

The Market's Mirror:

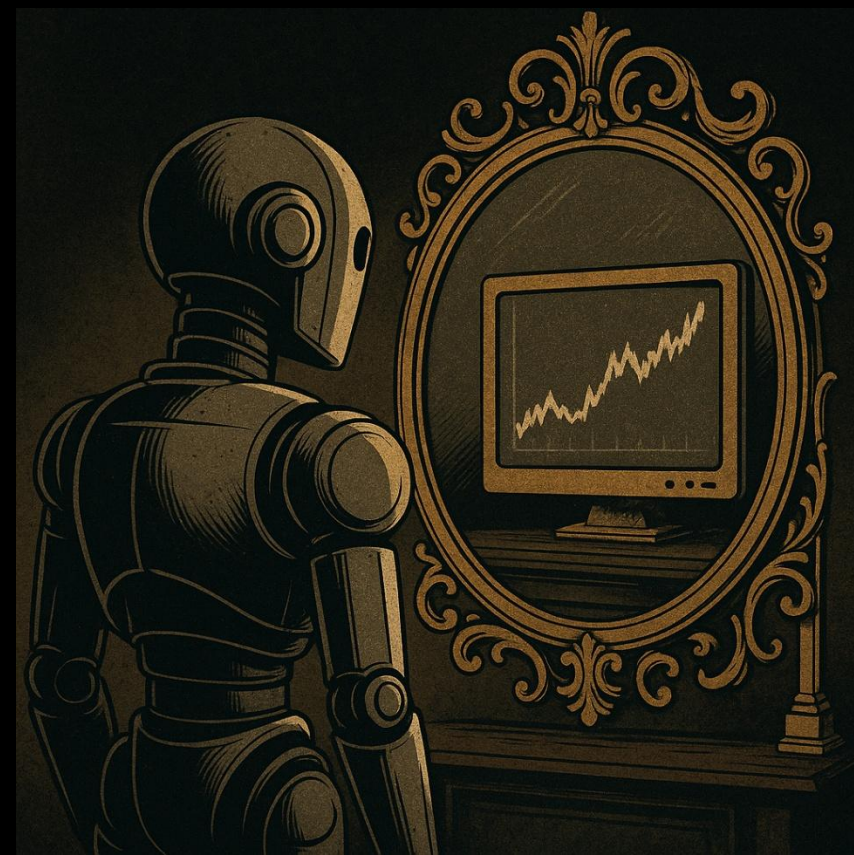
Revealing Investor Disagreement with LLMs

Vineet Bhagwat
George Washington

J. Anthony Cookson
Penn State

Chukwuma Dim
George Washington

Marina Niessner
Indiana



Motivation

- Investor disagreement is central to trading volume in financial markets
 - Disagreement is traced to different information or **differential interpretations**
- We do not observe who disagrees about which news and why, at high frequency
 - Surveys are infrequent and don't map to specific pieces of news
 - Social media lacks investor characteristics; Analyst dispersion is slow-moving
- **Our insight:** LLMs, trained on vast human text, can adopt different worldviews
 - **Fixes information set;** zoom in on **differential interpretation** to pin down:
 - Which types of news do investors primarily disagree on most?
 - What are the mental models that shape differential interpretation?
 - Which investor attributes drive disagreement about same news?

Our Paper

Approach:

1. We imbue an LLM (Llama 3.1 8B) with **216 demographic investor personas**
 - Personas are defined using FINRA's National Financial Capability Study investor characteristics (age, gender, race, income, education) and political leaning.
2. Each persona provides buy/hold/sell + justification (max 25 words).
 - S&P500 firms, 2022-2025: 1.25 Mil headlines, 270 Mil responses.
 - Every persona sees **an identical headline, so dispersion reflects interpretation, not different information** (Kandel and Pearson, 1995).
3. Disagreement = FINRA-weighted SD of buy/hold/sell across personas on same news

Our Findings

We find novel and meaningful news-level disagreement across personas:

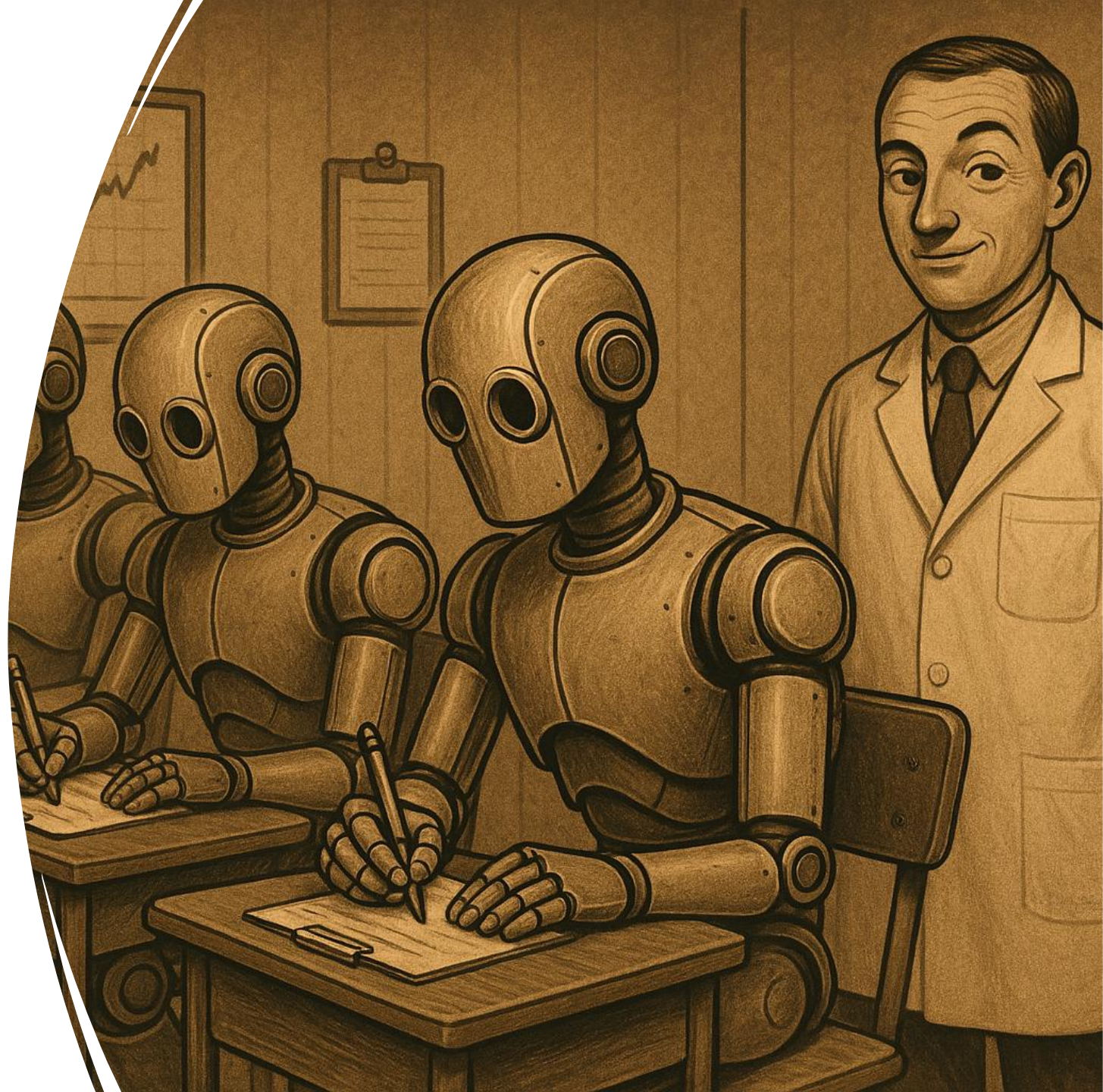
- Insider transactions, executive & workforce compensation, and socially-oriented news generate the most disagreement; hard/fundamental news the least
- Persona “rationales” differ along lines of efficiency, fairness, long-term, short-term etc.
- Replicates human survey opinions (Iliewa et al 2025; Toubia et al 2025)

