



*F*_M FinMMR: Make Financial Numerical Reasoning More Multimodal, Comprehensive, and Challenging

Zichen Tang Haihong E* Jiacheng Liu Zhongjun Yang Rongjin Li Zihua Rong
Haoyang He Zhuodi Hao Xinyang Hu Kun Ji Ziyan Ma Mengyuan Ji Jun Zhang
Chenghao Ma Qianhe Zheng Yang Liu Yiling Huang Xinyi Hu Qing Huang Zijian Xie

Shiyan Peng

*Corresponding author.



BUPT
Reasoning Lab

Reasoning Lab, Beijing University of Posts and Telecommunications

<https://bupt-reasoning-lab.github.io/FinMMR/>

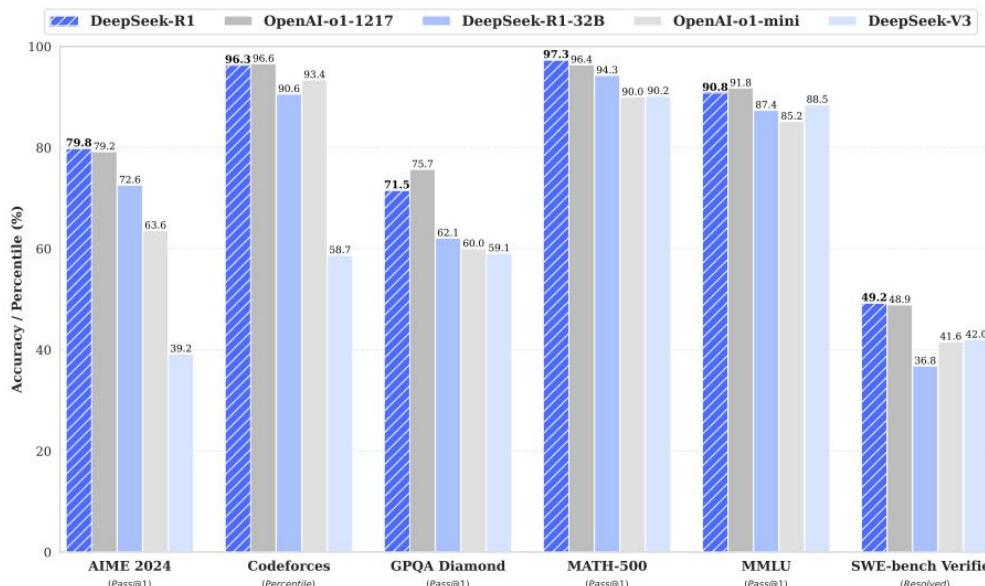
\mathcal{F}_M = Finance + Image & Text + Numerical Reasoning

FinMMR: Make Financial Numerical Reasoning

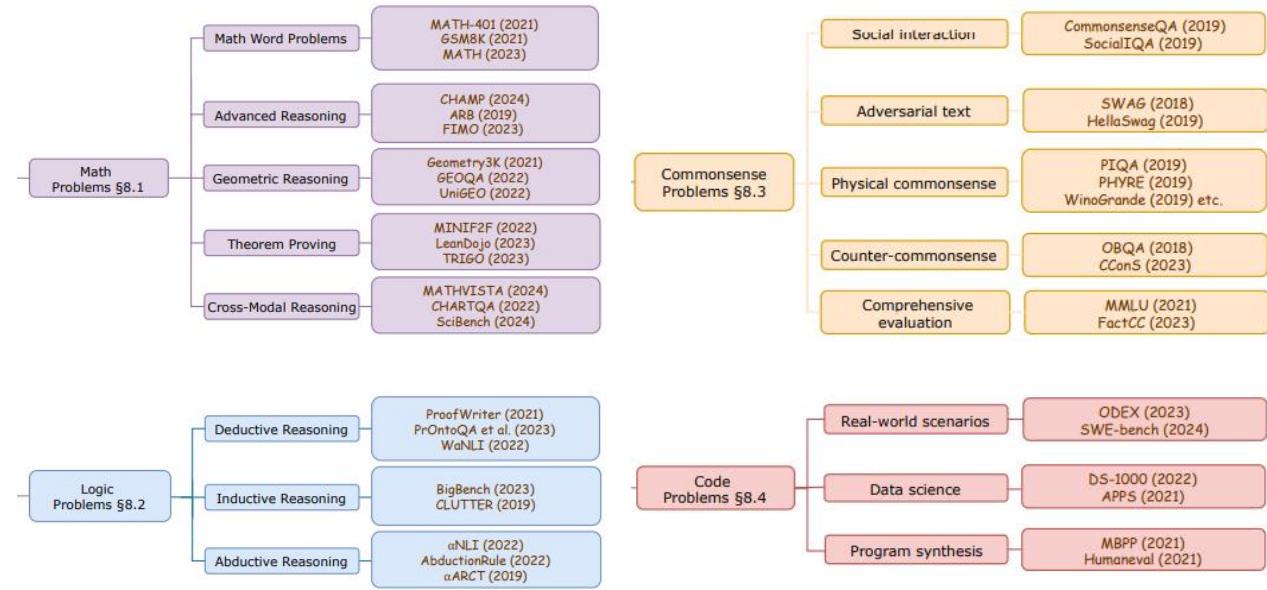
More Multimodal, Comprehensive, and Challenging

Research Motivation (1/2)

