

$$\begin{array}{c} p \\ p \rightarrow q \\ \hline \therefore q \end{array}$$

Improving the Reasoning of Large Language Models

William Marsan
University of Minnesota-Morris Senior Seminar
November 2024

Large Language Models (LLMs)

- Natural language processing technology
- Trained on vast amounts of text to understand patterns
- This allows them to predict the next most likely word given some text
 - When prompted “Who was the lead singer of Queen?” ChatGPT would likely respond, “Freddie Mercury”.
- Problem: They are good at generating linguistically plausible text but not mathematically precise reasoning

Reasoning Struggles

User prompt

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model response

A: The answer is 27.



Outline

- LLM architecture
 - Embedding
 - Transformer blocks
 - Next word prediction
- Chain of Thought Prompting
- Chain of Thought Fine Tuning
- Conclusion

Embedding input text



Transformer block

Self-attention

Multilayer perceptron

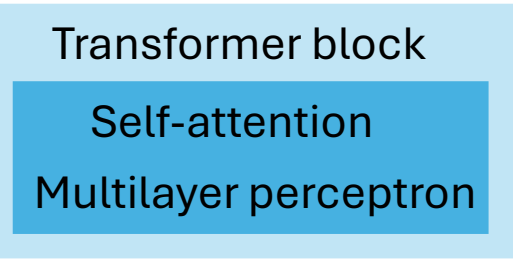


Next word prediction

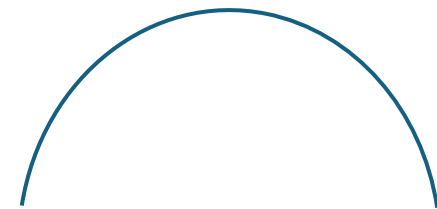
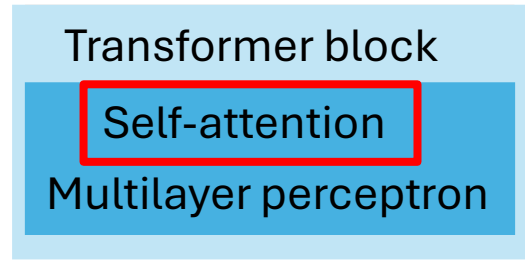
Embedding input text

	Who	was	the	lead	singer	of	Queen
Noun	1.2	-2.7	-3.3	4.3	9.4	-4.9	9.5
Verb	-3.6	8.4	-4.2	7.8	-2.5	-3.5	-9.8
Adjective	-6.2	-4.6	-3.9	3.6	-3.4	-5.2	-5.8
Royal	-0.3	-0.2	1.2	3.9	-1.7	-1.3	8.9
Musical	-0.5	-0.3	0.2	2.9	9.8	-0.6	4.7

Embedding input text



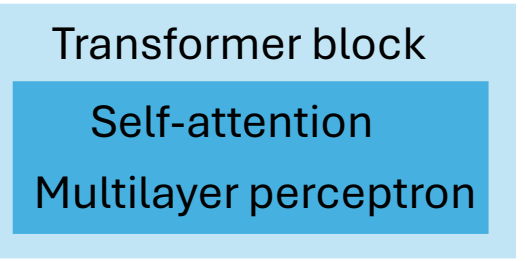
Next word prediction



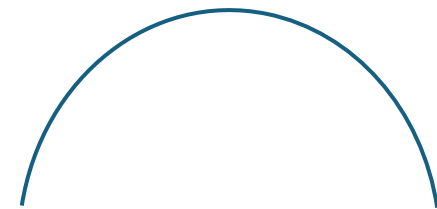
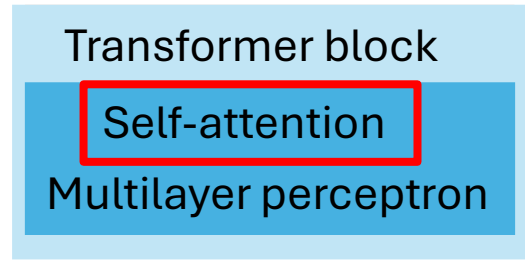
Who was the lead singer of Queen

	Who	was	the	lead	singer	of	Queen
Noun	1.2	-2.7	-3.3	4.3	9.4	-4.9	9.5
Verb	-3.6	8.4	-4.2	7.8	-2.5	-3.5	-5.8
Adjective	-6.2	-4.6	-3.9	3.6	-3.4	-5.2	-4.8
Royal	-0.3	-0.2	1.2	3.9	-1.7	-1.3	8.9
Musical	-0.5	-0.3	0.2	2.9	9.8	-0.6	4.7

Embedding input text



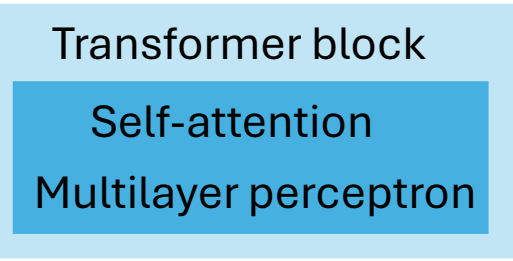
Next word prediction



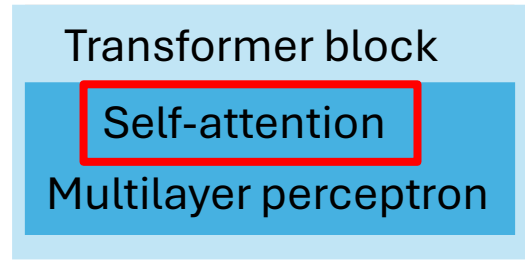
Who	was	the	lead	singer	of	Queen
------------	------------	------------	-------------	---------------	-----------	--------------

Noun	→	$\begin{bmatrix} 1.2 \\ -3.6 \\ -6.2 \\ -0.3 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} -2.7 \\ 8.4 \\ -4.6 \\ -0.2 \\ -0.3 \end{bmatrix}$	$\begin{bmatrix} -3.3 \\ -4.2 \\ -3.9 \\ 1.2 \\ 0.2 \end{bmatrix}$	$\begin{bmatrix} 4.3 \\ 7.8 \\ 3.6 \\ 3.9 \\ 2.9 \end{bmatrix}$	$\begin{bmatrix} 9.4 \\ -2.5 \\ -3.4 \\ -1.7 \\ 9.8 \end{bmatrix}$	$\begin{bmatrix} -4.9 \\ -3.5 \\ -5.2 \\ -1.3 \\ -0.6 \end{bmatrix}$	$\begin{bmatrix} 9.5 \\ -5.8 \\ -4.8 \\ 2.9 \\ 8.6 \end{bmatrix}$
------	---	---	---	--	---	--	--	---

Embedding input text



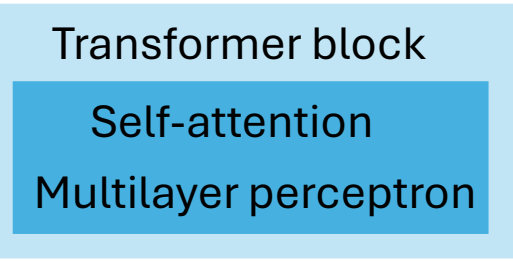
Next word prediction



Who	was	the	lead	singer	of	Queen
------------	------------	------------	-------------	---------------	-----------	--------------

Noun	→	$\begin{bmatrix} 1.2 \\ -3.6 \\ -6.2 \\ -0.3 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} -2.7 \\ 8.4 \\ -4.6 \\ -0.2 \\ -0.3 \end{bmatrix}$	$\begin{bmatrix} -3.3 \\ -4.2 \\ -3.9 \\ 1.2 \\ 0.2 \end{bmatrix}$	$\begin{bmatrix} 4.3 \\ 7.8 \\ 3.6 \\ 3.9 \\ 2.9 \end{bmatrix}$	$\begin{bmatrix} 9.4 \\ -2.5 \\ -3.4 \\ -1.7 \\ 9.8 \end{bmatrix}$	$\begin{bmatrix} -4.9 \\ -3.5 \\ -5.2 \\ -1.3 \\ -0.6 \end{bmatrix}$	$\begin{bmatrix} 9.5 \\ -5.8 \\ -4.8 \\ 2.9 \\ 8.6 \end{bmatrix}$
Verb	→							
Adjective	→							
Royal	→							
Musical	→							

