

Shuoling LIU, Liyuan CHEN, Jiangpeng YAN, Yuhang JIANG, Xiaoyu WANG, Xiu LI, Qiang YANG, 2025. When DeepSeek-R1 meets financial applications: benchmarking, opportunities, and limitations. *Frontiers of Information Technology & Electronic Engineering*, 26(10):1862-1870.
<https://doi.org/10.1631/FITEE.2500227>

When DeepSeek-R1 meets financial applications: benchmarking, opportunities, and limitations

Key words: Large language model; Reasoning; Financial application; Artificial intelligence; Machine learning

Qiang YANG

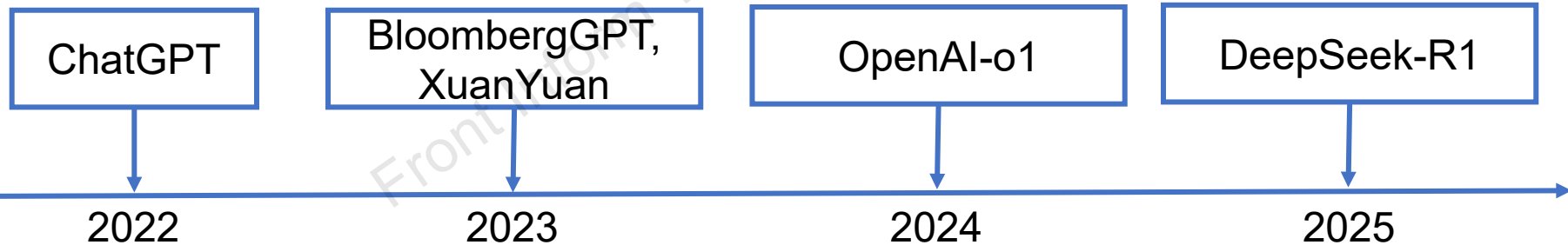
E-mail: qyang@cse.ust.hk

 ORCID: <https://orcid.org/0000-0001-5059-8360>

Motivation

Why reasoning LLMs for finance?

- ❑ Financial tasks require logic, multi-step reasoning, and reliability.
- ❑ Prior general-purpose and financial LLMs (e.g., ChatGPT[1], BloombergGPT[2], XuanYuan[3]) \approx strong NLP but weak reasoning.
- ❑ OpenAI-o1[4] demonstrated its reasoning power but is closed-source and costly.
- ❑ DeepSeek-R1[5]: first open-weight reasoning LLM with o1-level ability.



Timeline of the development of general-purpose and financial LLMs

Research Goals & Contributions

Objectives:

- ❑ Explaining DeepSeek-R1's technical foundation.
- ❑ Benchmarking it on public financial QA datasets.
- ❑ Discussing opportunities, limitations, and future directions.

Key Contributions:

- ❑ First quantitative study of a reasoning LLM in finance.
- ❑ Demonstrating reasoning-driven gains over generic LLMs.
- ❑ Proposing a roadmap for federated, collaborative, and multimodal financial AI.

Chain-of-Thought (CoT) Reasoning

Concepts:

- ❑ Prompting “think step by step” elicits explicit reasoning chains [6]
- ❑ Enhancing interpretability and accuracy
- ❑ Variants: self-consistent CoT, Tree-of-Thought, Graph-of-Thought

Impact on DeepSeek-R1:

- ❑ Integrating CoT into training via reinforcement learning
- ❑ Enabling systematic, transparent problem-solving

Future value of investment calculation example

Give me a financial question and answer it with and without chain-of-thought.

🔗 Financial question:

Question: “If you invest \$10,000 in a savings account with an annual interest rate of 5%, compounded annually, how much money will you have after 10 years?”

Answer without chain-of-thought:

After 10 years, you will have \$16,288.95.

Answer with chain-of-thought:

To calculate the future value of an investment with compound interest, we use the formula:

$$FV = PV \times (1+r)^n$$

Where:

- FV = Future value
- PV = Present value (initial investment) = \$10,000
- r = Annual interest rate = 5% = 0.05
- n = Number of years = 10

Plugging in the values:

$$FV = 10,000 \times (1 + 0.05)^{10}$$

$$FV = 10,000 \times (1.05)^{10}$$

$$FV = 10,000 \times 1.628\ 894\ 626\ 77$$

$$FV = 16,288.95$$

After 10 years, you will have \$16,288.95.