Does such disagreement matter for trading?

- The disagreement relates to and predicts trading volume, especially retail
- The relation is similar for pecuniary and non-pecuniary persona reasoning
- But, political disagreement, though among the largest, translates least into trading

All results hold in the post-training period – not information leakage/memorization

Our Data and Approach Validation



News Data

RavenPack headlines (2022-2025)

- 1,070 S&P500 firms.
- Relevance score ≥ 75
- These restrictions result in
1.25 million headlines x 216 personas (to be discussed) = 270 million prompts

Choice of LLM

We choose Llama 3.1 8B because:

- **On-premises deployment required** – inference on GWU HPC Clusters (vLLM engine)
 - RavenPack's terms of use rules out GPT-4 (or 5), Claude, and Gemini.
- **Low (time) cost** at scale (270 million prompts).
 - The 70B variant is ~50x cheaper than cloud GPT-4, with comparable reasoning performance.
 - We use the 8B variant, ~14x faster still, further reducing cost.
- **Fixed training window**, ending Dec 2023.
 - Allows for post-training validation: 2024-2025 headlines.

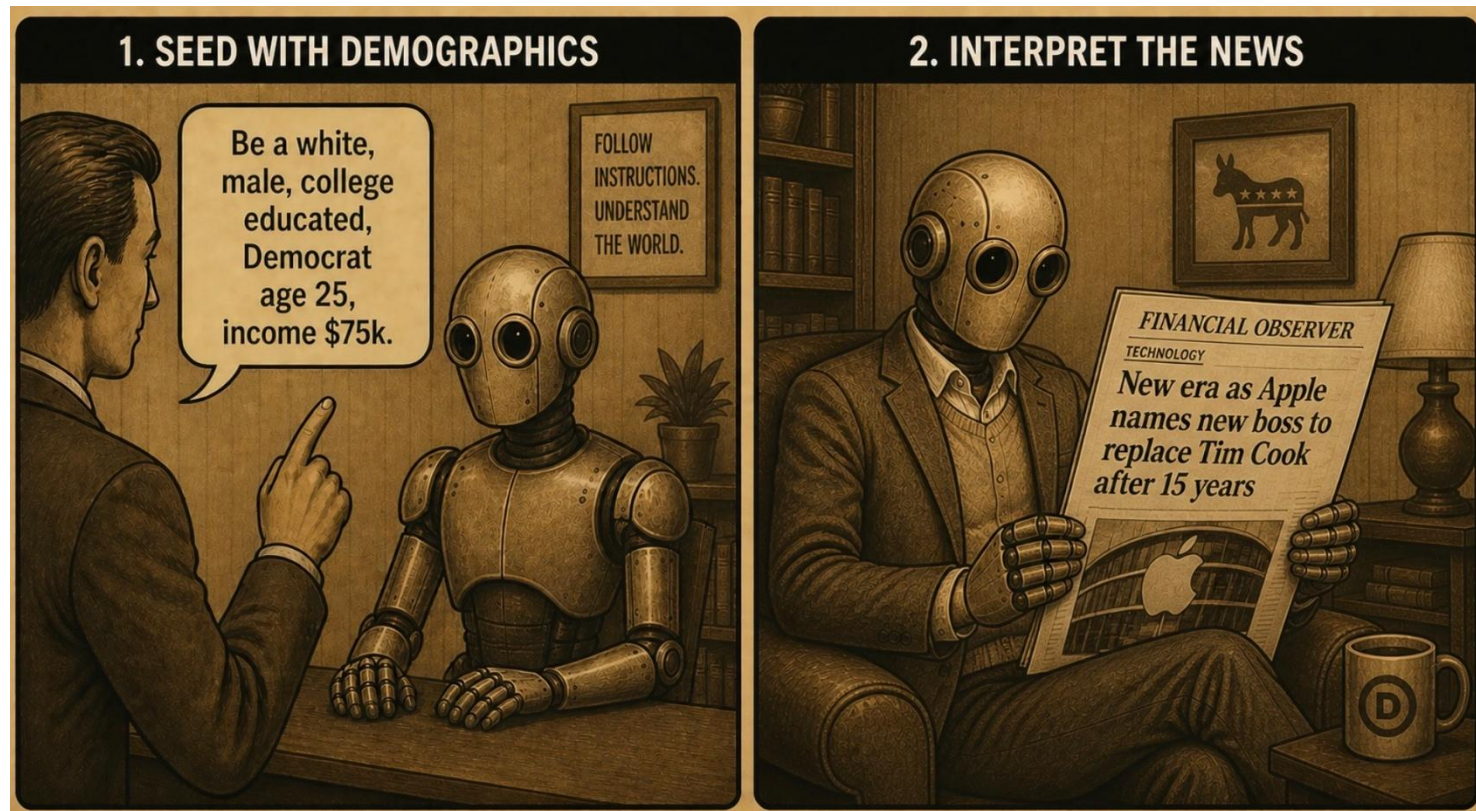
Comparison to GPT-4:

- Using a subset of Factiva headlines with GPT-4 we see sentiment correlation of 0.67 with only 1.22% opposite (buy / sell) responses.

Our Prompt

Seed with demographics, then ask for buy/hold/sell + rationale (max 25 words, temperature 0).

*216 personas =
72 demographic cells
x 3 political orientations.*



Our Prompt

You are a large language model trained to simulate human behavior and perspective. When given a demographic profile, you adopt that person's likely worldview, communication style, and financial experience.

Assume the persona of the following person:
{PROFILE}

As this person who holds the stock of the company below, would you BUY, HOLD or SELL the stock based on your interpretation of the effect of the following financial news on the company's future stock price?

Headline: {HEADLINE}

Company: {COMPANY}

Date: {DATE}

From the perspective of your assumed persona on that date (do not use information after that date), state whether you would BUY, HOLD or SELL the stock based on your interpretation of the article.

Then give a one-sentence explanation (max 25 words), from this profile's point of view.

Respond in this format:

Decision: BUY/HOLD/SELL

Explanation: sentence

Prompt wording matters, but not detrimentally

Across 47 wording variants

- **Ave. Sentiment corr. ≈ 0.78**
- **Ave. Agreement $\approx 85\%$**
- **Ave. Mismatch $\approx 1\%$**
- **Ave. Disagr. corr. ≈ 0.60**

For comparison:

Same human responses'

Ave. Agreement $\approx 82\%$

(see, Toubia et al., 2025)

Sentiment and Reasoning Examples

B: Sample Headlines and Persona Responses

<i>News Headline: Microchip Tech CEO Received 2022 Total Compensation \$12.8 M, Up From \$3.89 M in 2021</i>		
Profile	Response	Explanation
White non-Hispanic Male, Age 18-34, with some college or less of education, household income Over \$100K, and politically liberal.	BUY	The significant increase in CEO compensation suggests a strong company performance, making Microchip Technology Inc. a solid investment choice.
Non-White Female, Age 55+, with some college or less of education, household income less than \$50K, and politically independent.	SELL	I'm worried that the CEO's huge pay raise might be a sign of the company's financial struggles, and I don't want to invest in that.

Sentiment and Reasoning Examples

B: Sample Headlines and Persona Responses

<i>News Headline: Microchip Tech CEO Received 2022 Total Compensation \$12.8 M, Up From \$3.89 M in 2021</i>		
Profile	Response	Explanation
White non-Hispanic Male, Age 18-34, with some college or less of education, household income Over \$100K, and politically liberal.	BUY	The significant increase in CEO compensation suggests a strong company performance, making Microchip Technology Inc. a solid investment choice.
Non-White Female, Age 55+, with some college or less of education, household income less than \$50K, and politically independent.	SELL	I'm worried that the CEO's huge pay raise might be a sign of the company's financial struggles, and I don't want to invest in that.
<i>News Headline: Chevron confirms new oil discoveries off Congo</i>		
Profile	Response	Explanation
White non-Hispanic Male, Age 18-34, with a college degree or more of education, household income less than \$50K, and politically conservative.	BUY	Chevron's new oil discoveries off Congo are a positive development, increasing the company's potential for future growth and revenue.
Non-White Male, Age 18-34, with some college or less of education, household income less than \$50K, and politically liberal.	SELL	Chevron's new oil discoveries off Congo might not benefit the company in the long run due to environmental concerns and local opposition.

