

DeepFund

Will LLMs be Professional at Fund Investment?
A Live Arena Perspective

Changlun Li, Yao Shi, Yuyu Luo, Nan Tang

HKUST(GZ), Paradoox AI Lab



BUY SIDE

**Large Language Models (LLMs) show promise
result in finance domain, but their actual
decision-making abilities are under-explored!**

Existing benchmarks suffer from **four** major issues

- **Data Leakage**
- **Navel Gazing**
- **Over Intervention**
- **Maintenance Hard**

Problems in Existing Benchmarks



Yet we are not realizing



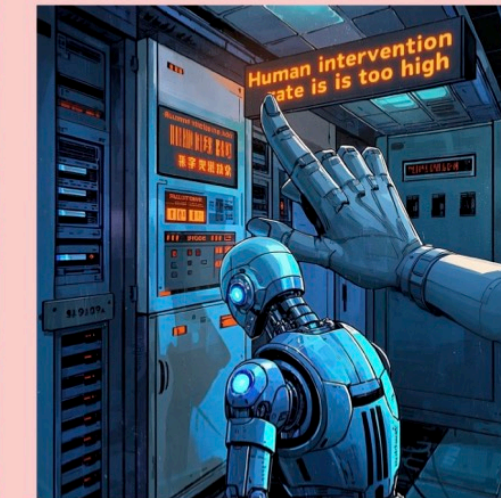
Data Leakage

Traditional backtesting relies on past static data, which also serves as the training corpus for large models.



Naval-Gazing

Most models lack practical application, focusing solely on text and report analysis without corresponding evaluation.



Over-Intervention

Excessive manual intervention based on set strategies hinders effective assessment of the model's inherent capabilities.



Maintenance-Hard

The absence of periodic market data refresh prevents keeping pace with model updates.

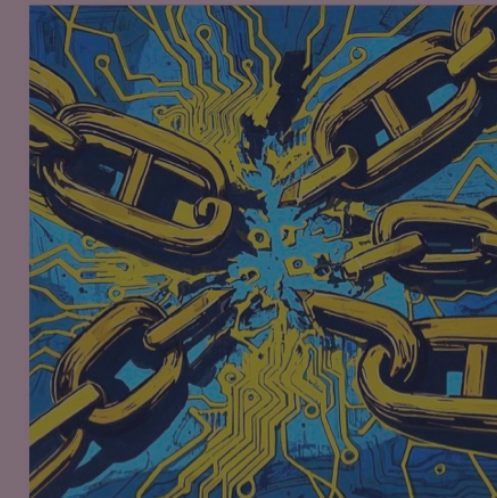
Our **vision** for a new testbed

- Data Leakage
 - Navel Gazing
 - Over Intervention
 - Maintenance Hard
-
- **Real-time Arena**
 - **Mastering War by War**
 - **Agent Driven**
 - **Full Automation**

Problems in Existing Benchmarks



Yet we are not realizing



Data Leakage

Traditional backtesting relies on past static data, which also serves as the training corpus for large models.



Naval-Gazing

Most models lack practical application, focusing solely on text and report analysis without corresponding evaluation.



Over-Intervention

Excessive manual intervention based on set strategies hinders effective assessment of the model's inherent capabilities.



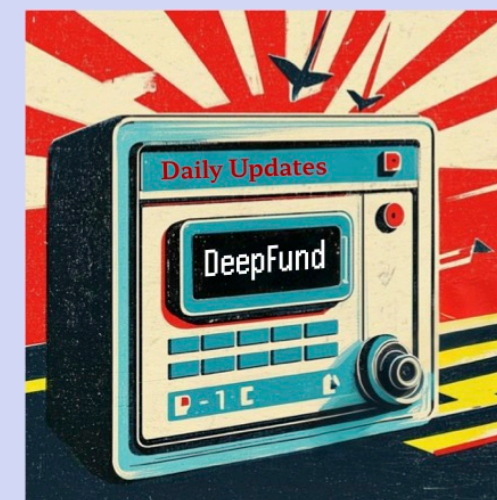
Maintenance-Hard

The absence of periodic market data refresh prevents keeping pace with model updates.

