-augmented-generation-rag/

Retrieval-Augmented Generation (RAG)

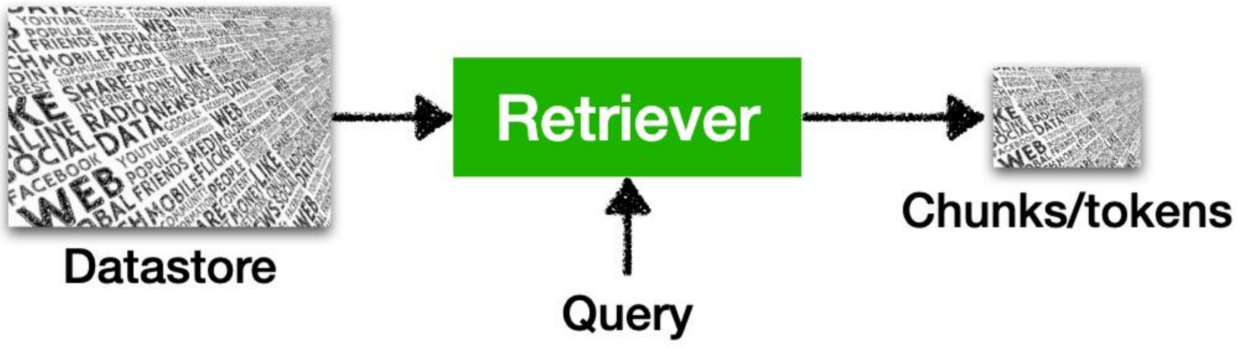


Retrieval-Augmented Generation (RAG)

Retrieval models and language models are trained **independently**

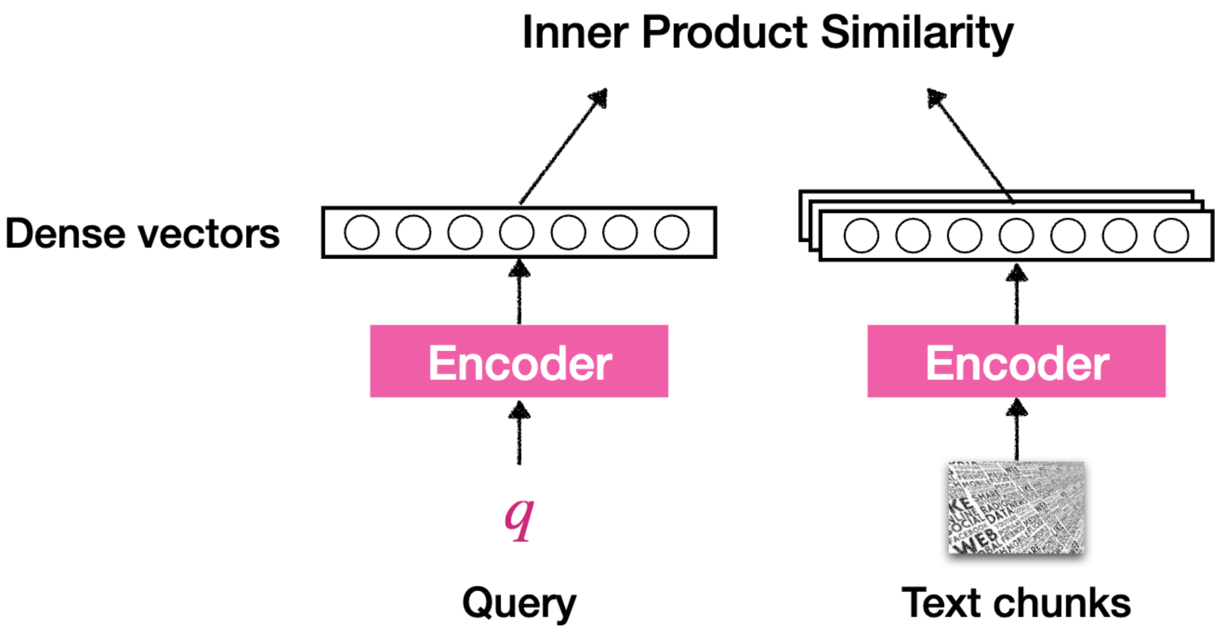
- Training language models

- Training retrieval models

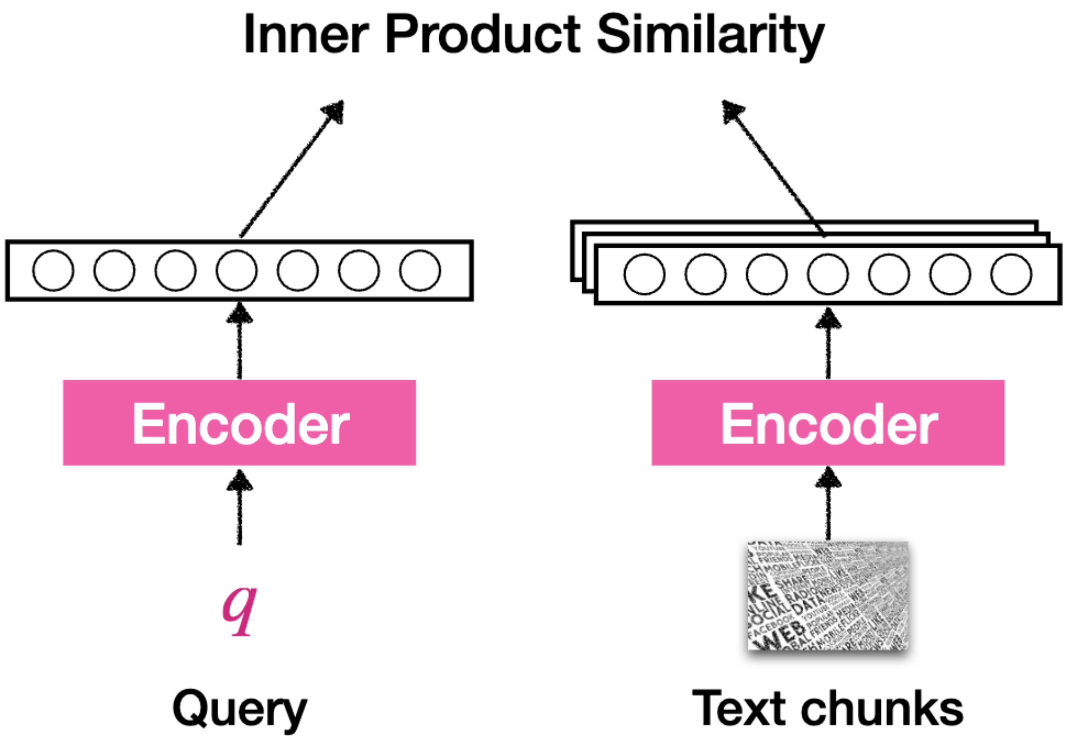


How to Train A Retriever?