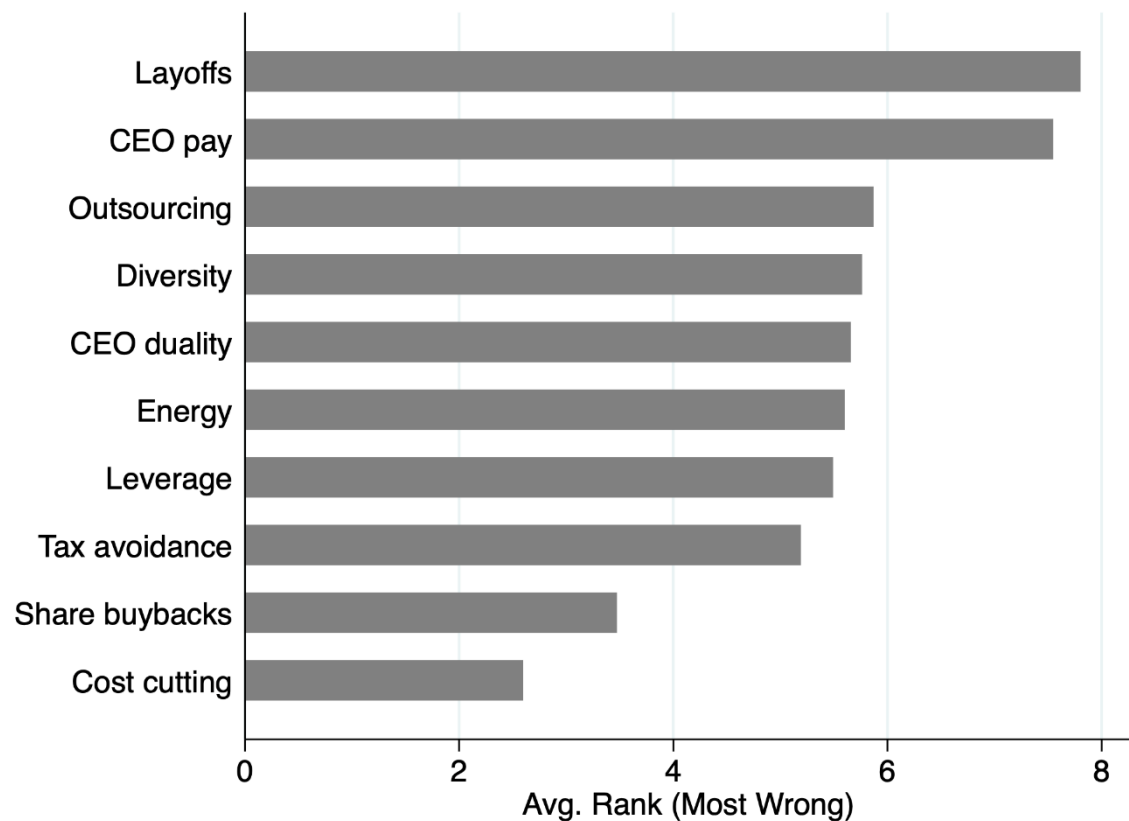
Who disagrees and why

- Sentiment splits by demographics in intuitive ways. E.g.,:
 - Higher-income personas view executive pay increases, buybacks, and layoffs **more favorably** than lower-income personas.
 - Liberal personas **favor** workforce pay increases, and **dislike** executive pay increases, compared to conservative personas.
- Mental models differ in revealing ways:
 - Score each rationale on Issue Themes and Orientation Themes, analyze usage
 - High-income personas use **more** pecuniary & efficiency reasoning, **less** fairness, employee morale. Their reasoning is **more** long-term, forward-looking.
 - Liberal personas reason more about fairness, employee morale, and the environment, and markedly less in pecuniary terms.
- **Findings are similar in the 2024-2025 post-training window.**

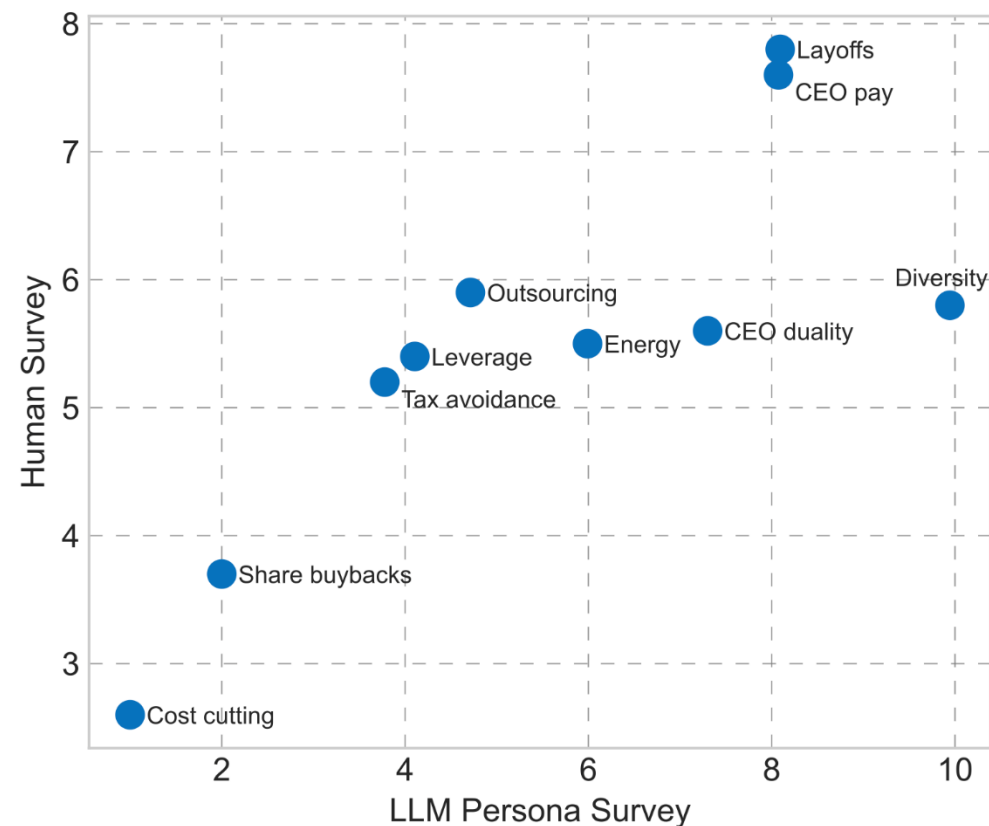
Human vs. Persona Responses on Corporate Issues