Embedding input text



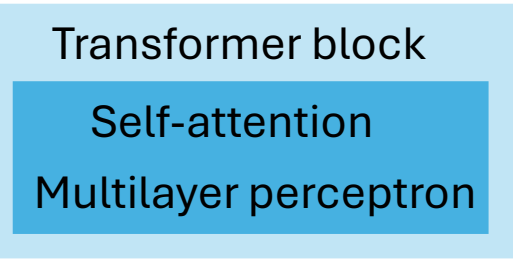
Next word prediction



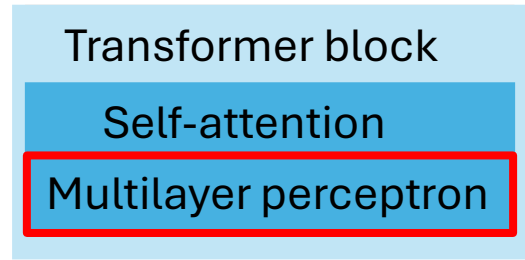
Who	was	the	lead	singer	of	Queen
------------	------------	------------	-------------	---------------	-----------	--------------

Noun	→	$\begin{bmatrix} 1.2 \\ -3.6 \\ -6.2 \\ -0.3 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} -2.7 \\ 8.4 \\ -4.6 \\ -0.2 \\ -0.3 \end{bmatrix}$	$\begin{bmatrix} -3.3 \\ -4.2 \\ -3.9 \\ 1.2 \\ 0.2 \end{bmatrix}$	$\begin{bmatrix} -4.4 \\ -6.8 \\ 8.6 \\ 3.9 \\ 2.9 \end{bmatrix}$	$\begin{bmatrix} 9.4 \\ -2.5 \\ -3.4 \\ -1.7 \\ 9.8 \end{bmatrix}$	$\begin{bmatrix} -4.9 \\ -3.5 \\ -5.2 \\ -1.3 \\ -0.6 \end{bmatrix}$	$\begin{bmatrix} 9.5 \\ -5.8 \\ -4.8 \\ 2.9 \\ 8.6 \end{bmatrix}$
Verb	→							
Adjective	→							
Royal	→							
Musical	→							

Embedding input text



Next word prediction



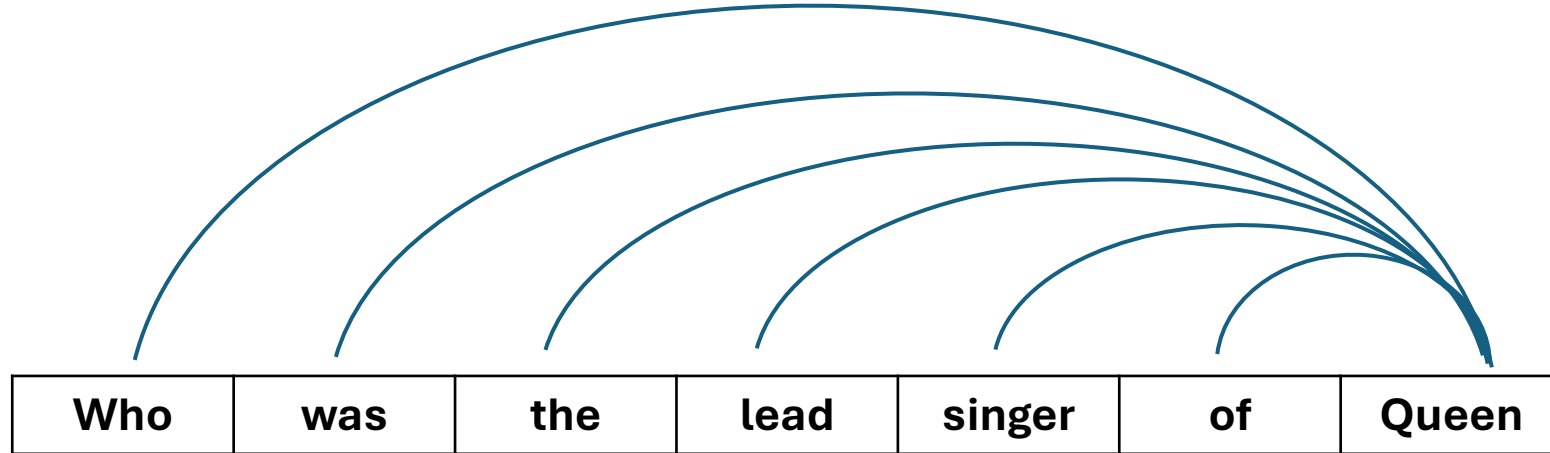
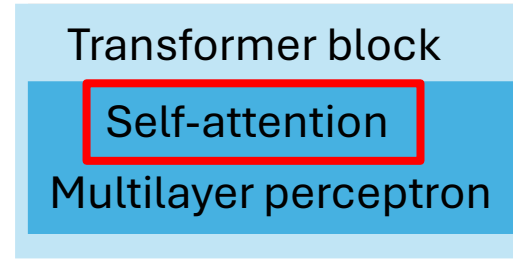
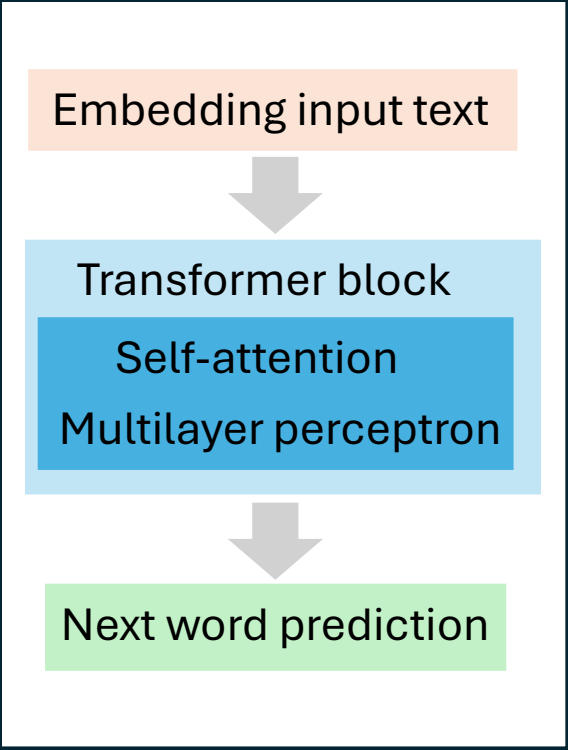
Queen

