

Open Source Reasoning Language Models

Fine-tune and Run DeepSeek, QwQ
and Other Open Reasoning Models on
Your Hardware





Learning Objectives

By the end of this course, you will understand:

- Understand the differences between normal LLMs and reasoning LLMs
- Learn how to download, use and prompt reasoning models
- Understand the challenges of reasoning models (e.g. evaluation time)
- Get insights into the training process of reasoning models
 - Finetuning (GRPO)
 - Reinforcement learning
 - Advancements by DeepSeek, Kimi & Co





Agenda

- Introduction to Reasoning Models
- Differences to “normal” LLMs
- Try Reasoning Models
 - Open Router
 - Local Computer (with/out GPU)
- Techniques behind Reasoning Models
 - Reinforcement Learning
 - GRPO
 - Advances by DeepSeek



Introduction

- About me (Christian Winkler)
 - Programming for more than 40 years
 - PhD in physics, working as a professor at a university of applied science
- About the course
 - LLMs and the technology can be intimidating
 - Difficult to keep up-to-date
 - Let's try to clarify the myths
 - Hands on experience





Introduction to Reasoning Models





Count letters

How many „r“ are in strawberry?

- Famous question
- Should be in training set
- Many models fail

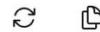
Problem

- Counting is easy, error detection is trivial
- But what happens with more complex problems?
- Idea: make models which challenge themselves
- Much better also for agentic tasks

How many "r" are in "strawberry"?



There are 2 "r"s in "strawberry".



Output of Llama-4-Scout-17B-16E-Instruct-Q4_K_M



Reasoning Approach

Expose process for insight generation

- Transparency
- Detect errors by understanding individual steps
- Get ideas to rephrase question

How many "r" are in "strawberry"?

Thought Process

The word "strawberry" is spelled **S-T-R-A-W-B-E-R-R-Y**. Breaking it down:

1. S
2. T
3. R (first "r")
4. A
5. W
6. B
7. E
8. R (second "r")
9. R (third "r")
10. Y

There are **three "r"s** in the word.

Answer: 3



Output of Qwen_Qwen3-1.7B-Q4_K_M



Reasoning Process (1)

Okay, let's see. The user is asking how many "r"s are in the word "strawberry". Hmm, first I need to make sure I spell the word correctly. "Strawberry" is spelled S-T-R-A-W-B-E-R-R-Y. Wait, let me check again. S-T-R-A-W-B-E-R-R-Y. So breaking it down letter by letter.

Starting from the beginning: S, T, R, A, W, B, E, R, R, Y. So the letters are S, T, R, A, W, B, E, R, R, Y. Now I need to count the number of "r"s. Let's go through each letter.

First letter is S, not an r. Second is T, no. Third is R, that's one. Then A, W, B, E, then R again, that's two. Then another R, making it three. So total of three r's? Wait, but wait, the word is "strawberry". Let me double-check the spelling. Sometimes people might mispronounce it, but the standard spelling is S-T-R-A-W-B-E-R-R-Y. So the letters are R, R, and then another R? Wait, no. Let me count again. Let's write it out:

S (1), T (2), R (3), A (4), W (5), B (6), E (7), R (8), R (9), Y (10). So positions 3, 8, and 9. So that's three R's. But wait, sometimes when people say "strawberry", they might say it with two R's, but the actual spelling has three? Wait, let me check another way. The word is "strawberry", which is from the French "abricot", but the spelling in English is S-T-R-A-W-B-E-R-R-Y. So yes, two R's in the middle, but the third R is at the end? Wait, no. Wait, the last letter is Y. So the letters are S-T-R-A-W-B-E-R-R-Y. So that's R, R. So two R's? Wait, but the user wrote "strawberry" with two R's? Wait, maybe I'm confused. Let me check the spelling again.



Reasoning Process (2)

Wait, "strawberry" is spelled S-T-R-A-W-B-E-R-R-Y. So the letters are S, T, R, A, W, B, E, R, R, Y. So there are two R's in the middle, but the third R is... no, wait, the third R is the second R. Wait, the letters are:

1. S
2. T
3. R
4. A
5. W
6. B
7. E
8. R
9. R
10. Y

So positions 3, 8, and 9. So three R's? Wait, but in the middle, after the "w", there's "b", "e", then "r", "r". So two R's? Wait, no, the third letter is R, then the next two are R again. So that's three R's? Wait, no, position 3 is the first R, then position 8 and 9 are the second and third R's? Wait, that's three R's. But sometimes people might think that "strawberry" has only two R's. Let me check an example.



Reasoning Process (3)

If I type "strawberry" into a search engine, it would show the spelling as S-T-R-A-W-B-E-R-R-Y. So yes, that's two R's in the middle, but the third R is actually the second one. Wait, no, the letters are S, T, R, A, W, B, E, R, R, Y. So there are two R's in the word. Wait, position 3 is the first R, then positions 8 and 9 are the second and third R's? No, position 8 is the second R, and position 9 is the third R? Wait, but the word is "strawberry" which is 10 letters. Let me count again:

1. S
2. T
3. R
4. A
5. W
6. B
7. E
8. R
9. R
10. Y



Reasoning Process (4)

So positions 3, 8, 9. That's three R's. Wait, but the word is "strawberry", which is 10 letters. Wait, but "strawberry" is spelled with two R's. Wait, maybe I'm making a mistake here. Let me think. The word "strawberry" is pronounced "straw-berry", so the spelling would be S-T-R-A-W-B-E-R-R-Y. So the R's are in the third and eighth positions, and the ninth is another R? Wait, that's three R's. But I thought that "strawberry" only has two R's. Maybe there's confusion with the word "strawberry" vs. "strawberry" with a different spelling. Wait, maybe the user is confused because they think "strawberry" has two R's, but actually it has three. Wait, but in the standard English spelling, "strawberry" is spelled with two R's. Let me check a dictionary. But since I can't access external sources, I need to rely on my memory.