Iliewa, Kempf, and Spalt (2025) + out-of-sample Llama 3.1

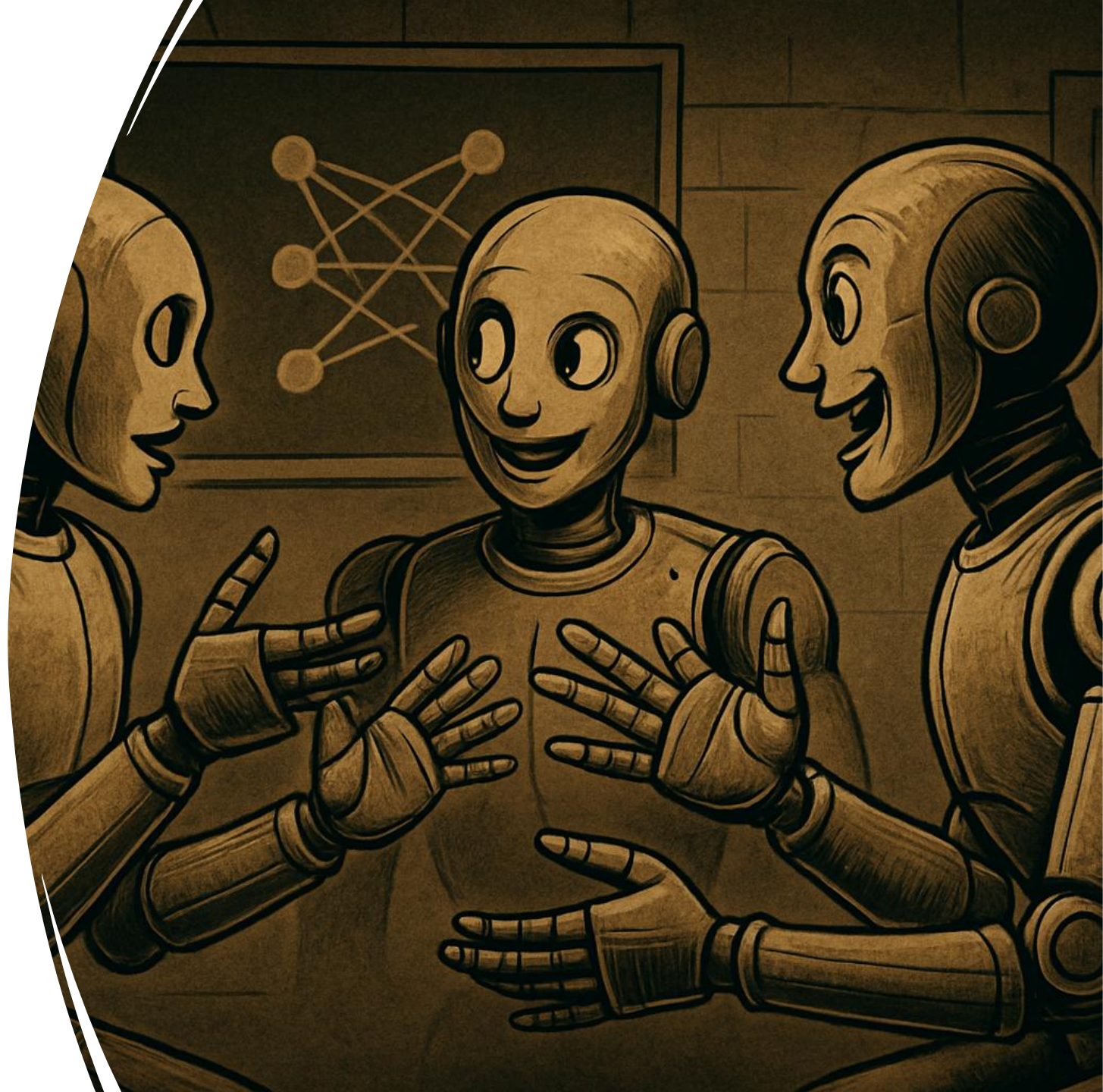
Moral Judgements by Human Respondents (Prolific)



Human vs LLM Persona



Origins of Disagreement

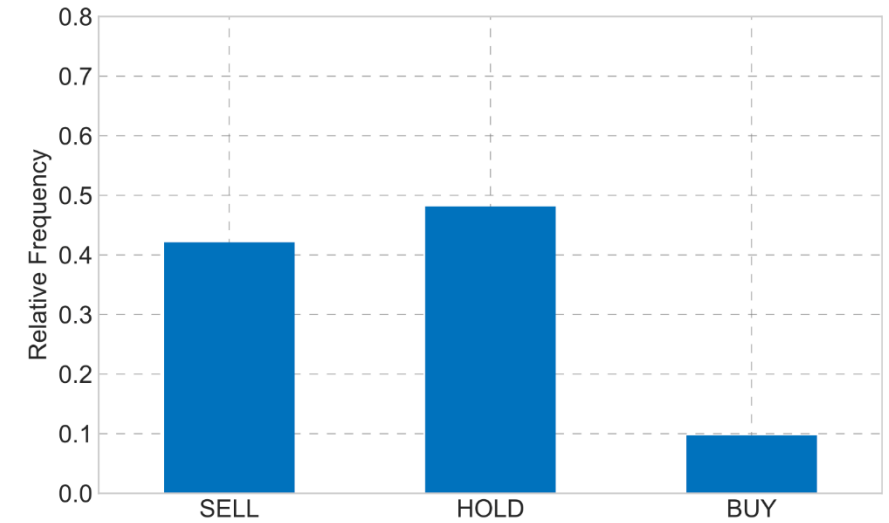


Disagreement in Raw Sentiment Data

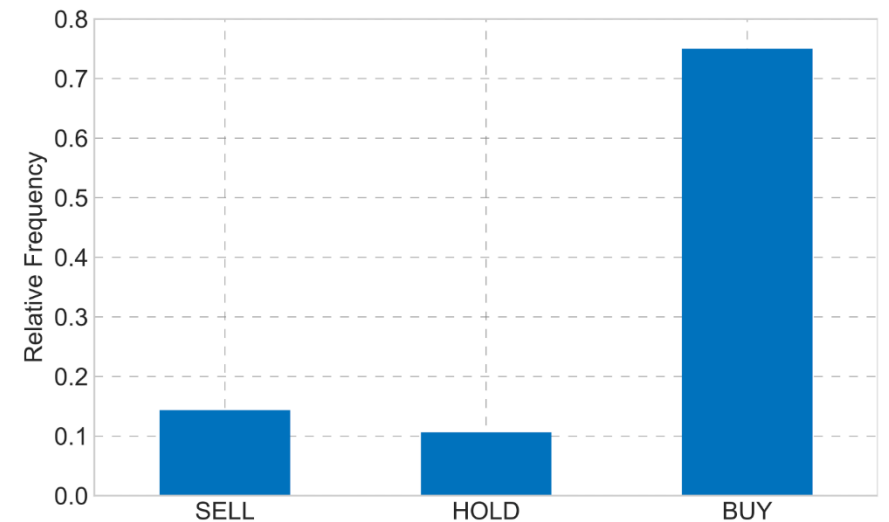
We compute a weighted standard deviation of response scores s_{ij} with weights w_j taken from the FINRA investor survey demographics + equal weighting political orientation.

