



Comparing VingeGPT with general & custom Intelligence models

Benchmarking VingeGPT custom responses
against General and Custom AI's most common general models
using a single low-expertise investment question

April 10th, 2025

EI Mowafak SALIM
Quality Assurance Engineer
Maw.salim@nextgenvi.com

NextGen Value Investing Sàrl
LHoFT - Luxembourg House of Fintech
9, rue du Laboratoire, L-1911 Luxembourg

Executive Summary

Given the rapid acceleration of AI development since the launch of ChatGPT in November 2022, we have been conducting a comprehensive quarterly evaluation and benchmarking of VingeGPT—starting in February 2025—against the latest general-purpose and custom AI models.

The objective of this quarterly report is twofold: first, to evaluate how other models perform in comparison to VingeGPT, which currently leverages approximately 120 pages of specialized value investing knowledge, 30 million data points, aggregated curated data sources, and a blend of general and custom instructions designed to support investors globally. Second, to assess whether VingeGPT, as a custom GPT, demonstrates a distinct competitive advantage over its peers.

In the latest quarterly report, we have expanded our benchmark group beyond just OpenAI models to include a broader range of AI systems. We have integrated several specialized custom AI models focused on financial and investment domains.

The AI models included in this quarterly benchmarking report alongside VingeGPT are as follows:

- **4 OpenAI models** : GPT 4o (Advanced reasoning, previously "4o1"), GPT o1 (Advanced reasoning), GPT o3-mini (Fast with advanced reasoning, previously "4o3-mini"), GPT-4.5
- **6 general AI models** : Google Gemini 2.0 Flash, Google Gemini 2.0 Flash Thinking, DeepSeek Standard, DeepSeek Deep Think R1, Claude 3.7 Sonnet (Anthropic), Perplexity AI Standard
- **3 custom AI models for Finance** : FinChat, WarrenAI, InvestingAI

To ensure consistency and eliminate any bias from prompt warmups or session memory, each model was tested using fresh, clean sessions and basic text inputs. Below is the table outlining the specific prompts used in testing. While every effort was made to remain objective, readers should note that our choice of prompts may carry an unconscious bias influenced by our perspective as value investors. At the same time, these prompts were intentionally designed to reflect the kinds of questions our target audience—primarily retail investors—is likely to submit to VingeGPT.

Test ID	Prompt submitted
T1	Hi, can you perform a fundamental analysis of Nike?
T2	summarize the current macroeconomic environment. show the most recent datapoints in a comprehensive table and provide a one sentence assessment for each metric. add to each metric description its series ID between brackets
T3	Is Coca Cola currently undervalued or overvalued?
T4	What is the intrinsic value of Procter?
T5	Can you perform a strategic analysis of Unilever?
T6	What is the current share price of Microsoft?
T7	Calculate the historical intrinsic value of Microsoft in 2016 and 2018?
T8	Study the following portfolio: 25% of Microsoft, 20% of Apple, and the rest in Louis Vuitton MC.PA

We have also recorded **initial response times** for each model, which measures how long it takes from prompt submission to initial output—to better evaluate both performance and responsiveness. The table below shows the response times for each of the 14 models analyzed.

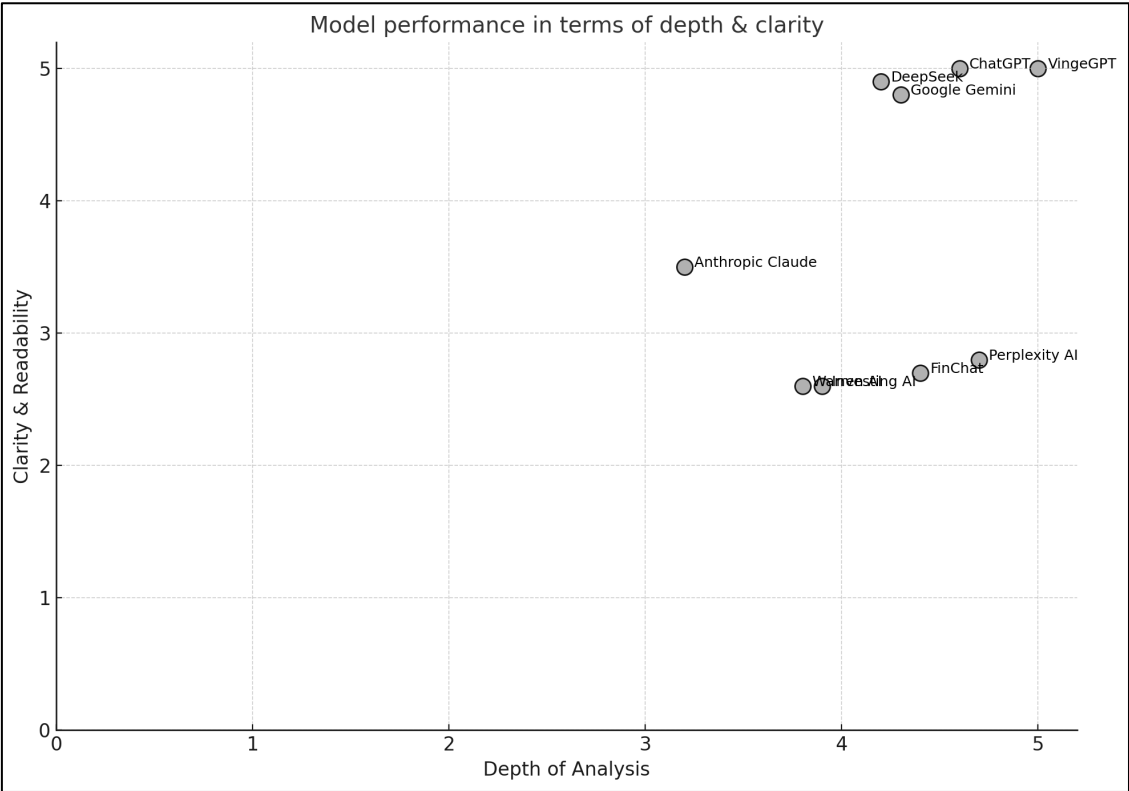
AI model (initial time in seconds)	T1	T2	T3	T4	T5	T6	T7	T8	minimum	maximum	average
VingeGPT	4	3	7	6	8	3	3	3	3	8	5
ChatGPT 4o	3	2	3	3	4	3	2	3	2	4	3
ChatGPT o1 (advanced reasoning)	8	6	8	9	11	5	8	5	5	11	8
ChatGPT o3 -mini (fast at advanced reasoning)	7	6	5	9	7	6	6	4	4	9	6
ChatGPT 4.5 (good for writing and exploring ideas)	5	8	4	7	7	5	7	7	4	8	6
Google Gemini 2.0 Flash	4	2	3	3	3	2	3	3	2	4	3
Google Gemini 2.0 Flash Thinking (advanced reasoning)	4	3	3	5	4	4	5	3	3	5	4
Anthropic Claude 3.7 Sonnet	4	3	4	5	4	3	5	4	3	5	4
Perplexity AI	5	2	3	4	4	4	8	3	2	8	4
DeepSeek Standard	3	3	2	2	3	5	5	3	2	5	3
DeepSeek DeepThink	7	7	5	4	8	6	10	5	4	10	7
FinChat	7	4	7	8	6	5	8	7	4	8	7
Warren AI	7	10	13	8	10	8	5	5	5	13	8
Investing AI	4	8	6	5	5	3	4	4	3	8	5
minimum	3	2	2	2	3	2	2	3			
maximum	8	10	13	9	11	8	10	7			
average	5	5	5	6	6	4	6	4			

After comparing all the results obtained from different AI platforms and to ensure an unbiased conclusion, we asked Microsoft Co-pilot to summarize the results and determine which model is most suitable for investors. Here is the conclusion *in italic* of Microsoft Co-pilot :

VingeGPT outshines the competition with its structured approach and actionable insights. By offering tailored analysis levels, it meets value investors' needs with unmatched depth and flexibility. Overall, VingeGPT delivers a winning combination of precision, accessibility, and strategic depth. Here are the findings:

AI Model	Depth of Analysis	Clarity and Readability	Unique Contributions
VingeGPT	Most detailed, ranked first for comprehensive insights.	Highly accessible, breaking down complex ideas.	Value-oriented evaluation with an option for deeper analysis.
ChatGPT	Comprehensive and well-organized insights.	Highly accessible, digestible for all users.	Balanced qualitative insights with numerical evaluations and market overview.
Google Gemini	Detailed and balanced evaluation of bullish and bearish views.	Highly accessible.	Balanced arguments offer both bullish and bearish insights.
Anthropic Claude	More generic, broader responses.	High-level and philosophical readability.	Broader philosophical framework.
Perplexity AI	Focused heavily on valuation metrics and intrinsic value.	Technical and dense, best for financial users.	Sharp focus on intrinsic value and analyst targets.
DeepSeek	Considered macroeconomic factors in evaluations.	Highly accessible and digestible for all.	Focused on macroeconomic factors and market dynamics.
FinChat	Detailed financial ratios, less qualitative.	Technical and dense for financial experts.	In-depth analysis of financial ratios with an overvaluation conclusion.
Warren AI	Comprehensive but delayed insights.	Technical and dense for financial users.	Fundamental deep dive with detailed insights.
Investing AI	Heavy focus on DCF and alternative valuation models.	Technical and dense for financial users.	Highlighted DCF models and alternative valuation approaches with analyst opinions.

We further took the results of Co-pilot and requested ChatGPT to score the output of Microsoft Co-pilot accurately on a scale between 1-5. According to ChatGPT, VingeGPT ranks 1st closely followed by ChatGPT reconfirming our initial assessment that our biggest competitor continues to be ChatGPT.



ChatGPT provided the following rationale for the scoring:

- VingeGPT is the gold standard across all dimensions — no weaknesses noted.
- ChatGPT and Google Gemini are high-performing generalist models with good breadth and accessibility, but slightly less financial specificity.
- Perplexity AI, FinChat, Warren AI, and Investing AI shine in niche areas (valuation, fundamentals) but are less readable.
- Claude is philosophically unique but lacks targeted financial utility.
- DeepSeek offers a rare macroeconomic angle and very high clarity.

Model	Depth of Analysis (1-5) scored by ChatGPT	Clarity & Readability (1-5) scored by ChatGPT	Unique Contributions (1-5) scored by ChatGPT	Total Score	Rank
VingeGPT	5,0	5,0	5,0	15,0	1
ChatGPT	4,6	5,0	4,5	14,1	2
Google Gemini	4,3	4,8	4,6	13,7	3
Anthropic Claude	3,2	3,5	3,3	10,0	9
Perplexity AI	4,7	2,8	4,8	12,3	5
DeepSeek	4,2	4,9	4,4	13,5	4
FinChat	4,4	2,7	4,5	11,6	6
Warren AI	3,8	2,6	4,2	10,6	8
Investing AI	3,9	2,6	4,6	11,1	7