Dense retrieval models: DPR (Karpukhin et al. 2020)



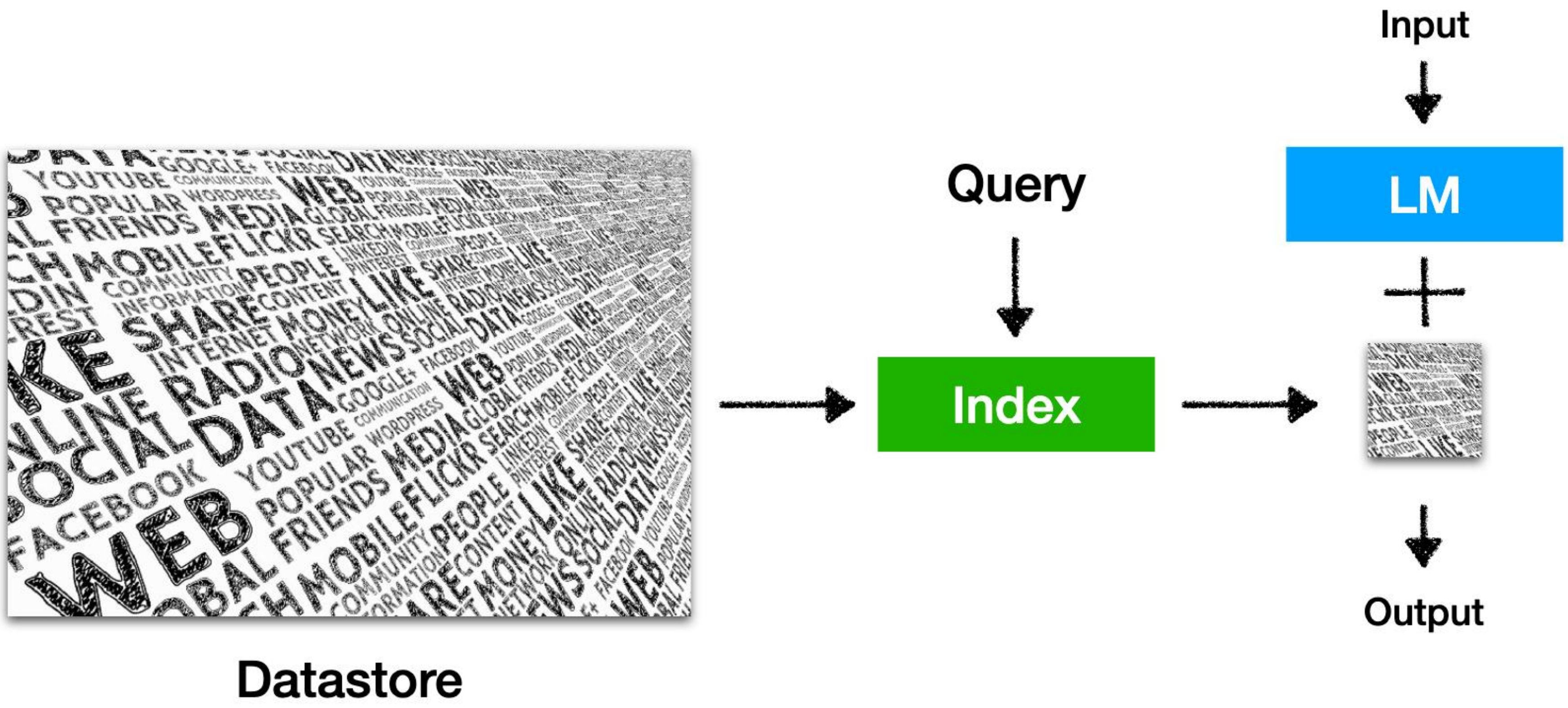
How to Train A Retriever?



$$L(q, p^+, p_1^-, p_2^-, \ldots, p_n^-) = -\log \frac{\exp(\mathrm{sim}(q, p^+))}{\exp(\mathrm{sim}(q, p^+)) + \sum_{j=1}^{n} \exp(\mathrm{sim}(q, p_j^-))}$$

Contrastive learning

Retrieval-Augmented Generation (RAG)



Why RAG?

- LLMs can’t memorize all (long-tail) knowledge in their parameters

List 5 important papers authored by Geoffrey Hinton

Geoffrey Hinton is a renowned computer scientist …
Here are five important papers authored by him:

1. "**Learning Internal Representations by Error Propagation**" (with **D. E. Rumelhart** and **R. J. Williams**) - This paper, published in **1986**, ..

2. "**Deep Boltzmann Machines**" (with **R. Salakhutdinov**) - Published in **2009**, ..

…

4. "**Deep Learning**" (with Y. Bengio and A. Courville) - Published as a book in **2016**,…

5. "**Attention Is All You Need**" (with V. Vaswani, N. Shazeer, et al.) - Published in **2017**, this paper introduced the Transformer model,…

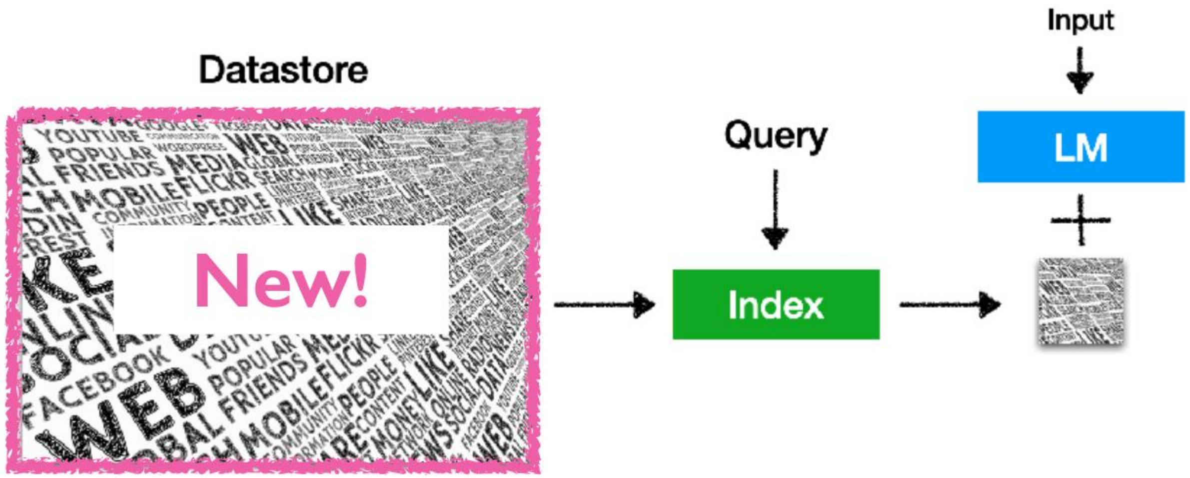
Why RAG?