- Domain-specific numerical reasoning in real-world tasks requires:
 - Applying domain expertise
 - Performing complex mathematical computation over image-text contexts



DeepSeek-R1 Excels in
Code/Math/Scientific Tasks^[1]



Current LLM Reasoning Benchmarks (Selected)
Math, Logic, Commonsense, Code^[2]

Question: *Can current MLLMs seamlessly integrate vision and text to perform domain-specific complex reasoning, matching the proficiency of LRM_s in pure text-based tasks?*

Research Motivation (2/2)

➤ Financial Multimodal Numerical Reasoning requires:

Fine-grained Visual Perception

Comprehend rich financial visuals with domain knowledge

Specialized Knowledge Reasoning

Apply financial knowledge and formulas for multi-step reasoning

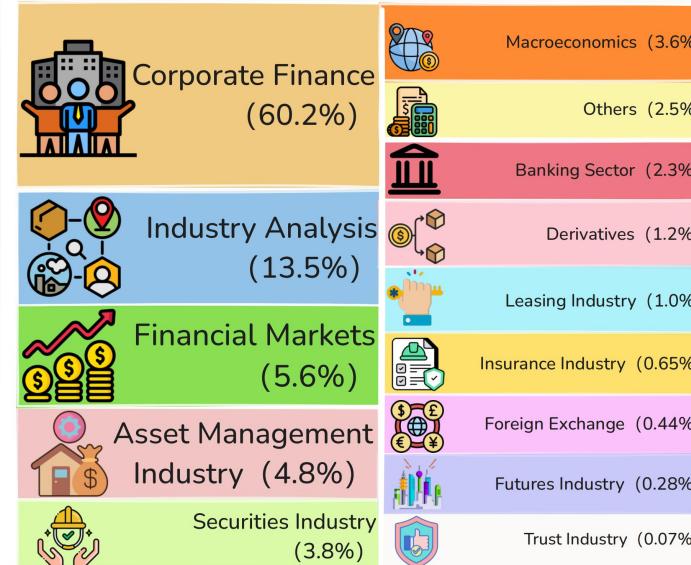
Precise Numerical Computation

Produce explainable reasoning with accurate, verifiable outputs

Rich Images



Comprehensive Subdomains



Complex Formulas

Calculate the percentage by which Samsung SDS directly holds more equity in MULTICAMPUS than the combined indirect holdings of Samsung SDI and Samsung Electronics in MULTICAMPUS. Round to two decimal places.

Program:

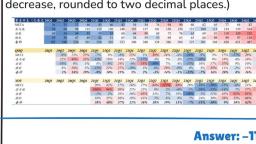
```
SAMSUNG Bio, Research & Education Companies' Equity Structure
def solution():
    # Define ownership percentages
    sds_ownership = 47.24
    # Samsung SDS's direct ownership of MULTICAMPUS
    sdi_ownership = 29.60
    # Samsung SDI's ownership of GLOBAL RESEARCH
    electronics_ownership = 29.80
    # Samsung Electronics' ownership of GLOBAL RESEARCH
    global_research_to_multicampus = 15.16
    # GLOBAL RESEARCH's ownership of MULTICAMPUS
    # Calculate SDI and Electronics' indirect ownership of MULTICAMPUS
    indirect_ownership = (sdi_ownership + electronics_ownership) * global_research_to_multicampus / 100
    # Calculate the difference
    difference = sds_ownership - indirect_ownership
    # Round to two decimal places
    answer = round(difference, 2)
    return answer
```

Answer: 38.23%

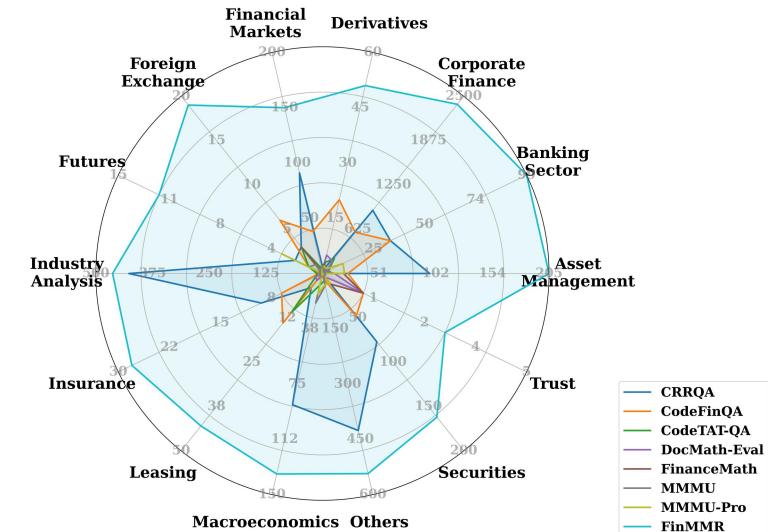
Visual-rich, Knowledge-intensive, Reasoning-heavy, Outcome-sensitive

Benchmark Construction

- 8.7K images (14 types) + 4.3K questions (14 subdomains): bridging finance and multimodality
- 2.15K English + 2.15K Chinese: converted / LLM-generated with expert verification, w. distractor images
- Complexity labels (Easy / Medium / Hard): Claude 3.7 Sonnet only 53.0% on *Hard* subset