Noun	→	9.5
Verb	→	-5.8
Adjective	→	-4.8
Royal	→	2.9
Musical	→	8.6

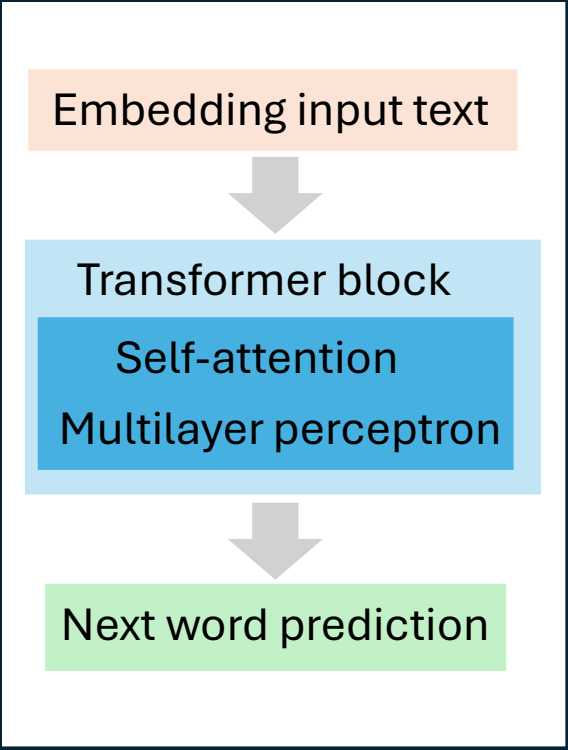


Queen

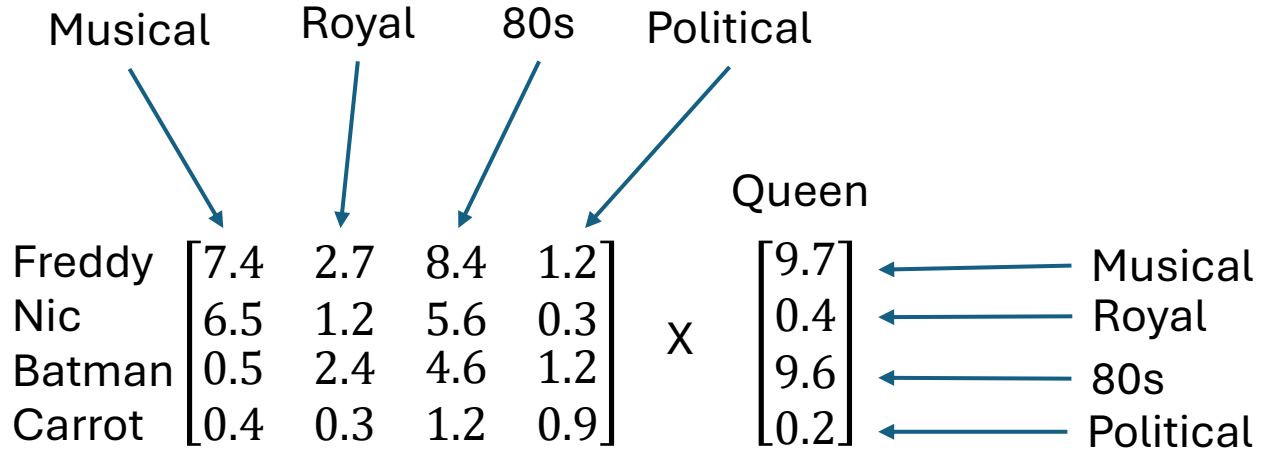
Noun	→	9.5
Verb	→	-5.8
Adjective	→	-4.8
Royal	→	2.9
Musical	→	8.6
80s	→	7.8
British	→	8.1
Feminine	→	-3.7
Political	→	-4.3

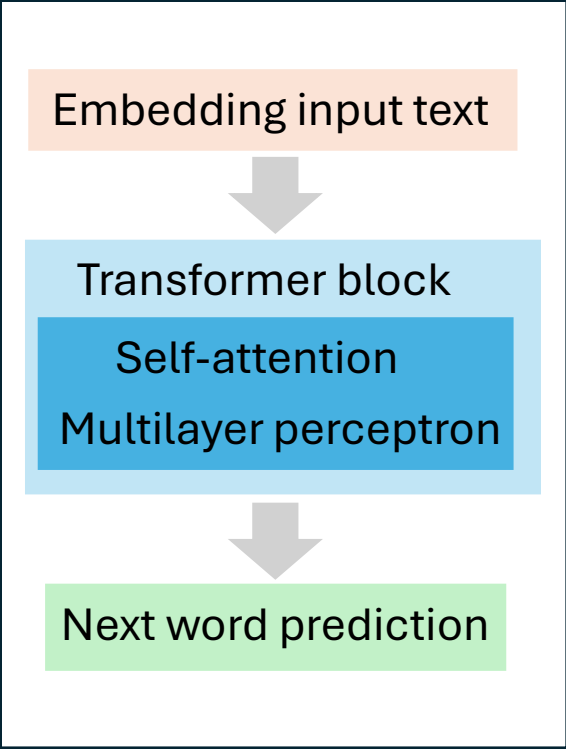


$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$

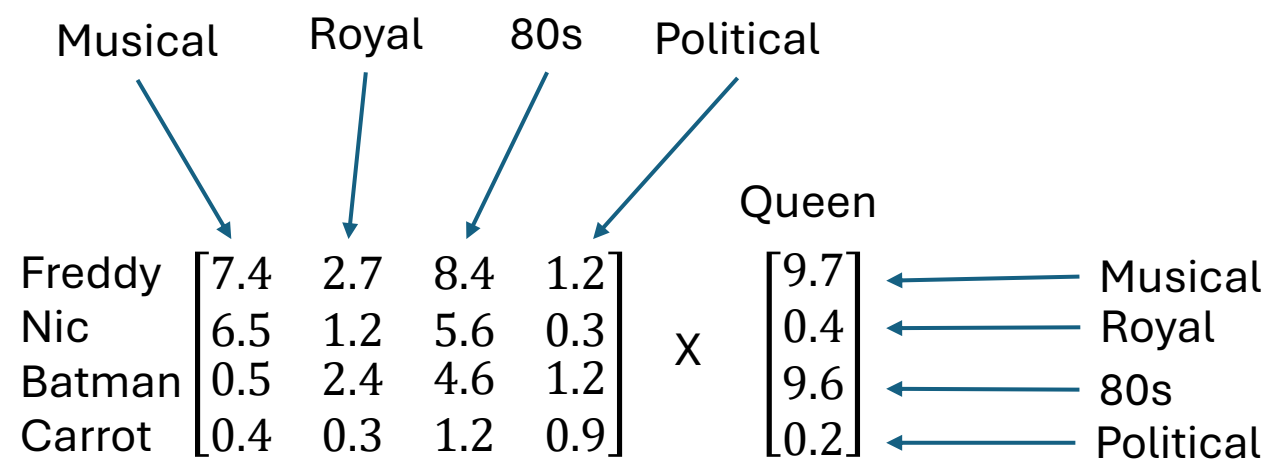


Next word prediction



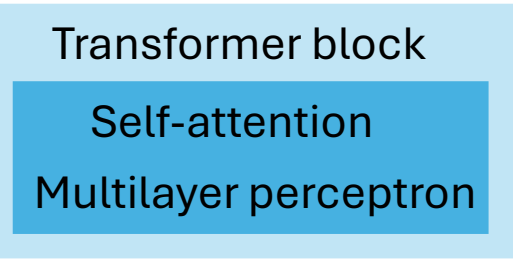


Next word prediction



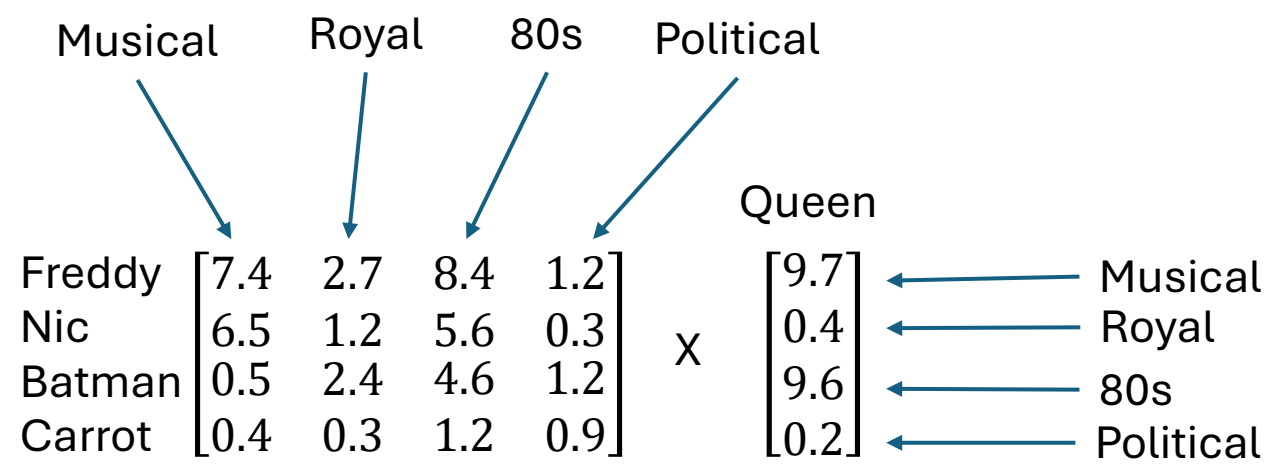
$$\text{Freddy} \left[(7.4 * 9.7) + (2.7 * 0.4) + (8.4 * 9.6) + (1.2 * 0.2) \right] = \left[153.74 \right]$$

