


































CoT &

-- “ ”














 o1






Chain of Thought
o1

1. 
2. 
 1. 
 2. 
 3. 
3. 
 1. 
4.  “ ” “ ” AGI
5.  **self attention** “ ”
6.  RL Test Time Scaling law Scaling Law
 RL Scaling Law Scaling Law
 S S Test Time
 S
7.  “ ”



1.  KV cache
2.  kv cache
3. 
4. 
 1. 
 2. 
5.  loss kv cache /



 LLM
test-time compute

LLM o3

/

- 1. CoT
 - 1. ☐ ☐ ☐ ☐
 - 2. ☐
 - 3. ☐ AI ☐
 - 4. ☐ AI ☐ token ☐
- 2. ☐
 - 1. ☐
 - 2. ☐ KV cache ☐
- 3. ☐ / ☐
 - 1. ☐
 - 2. ☐ LLM ☐
 - 3. ☐ CoT ☐
 - 1. o1-ioi ☐ c++ ☐
- 4. ☐
 - 1. ☐ CoT ☐ CoT ☐

CoT

- 1. ☐ CoT ☐ CoT ☐
- 2. ☐ CoT ☐
 - 1. ☐
 - 2. ☐ token ☐
- 3. ☐ “☐” ☐
 - 1. ☐ CoT ☐
 - 2. ☐
 - 3. ☐ “☐” + “☐” ☐

CoT

1. CoT
2. “ ” “ ”
3. “ ”



CoT

RL

https://mp.weixin.qq.com/s/rDjdgKS_qSfUwqCuoEjJsA



LLM

Scaling



- 1. [row of 10 squares]
- 2. [row of 8 squares]
- 3. [row of 20 squares]

Transformer

ReasonFlux

- 1. [row of 8 squares] 500 [row of 20 squares]
- 2. [row of 16 squares] CoT [row of 8 squares]
- 3. LLM [row of 10 squares]
- 3. [row of 4 squares] Scaling [row of 8 squares] Scaling [row of 8 squares] LLM [row of 4 squares]



- 1. [row of 4 squares] LLM [row of 16 squares]
- 2. [row of 12 squares] 500 [row of 12 squares]
- 3. [row of 12 squares] LLM [row of 20 squares]
- 4. [row of 8 squares] Scaling [row of 8 squares] Scaling [row of 20 squares]