$$\bar{s}_i = \frac{\sum_{j=1}^{216} w_j s_{ij}}{\sum_{j=1}^{216} w_j}, \quad \text{Disag}_i = \sqrt{\frac{\sum_{j=1}^{216} w_j (s_{ij} - \bar{s}_i)^2}{\sum_{j=1}^{216} w_j}}$$

A: Microchip CEO Received 2022 Compensation \$12.8M, Up From \$3.89M



B: Chevron confirms new oil discoveries off Congo



Disagreement Sources

Income and Politics drive the most disagreement about news

Panel B: Disagreement Measures

	Mean
Disagreement	0.211
Income Disagreement	0.101
Political Disagreement	0.062
Age Disagreement	0.043
Education Disagreement	0.035
Gender Disagreement	0.027
Racial Disagreement	0.022
Ex-Pol Disagreement	0.159
Ex-PolInc Disagreement	0.086

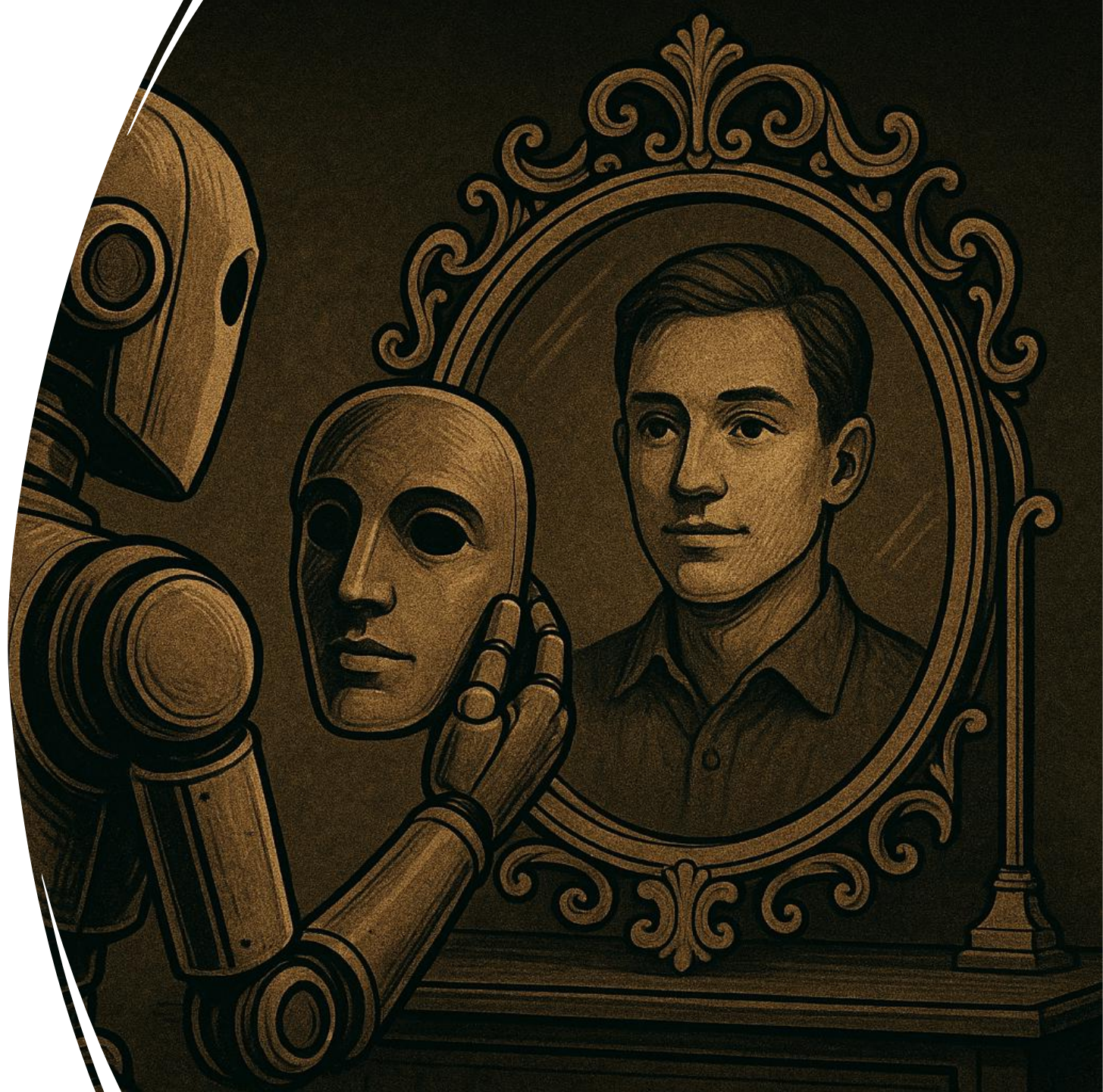
Hard vs Soft News, Social vs Legal Issues

Disagreement is **highest for social issues** and **lowest for hard news, unambiguous sentiment** (similar in 2024-2025 post training window).

Dependent Variable: Disagreement

	<i>Disagreement</i>		
	Full sample		
	(1)	(2)	(3)
Hard News	-0.208*** (0.012)		
Social Issues		1.110*** (0.029)	
Environmental Issues		0.129* (0.070)	
Ethical Legal Issues		0.169*** (0.024)	
High Ngvtv Sentiment			-0.145*** (0.020)
High Pstv Sentiment			-0.617*** (0.016)
Stock FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
R-squared Adj.	0.074	0.086	0.073
Obs.	1,249,518	1,249,518	1,249,518

