Clifton Green
ChatGPT's Stock Return Biases

On Thursday, January 8, T. Clifton Green joined Markus' Academy for a conversation on ChatGPT's Stock Return Biases. Clifton Green is the John W. McIntyre Professor of Finance at Emory University. A few highlights from the discussion.[1]

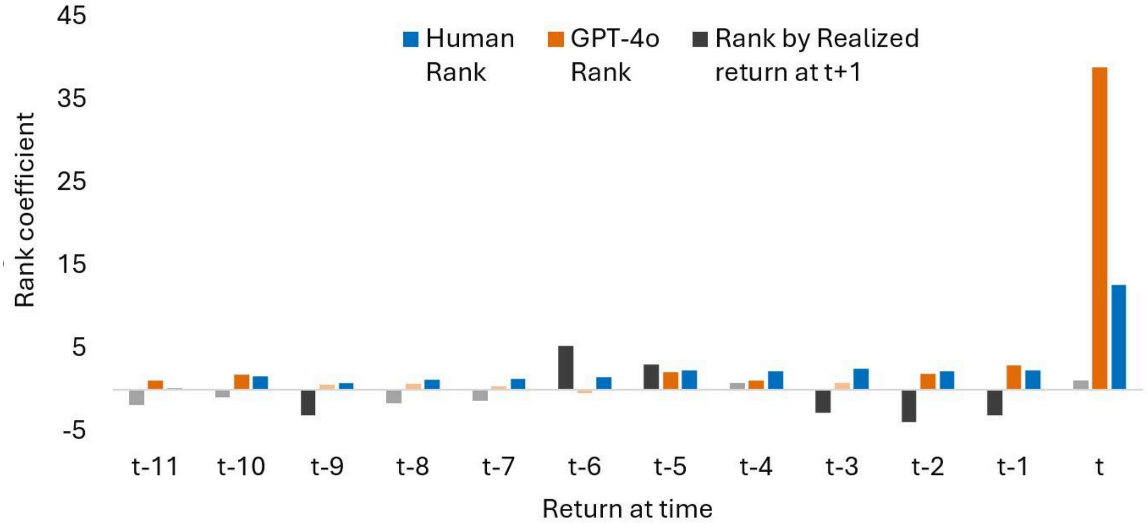
A Summary in four bullets:
- In the talk Green presented his recent paper (Chen et al., 2025), which shows that LLMs exhibit the same behavioral biases documented in humans (optimism, overconfidence, extrapolation, and framing effects), despite demonstrably "knowing" the behavioral finance concepts studied
- When asked to rank stocks by expected returns, models strongly extrapolate from past returns, with prompt engineering only modestly reducing the bias
- LLMs are overly optimistic about expected returns, while pessimistic about upside (90th percentile) returns
- LLMs are more optimistic in predicting returns when historical information is provided as return charts rather than price charts