Prompting DeepSeek-V3 to raise and answer a financial question with and without chain-of-thought

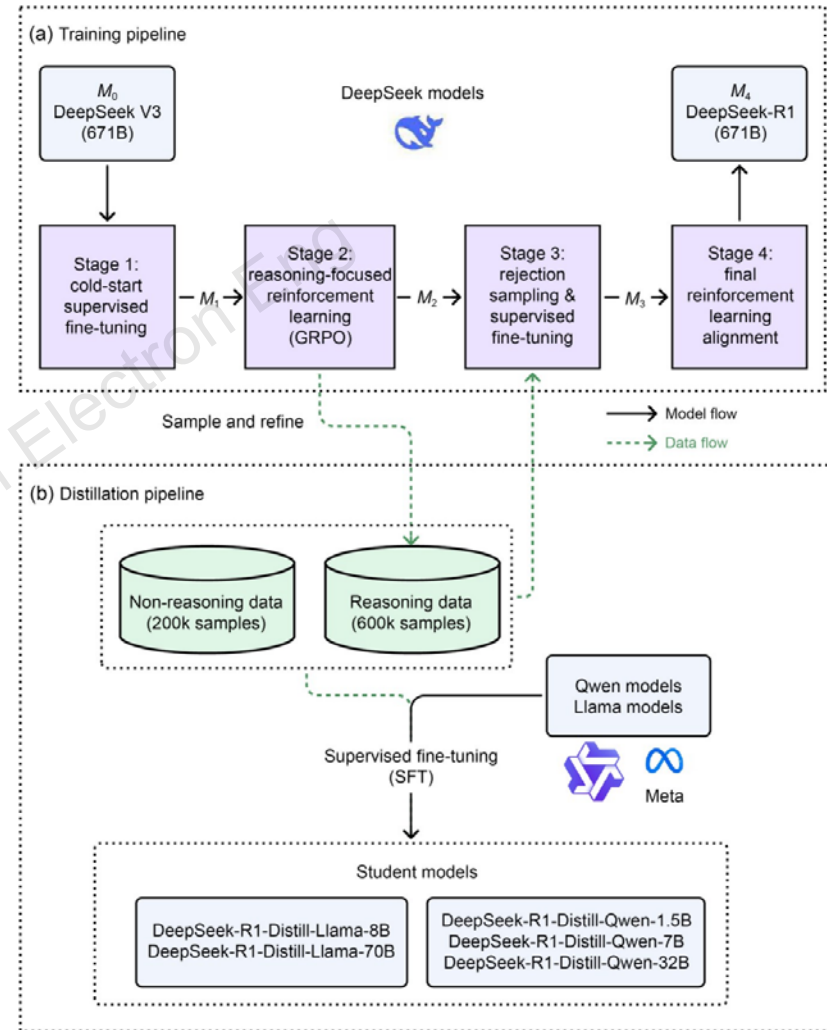
DeepSeek-R1 Training Pipeline

Four Stages of the Training Pipeline:

- ❑ Cold-start SFT → Model M_1
- ❑ Reasoning-focused RL (GRPO algorithm) → M_2
- ❑ Rejection Sampling + SFT → M_3
- ❑ Final RL Alignment → M_4 (DeepSeek-R1)

Highlights:

- ❑ 671B parameters, high-quality CoT data
- ❑ Likely including finance-related reasoning examples



Overview of the DeepSeek-R1 training pipeline: (a) **four-stage training steps**; (b) knowledge distillation pipeline

DeepSeek-R1 Knowledge Distillation

Goal:

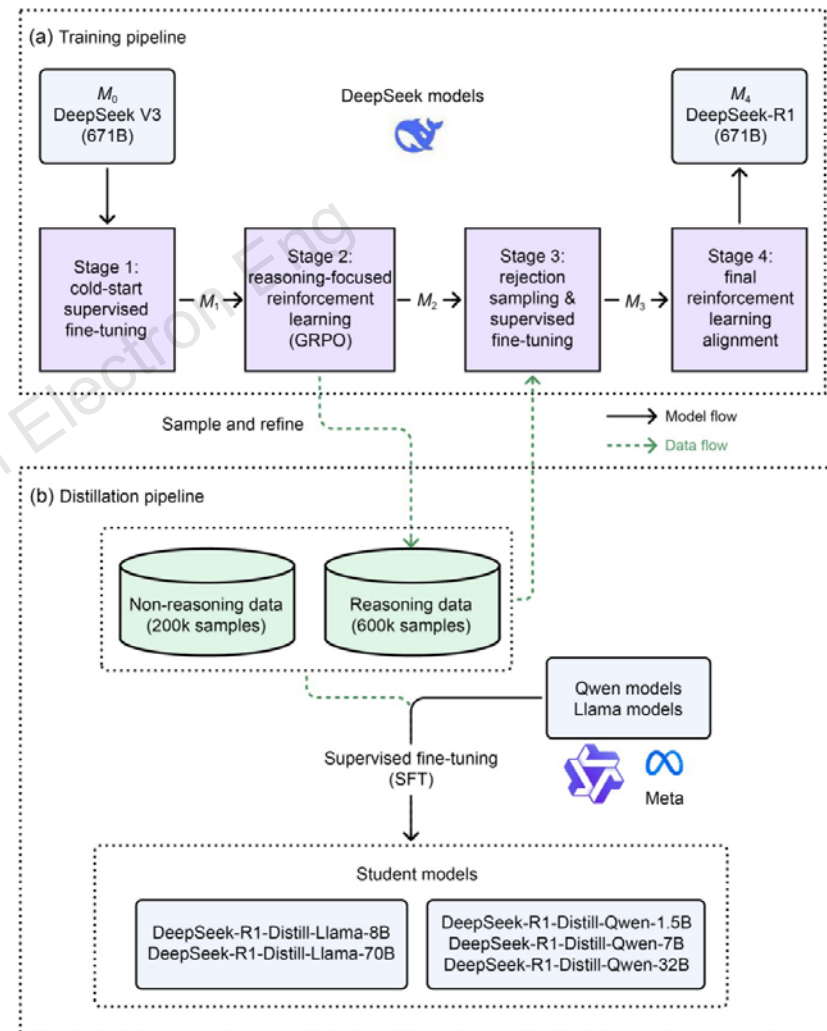
Deploying reasoning skills in smaller LLMs

Distilled Student Models:

Distilled from M2 to Qwen- and Llama-based models (1.5B–70B parameters)

Performance of Students:

- DeepSeek-R1-Distill-Qwen-32B \approx comparable accuracy on math/coding, demonstrating feasible reasoning transfer



Overview of the DeepSeek-R1 training pipeline: (a) four-stage training steps; (b) **knowledge distillation pipeline**

Financial QA Benchmarking

Datasets

- ❑ FinEval (1321 QAs): finance, economics, accounting, certification
- ❑ FinanceIQ (7173 QAs): 10 exams from different financial subjects

Results

Model	Accuracy (%)		
	FinEval	FinanceIQ	FinanceIQ's Financial Math
DeepSeek-V3	82.20	73.46	54.84
DeepSeek-R1	91.52	86.35	82.80
DeepSeek-R1-Distill-Qwen-32B	82.73	77.19	56.99
DeepSeek-R1-Distill-Qwen-7B	51.17	45.63	18.28
GPT-4	70.03	63.40	37.50
OpenAI-o1-preview	78.11	79.20	72.73


Best results are in bold

Findings

- ❑ Reasoning LLMs \gg standard LLMs.
- ❑ DeepSeek models are stronger in Chinese financial knowledge compared to OpenAI models.

Financial QA Benchmarking

DeepSeek-R1 can perform precise calculations and handle basic compound interest problems in finance. Nevertheless, because of its restricted advanced financial knowledge, it may be confused by irrelevant conditions in certain complex problems.

 Question

甲公司年初累计未分配利润1000万元，本年净利润500万元。按法律规定，净利润要提取10%的盈余公积金。预计明年需要进行长期投资800万元，公司的目标资本结构是债务资本占40%、权益资本占60%。公司采用剩余股利政策，本年应分配的股利是（）万元。[Company A had an accumulated undistributed profit of 10 million yuan at the beginning of the year and a net profit of 5 million yuan this year. According to the law, 10% of the net profit should be set aside as surplus reserve. It is estimated that the company will need to make a long-term investment of 8 million yuan next year. The company's target capital structure is 40% debt capital and 60% equity capital. The company follows a residual dividend policy. The dividends to be distributed this year are（）*10 thousand yuan.]
A. 480 B. 0 C. 20 D. 520

你需要一步一步思考，输出答案和思考过程，遵循如下JSON格式。[Please think step by step and provide the reasoning behind selecting the correct answer as the output. Your output should follow the JSON format.]

```
{
  "think": "你的思考过程，[Your step-by-step reasoning process]
  "answer": "你从A, B, C, D中选择的正确答案，[The correct choice from A, B, C, D]
}
```


Answer: C

 DeepSeek-R1

<think>\nOK, I am now going to solve this accounting question. (>1000 characters) ... In summary, the correct answer is Option B.</think>