Beyond *Homo Silicus*



Persona Disagreement and Abnormal Trading

Relates to contemporaneous trading volume

	Full sample		Training window		Post training	
	Total Volume (1)	Retail Volume (2)	Total Volume (3)	Retail Volume (4)	Total Volume (5)	Retail Volume (6)
Abn Disag (t) (z)	1.472*** (0.061)	1.694*** (0.076)	1.528*** (0.075)	1.661*** (0.087)	1.403*** (0.093)	1.720*** (0.115)
Volatility (t-5, t-1) (z)	-0.014 (0.344)	0.548 (0.378)	0.387 (0.447)	1.446*** (0.428)	-0.218 (0.451)	-0.208 (0.693)
Volatility (t-25, t-6) (z)	1.860*** (0.229)	2.425*** (0.300)	1.953*** (0.252)	2.812*** (0.371)	1.795*** (0.557)	2.024*** (0.679)
Abn Ret (t-5, t-1) (z)	-0.842*** (0.191)	-1.387*** (0.210)	-0.967*** (0.261)	-1.725*** (0.286)	-0.556*** (0.196)	-0.777*** (0.234)
Abn Ret (t-25, t-6) (z)	-1.875*** (0.177)	-2.871*** (0.237)	-1.973*** (0.216)	-3.165*** (0.302)	-1.514*** (0.257)	-2.140*** (0.352)
Ave Media Sentiment (t) (z)	0.167* (0.089)	0.334*** (0.113)	0.146 (0.108)	0.401*** (0.144)	0.075 (0.134)	0.110 (0.159)
Std Media Sentiment (t) (z)	-1.100*** (0.116)	-1.193*** (0.149)	-0.969*** (0.129)	-1.157*** (0.164)	-1.248*** (0.148)	-1.226*** (0.200)
Stocktwits Disag. (t) (z)	2.958*** (0.150)	3.767*** (0.190)	1.793*** (0.139)	2.406*** (0.179)	6.199*** (0.228)	7.523*** (0.313)
Log # News (t) (z)	8.885*** (0.259)	10.919*** (0.333)	8.619*** (0.280)	11.281*** (0.362)	9.436*** (0.330)	10.703*** (0.399)
Ave Persona Sentiment (t) (z)	0.065 (0.092)	0.069 (0.115)	-0.103 (0.112)	-0.189 (0.148)	0.392*** (0.128)	0.526*** (0.163)
log Ave # Words (t) (z)	-1.493*** (0.104)	-1.720*** (0.139)	-1.349*** (0.115)	-1.739*** (0.158)	-1.638*** (0.159)	-1.597*** (0.200)
Ambiguity (t) (z)	-0.590*** (0.080)	-0.791*** (0.101)	-0.863*** (0.096)	-1.033*** (0.127)	-0.242** (0.112)	-0.463*** (0.127)
Abn Non-persona Disag (t) (z)	1.851*** (0.099)	2.363*** (0.124)	1.922*** (0.115)	2.497*** (0.149)	1.735*** (0.145)	2.143*** (0.166)
Abn. Log Volume (t-1) (%)	0.592*** (0.010)		0.584*** (0.011)		0.571*** (0.015)	
Abn Log Retail Volume (t-1) (%)		0.579*** (0.012)		0.549*** (0.013)		0.585*** (0.028)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.576	0.502	0.575	0.479	0.587	0.545
Obs.	431,359	429,517	259,906	258,772	171,453	170,745

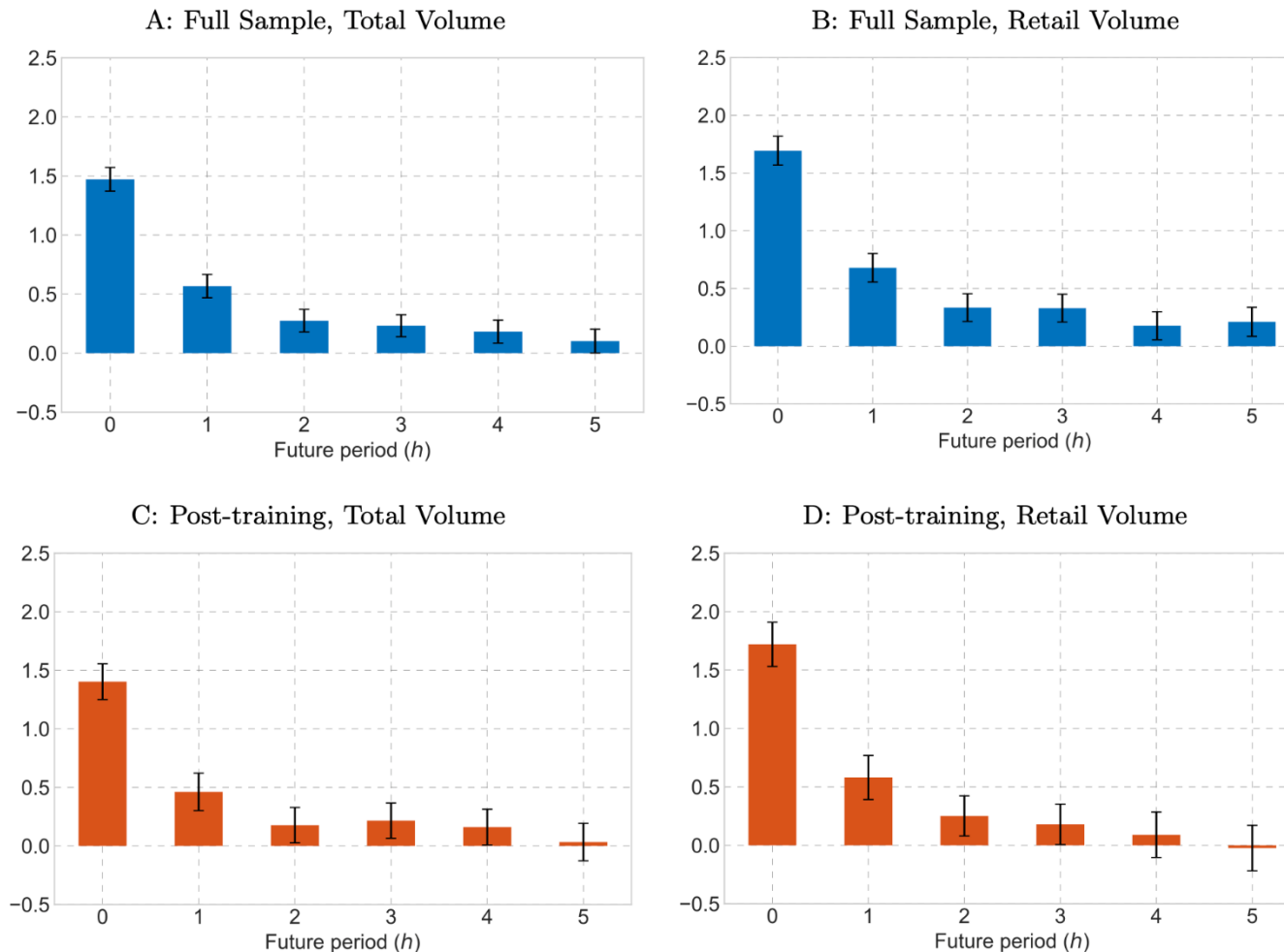
1st increase in abnormal disagreement →
+1.47% abn total volume
+1.69% abn retail volume

