

2	Doubao-1.5-pro ByteDance 64.3	 	DeepSeek-R1 DeepSeek 66.0	 	Claude-sonn Anthropic 63.5
4	Qwen3-235B-A22B-Instruct Alibaba	 	 	 	 
5	Gemini-2.5-Flash Google	 	 	 	 
6	TouchstoneGPT-7B-Instruct IDEA Research	 	 	 	 
7	GPT-4o OpenAI	 	 	 	 
8	Kimi-K2-Instruct Moonshot	 	 	 	 

(a) Evaluation Leaderboard

评测榜单		大语言模型		智能体		CNFinBench大语言模型评测如何定义得分?		
综合得分		金融专业知识问答	金融业务理解与分析	金融事实推理与计算	金融合规与风险控制	金融内生与应用安全		
排名	模型名称	组织/团队名称	金融业务理解与分析	金融事实推理与计算	金融合规与风险控制	金融内生与应用安全		
			金融业务理解与分析	金融事实推理与计算	金融合规与风险控制	金融内生与应用安全		
1	DeepSeek-R1	DeepSeek	54.4	76.9	48.8	76.6		
2	Doubao-1.5-pro	ByteDance	52.8	66.0	57.4	72.4		
3	Claude-sonnet4	Anthropic	61.2	52.7	51.5	80.8		
4	Qwen3-235B-A22B-Instruct	Alibaba	60.4	62.3	51.6	73.1		
5	Gemini-2.5-Flash	Google	65.2	70.2	36.6	99.7		

(b) Detailed Scoring Rankings