Embedding input text

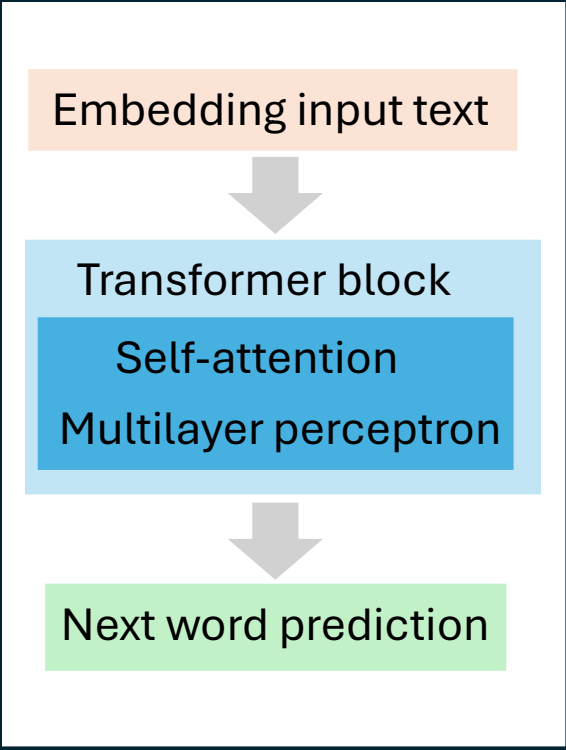


Next word prediction

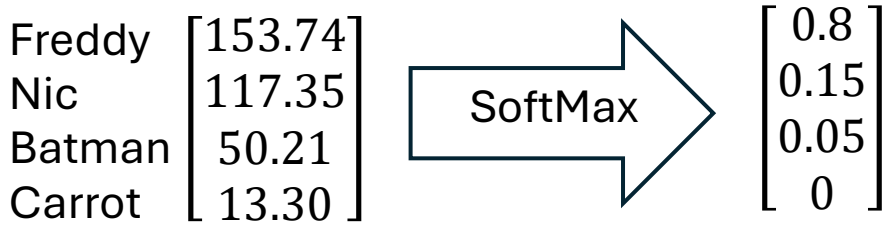
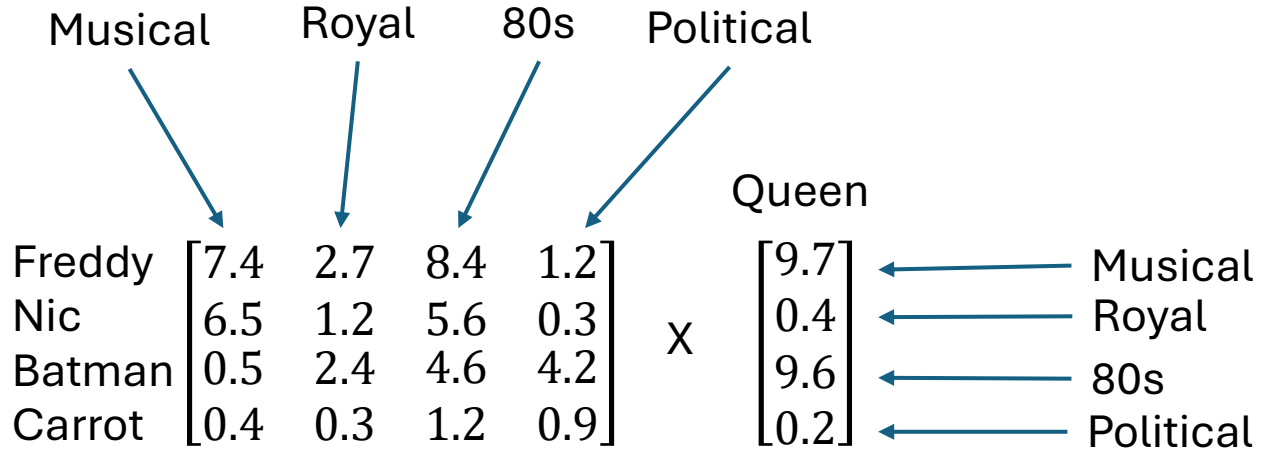
Next word prediction



$$\begin{array}{l}
 \text{Freddy} \\
 \text{Nic} \\
 \text{Batman} \\
 \text{Carrot}
 \end{array}
 \begin{pmatrix}
 (7.4 * 9.7) + (2.7 * 0.4) + (8.4 * 9.6) + (1.2 * 0.2) \\
 (6.5 * 9.7) + (1.2 * 0.4) + (5.6 * 9.6) + (0.3 * 0.2) \\
 (0.5 * 9.7) + (2.4 * 0.4) + (4.6 * 9.6) + (1.2 * 0.2) \\
 (0.4 * 9.7) + (0.3 * 0.4) + (1.2 * 9.6) + (0.9 * 0.2)
 \end{pmatrix}
 =
 \begin{pmatrix}
 153.74 \\
 117.35 \\
 50.21 \\
 13.3
 \end{pmatrix}$$



Next word prediction



Outline

- LLM architecture
 - Embedding
 - Transformer blocks
 - Next word prediction
- **Chain of Thought Prompting**
- Chain of Thought Fine Tuning
- Conclusion

Reasoning Struggles

User prompt

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model response

A: The answer is 27.



Reasoning Solution – Chain of Thought (CoT)

User prompt with CoT example

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model response

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9.



Why does CoT prompting work?

- Scientific theory: $_(_)_/_$
- Plausible hypothesis: System 1 vs system 2 thinking (Daniel Kahneman)

Why does CoT prompting work?

- Scientific theory: $_(_)_/_$
- Plausible hypothesis: System 1 vs system 2 thinking (Daniel Kahneman)

System 1 question:

What year were you born?

Why does CoT prompting work?

- Scientific theory: $_(_)_/_$
- Plausible hypothesis: System 1 vs system 2 thinking (Daniel Kahneman)

System 1 question:

What year were you born?

System 2 question (use system 1):

The cafeteria had 22 apples. If they used 10 to make lunch and bought 4 more, how many apples do they have?

Why does CoT prompting work?

- Scientific theory: $_(_)_/_$
- Plausible hypothesis: System 1 vs system 2 thinking (Daniel Kahneman)
 - LLMs usually try to solve system 2 problems with system 1 thinking
 - CoT prompting helps models break system 2 problems down into smaller system 1 problems which they are good at solving.