In the word "strawberry", the R's are in the third and eighth positions. Wait, no, let's see:

Breaking it down:

S-T-R-A-W-B-E-R-R-Y

So the R is the third letter, then after the W, it's B, E, R, R. So that's two R's. Wait, so positions 3, 8, and 9? Wait, position 3 is the first R, then after B, E, then R, then another R. So that's three R's? Wait, but the word is S-T-R-A-W-B-E-R-R-Y. So after the first R, which is position 3, then A, W, B, E, then R, R. So positions 3, 8, 9. So that's three R's. But that would make the word have three R's. However, when I think of "strawberry", the R's are at the end, but the spelling is S-T-R-A-W-B-E-R-R-Y. So the third R is after the W, which is at position 8, and the second R is at position 9.



Reasoning Process (5)

Wait, maybe I'm miscounting. Let me write it out:

1. S
2. T
3. R
4. A
5. W
6. B
7. E
8. R
9. R
10. Y

Yes, positions 3, 8, and 9. So that's three R's. Therefore, the answer would be three. But I'm a bit confused because sometimes people might think it's two. But according to the spelling, it's three. So the answer should be three.



Reasoning Process (final answer)

The word "strawberry" is spelled **S-T-R-A-W-B-E-R-R-Y**. Breaking it down:

1. S
2. T
3. R (first "r")
4. A
5. W
6. B
7. E
8. R (second "r")
9. R (third "r")
10. Y

There are **three "r"s** in the word.

Answer: 3



How does it work?

Models are *finetuned*

- Special reasoning dataset
- Happens after SFT and RLHF
- Originally with o1 mini (preview)

Finetuning process has been heavily optimized

- Many new ideas
- Reasoning finetuning is now also possible with moderate hardware
- Datasets are available

> What is the sum of 13 and 25?

`<think>`

Okay, let's see. The question is asking for the sum of 13 and 25. Hmm, sum means them together, right? So I need to add 13 and 25. Let me think about how to do this.

First, maybe I can break it down. 13 plus 25. Let me add the ones place first. 13 has 3 ones, and 25 has 5 ones. So 3 plus 5 is 8. Then the tens place: 1 ten from 13 and 2 tens from 25. So 1 + 2 is 3. So combining those, it's 38? Wait, is that right?

...

`</think>`

To find the sum of 13 and 25, we simply add the two numbers together:

$$13 + 25 = 38$$

This can be verified by breaking down the numbers:

- Add the ones place: $3 + 5 = 8$
- Add the tens place: $1 + 2 = 3$
- Combine the results: 38

Thus, the final answer is: **38**



Why are Reasoning Models Popular?

Model	Organization	Global	Reasoning	Coding	Agentic Coding	Mathematics	Data Analysis	Language	IF Average
		Average	Average	Average	Average	Average	Average	Average	Average
1 Claude 4.6 Opus Thinking High Effort	Anthropic	76.33	88.67	78.18	61.67	89.32	69.89	83.27	63.31
2 Claude 4.5 Opus Thinking High Effort	Anthropic	75.96	80.09	79.65	63.33	90.39	74.44	81.26	62.55
3 Claude 4.6 Sonnet Thinking Medium Effort	Anthropic	75.47	84.77	79.27	60	86.99	77.95	76.1	63.22
4 GPT-5.2 High	OpenAI	74.84	83.21	76.07	51.67	93.17	78.16	79.81	61.77
5 GPT-5.2 Codex	OpenAI	74.3	77.71	83.62	51.67	88.77	78.2	73.68	66.45
6 GPT-5.1 Codex Max High	OpenAI	73.98	83.65	80.68	53.33	83.22	70.12	76.48	70.38
7 Gemini 3 Pro Preview High	Google	73.39	77.42	74.6	55	81.84	74.39	84.62	65.85
8 Gemini 3 Flash Preview High	Google	72.4	74.55	73.9	40	84.17	74.77	84.56	74.86
9 GPT-5.1 High	OpenAI	72.04	78.79	72.49	53.33	86.9	69.61	79.26	63.9
10 GPT-5 Pro	OpenAI	70.48	81.69	72.11	51.67	86.17	57.04	80.69	63.96
11 Kimi K2.5 Thinking	Moonshot AI	69.07	75.96	77.86	48.33	84.87	61.36	77.67	57.41
12 GLM 5	Z.AI	68.85	69.11	73.64	55	83.46	67.9	77.53	55.33
13 GPT-5.1 Codex	OpenAI	68.61	81.98	71.78	53.33	79.58	60.75	69.48	63.39
14 Claude Sonnet 4.5 Thinking	Anthropic	68.19	77.59	80.36	53.33	79.31	56.97	76.45	53.35



Why are Reasoning Models Popular?

Rank	Rank Spread	Model	Organization	Score	Votes
1	1-3	claude-opus-4-6	Anthropic	1505±8	6212
2	1-3	claude-opus-4-6-thinking	Anthropic	1504±8	5350
3	1-3	gemini-3.1-pro-preview	Google	1500±9	4072
4	4-5	gemini-3-pro	Google	1486±4	37088
5	4-12	dola-seed-2.0-preview	Bytedance	1475±10	3379
6	5-9	gemini-3-flash	Google	1474±5	27875
7	5-9	grok-4.1-thinking	xAI	1474±4	36233
8	5-10	claude-opus-4-5-20251101-thinking-32k	Anthropic	1471±4	29183
9	5-12	claude-opus-4-5-20251101	Anthropic	1467±4	34111
10	8-16	grok-4.1	xAI	1463±4	40453
11	8-16	gemini-3-flash (thinking-minimal)	Google	1462±5	19226
12	10-19	gpt-5.1-high	OpenAI	1458±4	33100
13	7-25	claude-sonnet-4-6	Anthropic	1457±10	3255
14	10-25	glm-5	Z.ai	1455±8	5319



Why are Reasoning Models Popular?

Rank	Rank Spread	Model	Organization	Score	Votes
1	1-2	claude-opus-4-6	Anthropic	1561+14/-14	2524
2	1-3	claude-opus-4-6-thinking	Anthropic	1551+16/-16	1919
3	2-4	claude-sonnet-4-6	Anthropic	1524+20/-20	1021
4	3-4	claude-opus-4-5-20251101-thinking-32k	Anthropic	1501+8/-8	10620
5	5-9	gpt-5.2-high	OpenAI	1471+16/-16	1695
6	5-8	claude-opus-4-5-20251101	Anthropic	1469+8/-8	10730
7	5-13	gemini-3.1-pro-preview	Google	1461+15/-15	1831
8	5-13	glm-5	Z.ai	1456+14/-14	2243
9	7-14	gemini-3-pro	Google	1444+7/-7	16673
10	6-14	minimax-m2.5	MiniMax	1443+12/-12	3260
11	7-14	glm-4.7	Z.ai	1440+10/-10	5128
12	7-14	gemini-3-flash	Google	1440+8/-8	12350



Qwen/Qwen3.5-397B-A17B

Image-Text-to-Text • Updated 1 day ago • ↓ 133k • ⚡ • ❤ 813

MinimaxAI/Minimax-M2.5

Text Generation • Updated 5 days ago • ↓ 173k • ⚡ • ❤ 823

Nanbeige/Nanbeige4.1-3B

Text Generation • ∵ 4B • Updated 3 days ago • ↓ 130k • ⚡ • ❤ 640

Qwen/Qwen3-TTS-12Hz-1.7B-CustomVoice

Text-to-Speech • Updated 23 days ago • ↓ 933k • ❤ 1.12k

nineninesix/kani-tts-2-en

Text-to-Speech • ∵ 0.4B • Updated 2 days ago • ↓ 2.59k • ❤ 163

moonshotai/Kimi-K2.5

Image-Text-to-Text • Updated 17 days ago • ↓ 1.07M • ⚡ • ❤ 2.06k

jopendata/JoyAI-LLM-Flash

Text Generation • ∵ 49B • Updated 3 days ago • ↓ 807 • ❤ 131

deepgenteam/DeepGen-1.0

Text-to-Image • Updated about 11 hours ago • ↓ 24 • ❤ 128

nvidia/personaplex-7b-v1

Audio-to-Audio • Updated 6 days ago • ↓ 539k • ❤ 2.1k

zai-org/GLM-5

Text Generation • ∵ 754B • Updated 8 days ago • ↓ 177k • ⚡ • ❤ 1.4k

FireRedTeam/FireRed-Image-Edit-1.0

Image-to-Image • Updated 7 days ago • ↓ 2.15k • ⚡ • ❤ 228

unsloth/Qwen3.5-397B-A17B-GGUF

Image-Text-to-Text • ∵ 396B • Updated 5 days ago • ↓ 73.4k • ❤ 168

xgen-universe/Capybara

Any-to-Any • Updated 1 day ago • ❤ 144

OpenMOSS-Team/MOSS-TTS

Text-to-Speech • ∵ 8B • Updated 8 days ago • ↓ 41.2k • ❤ 290

unsloth/Minimax-M2.5-GGUF

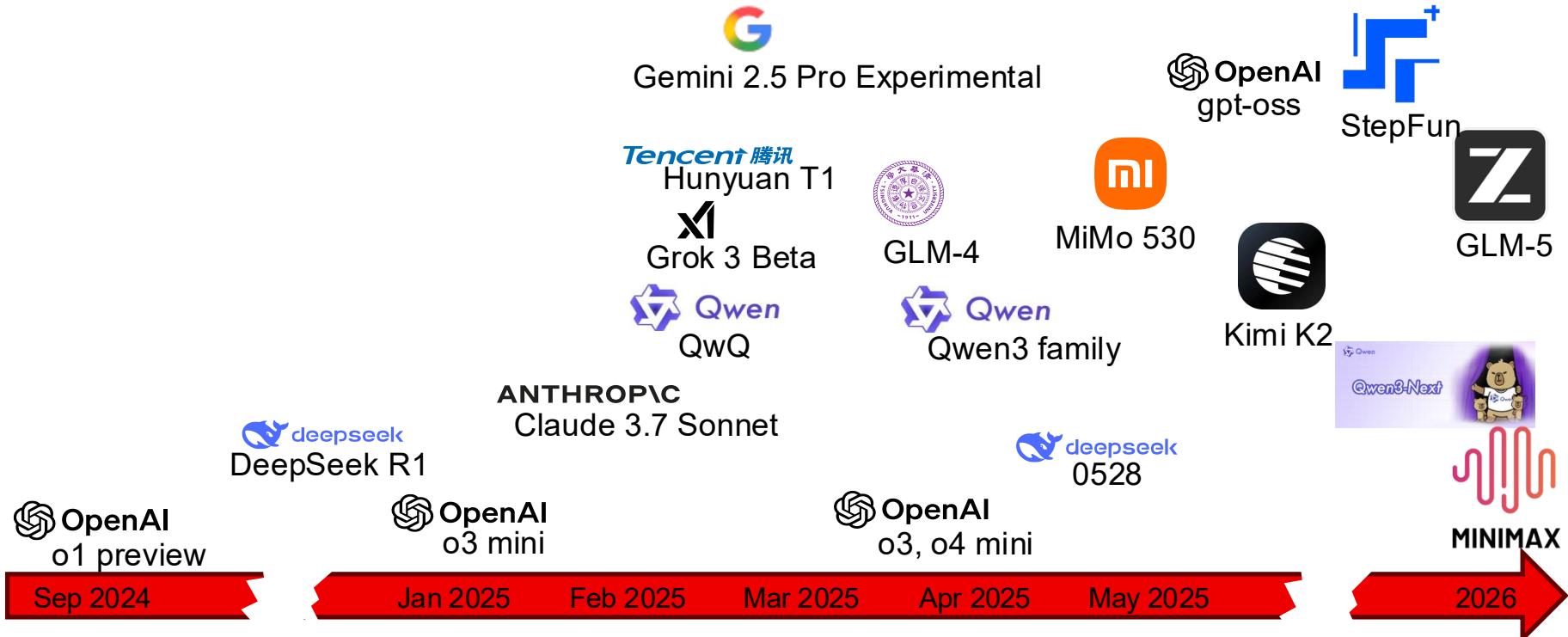
Text Generation • ∵ 229B • Updated 7 days ago • ↓ 64.5k • ❤ 137

Qwen/Qwen3-Coder-Next

Text Generation • Updated 18 days ago • ↓ 415k • ⚡ • ❤ 938



Booming Reasoning Models





Sudden Interest?

First reasoning model from OpenAI (o1 preview)

- Expensive and an experiment (preview)

Booster: Publication of DeepSeek R1

- Big surprise – big panic
- New ideas, *everything Open Source* (except for training data!)

Very surprising?

- Not really, DeepSeek Coder, DeepSeek V3 were there before

AI knowledge coming from nothing?

- No, AI was used for Hedge Fund (High-Flyer)
- Earned lots of money

Strategic AI competences for China





Q&A



Try Reasoning Models



Different Alternatives

Local Installation with GPU

- Hugging Face [transformers](#)
- [llama.cpp](#)

Local Installation without GPU

- [llama.cpp](#)
- [mlx](#) (Apple Silicone)

Service Provided by an API

- [DeepSeek](#)
- [Qwen](#)
- [Kimi](#)
- [MiniMax](#)
- [GLM](#)

Use [OpenRouter](#)

- Free services available
- Less (or no) tracking



DeepSeek-like models

- DeepSeek
 - Too large
 - Distills are possible
- Kimi K2, K2.5
 - Arguably most powerful open model
 - From Moonshot.ai
 - Copied DeepSeek architecture
 - Even larger!





Qwen3 models

- Model family
 - Different sizes
 - Different architectures
- Small models
 - Suitable for finetuning
- Medium models
 - Quite powerful
 - Working on GPUs and CPUs
- Big models
 - Mainly for cloud
 - New architecture: Qwen3-Next (one of the most powerful coding models)



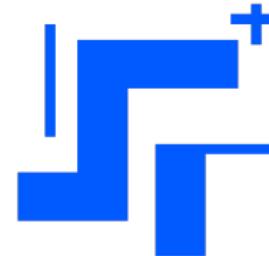


Other models

- GLM
 - Originally from Tsinghua University
 - Now from Z.ai
 - Very powerful (GLM 5), older models not too large
- MiniMax
 - Powerful
 - Small than GLM 5
- Stepfun
 - Powerful and moderately sized
- Many others, like very small Nanbeige



MINIMAX





Live Demo

Jupyter,

llama.cpp,

mlx





Q&A



Techniques behind Reasoning Models

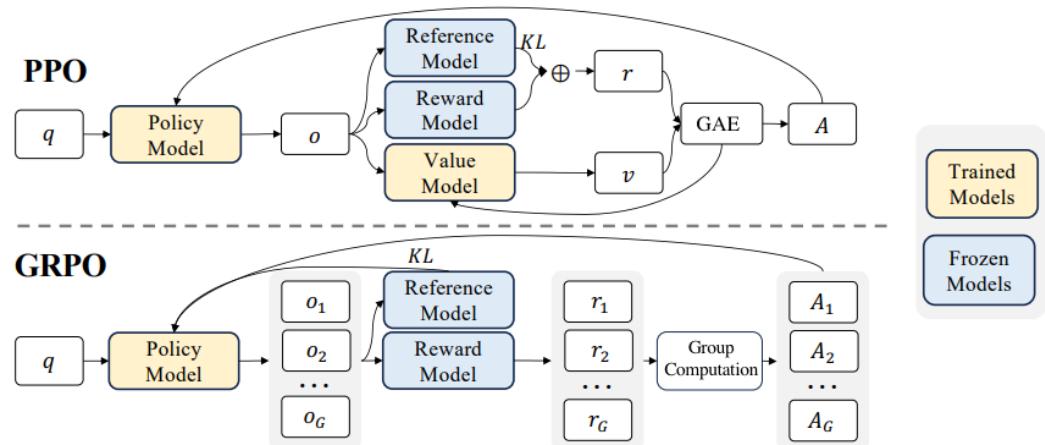
Innovations by DeepSeek (1)