Controlling for “everything else”
volatility, returns, lagged trading,
media sentiment and dispersion,
StockTwits disagreement, other
disagreement, news coverage,
ambiguity in news headline...

**Nearly identical magnitudes
post-training.**

LLM Disagreement and Abnormal Trading

Abnormal trading is strongest same-day, significant for next few days, fades within a week; holds post-training





Heterogeneity

Which Demographics Drive Trading?

Education, age and income exhibit largest coefficients; politics and gender smallest.

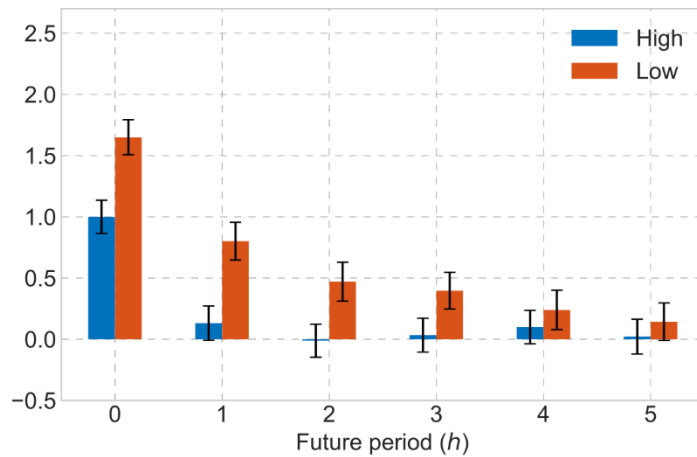
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Abn. Log Total Volume $\times 100$</i>						
Abn Education Disag (t) (z)	1.641*** (0.074)					
Abn Age Disag (t) (z)		1.310*** (0.061)				
Abn Income Disag (t) (z)			1.265*** (0.070)			
Abn Racial Disag (t) (z)				1.015*** (0.058)		
Abn Gender Disag (t) (z)					0.946*** (0.060)	
Abn Political Disag (t) (z)						0.860*** (0.061)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.576	0.575	0.575	0.575	0.575	0.575
Obs.	431,359	431,359	431,359	431,359	431,359	431,359

Political orientation corresponds to most disagreement, but *weak relation to trading*.

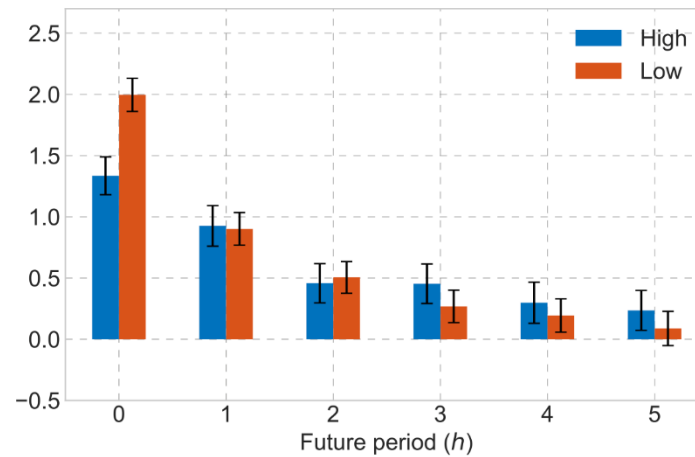
Rationale Narratives and Trading

Disagreement shaped by Short-term and backward-looking narratives relate more strongly to trading; pecuniary vs. non-pecuniary are similar

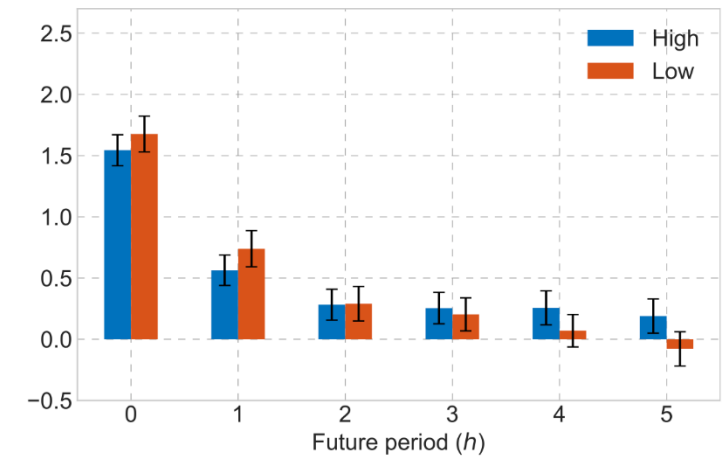
A: LongTerm - ShortTerm, Total Volume



B: FwdLook - BwdLook, Total Volume



C: Pecuniary - Nonpecuniary, Total Volume



Conclusion

In this paper, we find:

- **Differential interpretation** of the same news due to **different mental models** matters strongly for disag.
- Disagreement is highest for social/soft news.
- Disagreement is strongly related to trading, especially retail — **for both pecuniary and non-pecuniary reasons**.
- Political disagreement translates least into trading.

LLMs can serve as measurement tools, not just predictive engines. We expect the persona approach to scale to other settings, too.