Rich Images Challenge Visual Perception	Comprehensive Subdomains Challenge Knowledge Reasoning	Complex Formulas Challenge Numerical Computation
<p>問題：计算2016年至2024年Q2期间，上海银行和上市城商行的个人住房贷款年均复合增长率之差，结果以百分比表示，保留两位小数。（Calculate the difference in the compound annual growth rate (CAGR) of personal housing loans between Bank of Shanghai and listed city commercial banks from 2016 to Q2 2024. Present the result as a percentage, rounded to two decimal places.）</p>  <p>Answer: 2.17</p>	<p>Q: Table 19.3 shows a book balance sheet for the Wishing Well Motel chain. The company's long-term debt is secured by its real estate assets, but it also uses short-term bank loans as a permanent source of financing. It pays 10% interest on the bank debt and 9% interest on the secured debt. Wishing Well has 10 million shares of stock outstanding, trading at \$90 per share. The expected return on Wishing Well's common stock is 18%. Calculate Wishing Well's WACC. Assume that the book and market values of Wishing Well's debt are the same. The marginal tax rate is 21%. Answer as a percentage to single decimal place.</p> <p>Q: What is the total weighted average Cash and Payment-In-Kind (PIK) interest rate payable under the subordinated and senior notes portfolio in the year 2023?</p> <p>Q: What is the total weighted average Cash and Payment-In-Kind (PIK) interest rate payable under the subordinated and senior notes portfolio in the year 2023?</p> <p>Q: John oversees a fund, with the returns for the first three years displayed below: What will be the holding period return (expressed as a percentage)? Answer to three decimal places.</p> <p>Q: What is the average quarterly share price in 2021 assuming each quarter had a completely uniform price distribution in high and low bid prices, measured in dollars?</p>	<p>Q: What is the total weighted average Cash and Payment-In-Kind (PIK) interest rate payable under the subordinated and senior notes portfolio in the year 2023?</p> <p>Q: What is the total weighted average Cash and Payment-In-Kind (PIK) interest rate payable under the subordinated and senior notes portfolio in the year 2023?</p> <p>Q: John oversees a fund, with the returns for the first three years displayed below: What will be the holding period return (expressed as a percentage)? Answer to three decimal places.</p> <p>Q: What is the average quarterly share price in 2021 assuming each quarter had a completely uniform price distribution in high and low bid prices, measured in dollars?</p>
<p>Compound Average Growth Rate (CAGR)</p> <p>問題：请计算惠济新生对日照公司的间接持股比例，结果以百分比表示，保留两位小数。（Please calculate the indirect shareholding ratio of Huiji Xinsheng in Rizhao Company. Present the result as a percentage, rounded to two decimal places.）</p>  <p>Answer: 13.65</p>	<p>Weighted Average Cost of Capital (WACC)</p> <p>Q: For each of the investments shown in the following table, calculate the rate of return earned over the unspecified time period. What is the rate of return for Investment A? Answer as a percentage to the nearest integer.</p> <p>Q: John oversees a fund, with the returns for the first three years displayed below: What will be the holding period return (expressed as a percentage)? Answer to three decimal places.</p> <p>Q: John oversees a fund, with the returns for the first three years displayed below: What will be the holding period return (expressed as a percentage)? Answer to three decimal places.</p> <p>Q: What is the average quarterly share price in 2021 assuming each quarter had a completely uniform price distribution in high and low bid prices, measured in dollars?</p>	<p>Rate of Return (RoR)</p> <p>Q: Ricky is considering purchasing an apartment costing \$700,000. He will pay a 30% down payment and take out a mortgage for the remainder. Since he just got married and wants to save some money for future use, he will choose the plan with the lowest monthly payment. After visiting several banks, he received the following mortgage offers: What is the monthly payment for Bank A? Answer to two decimal places.</p> <p>Q: What is the average quarterly share price in 2021 assuming each quarter had a completely uniform price distribution in high and low bid prices, measured in dollars?</p>
<p>Indirect Shareholding Ratio</p> <p>問題：请计算2022年第四季度和2023年第一季度的总资本开支，并将其与2021年第四季度的资本开支进行比较，计算下降百分比，结果保留两位小数。（Calculate the total capital expenditure for Q4 2022 and Q1 2023 combined, and compare it to the capital expenditure for Q4 2021. Calculate the percentage decrease, rounded to two decimal places.）</p>  <p>Answer: -110.36</p>	<p>Capital Expenditure</p> <p>Q: Calculate the monthly payment for a \$100,000 loan with an interest rate of 5% over a 30-year period.</p> <p>Q: Calculate the monthly payment for a \$100,000 loan with an interest rate of 5% over a 30-year period.</p> <p>Q: Calculate the monthly payment for a \$100,000 loan with an interest rate of 5% over a 30-year period.</p>	<p>Monthly Payment</p> <p>Q: Calculate the monthly payment for a \$100,000 loan with an interest rate of 5% over a 30-year period.</p> <p>Q: Calculate the monthly payment for a \$100,000 loan with an interest rate of 5% over a 30-year period.</p>
<p>Average Quarterly Price</p>	<p>Average Quarterly Price</p>	<p>Average Quarterly Price</p>

Benchmark	Size (Fin)	Domain Coverage	Modalities	Question Type
MMMU [65]	11,550 (1,603)	10	T+I	MC
MMMU-Pro [66]	1,730 (286)	10	T+I, P.I.	MC
FinanceMath [67]	1,200 (1,200)	10	T	NUM
CodeTAT-QA [32]	3,144 (3,144)	6	T	NUM
CodeFinQA [32]	5,463 (5,463)	13	T	NUM
DocMath-Eval [68]	4,000 (4,000)	12	T	NUM
FAMMA [63]	1,758 (1,758)	8	T+I	MC, NUM
CRRQA (ours)	2,150 (2,150)	13	T+I, P.I.	NUM
FinMMR (ours)	4,300 (4,300)	14	T+I, P.I.	NUM