Reinforcement Learning

- Already used in older models (RLHF in SFT stage)
- Now in multiple stages

New method for RL: GRPO

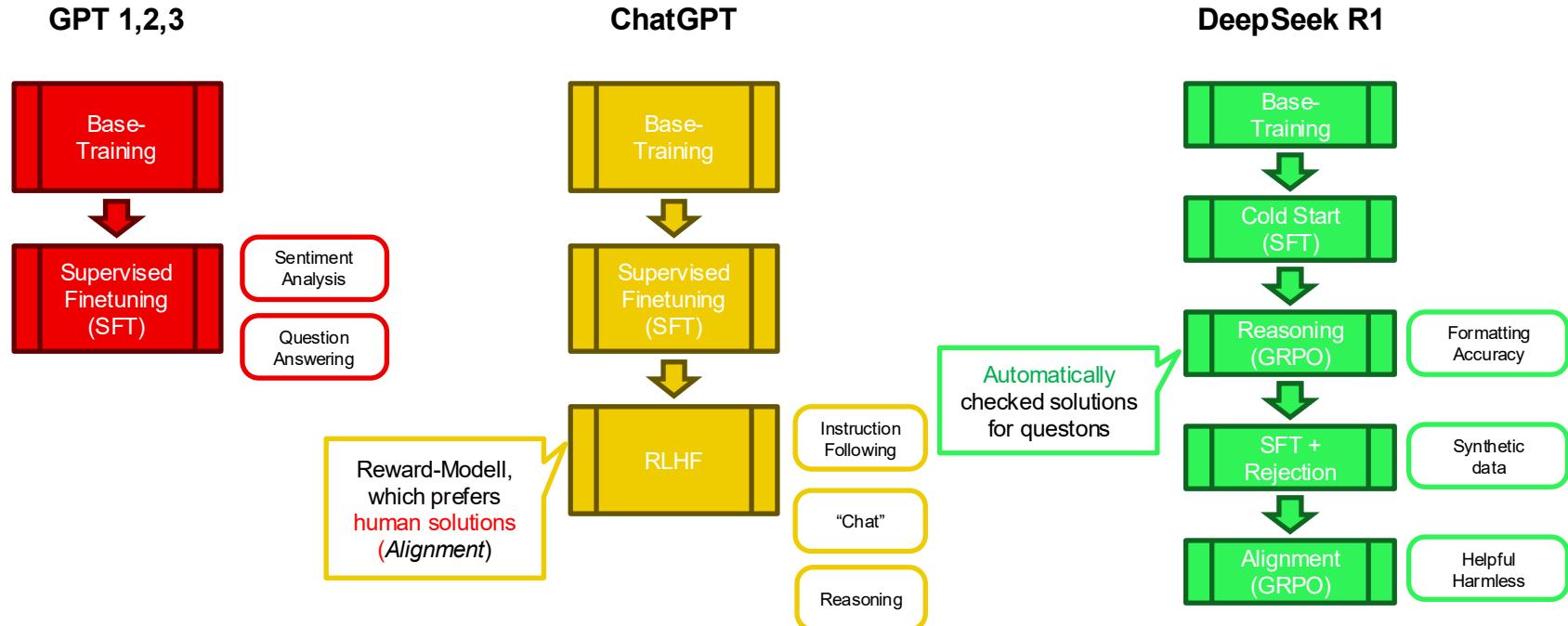
- Group Relative Policy Optimization
- Avoids value model
- Checks different alternatives
- Builds ranking



Source: <https://arxiv.org/pdf/2402.03300>



Focus: New Finetuning Process (GRPO, RL)





Innovations by DeepSeek (2)

Multi-head Latent Attention (MLA)

Multi-Token Prediction

Low-Level optimization (PTX)

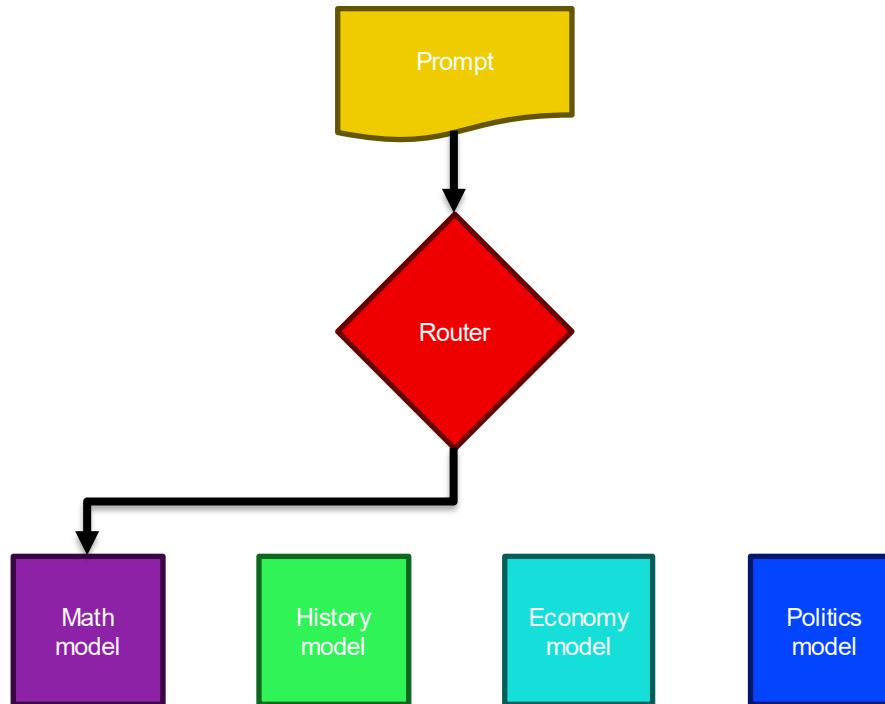
- Hardware export not allowed
- Only few(er) GPUs available
- Training more efficient and cheaper
- Much cheaper inference
(\$ per token)

Weights as FP8

- Saves 50% VRAM and doubles inference speed (bandwidth)
- More sophisticated training process (once)