- LLMs' knowledge is easily outdated and hard to update

Who is the CEO of Twitter?

ChatGPT

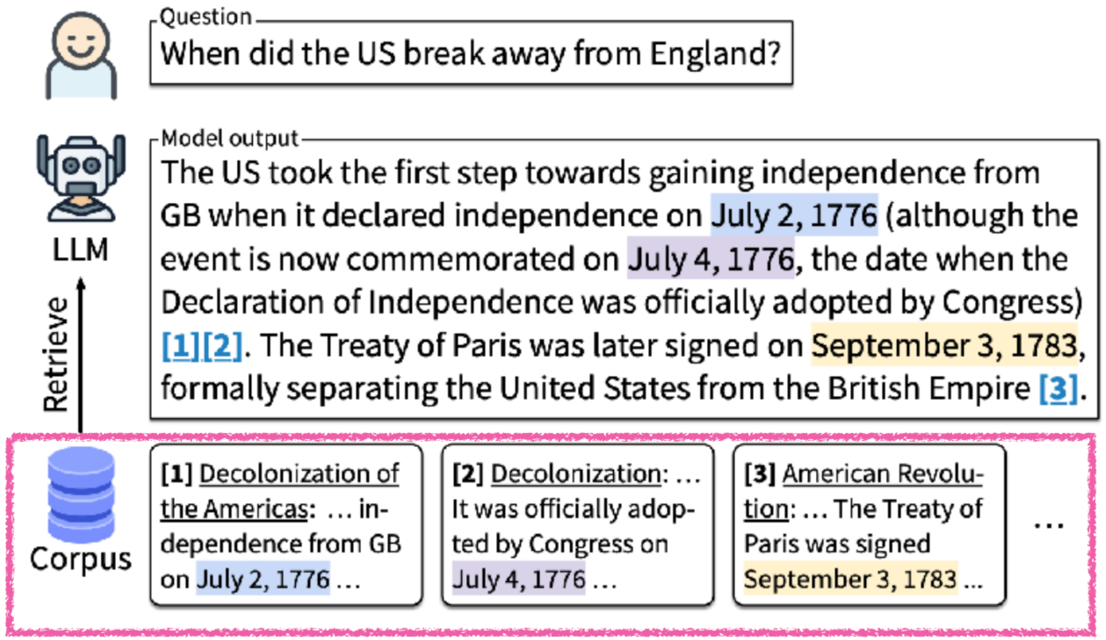
As of my **knowledge cutoff in September 2021**, the CEO of Twitter is **Jack Dorsey**….



Why RAG?

- LLMs' output is challenging to interpret and verify

Generating text with citations



(Nakano et al. 2021; Menick et al., 2022; Gao et al., 2023)

Why RAG?

- LLMs are shown to easily leak private training data

Extracting Training Data from Large Language Models

Nicholas Carlini[1] Florian Tramèr[2] Eric Wallace[3] Matthew Jagielski[4]
Ariel Herbert-Voss[5,6] Katherine Lee[1] Adam Roberts[1] Tom Brown[5]
Dawn Song[3] Úlfar Erlingsson[7] Alina Oprea[4] Colin Raffel[1]
[1]*Google* [2]*Stanford* [3]*UC Berkeley* [4]*Northeastern University* [5]*OpenAI* [6]*Harvard* [7]*Apple*

Category	Count
US and international news	109
Log files and error reports	79
License, terms of use, copyright notices	54
Lists of named items (games, countries, etc.)	54
Forum or Wiki entry	53
Valid URLs	50
Named individuals (non-news samples only)	46
Promotional content (products, subscriptions, etc.)	45
High entropy (UUIDs, base64 data)	35
Contact info (address, email, phone, twitter, etc.)	32
Code	31
Configuration files	30
Religious texts	25
Pseudonyms	15
Donald Trump tweets and quotes	12
Web forms (menu items, instructions, etc.)	11
Tech news	11
Lists of numbers (dates, sequences, etc.)	10

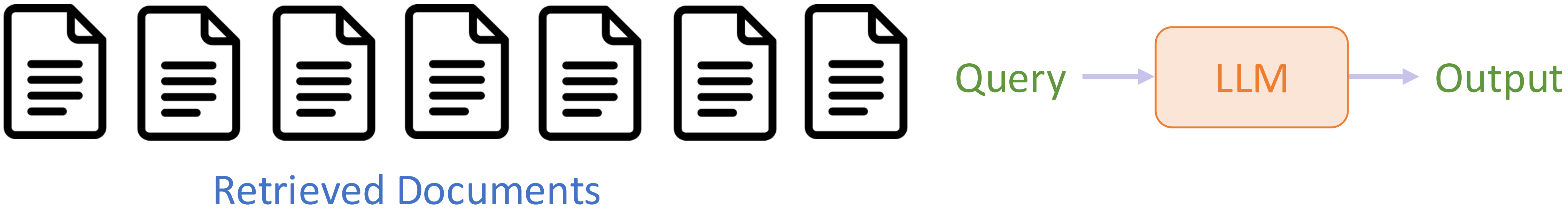
Why RAG?

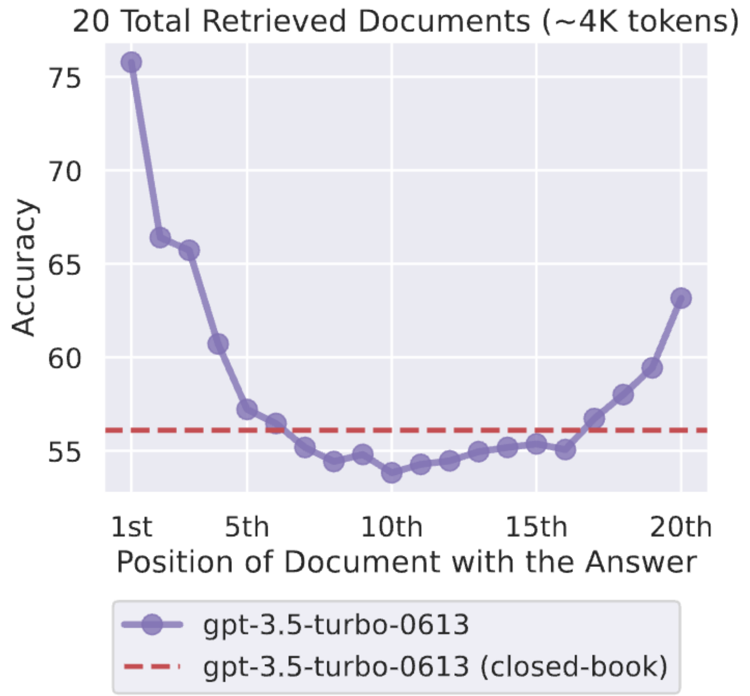
- Potentially leverage other modalities
  - Knowledge base
  - Tabular data
  - …

Challenges with RAG

- Longer input text
  - Length generalization
  - KV cache
- The lost-in-the-middle problem

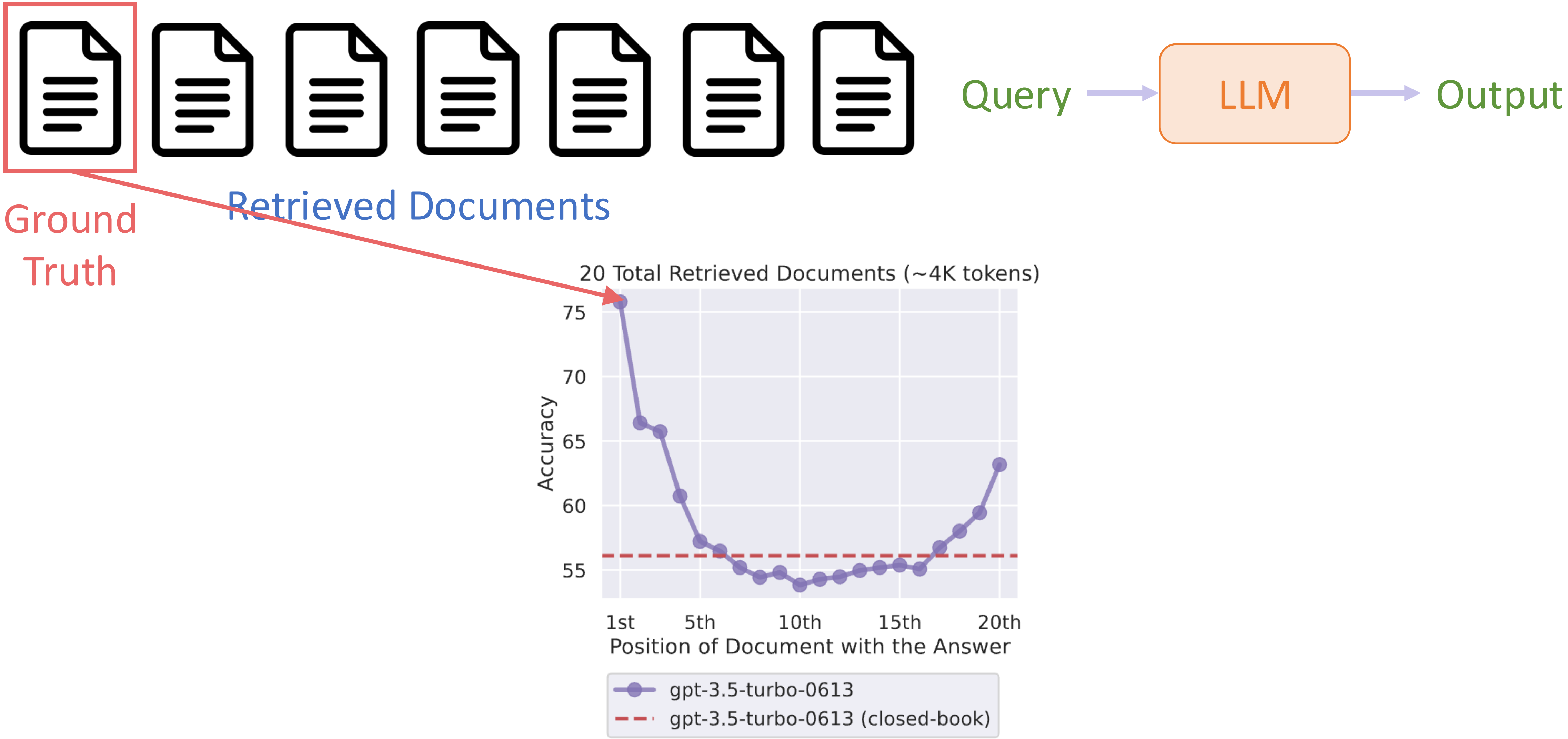
The Lost-in-the-Middle Problem





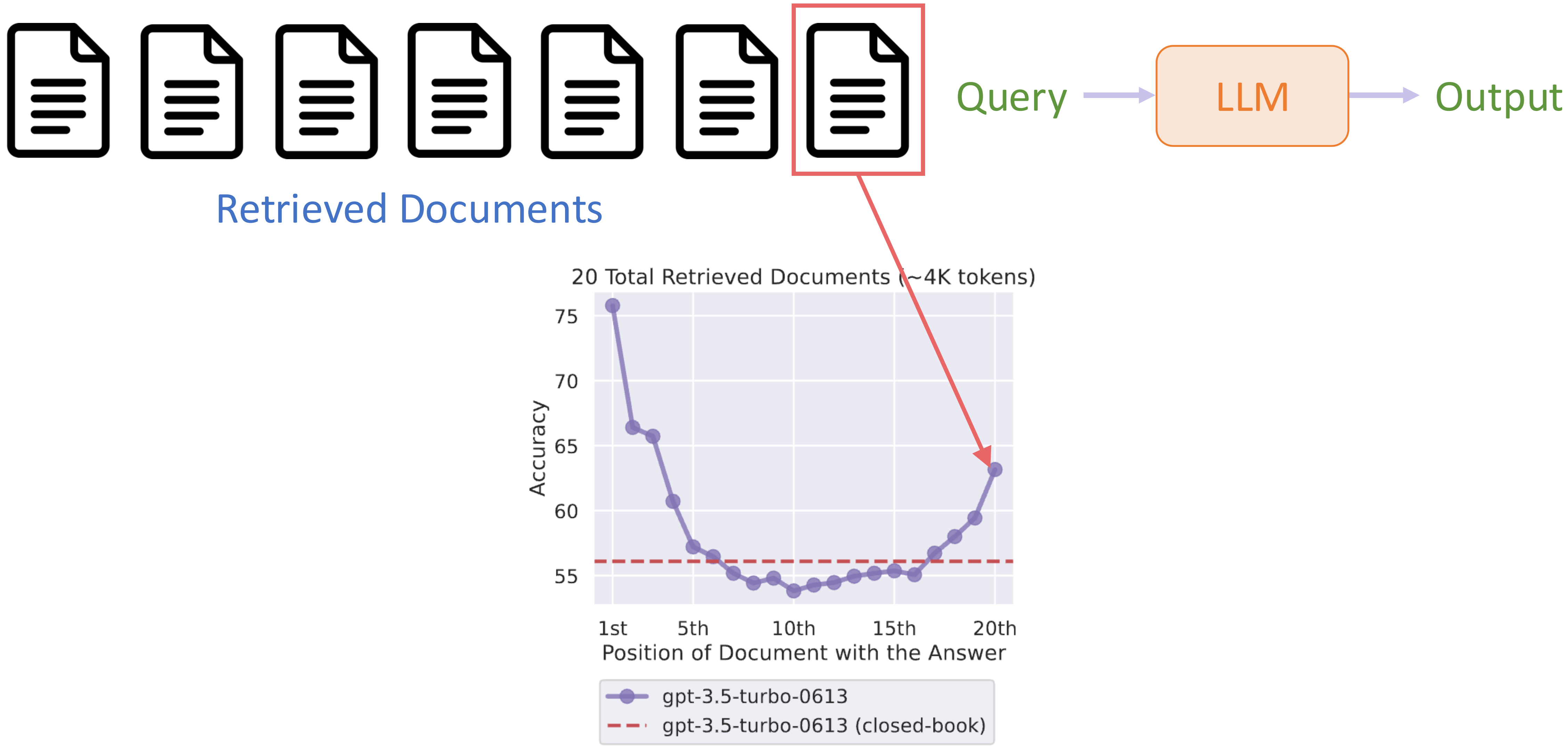
Lost in the Middle: How Language Models Use Long Contexts, 2023

The Lost-in-the-Middle Problem

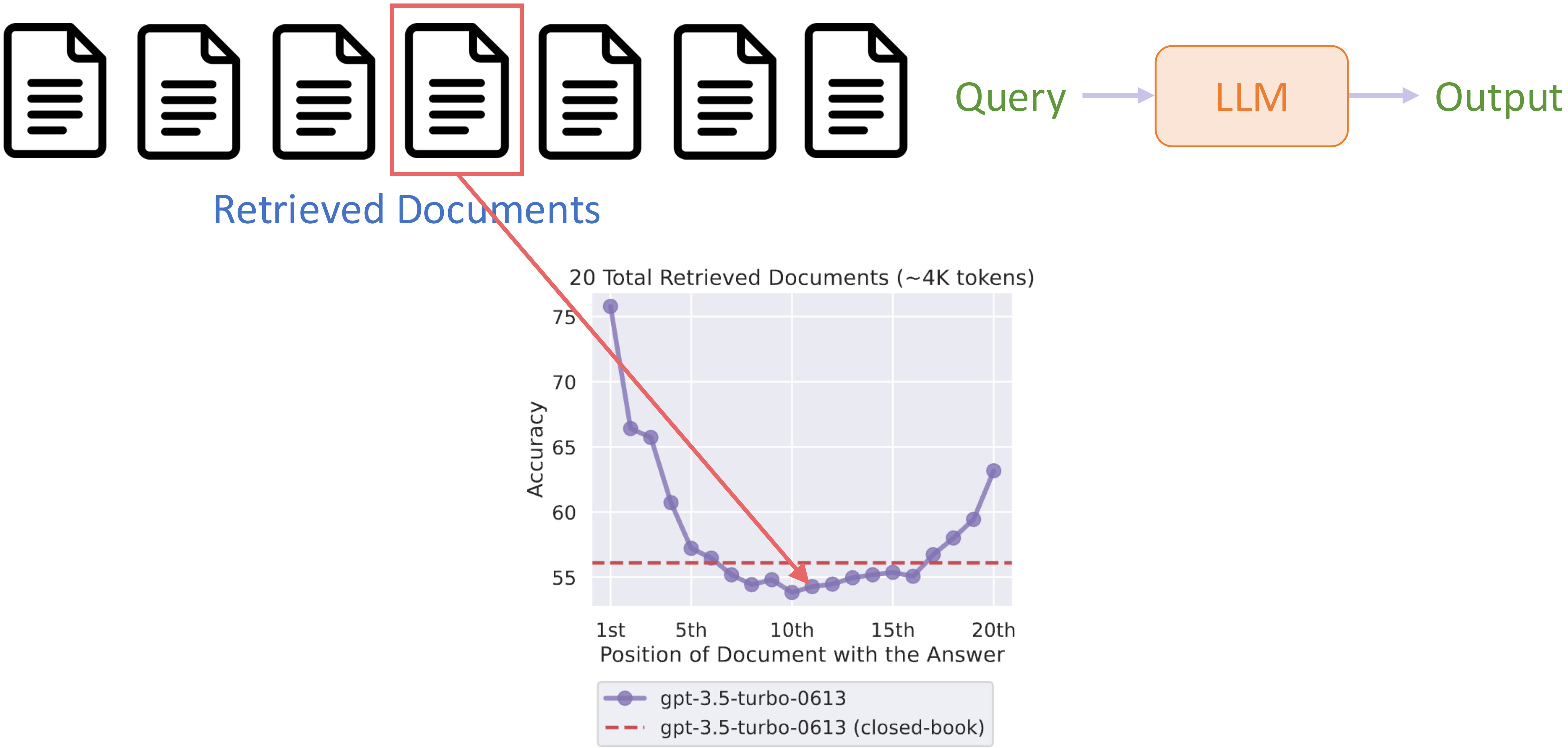


Lost in the Middle: How Language Models Use Long Contexts, 2023

The Lost-in-the-Middle Problem



The Lost-in-the-Middle Problem



Lost in the Middle: How Language Models Use Long Contexts, 2023

Reasons for Positional Bias: Pre-Training Data

Introduction

First Main Point

Second Main Point

Third Main Point

Conclusion

The 5 Paragraph Essay Outline

Topic sentence. xxxx xxxx xxx xx xxxx xx xxx xxxxxxx xx xx x x xxxxx xxxx xx xxxxx xx xxx xx xx xxx xxx x xxxx xxx.

Topic sentence. xxxx xxxx xxx xx xxxx xx xxx xxxxxxx xx xx x x xxxxx xxxx xx xxxxx xx xxx xx xx xxx xxx x xxxx xxx.

Topic Sentence

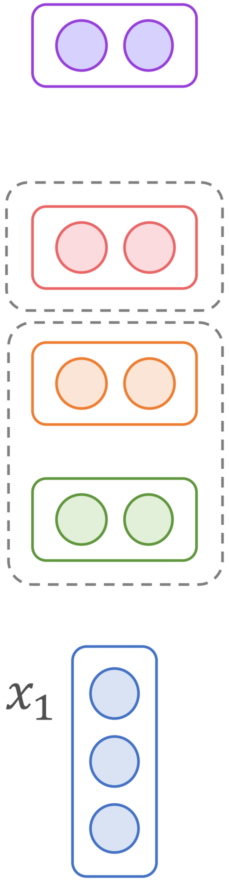
Reasons for Positional Bias: Attention Mechanism

Output

$$q_i = W^Q x_i$$

$$k_i = W^K x_i$$

$$v_i = W^V x_i$$



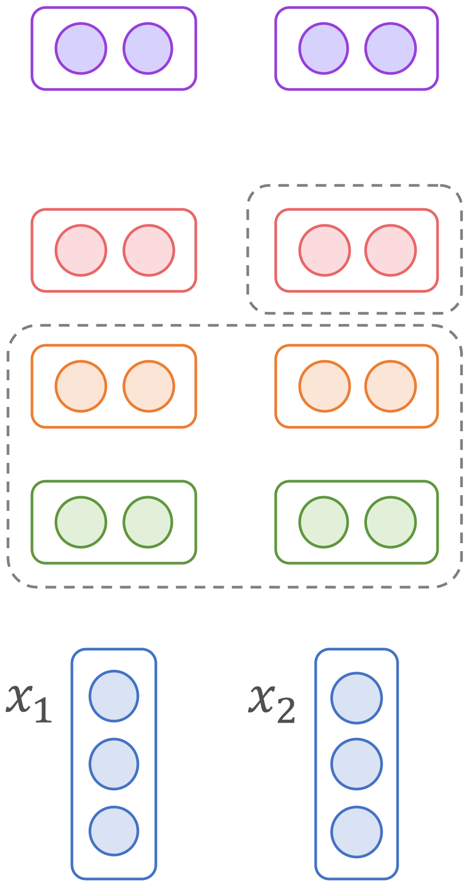
Reasons for Positional Bias: Attention Mechanism

Output

$q_i = W^Q x_i$

$k_i = W^K x_i$

$v_i = W^V x_i$



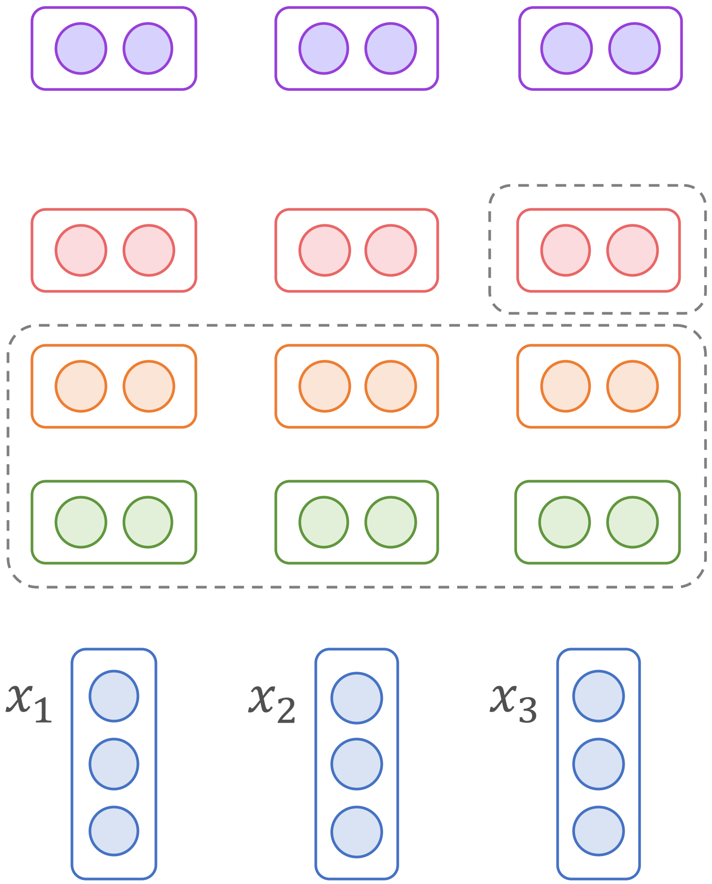
Reasons for Positional Bias: Attention Mechanism