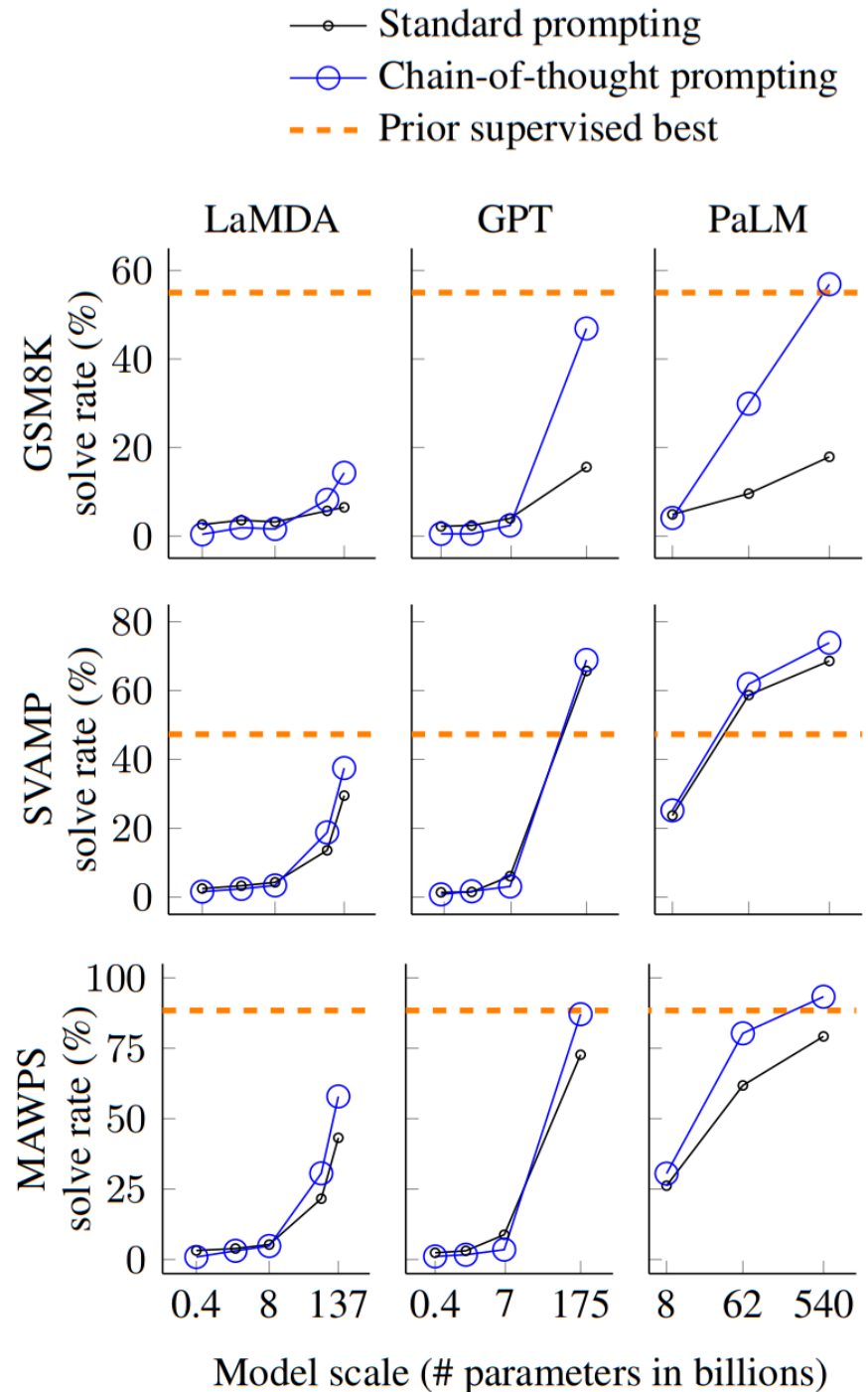
System 1 question:

What year were you born?

System 2 question (use system 1):

The cafeteria had 22 apples. If they used 10 to make lunch and bought 4 more, how many apples do they have?

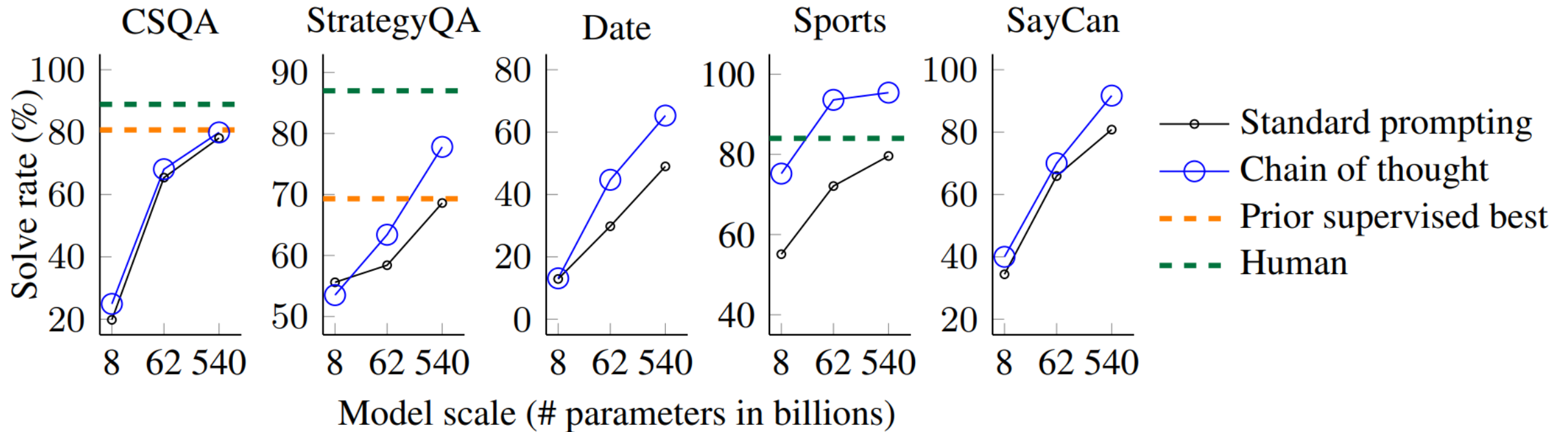
Experiments on CoT Prompting Arithmetic Reasoning



Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems* (2022).

Experiments on CoT Prompting

Common Sense Reasoning



Larger Models → Better Reasoning

- This could be due to the fact that models tend to store low resolution linguistic information such as syntax in earlier layers and only begin storing high level information like reasoning patterns in the later layers.
- However, large models require vast amounts of data and energy to train, both of which are costly.

Is there a way to improve reasoning in smaller models?

Outline

- LLM architecture
 - Embedding
 - Transformer blocks
 - Next word prediction
- Chain of Thought Prompting
- Chain of Thought Fine Tuning
- Conclusion

Fine Tuning models on CoT

- Fine tuning is essentially a continuation of the training process but with more specific data.
 - Python
 - Shakespeare
 - CoT reasoning

Fu, et al. “Specializing Smaller Language Models towards Multi-Step Reasoning.” *Proceedings of the 40th International Conference on Machine Learning (2023)*.

Methodology

Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?

Desired model output:
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

Methodology

Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?

Desired model output:
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

Desired next output: 48

Methodology

Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May? 48

Desired model output:
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

Desired next output: /

Methodology

Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May? 48/

Desired model output:
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

Desired next output: 2

Methodology

Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May? $48/2$

Desired model output:
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

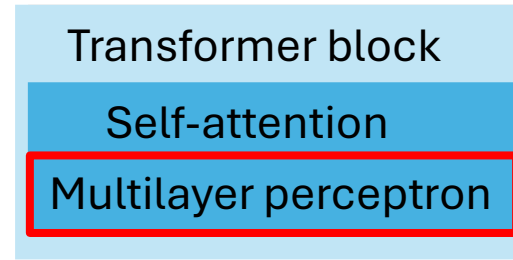
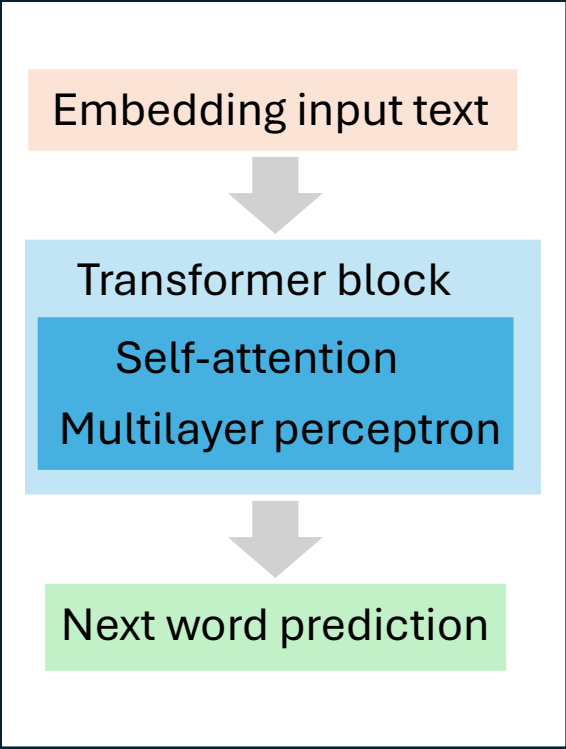
Desired next output: =