— CRRQA
— CodeFinQA
— CodeTAT-QA
— DocMath-Eval
— FinanceMath
— MMMU
— MMMU-Pro
— FinMMR

Comprehensive Evaluation

- **15 leading MLLMs (as of May 15, 2025):** comprehensive evaluation system built
- **Claude 3.7 Sonnet (64K thinking):** highest overall performance
- **Reasoning-enhanced vs. non-reasoning:** better accuracy but $\sim 12 \times$ token cost

Model	Size	Extended thinking	Hard			Medium		Easy		Avg.		Token (M)	
			IO	CoT	PoT	CoT	PoT	CoT	PoT	CoT	PoT	CoT	PoT
Proprietary MLLMs													
Claude 3.7 Sonnet		✓ (64K)	53.00	51.00	51.40	62.50	62.17	78.50	78.50	64.00	64.02	8.51	11.25
Claude 3.7 Sonnet		✗	49.80	50.80	48.50	62.25	58.83	77.00	76.92	63.35	61.42	0.99	0.89
OpenAI o1		✓	48.00	48.40	44.70	—	—	—	—	—	—	2.52	2.12
GPT-4o		✗	—	45.40	47.80	63.33	59.92	78.00	76.00	62.24	61.24	0.85	0.41
Gemini 2.0 Pro		✗	—	46.50	47.30	60.58	57.92	75.50	75.67	60.86	60.30	0.85	0.45
Gemini 2.0 Flash Thinking		✓	—	46.00	46.00	60.75	56.58	77.17	74.17	61.31	58.92	1.30	0.48
Gemini 2.0 Flash		✗	—	44.40	45.90	57.83	53.42	74.92	73.75	59.05	57.69	0.79	0.43
Grok 2 Vision		✗	—	27.80	25.50	41.50	35.83	73.08	72.83	47.46	44.72	1.13	0.60
Qwen-Omni-Turbo		✗	—	17.50	27.30	35.83	48.00	57.50	61.67	36.94	45.66	0.90	0.42
Open-source MLLMs													
Llama 4 Maverick	17B	✗	—	48.70	47.80	63.25	59.17	77.83	77.83	63.26	61.60	0.88	0.47
Qwen2.5-VL-72B	72B	✗	—	43.30	46.20	63.42	64.17	77.42	75.83	61.38	62.07	1.05	0.44
InternVL2.5-78B	78B	✗	—	37.40	44.00	60.50	61.17	70.92	70.58	56.27	58.58	—	—
QVQ-72B-Preview	72B	✓	43.30	40.30	6.20	55.67	9.67	75.42	12.42	57.13	9.43	5.43	5.70
Pixtral Large	124B	✗	—	19.70	25.00	39.83	39.75	70.00	70.17	43.18	44.97	1.15	0.75
Gemma 3 27B	27B	✗	—	23.40	22.30	45.17	36.42	69.08	61.58	45.88	40.10	0.97	0.47
Mistral Small 3.1	24B	✗	—	19.70	15.20	38.42	29.75	67.67	49.42	41.93	31.46	1.15	0.60

Error Analysis

Fine-grained Visual Perception (30%)

Question: What is the increase in the percentage of the total contract assets proportion to the total accounts receivable from December 31, 2019 to June 30, 2020 in percentage? Rounded to 2 decimal places.

Accounts Receivable		
The components of accounts receivable are as follows:		
	As of	
Trade receivables	\$ 4,592	December 31, 2019 6,425
Unbilled receivables	1,763	1,857
Other short-term receivables	19	7
Allowance for doubtful accounts	(716)	(1,816)
Total accounts receivable	\$ 5,658	\$ 6,473

Context: The timing of revenue recognition, billings and cash collections results in billed accounts receivable, unbilled receivables (contract assets), and customer advances and deposits (contract liabilities) on the consolidated balance sheets. In our Air Pollution Control Technology segment, amounts are billed as work progresses in accordance with agreed-upon contractual terms. Generally, billing occurs subsequent to revenue recognition, resulting in contract assets. These assets are reported on the consolidated balance sheet on a contract-by-contract basis at the end of each reporting period. At June 30, 2020 and December 31, 2019, contract assets were approximately \$1,763 and \$1,857, respectively, and are included in accounts receivable on the consolidated balance sheets.

```
def solution():
    # Define variables name and value
    contract_assets_june_2020 = 1763
    total_accounts_receivable_june_2020 = 5058 (x) 5058 (x) — 5658 (v)
    contract_assets_dec_2019 = 1857
    total_accounts_receivable_dec_2019 = 6473

    # Calculate the proportion of contract assets to total accounts
    # receivable
    proportion_june_2020 = contract_assets_june_2020 /
    total_accounts_receivable_june_2020
    proportion_dec_2019 = contract_assets_dec_2019 /
    total_accounts_receivable_dec_2019

    # Compute the percentage increase in proportion
    percentage_increase = ((proportion_june_2020 - proportion_dec_2019) /
    proportion_dec_2019) * 100

    # Return answer rounded to 2 decimal places
    return round(percentage_increase, 2)
```

Answer: 21.5

Ground Truth: 2.47

Specialized Knowledge Reasoning (38%)

Question: For each of the following \$1,000-parvalue bonds, assuming annual interest payment and a 40% tax rate, calculate the after-tax cost to maturity using the approximation formula. What is the after-tax cost to maturity for Bond C? Answer as a percentage to two decimal places.

Bond	Life (years)	Underwriting fee	Discount (-) or premium (+)	Coupon interest rate
A	20	\$25	-\$20	9%
B	16	40	+10	10
C	15	30	-15	12
D	25	15	Par	9
E	22	20	-60	11

```
def solution():
    # Define bond parameters
    par_value = 1000 # Par value of the bond
    discount = -15 # Discount for Bond C ($15)
    life = 15 # Bond life in years
    coupon_rate = 0.12 # Coupon interest rate (12%)
    tax_rate = 0.40 # Tax rate (40%)
    underwriting_fee = 30 # Underwriting fee ($30)
    annual_interest = par_value * coupon_rate
    after_tax_interest = annual_interest * (1 - tax_rate) (x)
    net_proceeds = par_value + discount - underwriting_fee
    numerator = after_tax_interest + ((par_value - net_proceeds) / life) (x)
    denominator = (par_value + net_proceeds) / 2
    after_tax_cost = (numerator / denominator) * 100 # Convert to
    percentage
    return round(after_tax_cost, 2)
```

Error Analysis: The error occurred because the tax rate was applied prematurely to the interest payment, while the amortized issuance cost was left untaxed, leading to an incorrect calculation. The correct approach is to first compute the pre-tax cost and then apply the tax rate to the entire cost to ensure consistency.

Ground Truth: 7.55

Precise Numerical Computation (32%)

Question: On February 17, 2017, what was the company's market capitalization as reported on the NYSE? Answer to three decimal places.

	2016	High	Low
Quarter ended March 31	\$102.93	\$83.07	
Quarter ended June 30	113.63	101.87	
Quarter ended September 30	118.26	107.57	
Quarter ended December 31	118.09	99.72	
2015	High	Low	
Quarter ended March 31	\$101.88	\$93.21	
Quarter ended June 30	98.64	91.99	
Quarter ended September 30	101.54	86.83	
Quarter ended December 31	104.12	87.23	

Context: Part II, Item 5. Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities. The following table presents the reported quarterly high and low per share sale prices of our common stock on the NYSE for the years 2016 and 2015. On February 17, 2017, the closing price of our common stock was \$108.11 per share as reported on the NYSE. As of February 17, 2017, we had 427,195,037 outstanding shares of common stock and 153 registered holders.

```
def calculate_market_cap():
    closing_price_per_share = 108.11
    outstanding_shares = 427_195_037
    market_capitalization = closing_price_per_share *
    outstanding_shares
    market_cap_billions = round(market_capitalization /
    1_000_000_000, 3) (x)
    return market_cap_billions
print(calculate_market_cap())
```

Error Analysis: The model arbitrarily converted the market capitalization into billions, despite the absence of any such requirement in the problem statement. This reflects a misalignment between the model's assumptions and the explicit task constraints.

Ground Truth: 46184055450.078

Reasoning Enhancement (1/3)

➤ Filtering for Reasoning:

- Eliminate interference from irrelevant images before inference
- Irrelevant images cause >10% drop, even without overlapping content

Question: 请根据图中数据，计算呼吸在第五年的销售额，并将其与第三年进行比较，计算增长率；请保留两位小数。

(Calculate the sales revenue of the “Respiratory” category in the fifth year based on the data in the table. Then, compare it with the third year and calculate the growth rate. Please round the result to two decimal places.)

Image 1:

	2021	2022	2023	2024E	2025E	2026E
总收入	1118.57	1363.35	1387.74	1542.15	1818.67	2117.58
YOY	-1.07%	21.88%	1.79%	11.13%	17.93%	16.44%
毛利率	42.36%	43.19%	45.89%	43.76%	43.94%	44.42%
归母净利润	106.09	166.57	192.49	220.85	269.86	323.18
YOY	-30.71%	57.00%	15.56%	14.74%	22.19%	19.76%
1.麻醉	349.08	409.45	467.73	493.46	592.15	710.58
YOY	8.67%	17.30%	14.23%	5.50%	20.00%	26.00%
毛利率	53.27%	54.81%	56.41%	56.00%	56.00%	55.00%
占比	31.21%	30.03%	33.70%	32.00%	32.56%	33.56%
2.导尿	290.40	424.19	382.21	443.36	523.10	627.80
YOY	5.24%	46.07%	-9.90%	16.00%	18.00%	20.00%
毛利率	27.16%	28.05%	29.27%	30.00%	30.00%	30.00%
占比	25.96%	31.11%	27.54%	28.75%	28.77%	29.65%
3.泌尿外科	169.56	196.04	203.44	189.20	223.26	263.44
YOY	27.84%	15.62%	3.77%	-7.00%	18.00%	18.00%
毛利率	80.35%	80.15%	75.90%	76.00%	76.00%	76.00%
占比	15.16%	14.38%	14.66%	12.27%	12.28%	12.44%
4.护理	134.02	159.44	135.13	189.18	227.01	215.66
YOY	39.52%	18.96%	-15.25%	40.00%	20.00%	-5.00%
毛利率	19.17%	23.20%	28.24%	25.00%	25.00%	25.00%
占比	11.98%	11.69%	9.74%	12.27%	12.48%	10.18%
5.呼吸	61.87	64.33	87.73	100.01	120.01	150.01
YOY	-9.87%	3.98%	36.36%	14.00%	20.00%	25.00%
毛利率	31.58%	37.69%	38.31%	37.00%	37.00%	37.00%
占比	5.53%	4.72%	6.32%	6.48%	6.60%	6.08%
6.血液透析	73.11	81.02	68.46	75.30	82.85	91.15
YOY	3.74%	10.82%	-15.51%	10.00%	10.00%	10.00%
毛利率	26.41%	24.46%	29.51%	28.00%	28.00%	28.00%
占比	6.54%	5.94%	4.93%	4.88%	4.58%	4.30%