Output

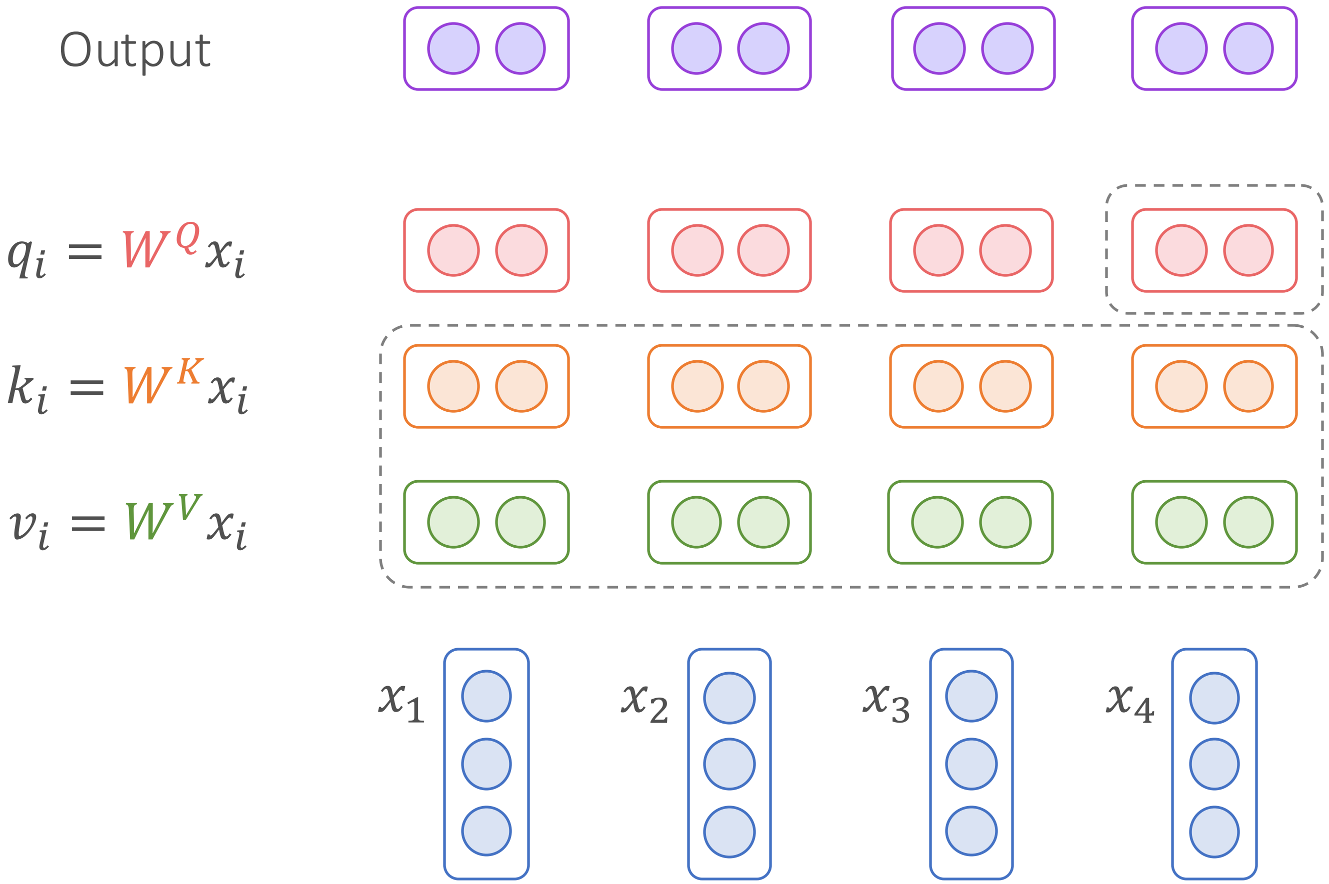
$q_i = W^Q x_i$

$k_i = W^K x_i$

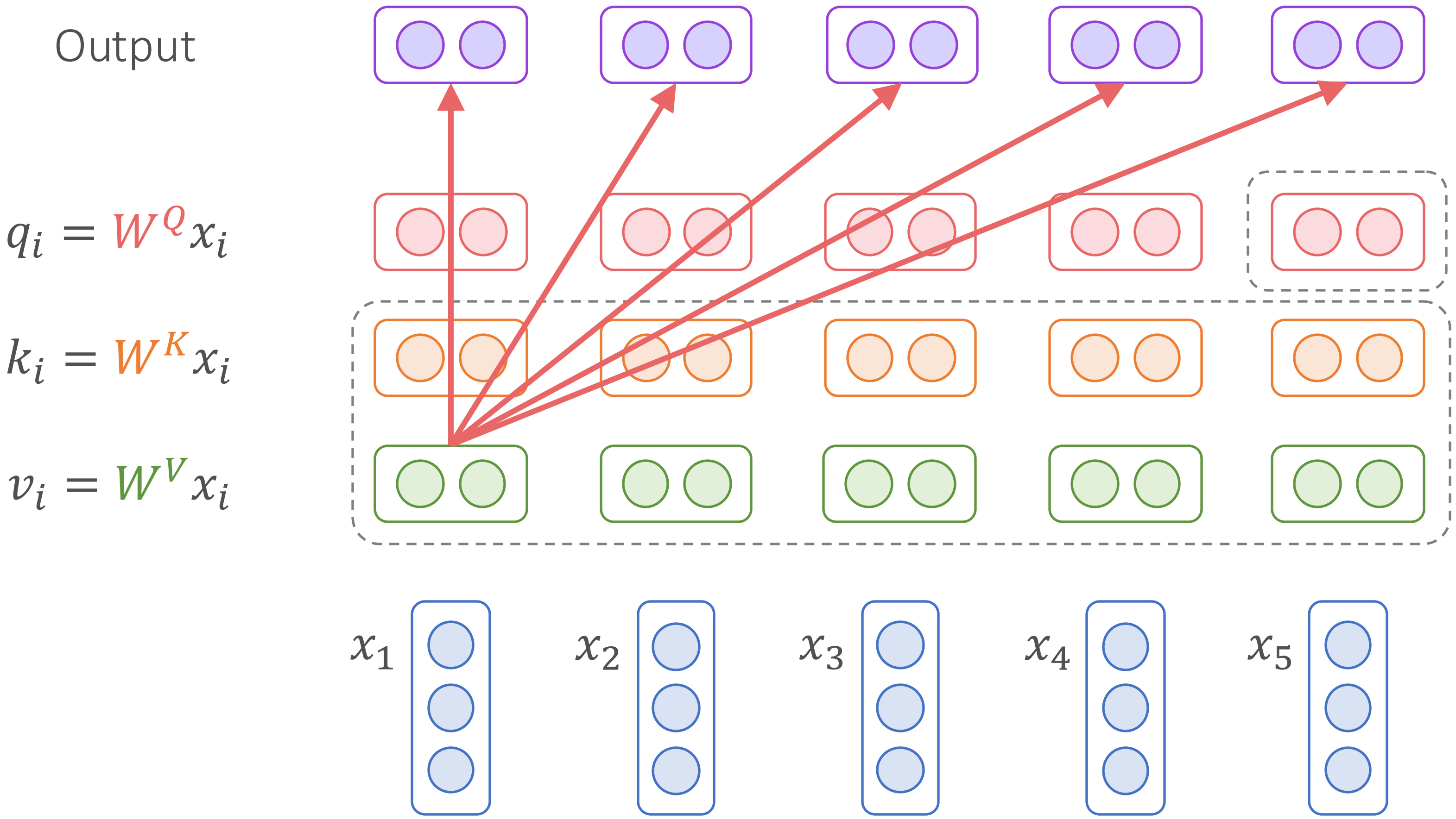
$v_i = W^V x_i$

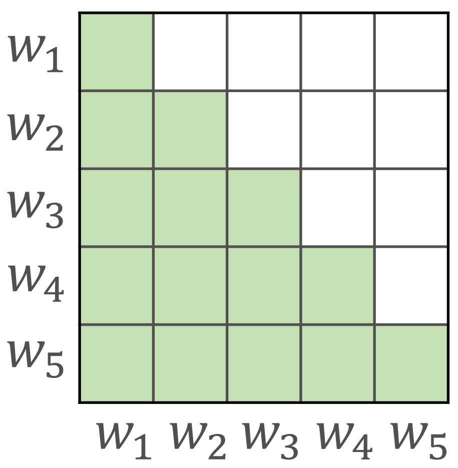


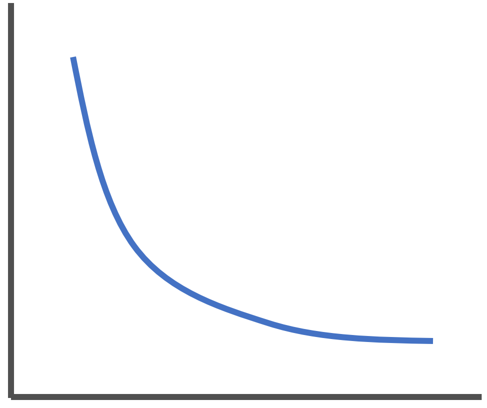
Reasons for Positional Bias: Attention Mechanism