Methodology

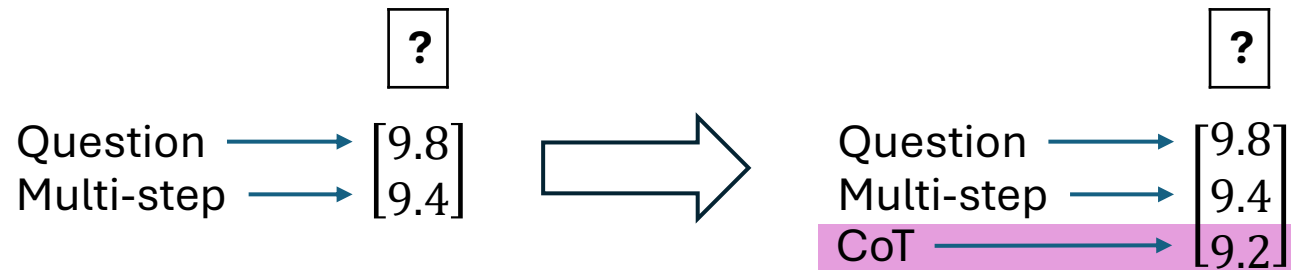
Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May? $48/2 =$

Desired model output :
 $48/2 = 24$, $48+24=72$. Natalia sold 72 clips altogether.

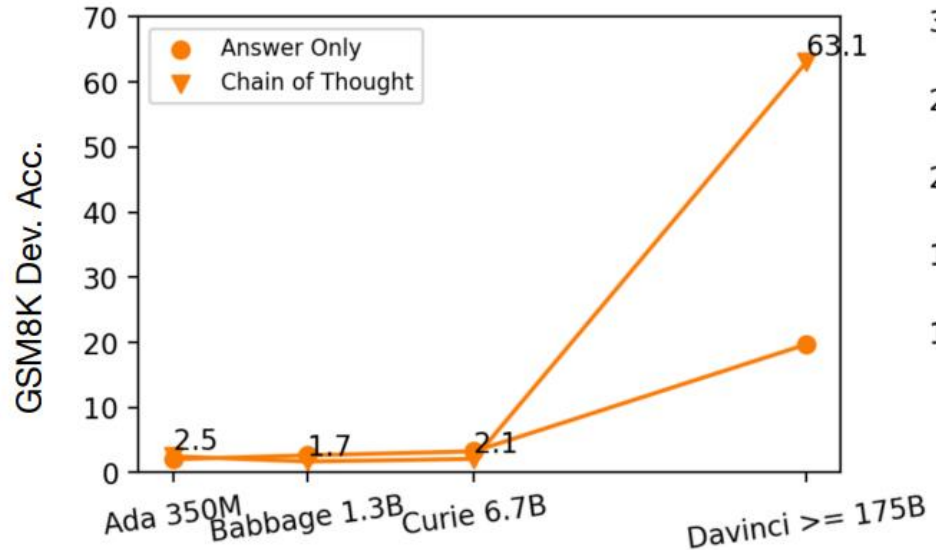
Desired next output: 24



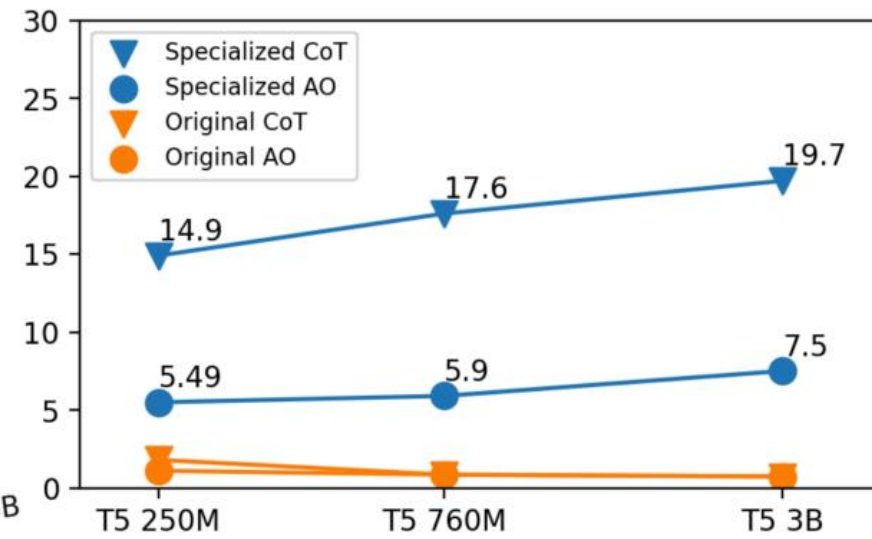
Prompt: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?



Math Results



A. GPT phase change curve, almost flat in small scale



B. Specialized T5 exhibit log-linear scaling curve

Downsides to CoT Fine Tuning

In a golf tournament, there were seven golfers: Ana, Eve, Ada, Dan, Rob, Amy, and Joe. Dan finished third. Ana finished above Ada. Amy finished last. Dan finished below Rob. Eve finished below Ada. Rob finished below Joe.

- (A) Ana finished third
- (B) Eve finished third
- (C) Ada finished third
- (D) Dan finished third
- (E) Rob finished third
- (F) Amy finished third
- (G) Joe finished third

Downsides to CoT Fine Tuning

In a golf tournament, there were seven golfers: Ana, Eve, Ada, Dan, Rob, Amy, and Joe. Dan finished third. Ana finished above Ada. Amy finished last. Dan finished below Rob. Eve finished below Ada. Rob finished below Joe.

- (A) Ana finished third
- (B) Eve finished third
- (C) Ada finished third
- (D) Dan finished third
- (E) Rob finished third
- (F) Amy finished third
- (G) Joe finished third

Performance dropped to near 0% on common sense questions like the one above.

Outline

- LLM architecture
 - Embedding
 - Transformer blocks
 - Next word prediction
- Chain of Thought Prompting
- Chain of Thought Fine Tuning
- Conclusion

Conclusion

- LLMs struggle with complex reasoning due to their architecture which only allows them to predict the next word given some input.
- CoT prompting allows them to break complex problems down into smaller chunks which they are better at solving.
- CoT fine tuning unlocks this ability at smaller scales.

References

- Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems* (2022).
- Fu, Yao, et al. "Specializing Smaller Language Models towards Multi-Step Reasoning." *Proceedings of the 40th International Conference on Machine Learning* (2023).
- Sanderson, Grant. "Neural Networks" series. *YouTube*, uploaded by 3Blue1Brown, (2024)
https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi.
- OpenAI. [ChatGPT-4o] (November 2024) <https://chat.openai.com>

Acknowledgements

Thank you to Nic McPhee and Elena Machkasova for advising me on this project.

Thanks you as well to Peter Dolan and Mark Collier for fostering my interest in AI.

Any questions?