USEFUL

Image 2:

	2021	2022	2023	2024E	2025E	2026E
总收入	1118.57	1363.35	1387.74	1542.15	1818.67	2117.58
YOY	-1.07%	21.88%	1.79%	11.13%	17.93%	16.44%
毛利率	42.36%	43.19%	45.89%	43.76%	43.94%	44.42%
归母净利润	106.09	166.57	192.49	220.85	269.86	323.18
YOY	-30.71%	57.00%	15.56%	14.74%	22.19%	19.76%
1.麻醉	349.08	409.45	467.73	493.46	592.15	710.58
YOY	8.67%	17.30%	14.23%	5.50%	20.00%	26.00%
毛利率	53.27%	54.81%	56.41%	56.00%	56.00%	55.00%
占比	31.21%	30.03%	33.70%	32.00%	32.56%	33.56%
2.导尿	290.40	424.19	382.21	443.36	523.10	627.80
YOY	5.24%	46.07%	-9.90%	16.00%	18.00%	20.00%
毛利率	27.16%	28.05%	29.27%	30.00%	30.00%	30.00%
占比	25.96%	31.11%	27.54%	28.75%	28.77%	29.65%
3.泌尿外科	169.56	196.04	203.44	189.20	223.26	263.44
YOY	27.84%	15.62%	3.77%	-7.00%	18.00%	18.00%
毛利率	80.35%	80.15%	75.90%	76.00%	76.00%	76.00%
占比	15.16%	14.38%	14.66%	12.27%	12.28%	12.44%
4.护理	134.02	159.44	135.13	189.18	227.01	215.66
YOY	39.52%	18.96%	-15.25%	40.00%	20.00%	-5.00%
毛利率	19.17%	23.20%	28.24%	25.00%	25.00%	25.00%
占比	11.98%	11.69%	9.74%	12.27%	12.48%	10.18%
5.呼吸	61.87	64.33	87.73	100.01	120.01	150.01
YOY	-9.87%	3.98%	36.36%	14.00%	20.00%	25.00%
毛利率	31.58%	37.69%	38.31%	37.00%	37.00%	37.00%
占比	5.53%	4.72%	6.32%	6.48%	6.60%	6.08%
6.血液透析	73.11	81.02	68.46	75.30	82.85	91.15
YOY	3.74%	10.82%	-15.51%	10.00%	10.00%	10.00%
毛利率	26.41%	24.46%	29.51%	28.00%	28.00%	28.00%
占比	6.54%	5.94%	4.93%	4.88%	4.58%	4.30%

USELESS

Image 3:

公司名称	代码	收盘价(元)	归母净利润(亿元)	23A	24E	25E	26E	24E	25E	26E	PE
康德莱	603987.SH	7.46	2.31	2.24	2.48	2.81	3.24	14.23	15.44	17.12	
三鑫医疗*	300453.SZ	7.74	2.07	2.45	2.86	3.36	3.66	16.60	17.30	12.17	
邦和医疗	605369.SH	28.72	109.13	193.66	251.00	308.57	357.00	18.51	18.55	15.11	
华邦健康								18.31	15.24	12.87	
维力医疗	603309.SH	11.74	1.92	2.21	2.70	3.23	3.54	12.67	10.58		

Qwen2.5-VL-72B: **73.4% filtering accuracy**
accuracy rises **64.7% → 71.6%**
near **77.8% (no distractors)**

Original Output:

```
def solution():
    # Define variables name and value
    year_3_sales = 87.73
    year_5_sales = 150.01 (x)
    # Do math calculation to get the answer
    growth_rate = ((year_5_sales - year_3_sales) / year_3_sales)
    * 100
    # return answer
    return round(growth_rate, 2)
    # Call the function and print the result
    print(solution())
```



Augmented Output:

```
def solution():
    # Define variables name and value
    year_3_sales = 87.73
    year_5_sales = 120.01
    # Do math calculation to get the answer
    growth_rate = ((year_5_sales - year_3_sales) / year_3_sales)
    * 100
    answer = round(growth_rate, 2)
    # return answer
    return answer
```



Ground Truth: 36.79

Reasoning Enhancement (2/3)

➤ Structured Knowledge Augmentation:

- Knowledge base of 3,133 financial Python functions
- Each with description, parameters, returns, constraints, code & comments

Financial Function

```
def calc_net_return(init_investment: float,
                    growth: float, fee_rate: float, inc_rate:
                    float, hurdle: float) -> float:
    """
    Calculate the net return for an investor in
    a hedge fund given various parameters.

    Args:
        initial_investment (float): The initial
        amount invested in the hedge fund...

    Returns:
        net_return (float): The net return for the
        investor after fees, in millions.
    """
    end_value = init_investment * (1 + growth)
    fee = end_value * fee_rate
    net_value = end_value - fee
    hurdle_value = init_investment * (1 + hurdle
    )
    inc_fee = max(0, (net_value - hurdle_value)
    * inc_rate)
    net_return = end_value - (fee + inc_fee) -
    init_investment
    return round(net_return, 2)
```