[{"think": "首先，甲公司需要从上一财年的净利润500万元中提取10%的法定盈余公积金，即50万元，剩余450万元。[First, Company A needs to set aside 10% of the net profit from the previous fiscal year, which is 5 million yuan, as the statutory surplus reserve, that is, 0.5 million yuan, leaving 4.5 million yuan.] 根据剩余股利政策，公司需优先满足投资所需的权益资本。本年度计划增加800万元长期资本，权益部分为800×60%=480万元。[The company needs to prioritize meeting the equity capital required for investment. The company plans to increase long-term capital by 8 million yuan this year, with the equity portion being 8 million yuan × 60% = 4.8 million yuan] 由于当年的可分配利润（450万元）不足以满足权益需求（480万元），公司将全部450万元用于投资，剩余股利为0。"[Since the distributable profit for the year (4.5 million yuan) is not sufficient to meet the equity requirement (4.8 million yuan), the company needs to use the entire 4.5 million yuan for investment, and the residual dividend is 0.] "Answer": "B"]

Comment: 该回答考虑了题目中给出的所有条件，但它忽略了在剩余股利政策下，只需要考虑投资资本的成本，不需要考虑盈余公积。因此，它多扣减了一些不必要的项目。由于缺乏特定领域的金融知识，模型回答错误。实际上，我们使用净利润500万元减去权益资本成本480万元，得到的20万元即为股利，正确答案为C选项。[The response takes into account all the conditions given in the question, but it overlooks the fact that under the residual dividend policy, only the cost of investment capital needs to be considered, and there is no need to consider the surplus reserve. Therefore, it deducts more items than necessary. The lack of specific financial knowledge leads to the incorrect answer. Actually, we subtract the cost of equity capital, which is 4.8 million yuan, from the net profit of 5 million yuan. The resulting 200,000 yuan is the dividend. So the correct answer is Option C.]


 Question

A打算从一家企业那里借入10000元，借款的期限为二年。第一年的年实际贴现率为6%，而第二年则是以8%的年名义贴现率进行两次贴现。请问，A在年初时实际能借到的金额是多少元？[A plans to borrow 10,000 yuan from a company for a period of two years. The actual annual discount rate for the first year is 6%, while in the second year, the discount is applied twice based on an annual nominal discount rate of 8%. What is the actual amount that A can borrow at the beginning of the year? () yuan.]
A:9665 B:9660 C:8663 D:8667

你需要一步一步思考，输出答案和思考过程，遵循如下JSON格式。[Please think step by step and provide the reasoning behind selecting the correct answer as the output. Your output should follow the JSON format.]

```
{
  "think": "你的思考过程，[Your step-by-step reasoning process]
  "answer": "你从A, B, C, D中选择的正确答案，[The correct choice from A, B, C, D]
}
```

Answer: C

 DeepSeek-R1

<think>\nOK, I am now going to solve this financial-math question. (>1000 characters) ... In summary, the correct answer is Option C.</think>

[{"think": "借款两年后需偿还10000元。[The loan must be repaid in full after two years with a total repayment amount of 10,000 yuan.] 第二年贴现计算：名义贴现率8%分两次（每半年一次），每次贴现率4%。第二年末的10000元贴现到第二年初的现值为：[Discount Calculation for the Second Year: The nominal discount rate of 8% is applied twice (every six months), with each discount rate being 4%. The present value of 10,000 yuan at the end of the second year, discounted to the beginning of the second year, is] $10000 \times (1 - 0.04)^2 = 9216$ 。 第一年贴现计算：将第二年初的9216元以第一年实际贴现率6%贴现到当前。[Discount Calculation for the First Year: Discount the 9,216 yuan at the beginning of the second year back to the present using the actual discount rate of 6% for the first year:] $9216 \times (1 - 0.06) = 9216 \times 0.94 = 8663.04$ " "Answer": "C"]

Comment: 该回答将问题拆解为具体的计算步骤，并最终得出了正确答案。[The response breaks down the problem into specific calculation steps and ultimately arrives at the correct answer.]

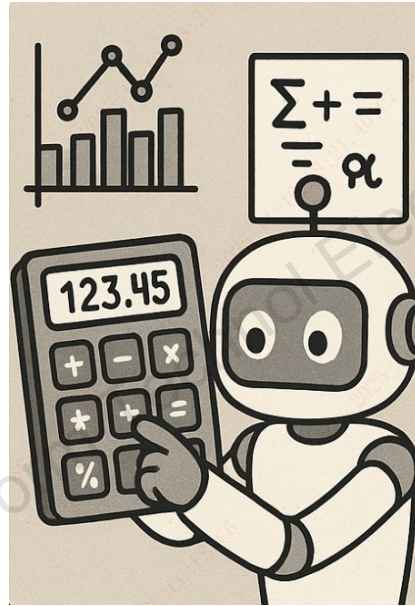
Two cases of DeepSeek-R1's answers to finance math questions. The original Chinese and translated English contents are shown below for readability. DeepSeek-R1's reasoning contents (>1000 characters) are in gray and shortened for space limitation. The same prompting template is used for all LLMs.

Opportunities in Financial Applications

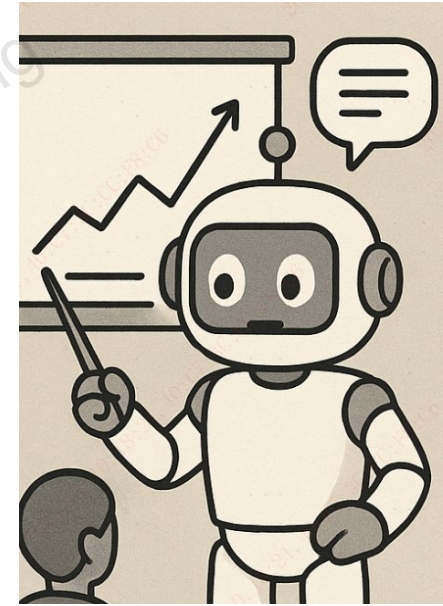
Three Emerging Roles



Financial Text Processor:
summaries, regulation interpretation, and compliance review



Numerical Analysis Aide:
precise calculations and reasoning for financial math



Financial Educator:
step-by-step explanations for training and certification

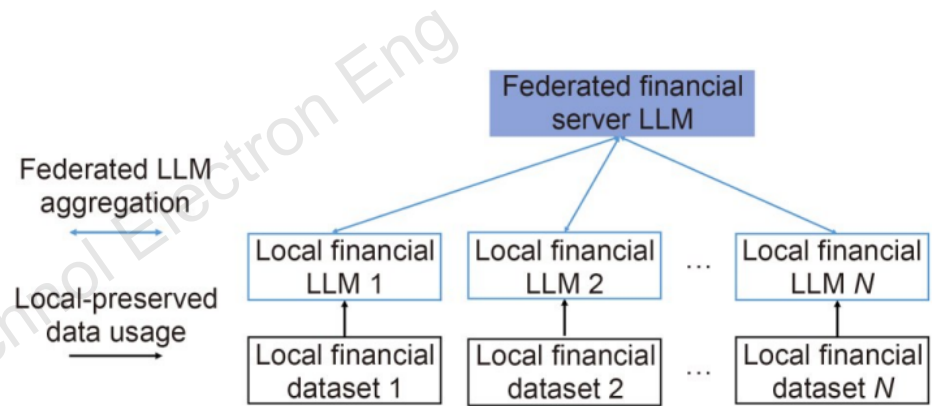
Limitations & Future Directions

Current Limitations:

- ❑ Small-scale benchmarks only
- ❑ Residual hallucination risk
- ❑ Inefficient “slow-thinking-only” mode
- ❑ Lack of multimodal (financial charts/time series) capability

Future Research:

- ❑ Federated Financial Reasoning LLMs – privacy-preserving training
- ❑ Fast & Slow Thinking Collaboration – hybrid System-1/2 models
- ❑ Multimodal Financial Agents – charts + text + audio integration



An implementation of federated financial LLM training frameworks

Conclusions & Key References

DeepSeek-R1

- ❑ brings reasoning LLMs into finance research
- ❑ outperforms prior models in accuracy and explainability
- ❑ opens new avenues for trustworthy, efficient, and federated financial AI

Key References:

- [1] Zhou J, Ke P, Qiu XP, et al., 2024. ChatGPT: potential, prospects, and limitations. *Front Inform Technol Electron Eng*, 25(1):6-11. <https://doi.org/10.1631/FITEE.2300089>
- [2] Wu SJ, Irsoy O, Lu S, et al., 2023. BloombergGPT: a large language model for finance. <https://arxiv.org/abs/2303.17564>
- [3] Zhang XY, Yang Q, 2023. XuanYuan 2.0: a large Chinese financial chat model with hundreds of billions parameters. Proc 32nd ACM Int Conf on Information and Knowledge Management, p.4435-4439. <https://doi.org/10.1145/3583780.3615285>
- [4] Jaech A, Kalai A, Lerer A, et al., 2024. OpenAI o1 system card. <https://arxiv.org/abs/2412.16720>
- [5] Guo DY, Yang DJ, Zhang HW, et al., 2025. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning. <https://arxiv.org/abs/2501.12948>
- [6] Wei J, Wang XZ, Schuurmans D, et al., 2022. Chain-of-thought prompting elicits reasoning in large language models. Proc 36th Int Conf on Neural Information Processing Systems, Article 1800.