- LLM architecture
 - Embedding
 - Transformer blocks
 - Next word prediction
- Chain of Thought Prompting
- Chain of Thought Fine Tuning
- Conclusion

Slides after this are garbage but I'm still emotionally attached to them

Further Optimizations: Reinforcement Learning

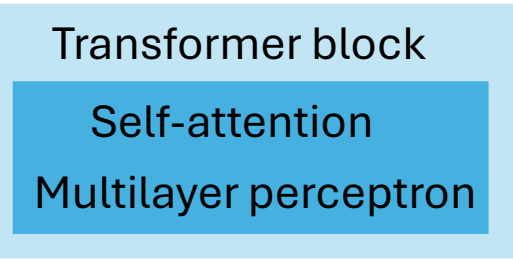
- LLM generates CoT solutions and a “Process Reward Model” rewards the LLM for steps that are more likely to lead to correct answers.
- Differs from fine tuning in that it does not use labeled data to calculate reward/punishment but instead uses an outside source.

Model	Base	Params	GSM8k	MATH
Proprietary models				
GPT-o1 (OpenAI, 2023)	-	-	-	94.8
GPT-o1-mini	-	-	-	90.0
Gemini-1.5 002	-	-	-	86.5
Claude 3.5 Sonnet (Bai et al., 2022)	-	-	96.4	71.1
GPT-4o-2024-0513	-	-	96.1	76.6
Mistral-v0.1 (Jiang et al., 2023)	-	7B	42.9	12.9
WizardMath-Mathstral	Mathstral-v0.1	7B	93.8	70.9
GPT-2-Small (Brown et al., 2020c)	-	0.1B	6.9	5.4
GPT-2-Medium (Brown et al., 2020c)	-	0.3B	11.2	6.2
GPT-2-Large (Brown et al., 2020c)	-	0.7B	13.6	6.4
GPT-2-XL (Brown et al., 2020c)	-	1.5B	15.4	6.9
WizardMath-GPT	GPT-2-Small	0.1B	26.4	12.3
WizardMath-GPT	GPT-2-Medium	0.3B	38.7	15.6
WizardMath-GPT	GPT-2-Large	0.7B	50.1	21.2
WizardMath-GPT	GPT-2-XL	1.5B	58.9	25.4

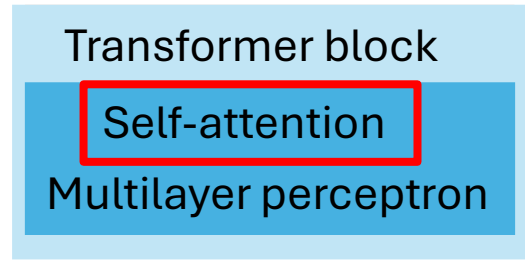
Further Optimization: Reinforcement Learning

- Studies are not directly comparable as researchers begin with different base-LLMs
- However, one average fine tuning increases math performance by factor of less than 2. Reinforcement learning increased math performance on average by a factor of greater than 2 and for some models, by a factor of 5.

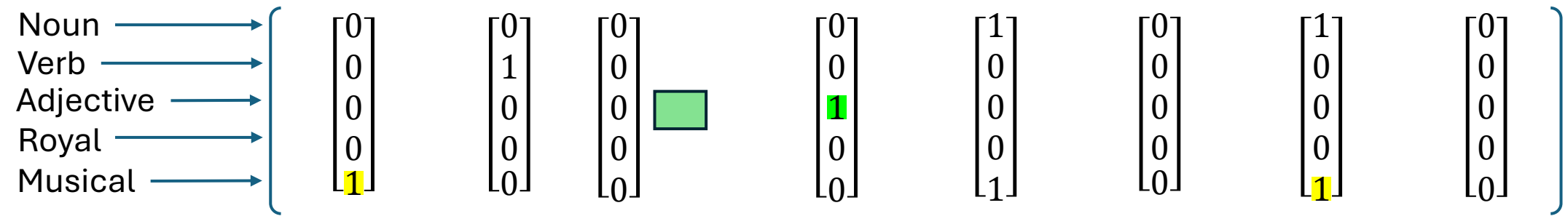
Embedding input text



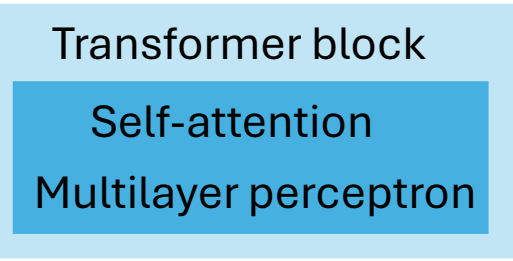
Next word prediction



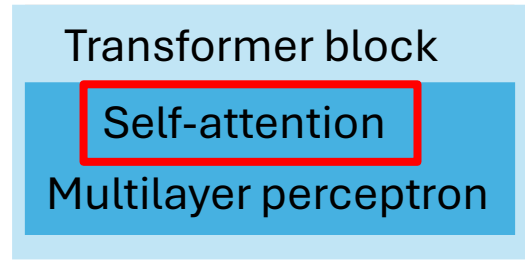
Who	was	the	lead	singer	of	Queen	?
-----	-----	-----	------	--------	----	-------	---



Embedding input text



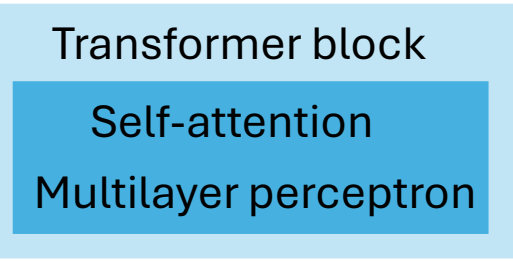
Next word prediction



Who	was	the	lead	singer	of	Queen
-----	-----	-----	------	--------	----	-------

Noun	→	[0]	[0]	[0]	[0]	[1]	[0]	[1]
Verb	→	[0]	[1]	[0]	[0]	[0]	[0]	[0]
Adjective	→	[0]	[0]	[0]	[1]	[0]	[0]	[0]
Royal	→	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Musical	→	[1]	[0]	[0]	[0]	[1]	[0]	[1]
80s	→	[1]	[0]	[0]	[0]	[1]	[0]	[0]
British	→	[1]	[0]	[0]	[0]	[1]	[0]	[0]
Feminine	→	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Political	→	[0]	[0]	[0]	[0]	[0]	[0]	[0]