Setting	PoT	RAG with PoT
Gemini 2.0 Flash Thinking	78.71	83.02 (+4.31)
GPT-4o	80.60	83.62 (+3.02)
Claude 3.7 Sonnet	81.21	85.43 (+4.22)
Claude 3.7 Sonnet (64K)	83.53	86.29 (+2.76)

On 1,160 tabular QA problems, knowledge augmentation consistently improves performance:

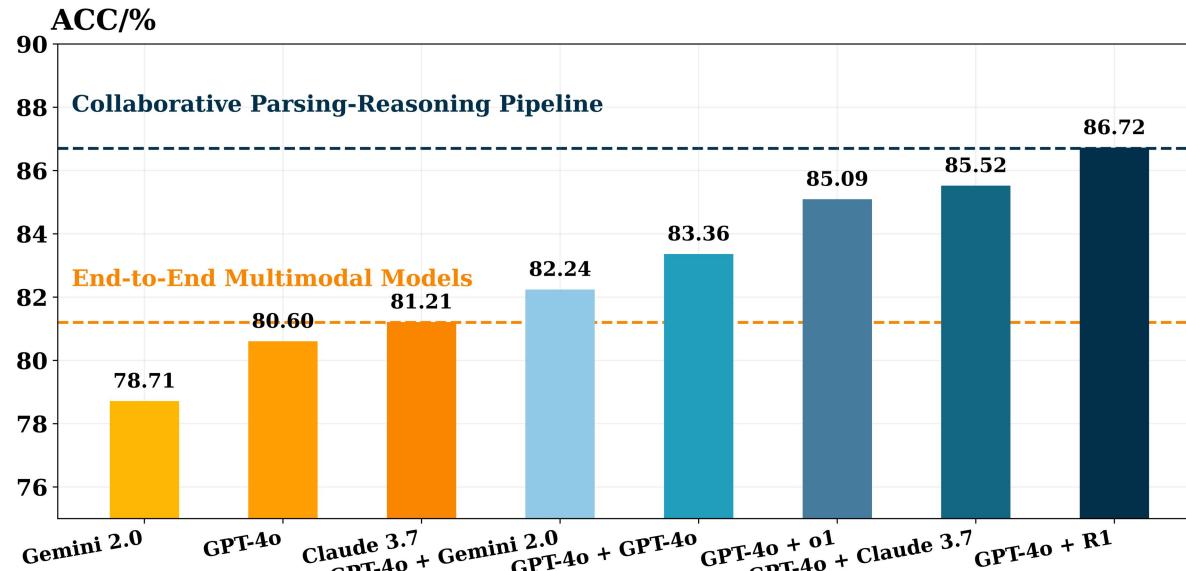
Claude 3.7 Sonnet: 83.5% → 86.3%

Gemini 2.0 Flash Thinking: 78.7% → 83.0%
(approaching Claude's level)

Reasoning Enhancement (3/3)

➤ Model Collaboration:

- Multimodal model (e.g., GPT-4o) as microscope for visual parsing
- Reasoning model (e.g., DeepSeek-R1) as expert for complex inference



On tabular QA tasks, model collaboration boosts accuracy from 80.6% → 86.7% (surpassing Claude 3.7 Sonnet)

Question: What is the anticipated portfolio return made by an investment company analyst named Maud, under two scenarios of portfolio returns in various economic conditions (expressed as a percentage) ? Answer to a single decimal place.

Image:

Scenario	Probability of scenario (%)	Portfolio return	Probability of return (%)
good economic situation	70	20%	50
good economic situation	70	10%	50
bad economic situation	30	5%	60
bad economic situation	30	-10%	40

Structured Data:

```
| Scenario | Probability of scenario (%) | Portfolio return | Probability of return (%) |  
| good economic situation | 70 | 20% | 50 |  
| good economic situation | 70 | 10% | 50 |  
| bad economic situation | 30 | 5% | 60 |  
| bad economic situation | 30 | -10% | 40 |
```

References



BUPT
Reasoning Lab

- [1] Guo D, Yang D, Zhang H, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning[J]. arXiv preprint arXiv:2501.12948, 2025.
- [2] Xu F, Hao Q, Zong Z, et al. Towards Large Reasoning Models: A Survey of Reinforced Reasoning with Large Language Models[J]. arXiv preprint arXiv:2501.09686, 2025.