Reasons for Positional Bias: Attention Mechanism





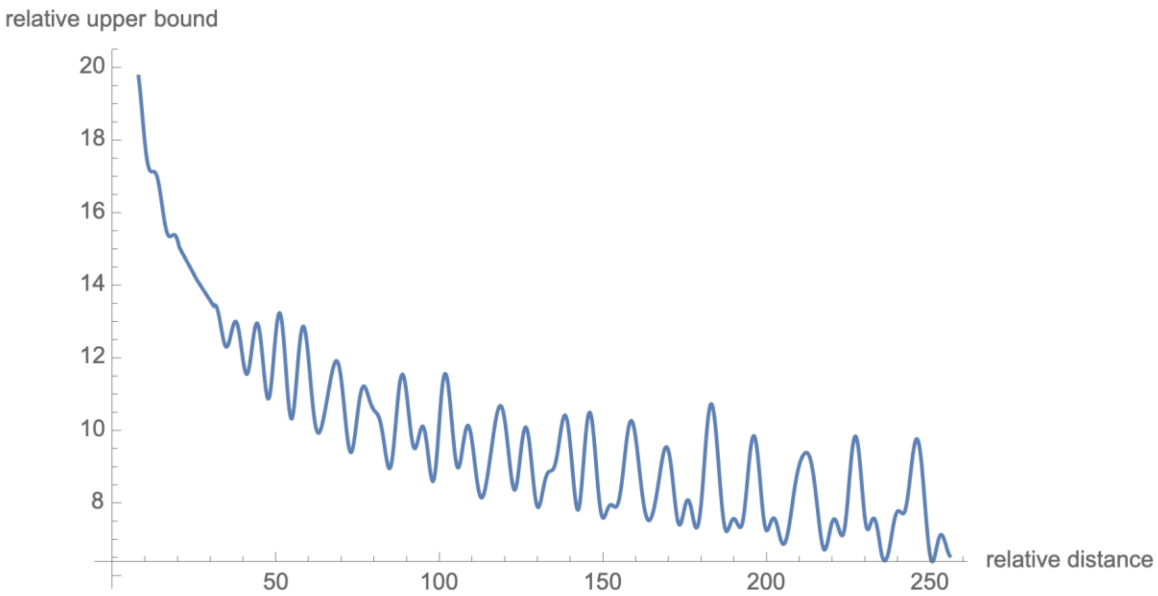


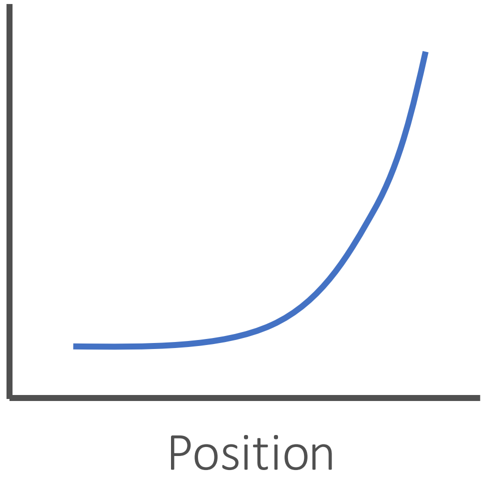
Reasons for Positional Bias: Positional Encoding

Rotary Position Embedding (RoPE)

$$\boldsymbol{q}_m = f_q(\boldsymbol{x}_m, m)$$
$$\boldsymbol{k}_n = f_k(\boldsymbol{x}_n, n)$$
$$\boldsymbol{v}_n = f_v(\boldsymbol{x}_n, n)$$

$$f_q(\boldsymbol{x}_m, m) = (\boldsymbol{W}_q \boldsymbol{x}_m) e^{im\theta}$$
$$f_k(\boldsymbol{x}_n, n) = (\boldsymbol{W}_k \boldsymbol{x}_n) e^{in\theta}$$
$$\langle f_q(\boldsymbol{x}_m, m), f_k(\boldsymbol{x}_n, n) \rangle = \mathrm{Re}[(\boldsymbol{W}_q \boldsymbol{x}_m)(\boldsymbol{W}_k \boldsymbol{x}_n)^* e^{i(m-n)\theta}]$$





RoFormer: Enhanced Transformer with Rotary Position Embedding, 2021

Combine All Together

Introduction

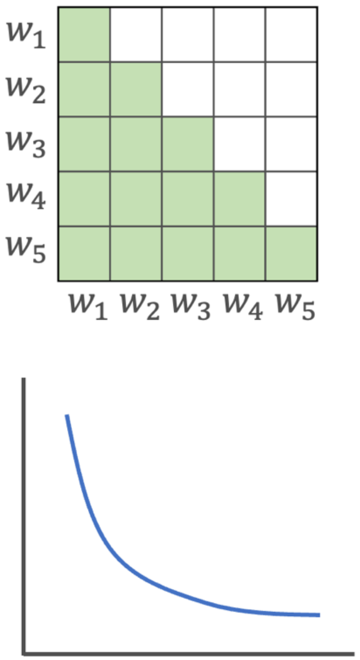
First Main Point

Second Main Point

Third Main Point

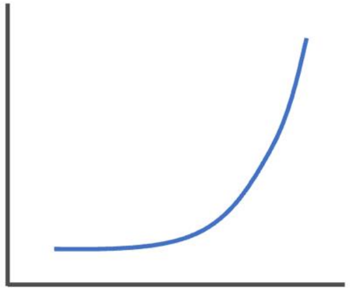
Conclusion

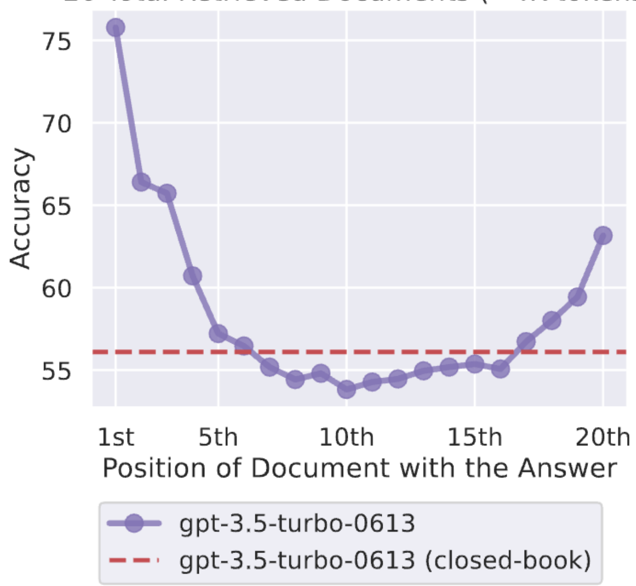
Causal Attention Mask



Rotary Position Embedding (RoPE)

$$\boldsymbol{q}_m = f_q(\boldsymbol{x}_m, m)$$
$$\boldsymbol{k}_n = f_k(\boldsymbol{x}_n, n)$$
$$\boldsymbol{v}_n = f_v(\boldsymbol{x}_n, n)$$





Lecture Plan

- Human Preference Optimization
  - Simple Preference Optimization
  - Group Relative Policy Optimization
- Text Similarity
  - Sentence-BERT
  - SimCSE, DIffCSE, DPR
- Retrieval-Augmented Generation