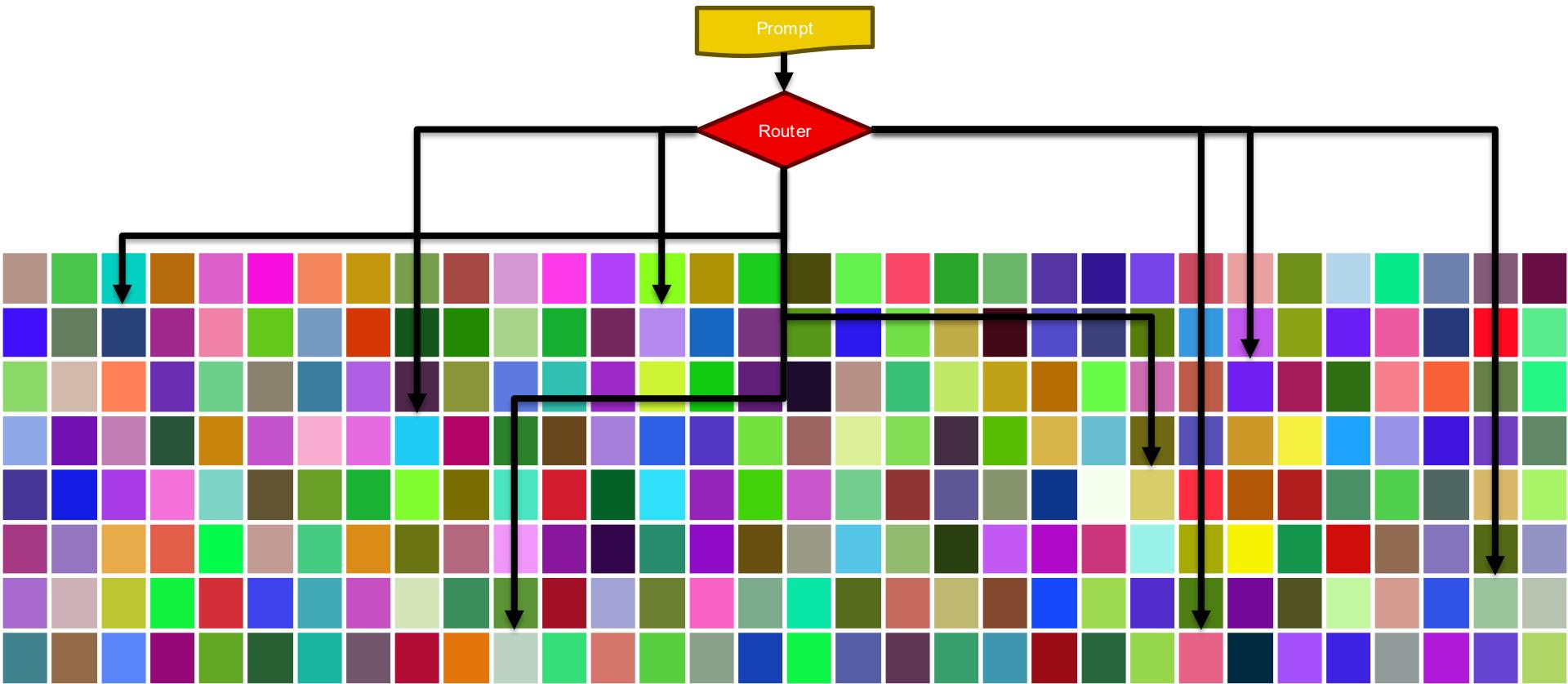
**Architecture: Sparse Mixture of Experts
with Auxiliary-Loss-Free Load Balancing**

Focus: Sparse Mixture of Experts (SMoE, traditional)



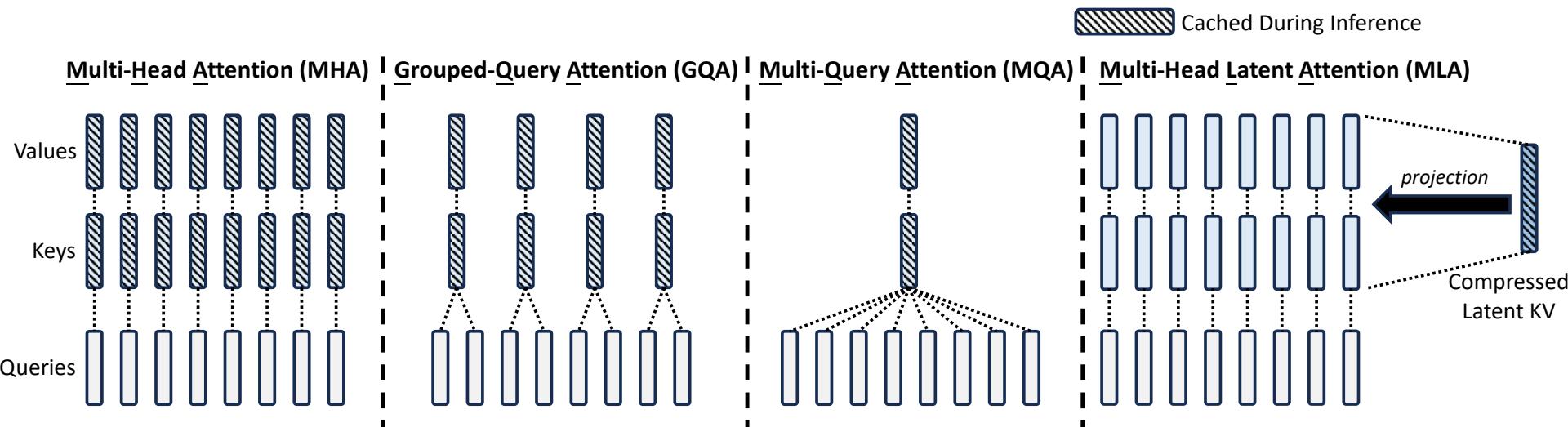


Focus: SMoE (DeepSeek 671B-37A)