Embedding input text

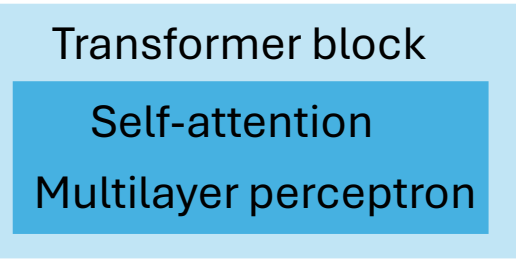


Next word prediction

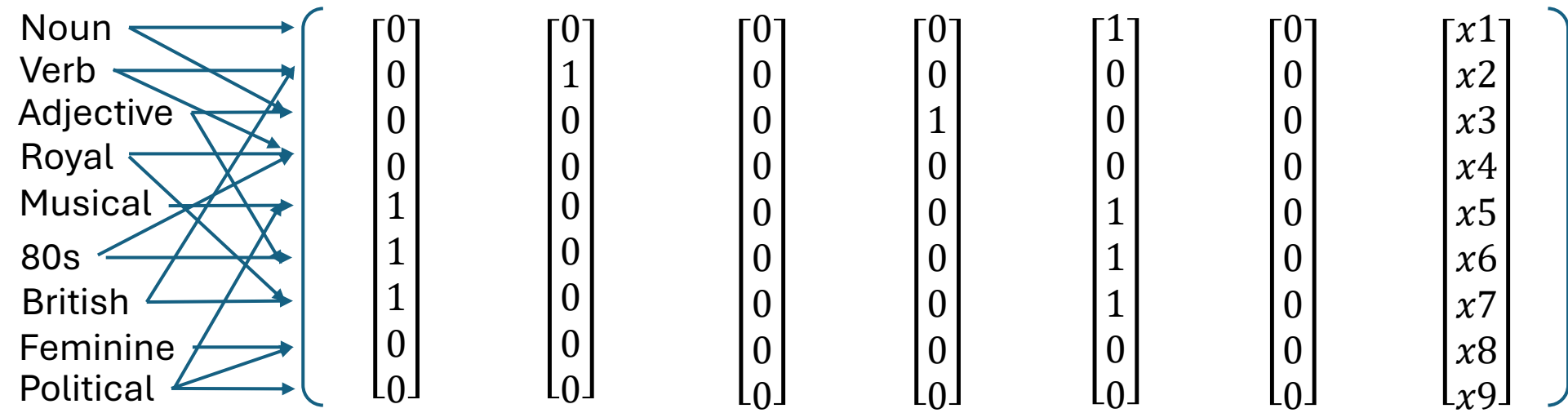
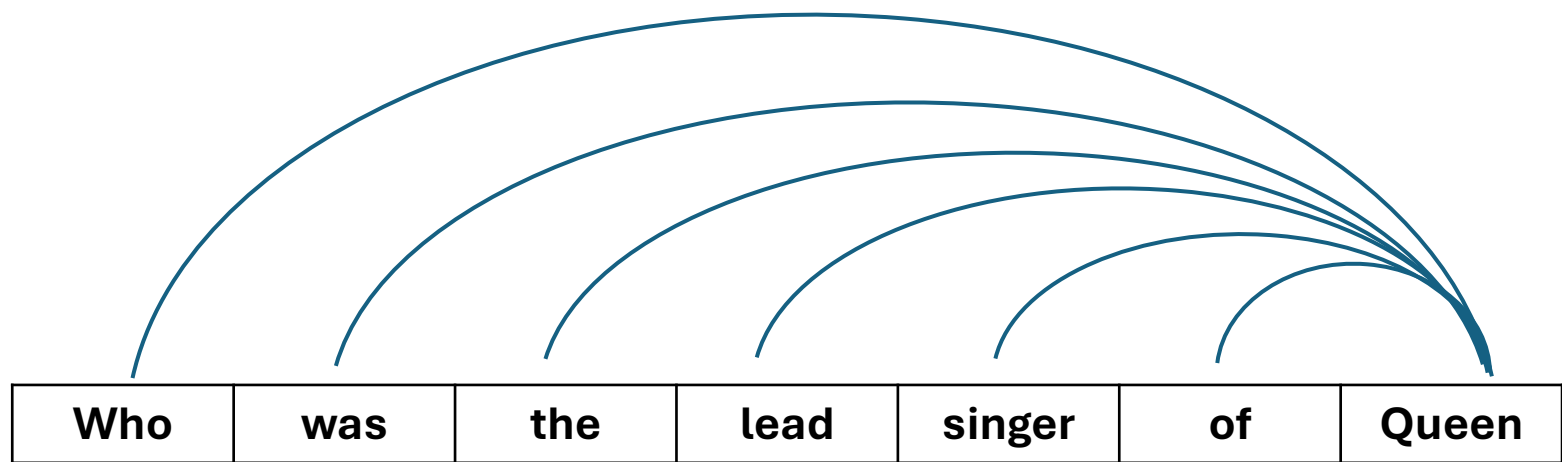
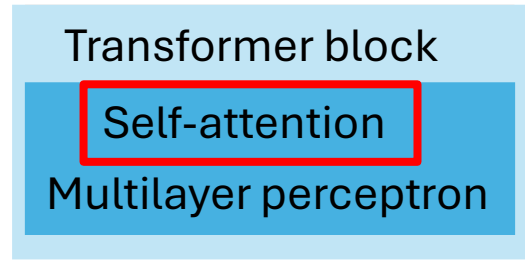
Next word prediction

$$\begin{matrix}
 & \boxed{\text{Vocabulary}} & & \boxed{?} & & \\
 50,000 \text{ rows } \} & \begin{bmatrix} 6.6 & \dots & 9.2 \\ \vdots & \ddots & \vdots \\ 5.1 & \dots & 5.9 \end{bmatrix} & \times & \begin{bmatrix} 8.3 \\ \vdots \\ 4.4 \end{bmatrix} & \{768 \text{ rows} & = & \begin{bmatrix} 2.2 \\ \vdots \\ 9.8 \end{bmatrix} \{50,000 \text{ rows} \\
 & \underbrace{\hspace{1.5cm}} & & & & & & \\
 & 768 \text{ columns} & & & & & &
 \end{matrix}$$

Embedding input text

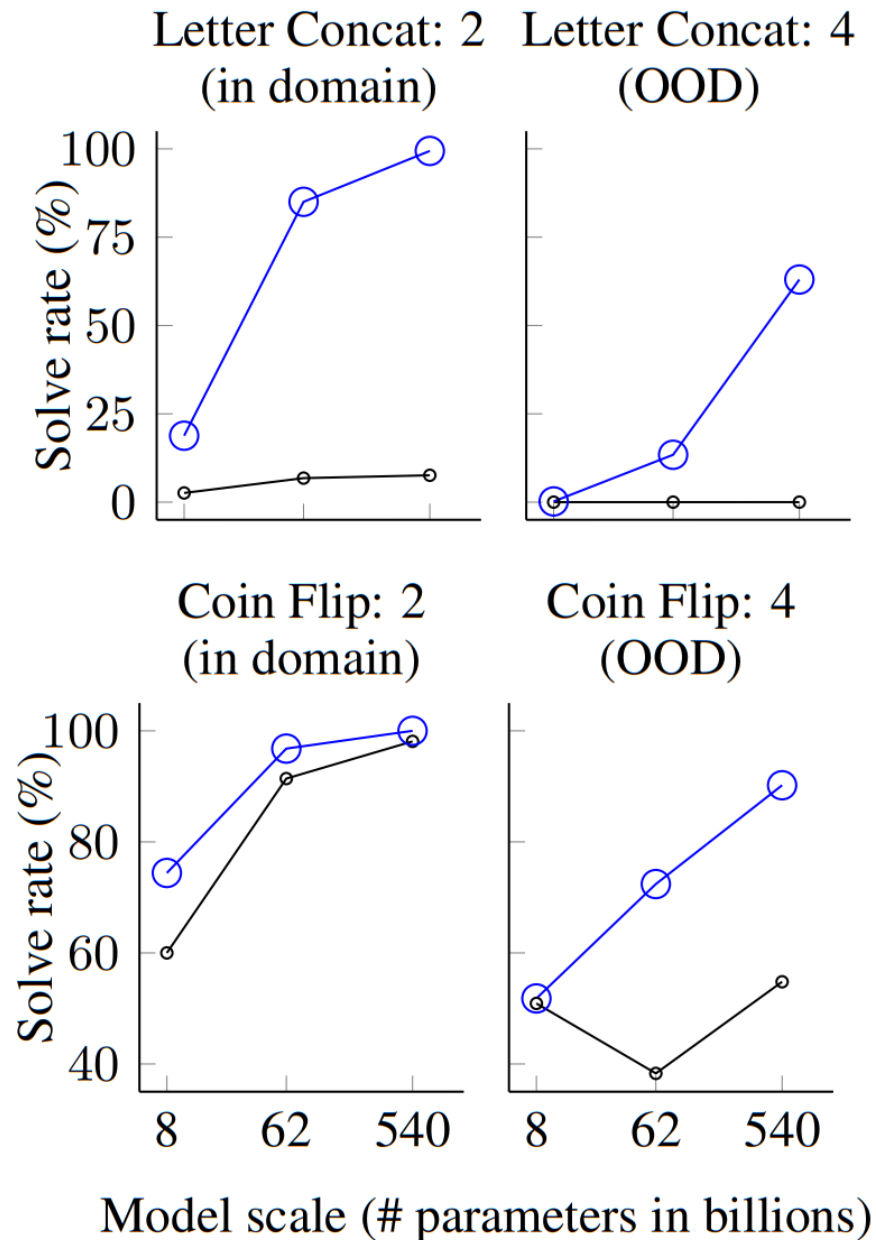


Next word prediction



Experiments on CoT Prompting Symbolic Reasoning

- Standard prompting
- Chain-of-thought prompting



Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems* (2022).