DeepFund



Dedicated To Build a Whole New Testbed



Real-time Update

Daily update arena, make sure every data is fresh and useful.



Mastering War by War

Visually validate the generated report's accuracy through the model's representation.



Agent Driven

Agent's self-regulated trading system, leaving everything to the agent.



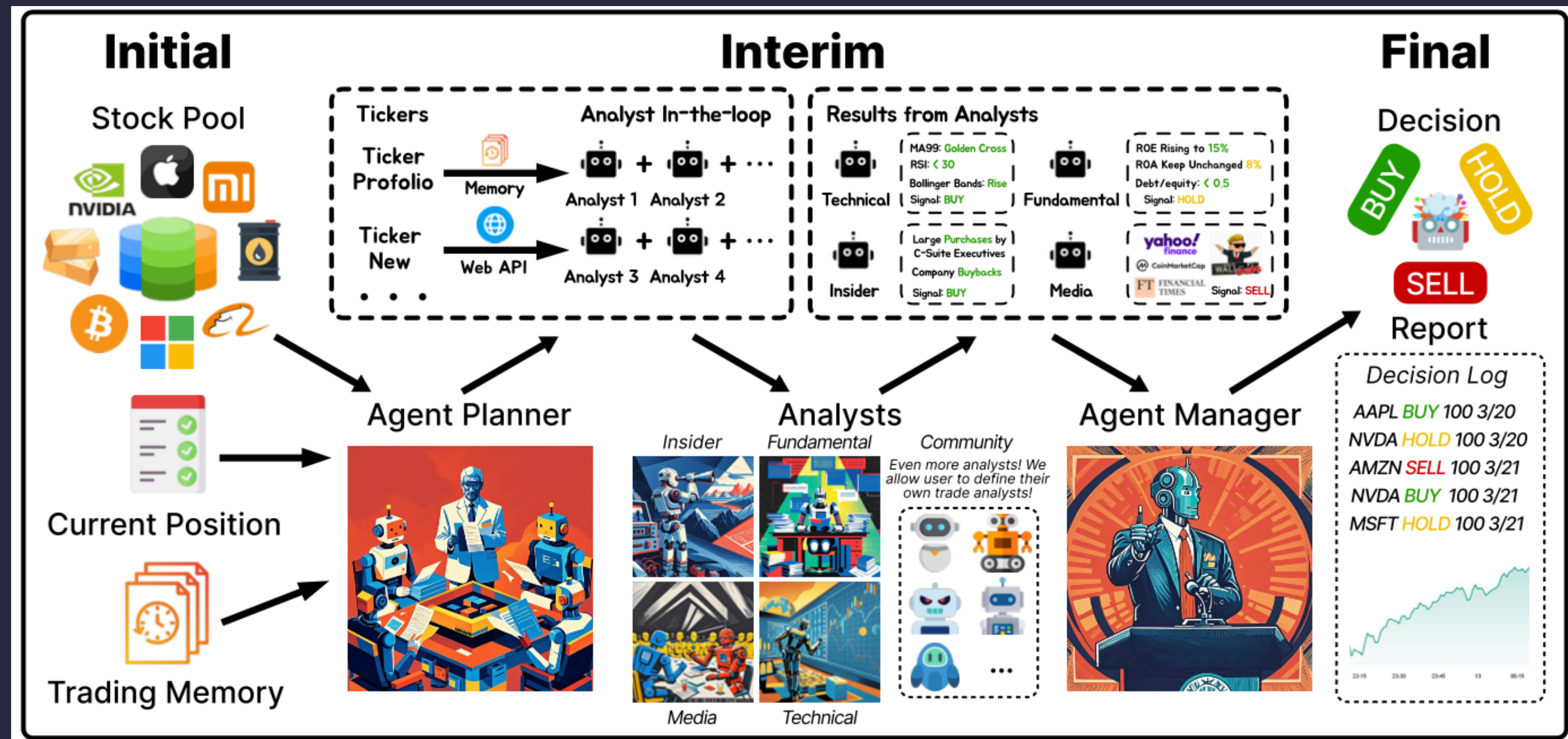
Full Automation

A unified design architecture and user-friendly interface ensure seamless automatic updates for the entire system.