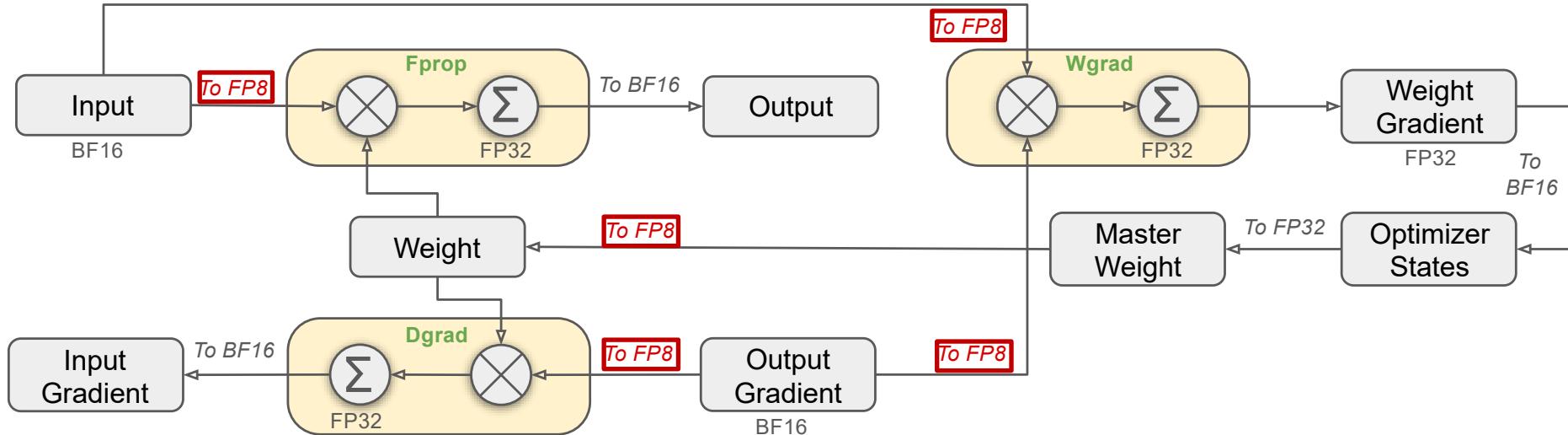
Focus: Multi-head latent attention



Source: <https://arxiv.org/pdf/2405.04434>

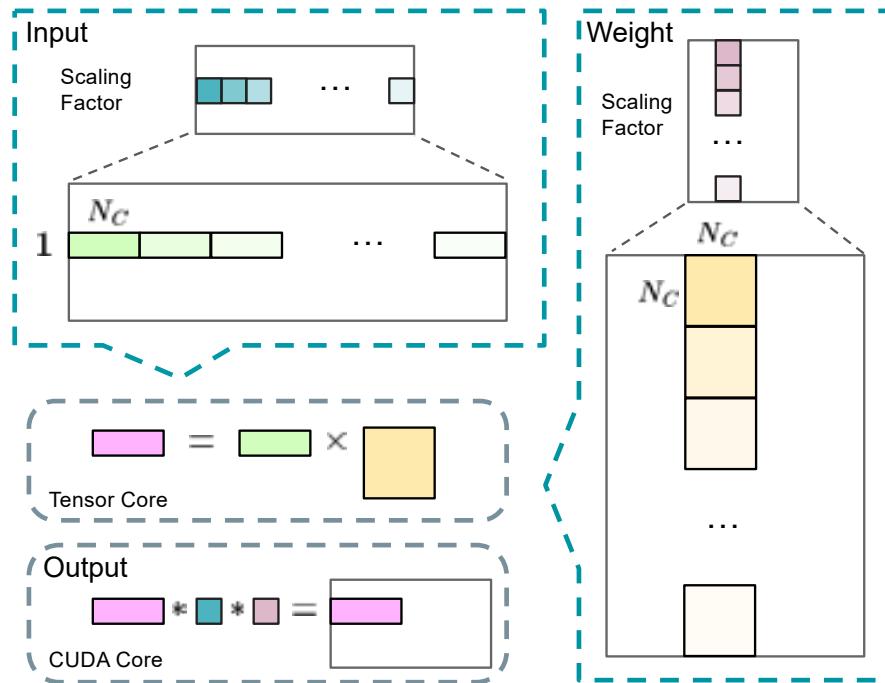


Focus: FP8 training (mixed precision)

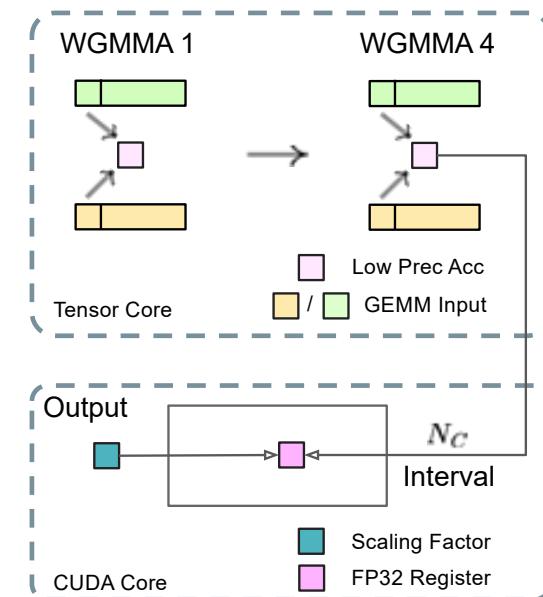


Source: <https://arxiv.org/abs/2412.19437>

Focus: FP8 training (accumulation)

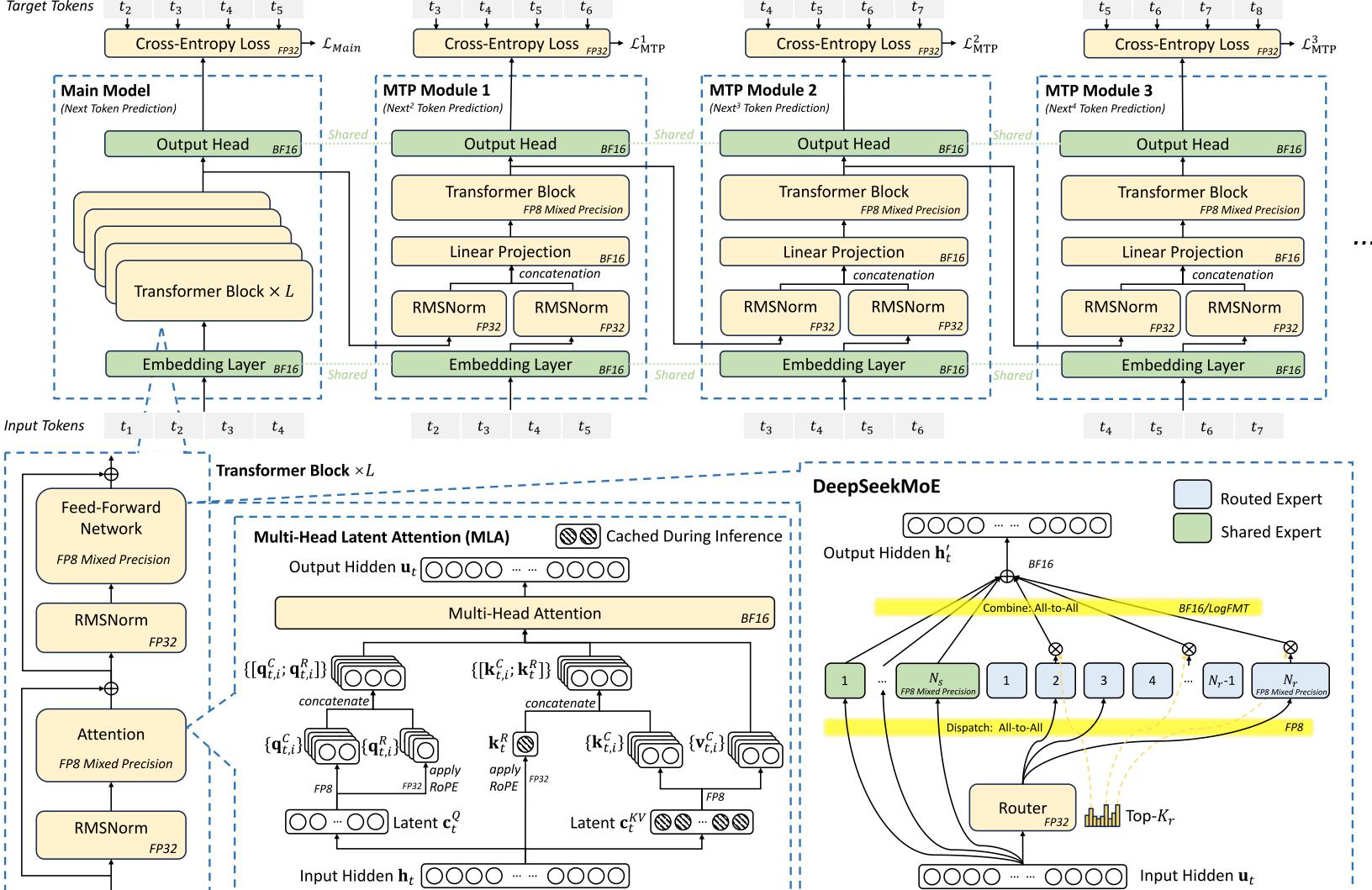


(a) Fine-grained quantization



(b) Increasing accumulation precision

Source: <https://arxiv.org/abs/2412.19437>



Is DeepSeek an Open Model?

Much more open compared to OpenAI

- Architecture and methods published
- Weights publically available
- Can run on own hardware

Much information missing

- Training data (objectivity, Open AI → Google?)
- Heavily censored for China-critical topics





Live Demo

Train your own reasoning model





Q&A

Summary & discussion





Group Discussion

- Any questions left?
- What would you like to achieve with this technology?



Summary & Outlook

Deep Research helps science

- Basis: reasoning (combined with RAG)
- Existing applications like Manus



New interesting scenarios are explored

- Find flaws in scientific articles
- Generate new ideas

Methods and technologies from DeepSeek

- RL, GRPO, MoE standard in new models
- New: Reinforcement Learning with Verifiable Rewards

Size of models decreasing

- Qwen3 has reasoning <1B parameters

Models are getting faster

- MoE allows fast inference for large models
- Important because of long *thinking* process

 **deepseek**  **Tencent** 腾讯



→ **Democratizing reasoning models**

→ **New (local) applications possible**



Review Course Outcomes

By the end of this course, you will understand:

- Understand the differences between normal LLMs and reasoning LLMs
- Learn how to download, use and prompt reasoning models
- Understand the challenges of reasoning models (e.g. evaluation time)
- Get insights into the training process of reasoning models
 - Finetuning (GRPO)
 - Reinforcement learning
 - Advancements by DeepSeek



Resources

GitHub repository

- <https://github.com/datanizing/oreilly-open-source-reasoning>

General Discussion

- <https://www.reddit.com/r/LocalLLaMA/>
- Mixed audience, be careful!

Used programs and libraries

DeepSeek models and architecture

- <https://huggingface.co/deepseek-ai>
- <https://arxiv.org/pdf/2402.03300>
- <https://arxiv.org/pdf/2501.12948>
- <https://arxiv.org/pdf/2505.09343>

Qwen models and architecture

- <https://huggingface.co/Qwen>
- <https://qwenlm.github.io/blog/qwen3/>

O'REILLY®