Additional Resources



BUPT
Reasoning Lab

FinMMR

Make Financial Numerical Reasoning More Multimodal, Comprehensive, and Challenging

Zichen Tang, Haihong E*, Jiacheng Liu, Zhongjun Yang, Rongjin Li, Zihua Rong, Haoyang He, Zhuodi Hao, Xinyang Hu, Kun Ji, Ziyan Ma, Mengyuan Ji, Jun Zhang, Chenghao Ma, Qianhe Zheng, Yang Liu, Yiling Huang,

Xinyi Hu, Qing Huang, Zijian Xie, Shiyao Peng

Beijing University of Posts and Telecommunications, Beijing, China

ICCV 2025

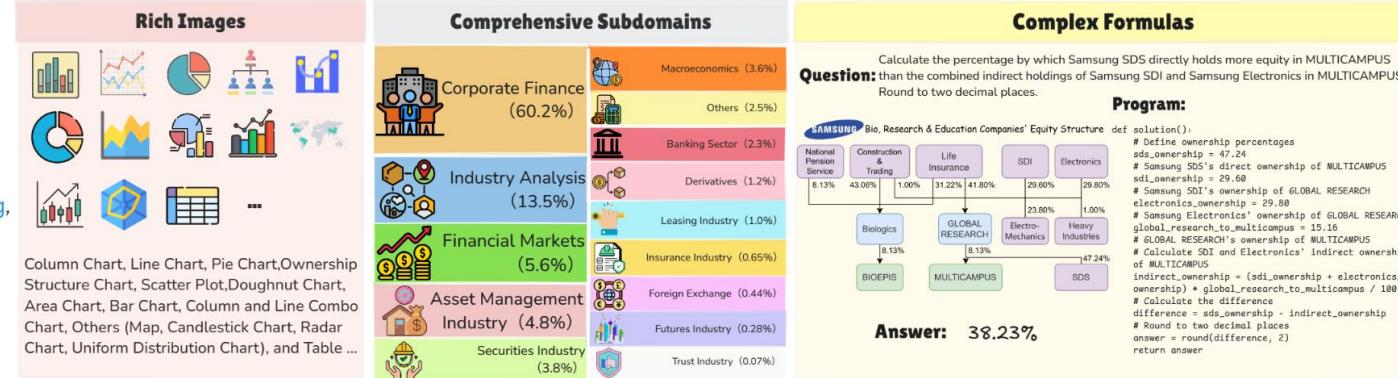
*Corresponding author.

Paper

Code

arXiv

Dataset



Overview of the FinMMR dataset. FinMMR presents three challenges: (1) visual perception: 8.7K financial images of 14 categories; (2) knowledge reasoning: 4.3K financial questions of 14 subdomains; (3) numerical computation: multi-step precise calculation.

Project Page

<https://bupt-reasoning-lab.github.io/FinMMR>

Arxiv Preprint

<https://arxiv.org/abs/2508.04625>

Code Repository

<https://github.com/BUPT-Reasoning-Lab/FinMMR>

Open Dataset

<https://huggingface.co/datasets/BUPT-Reasoning-Lab/FinMMR>

Contact with Us

Wechat: StarLight2224



Official Account



BUPT Reasoning

北京

Homepage: <https://zichentang.github.io>
Email: tangzichen@bupt.edu.cn



Zichen Tang
汤子辰

MS. Student @
BUPT

Major in Computer Science & Technology. Focusing on Domain-specific Knowledge Reasoning & Multimodal Large Language Model.

- 📍 Beijing, China
- ✉ Email
- 🐙 Github
- 🎓 Google Scholar

>About Me

I'm currently a MS. Student from [School of Computer Science \(SCS\), Beijing University of Posts and Telecommunications \(BUPT\)](#), supervised by [Prof. Haihong E](#). Before that, I received the bachelor degree at [School of Computer Science \(SCS\), Beijing University of Posts and Telecommunications \(BUPT\)](#) in 2023.

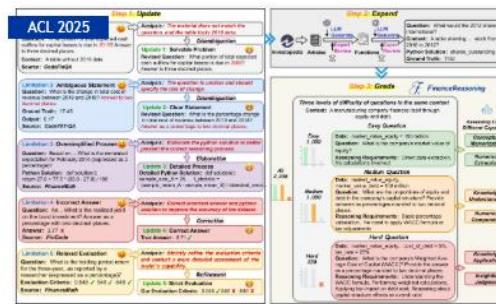
My research interest includes domain-specific knowledge reasoning and multimodal large language model. I have published 6 papers at the top international AI conferences such as ACL, ICCV, NeurIPS, AAAI, with total [google scholar citations 100+](#).

If you are seeking any form of **academic cooperation**, please feel free to email me with tangzichen@bupt.edu.cn.

News

- 2025.06: 🎉 One paper is accepted by ICCV 2025.
- 2025.05: 🎉 One paper is accepted by ACL 2025.

Publications



[FinanceReasoning: Benchmarking Financial Numerical Reasoning More Credible, Comprehensive and Challenging](#)

Zichen Tang, Haihong E*, Ziyan Ma, Haoyang He, Jiacheng Liu, Zhongjun Yang, Zihua Rong, Rongjin Li, Kun Ji, Qing Huang, Xinyang Hu, Yang Liu, Qianhe Zheng

Project

- We present FinanceReasoning, a new benchmark for evaluating large reasoning models (LRMs) in financial numerical reasoning, featuring enhanced rigor and an open-source function library.



Thank You!

F FinMMR: Make Financial Numerical Reasoning
F_M More Multimodal, Comprehensive, and Challenging

Zichen Tang Haihong E* Jiacheng Liu Zhongjun Yang Rongjin Li Zihua Rong
Haoyang He Zhuodi Hao Xinyang Hu Kun Ji Ziyan Ma Mengyuan Ji Jun Zhang
Chenghao Ma Qianhe Zheng Yang Liu Yiling Huang Xinyi Hu Qing Huang Zijian Xie

Shiyan Peng

Reasoning Lab, Beijing University of Posts and Telecommunications **BUPT**
<https://bupt-reasoning-lab.github.io/FinMMR/> **Reasoning Lab**