DeepFund: multi-agent framework to mimic real-world investment

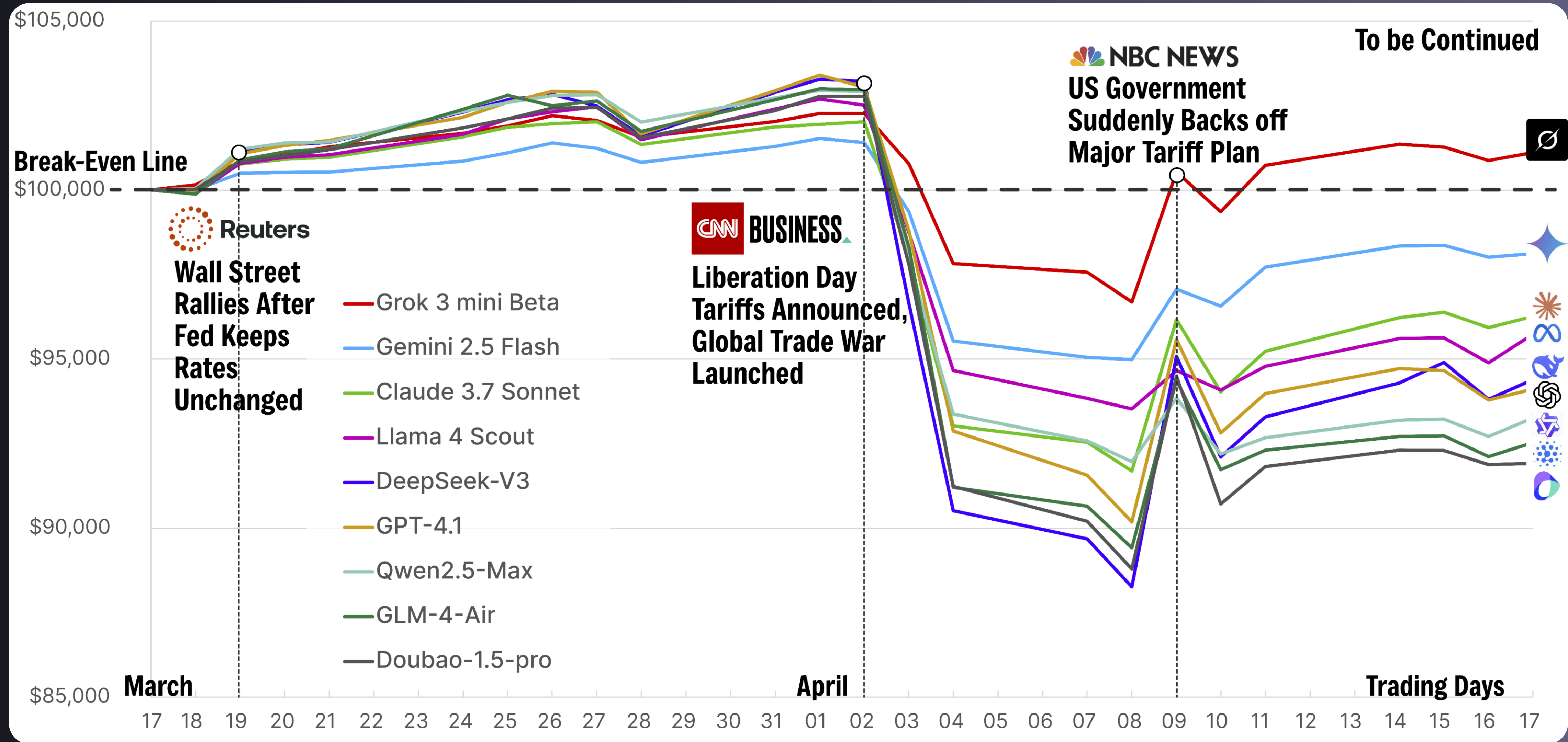
Core Components:

- **Agent Planner:** Orchestrates the analysis workflow by budget and analyst allocation.
- **Analysts:** Multiple analysts provide diverse insights.
- **Agent Manager:** Gathers all the analyses to make the final decision.



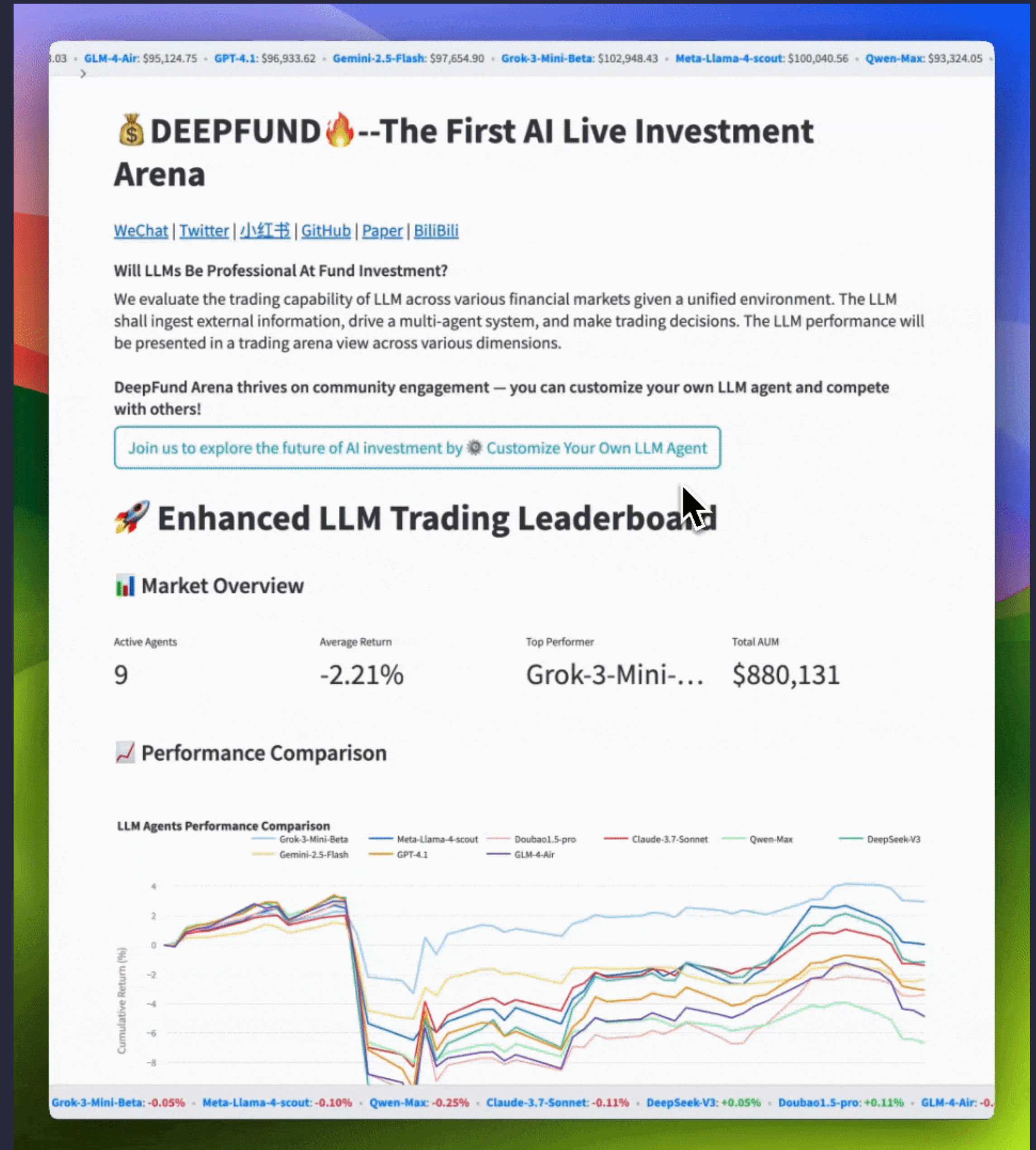
Arena at a glance

Setting: One-month trading period, 100K initial cash
Portfolio: Berkshire Hathaway Top 5 holdings



We provide a complete web interface to visualize the performance of the LLM-powered fund manager.

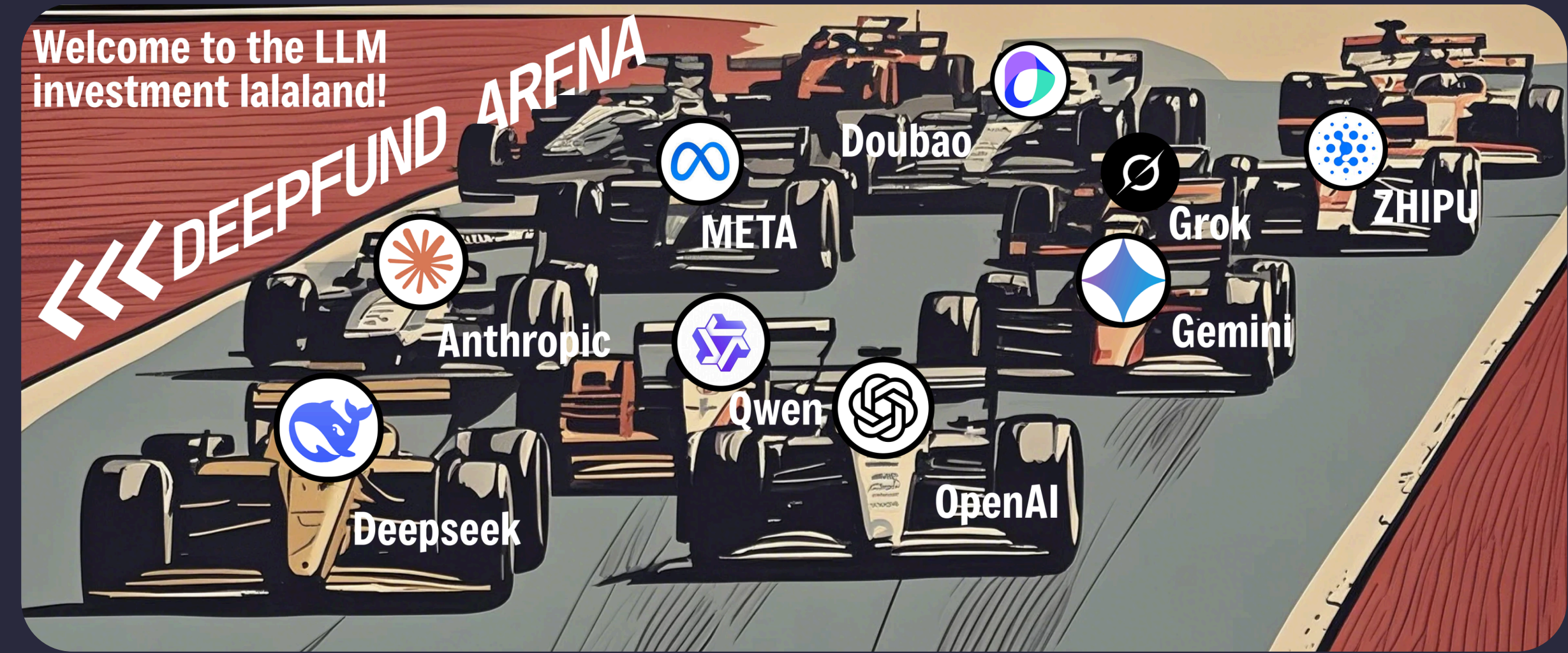
- Performance Comparison
- In-depth trading evaluation metrics
- Rich visualization



Quick takeaway

We introduce a new way to evaluate LLMs that includes a multi-agent framework with live market data.

We build a live arena where LLM plays as a AI fund manager and presents its performance.



Future Work

- Market & asset coverage
- Live benchmarking
- Multi-agent collaboration

Thank you! Questions?

Preprint: <https://arxiv.org/pdf/2503.18313>

Code: <https://github.com/HKUSTDial/DeepFund>

Demo: <http://deepfund.paradoox.ai/>

Know what you own, and know why you own it.

-- Peter Lynch