



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

COMP 4901B

Large Language Models

Deep Reasoning Models

Junxian He

Oct 31, 2025

Recap: Self-Improving

even's former

short answer

no way to spend more compute

Test-Time Scaling

Long CoT opens a new era and paradigm shift (e.g., o1, o3, GPT-5, DeepSeek-R1)

How to Train Models to Reason Longer?

How to Train Models to Reason Longer?

- Collect Human Data?

How to Train Models to Reason Longer?

- Collect Human Data? Human does not write down the inherent thinking

How to Train Models to Reason Longer?

- Collect Human Data? Human does not write down the inherent thinking
- Synthesize/Distill Data from another model?

~~think~~
~~/think~~

inherent thinking
chain of thought

Summary

Recap: Policy Gradient