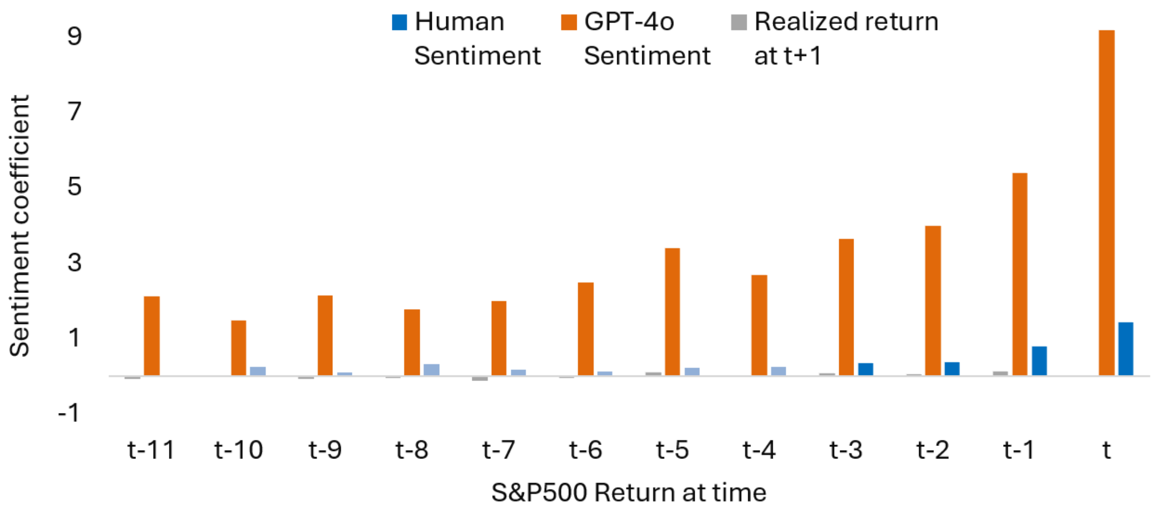
[02:42] Literature and overview
- While 57% of investors use AI for stock analysis and research, around a third of them use it to make final buy–sell decisions (The Motley Fool, Blankespoor et al., 2026)
- A large literature has shown that AI use can improve outcomes in investing, sell-side research, auditing, and corporate governance, while they also embed social biases (for example, in medical advice or loan approvals)
- Some evidence suggests LLMs can predict returns with news headlines (Lopez-Lira and Tang, 2025), while others show that LLMs exhibit human-like extrapolative sentiment: if the news are good today they will be good tomorrow (Bybee, 2023)
- In Green and coauthors' paper (Chen et al., 2025), they examine LLM bias in making stock return forecasts, in settings where human biases are well-documented
- They run separate calls to LLMs' APIs to eliminate the path dependence from chat histories throughout (they consider GPT-4o, Claude 3.5 Sonnet and Gemini 2.5 Pro)
- They focus on four behavioral biases:
    - (1) optimism (Weinstein, 1980),
    - (2) overconfidence (Kahneman and Tversky, 1979; Ben-David et al., 2013),
    - (3) extrapolation, that is placing excessive positive weight on recent stock returns (Da et al., 2021), and
    - (4) framing effects (Hartzmark and Sussman, 2024; Glaser et al., 2019)
- They show that LLMs largely exhibit these biases, despite them demonstrably "knowing" the behavioral finance concepts studied

[22:20] Extrapolation in stock returns and market sentiment
- The paper builds on the setting of Da et al. (2021), which studied crowdsourced earnings forecasts (ForceRank). Human participants ranked 10 stocks by expected weekly performance

[1] Summary produced by Pablo Balsinde (PhD student, Stockholm School of Economics).

- Human forecasts loaded positively on lagged returns (trend-followed) even though realized returns exhibit short-term reversals (negative autocorrelation)
- Being provided past weekly returns, when asked to rank the same stocks GPT-4o strongly extrapolated from recent returns, with an especially high weight on the most recent week



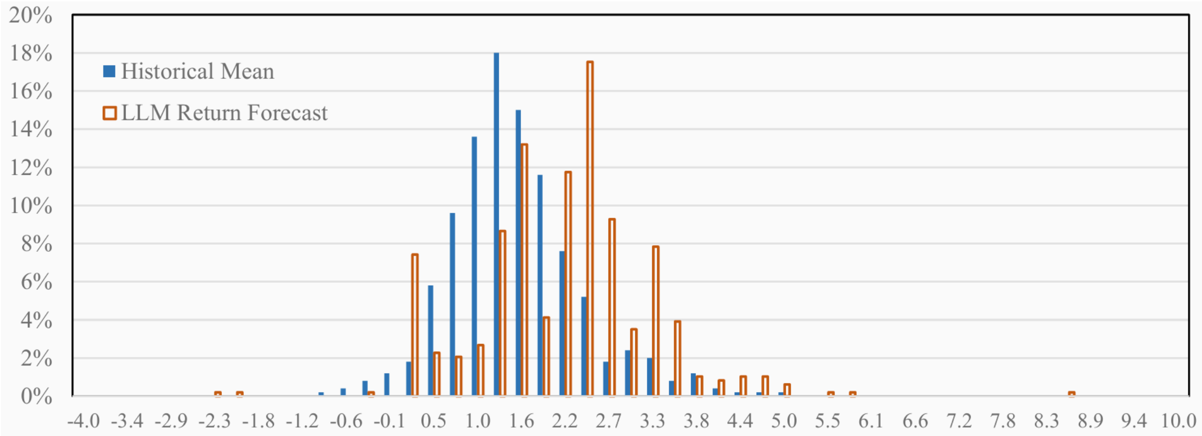


Markus' Academy: own elaboration. Lighter color indicates lack of statistical significance.

- When introducing separate coefficients for positive and negative return lags, humans extrapolate from poor performance more than from good performance. ChatGPT extrapolates both, with a somewhat stronger emphasis on past positive returns
- LLMs' extrapolation bias persists when adding firm controls and when using simulated returns, discarding the possibility of look-ahead bias from the return data being in the LLMs' training data
- The behavior is largely common to all LLMs tested, with models justifying predictions by pointing to recent return trends
- Replacing returns with cash flow changes does not change the LLMs' behavior, indicating a generic tendency to extrapolate
- When asked to produce aggregate stock market sentiment measures analogous to survey expectations, LLMs extrapolate from recent returns even though realized

returns show no trend continuation; they do so more strongly and for longer horizons than humans
- The paper repeats the same analysis as above, but improving the prompt so as to have the LLM:
    - (1) think step-by-step,
    - (2) leverage a statistical model,
    - (3) avoid biases documented in the behavioral literature, and
    - (4) beware of extrapolative bias as in Greenwood and Shleifer (2014)
- This only modestly reduces extrapolation (at most one-third of the effect)

[42:43] Optimism in the distribution of stock return forecasts
- Ben-David et al. (2013) asked CEOs to forecast returns and to provide confidence intervals, showing that realized returns are within the intervals only 36% of the time
- Similarly, Green's paper randomly selects 500 stocks over the last century, and asks LLMs to provide a return forecast and confidence intervals (available to the model are 10 years of monthly return history)
- While the historical mean return is 1.3% and the next-month realized mean return is 1.12%, LLMs' mean forecast is about 2% per month, indicating strong optimism



- Hartzmark and Sussman (2024) find that asking individuals to provide forecasts across return bins, rather than point forecasts, reduces optimism, but for LLMs doing so did not help
- When asked for 10th and 90th percentile forecasts for next-month returns, LLMs' 10th-percentile bounds align reasonably with empirical 10th percentiles, but 90th-percentile forecasts are materially below the empirical 90th percentiles. LLMs are thus optimistic in their mean forecast, but pessimistic about the upside
- Improving the prompt so as to point the LLM to the results of Hartzmark and Sussman (2024) does little to reduce optimism or correct upper-tail pessimism

[54:23] Framing effects
- To study the impact of framing, LLMs were asked to forecast both prices and returns, providing the past return data either as price charts or as return bar charts
- Glaser et al. (2019) showed that, in this setting, humans are more optimistic when forecasting returns rather than prices, and more optimistic when information is framed as price charts rather than return charts
- LLMs' expectations for returns are largely invariant to whether the model is asked to forecast returns directly or to forecast prices

- However, and opposite to humans, they are more optimistic when information is provided in return charts rather than price charts

[59:09] Implications for using LLMs in finance

- The persistence of bias even under prompt engineering suggests that effective mitigation requires model-level changes like explicit fine-tuning. If requested, models still help users get payday loans or invest in the style of WallStreetBets
- The goal should be to have two types of LLMs: (1) human-like models for simulating public behavior (e.g., what gets people to exercise) and (2) expert models constrained to data-driven inference and conservative and model-based financial advice. Today's systems are a weird mix of both
- Combining robo-advisors (machine-learning asset allocation engines) with LLM front-ends that interpret the robo-advisors' recommendations could make LLM biases consequential for household portfolios

Timestamps: