The Market's Mirror:
Revealing Investor Disagreement with LLMs*

Vineet Bhagwat† J. Anthony Cookson‡

Chukwuma Dim§ Marina Niessner¶

January 2026

Abstract

Large language models (LLMs) can emulate human perspectives. Leveraging this idea, we study how investor disagreement emerges in response to firm news. We endow an LLM with 216 representative investor personas and elicit buy, hold, or sell responses to S&P500 firm news headlines from 2010-2025. Dispersion in responses yields article-level disagreement for 5.5 million headlines and sheds light on its sources. Disagreement is highest for socially-oriented news and lowest for fundamentals. Persona responses reflect non-pecuniary rationales and align with human-survey benchmarks. Disagreement predicts elevated same- and next-day abnormal trading volume, especially retail, and results persist beyond the LLM's training cutoff.

Keywords: Investor Disagreement, LLMs, Beliefs, Sentiment, Narratives, Trading

*We gratefully acknowledge the computing resources provided on the High Performance Computing Cluster operated by Research Technology Services at the George Washington University and Research Technologies at Indiana University. Cookson is indebted to financial support provided by the Michael A. Klump Professorship and the Leeds School of Business. We thank Jess Cornaggia, Mehmet Canayaz, Joey Engelberg, Clifton Green, Matthew Gustafson, Jerry Hoberg, Sophia Kazinnik, Will Mullins, Marcus Painter, Raghavendra Rau and Bryan Routledge for their helpful comments and suggestions, as well as seminar and workshop participants at the AFA's Junior Faculty Mentoring Workshop, American University, Big Data and Asset Management Conference, Carnegie Mellon University, International Behavioural Finance Conference, Monash University, Northeastern University, Oxford-Man Institute Big Data and Finance Workshop, Penn State University, Santiago Finance Workshop, University of Colorado Boulder, University of Melbourne, University of New South Wales, and University of Nevada Reno.

†George Washington University ✉ vineetb@gwu.edu.

‡University of Colorado Boulder ✉ tony.cookson@colorado.edu.

§George Washington University ✉ cdim@gwu.edu.

¶Indiana University ✉ mniessne@iu.edu.

"We see things not as they are, but as we are."

Anaïs Nin, *Seduction of the Minotaur, 1961*

1 Introduction

Personal characteristics shape financial behavior and portfolio choice.[1] It is therefore natural to expect heterogeneity in how investors interpret firm news. This differential interpretation can be an important driver of disagreement and trading (Miller, 1977; Kandel and Pearson, 1995; Hong and Stein, 1999). But measuring which demographic dimensions drive disagreement about *specific* news at high frequency is difficult: surveys are infrequent and rarely map cleanly to article-level belief updates. In this paper, we deploy a local LLM at scale to address these questions.

Large language models (LLMs), trained on a multitude of internet sources, can be conditioned to produce responses consistent with particular demographic profiles. Recent work has shown that LLM forecasts display human-like biases (Bybee, 2023; Qin, Huang, and Ding, 2024) and can flexibly adopt distinct viewpoints (Fedyk, Kakhbod, Li, and Malmendier, 2024). This flexibility suggests that LLMs can act as virtual surrogates for human perspectives, adopting the worldviews of specific demographic groups when interpreting information (e.g., Horton, 2023; Geiecke and Jaravel, 2024). By harnessing the latent "understanding" encoded in AI models, we systematically elicit and quantify how distinct investor profiles react to firm-level news, providing novel insight into the factors that shape investor disagreement.

We deploy Llama 3.1 8B, a local large language model, to generate 216 synthetic investor personas and elicit their reactions to news headlines about S&P 500 firms from 2010-2025. The dispersion of responses across personas for the same news headline yields a disagreement measure that mirrors known demographic patterns, scales easily, and provides information beyond existing daily measures. Crucially, this LLM-derived disagreement measure is associated with market behavior: it is strongly related to same-day and next-day abnormal trading volume beyond standard controls and alternative proxies for disagreement drawn from social and tradi-

[1] There is a multifaceted literature showing these differences (e.g., Madrian and Shea, 2001; Barber and Odean, 2001; Lusardi and Mitchell, 2008; Bailey, Cao, Kuchler, and Stroebel, 2018; Goldsmith-Pinkham and Shue, 2023; Gupta, Hansman, and Mabille, 2024)

tional media. These findings hold outside of Llama's training window, which ends in December 2023.

We construct our set of 216 investor personas by crossing six demographic attributes.[2] For each RavenPack headline, we first prompt Llama 3.1 to assume one of these personas, endowing it with that group's LLM-learned worldview, and then ask it to give a buy, hold or sell response (Figure 1 gives the full prompt). Applying this prompt to every news headline-persona pair yields 216 synthetic opinions per news headline, totaling 1.188 billion elicitations (5.5 million news articles $\times$ 216 personas). We measure "LLM-derived disagreement" as the within-headline weighted standard deviation of buy/hold/sell responses, with the persona weights sourced from the FINRA investor survey. For each prompt, we also elicit a concise rationale, which helps interpret the sources of disagreement. To validate our choice of model, we compare Llama 3.1 and GPT-4 responses on a matched sample of news articles and find a high correlation (0.67) . Because RavenPack provides headlines but not full article text, we also validate that headline-based and full-text-based responses are highly correlated (0.53) in a sample of *Wall Street Journal* articles, with outright disagreement (buy vs. sell) occurring in fewer than 4% of cases.

We first validate the disagreement measure by showing it mirrors real-world heterogeneity in investor beliefs. Across personas, high-income and conservative personas respond more positively to news about CEO pay increases and oil discoveries, while low-income and liberal personas respond more negatively to the same headlines, citing concerns about greed or environmental harm. More generally, income and political orientation generate the largest disagreement across personas, with other characteristics contributing less.

Next, using RavenPack's news categories, we exploit the article-level variation to understand how demographic disagreement responds to different types of news. Disagreement varies sharply across event types: headlines about insider transactions, executive pay, shareholder rights, and

[2]We consider all combinations of age (18–34, 35–54, 55+), gender (male, female), race (White non-Hispanic, non-White), income (<\$50 K, \$50–100 K, >\$100 K), education (some college or less, college+), and political orientation (conservative, independent, liberal). The first five attributes come directly from the FINRA Foundation's 2021 National Financial Capability Study (NFCS), Questions 102 to 106. We add political orientation because of its importance in financial decision-making (Cookson, Engelberg, and Mullins, 2020; Kempf and Tsoutsoura, 2021; Meeuwis, Parker, Schoar, and Simester, 2022; Engelberg, Guzman, Lu, and Mullins, 2022). The FINRA survey categories provide natural sample weights for each investor persona, facilitating aggregation to a representative population of investors. For political orientation, we assign equal weight to each category.

workforce compensation generate the greatest dispersion, while shareholder meetings and the deaths of executives and board members elicit near consensus. LLM disagreement is markedly greater for "soft" news than for "hard" news about firm fundamentals, and disagreement is highest for socially-oriented news headlines. Disagreement is also lower when the sentiment of the news has a clear direction (i.e., very positive or very negative RavenPack sentiment). We complement this analysis with a word-embedding topic analysis of the elicited rationales from our prompting. The emergent topics reflect the human-like nature of the sources of disagreement: liberal personas emphasize fairness, employee morale, and environmental responsibility; conservative personas focus more on efficiency and profitability; and higher-income personas emphasize employee morale and efficiency. This contextual analysis reinforces our interpretation that LLM disagreement reflects systematic differences in interpreting news.

To address the concern that the model memorized demographic differences from its training data, we replicate our analyses before and after Llama 3.1's training cutoff (December 2023). Both the dimensions of disagreement and the news types that generate high disagreement remain stable in the post-training window (2024-2025). This persistence supports the interpretation that LLM disagreement reflects genuine belief dispersion rather than memorized interpretations.

Having examined these underpinnings of LLM disagreement, we next relate it to established measures of uncertainty, disagreement, and volatility. Persona disagreement correlates significantly with volatility and daily StockTwits disagreement, but bears a weak relation to analyst forecast dispersion. Despite exhibiting statistically significant positive correlations, these variables collectively explain less than 10% of the variation in LLM-derived disagreement, indicating that our measure captures meaningful residual variation relative to existing measures. The relatively weak explanatory power of social media disagreement for LLM persona disagreement resonates with other work showing a limited transmission of traditional news information into social media signals (Cookson, Lu, Mullins, and Niessner, 2025). When we examine disagreement over time, we find that it was stable through most of the 2010s, then rose markedly after 2020, coinciding with greater polarization around firms' social and workplace actions, political issues, and the Covid-19 pandemic.

To evaluate whether LLM persona-generated patterns resemble human judgments, we use Llama 3.1 to replicate two human surveys released after Llama 3.1's training window. These survey responses provide a direct benchmark for evaluating our personas against human subjects for a wide range of behaviors relevant to interpreting news. First, we reproduce the survey of Iliewa, Kempf, and Spalt (2024), which asked a representative sample of U.S. adults to rate the moral acceptability of ten common corporate actions. The results show that LLM personas align strongly with human responses. Both personas and humans view layoffs and CEO pay increases as the most objectionable corporate actions, while viewing buybacks, cost cutting, and outsourcing as less objectionable. This correspondence holds across political, income, and gender subgroups. Second, we simulate human responses published in a recent "digital twins" database report (Toubia, Gui, Peng, Merlau, Li, and Chen, 2025), which elicited a wide array of behavioral biases in a benchmark human survey. Llama 3.1 seeded only with demographic information achieves significantly better accuracy than random guessing, while adding full respondent transcripts improves accuracy only modestly. These results confirm that demographics alone provide sufficient context for the model to produce human-like responses.

Finally, we ask whether LLM disagreement provides a useful lens into human disagreement by relating it to trading in financial markets. At the firm-day level, higher LLM disagreement is strongly associated with greater trading volume. A one–standard-deviation increase in LLM disagreement corresponds to a 0.42% rise in same-day abnormal trading volume (and a 0.51% rise in same-day abnormal retail trading volume), even after we control for other well-known drivers of trading and alternative disagreement measures. Moreover, volume spikes around high LLM disagreement days are short-lived, dissipating within a week. This pattern is inconsistent with the interpretation that the LLM merely learned which firms are most polarizing on average. These patterns hold across both the LLM's training (2010–2023) and post-training (2024–2025) windows. The results also hold under equal persona weighting and when excluding the income or political dimensions.

Decomposing disagreement by demographic source reveals meaningful heterogeneity in the disagreement-trading relation. Abnormal trading (and abnormal retail trading) relates most strongly to education-, age-, and income-based disagreement, but the relationship is more muted

for political and gender-based disagreement. Given that political differences give rise to the most disagreement in our setting, these findings indicate that certain characteristics can lead to polarized opinions without necessarily leading to more trading, echoing a theme in recent work on the connection between beliefs and trading (Giglio, Maggiori, Stroebel, and Utkus, 2021).

Our primary contribution is to the literature using LLMs as measurement tools in economics and finance.[3] Rather than focusing on prediction, we use LLMs to *reveal* human disagreement by conditioning on demographic personas and eliciting reactions to specific news. This "LLM-as-lens" approach complements recent work that uses generative models to approximate human judgments and forecasts. Our focus is in the spirit of Horton (2023) which shows AI's potential as a *computational model of human behavior*. Complementing this perspective, Hansen, Horton, Kazinnik, Puzzello, and Zarifhonarvar (2024) demonstrate that large-language-model "forecaster personas" perform well relative to human predictions in the Survey of Professional Forecasters. Research that utilizes simulated agents has a long history in economics,[4] but applications remained limited until recent advances in generative AI agents, which have begun to approximate human behavior in useful ways (Geiecke and Jaravel, 2024; Kozlowski and Evans, 2024). For example, recent research has shown that AI responses can reliably replicate studies conducted on human subjects (Qin, Huang, and Ding, 2024). Relative to this work, which replicates controlled experiments, our contribution is to produce a large-scale, representative disagreement measure linked to financial market outcomes.

In the financial market domain, our work is closely related to Bybee (2023), Chen, Green, Gulen, and Zhou (2024), and Goetzmann, Kim, and Shiller (2024) who show how LLM forecasts inherit behavioral biases and emotional content when trained on human texts. We also relate to Fedyk, Kakhbod, Li, and Malmendier (2024) who demonstrate that default LLM outputs are non-representative, reflecting the viewpoints of younger high-income people; however, seeding

[3]This burgeoning literature on AI in financial markets is multifaceted (Babina, Fedyk, He, and Hodson, 2024). Existing work has studied AI's impact on the real economy (Noy and Zhang, 2023; Brynjolfsson, Li, and Raymond, 2025; Eisfeldt, Schubert, and Zhang, 2023), the role of AI in return prediction (Kelly, Malamud, and Zhou, 2024; Lopez-Lira and Tang, 2023) and other predictive aspects and concerns of AI models (Engelberg, Manela, Mullins, and Vulicevic, 2025; Lopez-Lira, Tang, and Zhu, 2025; Chen, Green, Gulen, and Zhou, 2024).

[4]There are many such examples, but two early contributions are notable. Schelling (1971)'s "checkerboard" experiments show how segregation can emerge from dynamic choices, even if agents have only a small preference. Second, Axelrod and Hamilton (1981)'s work on the emergence of cooperation as a winning strategy in tournaments that play repeated Prisoner's Dilemma.

the LLM prompt with demographics yields viewpoints aligned with the target group's preferences. Like this research, our paper harnesses LLMs' ability to emulate real world personas. By seeding the model with investor demographics and eliciting investors' reactions to specific firm news, we demonstrate that this idea can scale. AI can serve as a powerful lens into financial markets.[5]

Our paper also contributes to the literature on investor disagreement, which has been of longstanding interest (Aumann, 1976; Varian, 1985; Karpoff, 1986). An important strand of the disagreement literature has sought to understand the *origins* of disagreement.[6] Our study contributes two distinct perspectives to this literature. First, our analysis is unique in providing a daily measure of disagreement about *traditional news*. This contribution is useful given the extensive literature on news media and financial markets (Tetlock, 2007, 2011; Dougal, Engelberg, Garcia, and Parsons, 2012; Garcia, 2013). Relatedly, Goldman, Gupta, and Israelsen (2024) document disagreement about firms in the content of news (*Wall Street Journal* versus the *New York Times*). Our evidence is distinct: we illustrate divergent viewpoints among investor personas on the same news headline. Second, although certain demographics have been shown to create important wedges in investor beliefs (e.g., political leanings in Cookson, Engelberg, and Mullins (2020) and Meeuwis, Parker, Schoar, and Simester (2022)), empirical evidence on which demographics matter most remains scarce because it is difficult to match demographics to sufficiently high-frequency belief updates (for instance, Giglio, Maggiori, Stroebel, and Utkus (2021) survey investors only every six weeks). Our approach of deriving investor beliefs from LLMs permits us to directly link demographic groups to the news they disagree about. Moreover, by studying the characteristics used in the FINRA investor survey, we can identify which characteristics matter most for investor disagreement.

[5]This perspective of using generative AI as a "lens" is akin to the use of social media as a lens into human behavior, an approach adopted widely in the social media and finance literature (e.g., Bailey, Cao, Kuchler, Stroebel, and Wong, 2018; Cookson, Mullins, and Niessner, 2024).

[6]In theory, investor disagreement stems either from differences in the models investors apply to interpret financial information (Kandel and Pearson, 1995; Daniel, Hirshleifer, and Subrahmanyam, 1998) or from differences in their information sets (Hong and Stein, 2007). Recent empirical work has confirmed the importance of both channels (Cookson and Niessner, 2020), finding that disagreement varies systematically with investment horizon, geographic location, and partisan orientation (Cookson, Dim, and Niessner, 2024; Gerken and Painter, 2023; Cookson, Engelberg, and Mullins, 2020). To date, however, most evidence has been derived from social media settings or studies of financial analyst belief updating (Kandel and Pearson, 1995; Kapons and Kelly, 2023).

2 Setting and Data

We construct our investor disagreement measures by eliciting investor beliefs about news headlines. For each news headline, we query the Llama 3.1 8B model for buy/hold/sell responses across 216 distinct investor personas. In this section, we describe our data sources, detail the procedures we use to elicit investor beliefs, contrast our approach with related approaches, and illustrate how we use these elicited investor beliefs to construct market-relevant measures of disagreement.

2.1 RavenPack News Headlines Data

We use RavenPack 1.0 (Dow Jones & PR Edition) headlines from January 2010 through April 2025 for current or former S&P 500 constituents. We keep headlines with RavenPack relevance scores of at least 75, which ensures relevance to the firm. To keep inference tractable, we cap coverage at five headlines per firm-day by randomly sampling when more headlines are available; otherwise we keep all headlines.[7] Our final dataset comprises 5.5 million headlines across 1,070 unique firms. This delivers comprehensive, high-quality coverage of firm-level news articles over a 15-year window. For each headline, we record the date, ticker, headline text, RavenPack's sentiment score, and the RavenPack news type.

2.2 Choosing a Large Language Model

The landscape of LLMs offers a wide range of capabilities and deployment options. Two key aspects guide our choice of Llama 3.1 8B, an open-source model that we can run locally. First, it is markedly more cost-effective than querying cloud-based services such as GPT-4. Second, local deployment ensures that all RavenPack content remains on premises, which is crucial to comply with RavenPack's Terms of Service. In contrast, GPT-4, Claude, and Gemini have usage-based pricing and offer limited transparency into model internals such as architecture, weights, tokenizers, and vocabulary, and they remain accessible only via proprietary cloud APIs.[8]

[7] Because the median and 75th percentiles of the number of news items per firm-day prior to sampling are 2 and 5, respectively, the sampling affects only the minority of stock-days with high news coverage.

[8] We obtain highly similar responses using GPT-4 on a matched set of articles from Factiva. See Section 2.3.2.

Llama 3.1 8B was released by Meta in mid-2024, with the training window ending in December 2023. Although more advanced LLMs exist, Llama 3.1 is comparable to top-tier LLMs in reasoning tasks like the investor beliefs elicitation we use. From independent evaluations of the 70B variant, Llama 3.1 is up to 50 $\times$ cheaper and 10 $\times$ faster in inference throughput compared to cloud-hosted GPT-4 APIs, making it attractive for large-scale or latency-sensitive deployments.[9] We use the smaller 8B variant, which is substantially more cost-effective while still yielding realistic elicitation of investor beliefs about news. To illustrate the substantial computational cost difference between the 8B and 70B Llama 3.1 variants: prompting an instance of Llama 3.1 8B running on a single node of a high-performance computing (HPC) cluster equipped with four Nvidia Tesla V100 SXM2 16GB GPUs takes $\approx$1.2 seconds. The same prompt on an instance of the 70B variant running on the same HPC node takes $\approx$17 seconds, roughly 14 times longer. Scaling up this difference to 1.188 billion prompts (= 216 $\times$ 5.5 million) amounts to a computational cost savings of roughly 219,444 GPU days by using Llama 3.1 8B.

2.3 Prompting Investor Personas

We elicit synthetic beliefs from Llama 3.1 in two stages, as shown in Figure 1, Panel A. First, we seed the model with a demographic persona by giving it a standardized instruction block that defines its role and the demographic profile attributes (age, income, education, etc.). This top block tells the model: "*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.*"

Second, immediately below this, we tell the LLM to assume a specific profile that we draw from the FINRA Foundation's 2021 National Financial Capability Study (NFCS) survey (e.g., "You are a White non-Hispanic male, age 18-34, with some college or less of education, and household income less than $50K"), plus a political view (conservative, independent, or liberal), anchoring each investor persona in a real-world investor segment. The prompt also specifies the headline, company name, and publication date, and instructs the LLM persona to state

[9] For example, an analysis by Ultralytics commented that Llama 3.1 is "making significant strides in the AI landscape, rivaling the capabilities of top-tier models like OpenAI's Chat GPT-4o and Anthropic's Claude 3.5 Sonnet." See the full comparison of Llama 3.1 class models here, which reports performance on a wide array of tasks, broken down by parameterization (link).

whether it would buy, hold, or sell, without using information after the publication date. We further instruct the LLM to provide a brief (max 25 words) explanation for its response. This gives a contextual rationale for the buy/hold/sell decision. For all prompts, we set the sampling temperature to zero to minimize variation across runs and to ensure replicability.

An important aspect of our prompting design is to anchor the LLM in realistic investor demographics. We consider the 72 distinct combinations of demographics from FINRA's 2021 NFCS, which asks questions about gender, age, race, education and income (Questions 102 to 106). We also include three possibilities for political orientation (conservative, independent, liberal), which gives us a total of 216 distinct demographic investor profiles. This approach allows us to trace how different segments of investors interpret news and to assess whether LLM-driven disagreement mirrors the investor heterogeneity observed in actual markets. For our analysis, we map the persona responses into a numerical measure by coding BUY as $+1$, HOLD as 0, and SELL as -1.

2.3.1 Illustrative Examples of Persona Disagreement

Our prompting procedure yields economically meaningful disagreement among the investor personas. Figure 1, Panel B illustrates how personas interpret the same headline differently. For an executive-compensation headline, higher-income personas tend to emphasize firm performance and buy, while lower-income personas more often cite fairness concerns and sell. For an oil-discovery headline, disagreement reflects different tradeoffs between expected profitability and environmental concerns.

To show how well these examples generalize, Table A1 reports the ten most salient bigrams, ranked by TF-IDF, from persona-specific response rationales for four demographic segments (conservative vs. liberal and low-income vs. high-income) across three news categories among those with the highest disagreement. For each category, we first identify the top ten news articles with the largest weighted standard deviation in responses, and then we extract the most distinctive bigrams from the response rationales. By averaging TF-IDF scores within each political affiliation-income subgroup, this analysis reveals the themes and concerns each group finds most salient in the articles that trigger the most pronounced disagreement.

The themes from the bigram analysis highlight systematic differences between political affiliations and income levels. For articles about executive compensation in Panel A, low-income conservatives focus on costs (e.g., "increased costs," "excessive compensation"), whereas high-income conservatives emphasize performance metrics and investment potential (e.g., "investment opportunity," "strong commitment"). Low-income liberals highlight inequality (e.g., "regular people," "corporate greed"), while high-income liberals also reference performance but note excessive pay (e.g., "excessive compensation," "future growth"). Similar divergences appear in resource discovery news in Panel B: conservatives (both income groups) focus on production and profit expectations (e.g., "expected growth," "future prospects"), whereas liberals underscore environmental risk and sustainability (e.g., "environment impact," "worried environment"). Panel C displays similar differences for articles about increases in workforce compensation: conservatives cite cost burdens (e.g., "hurt profits',' "labor costs"), while liberals emphasize corporate responsibility (e.g., "socially responsible," "demonstrates socially").

2.3.2 Comparison to GPT-4

We use Llama 3.1 8B for both computational reasons and to comply with RavenPack's Terms of Service. However, we can still compare the responses we obtain from Llama 3.1 with those from a proprietary model like GPT-4 using the news headlines from Factiva, which does not prohibit the use of proprietary cloud-based models. For this validation exercise, we use Factiva headlines from major U.S. and international outlets (2001–2019).[10]

We match articles in RavenPack to those in Factiva by applying a fuzzy matching algorithm to headline text and requiring the publication dates to differ by at most one day. When we restrict matches to those with at least an 85% similarity score, we obtain a sample of 78,310 articles common to both databases, yielding over 16.9 million persona-article combinations. For computational efficiency, we use a 10% random sample of these matches and follow the same prompt construction and procedure as before to elicit buy/hold/sell responses from GPT-4.1 mini.

[10]The data contain articles published in nine of the largest news outlets: Barron's, Financial Times (FT), The New York Times (NYT), Reuters, The Times, USA Today (USAT), the Washington Post (WP), The Wall Street Journal (WSJ), and the Wall Street Journal Online (WSJO)

The correlation between the persona-article level buy/hold/sell responses from Llama 3.1 and GPT-4 is 0.67, which gives us confidence that using an open-source model does not compromise our elicited belief measure. Furthermore, the two LLMs agree on the buy/hold/sell response for 74.18% of persona-article combinations, and only generate the polar opposite response (one saying "buy" while the other says "sell") in 1.22% of the cases.

To further validate comparability across models, we estimate the incremental R-squared that GPT-4 buy/hold/sell responses contribute when explaining the Llama 3.1 buy/hold/sell responses, beyond what can be explained by firm fixed effects alone. Specifically, we regress Llama 3.1 buy/hold/sell responses on (i) firm fixed effects alone, (ii) GPT-4 buy/hold/sell responses alone, and (iii) both together. The resulting R-squared values are 0.206, 0.449, and 0.512, respectively. Thus, the responses from GPT-4 alone explain close to half the variation in the responses for the same persona and article obtained from Llama 3.1, and the R-squared is more than double the value from using firm fixed effects alone (0.449 versus 0.206). This provides further support for the idea that using an open-source model like Llama 3.1 yields responses similar to those from a proprietary cloud-based model like GPT-4.

Because prompts and model families differ, we do not expect perfect agreement. We view cross-model comparisons as analogous to correlations across established human expectation surveys. In that context, a 0.67 correlation indicates strong alignment. As a benchmark in human samples, Greenwood and Shleifer (2014) examine six different human surveys that elicit investor expectations, finding an average pairwise correlation of 0.43, with a maximum correlation of 0.77 between the Gallup and Graham-Harvey surveys.

2.3.3 Comparison of Article Headlines to Full Text

A natural concern with our main approach is that we prompt LLM personas using only article headlines from RavenPack, which could introduce measurement error if human sentiment reflects a deeper reading of the article. This limitation arises because our RavenPack sample does not include the full text of news articles. To evaluate the importance of this limitation, we collect a focused sample of full-text *Wall Street Journal* articles from 2004-2020 covering the firms in our sample from the WSJ News Archive (available online). We subject this sample of full-text

articles to our prompting procedure, and as a direct comparison, we run a parallel set of prompts using just the headlines in this sample. This paired design allows us to directly compare BUY, HOLD, and SELL responses that we elicit from headlines with those we elicit from full articles.

Prompting this corpus of WSJ news with our 216 personas yields a sample of 9,636,408 article-persona pairs, for each of which we obtain both a headline-based sentiment and a full-text-based sentiment. When we map sentiment into numeric values (i.e., $-1, 0, +1$), the correlation between article and headline sentiment is 0.53, indicating a close correspondence that is comparable in magnitude to the human survey cross-correlations in Greenwood and Shleifer (2014), cited above. Beyond this strong positive correlation, we observe that headline and article sentiment disagree (i.e., one is buy while the other is sell) in fewer than 4% of cases. Although the correlation is not perfect, its magnitude, together with the low rate of outright disagreement, provides support for our use of headlines rather than the full article text in our more comprehensive RavenPack sample.

2.4 Constructing Disagreement from Belief Elicitations

Our prompting procedure yields 216 responses for each headline, one for each investor persona. To measure disagreement at the article level, we compute a weighted standard deviation of the response scores. Specifically, for each headline i, we combine the 216 scores s_{ij} into a single dispersion measure by calculating the weighted standard deviation:

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

where w_j denotes the FINRA survey weight for persona j.[11] Weighting the underlying buy/hold/sell responses by the empirical FINRA survey frequencies ensures that the resulting Disag_i is both representative of the demographic composition of U.S. retail investors and captures the full breadth of persona views on headline i.

[11] For political orientation, we assign equal weight to each possibility. The results are robust to equal weighting across all personas.

To give a visual depiction of the responses underlying our disagreement measure, Figure 2 displays histograms of the 216 responses for the two illustrative headlines in Figure 1, Panel B. Because these articles are examples of particularly pronounced disagreement, the responses heavily load on BUY and SELL. These figures illustrate the heterogeneity in persona responses that our disagreement measures capture.

To disentangle which demographic attributes primarily drive overall dispersion, we calculate "dimension-specific" disagreement by averaging over the other five dimensions and computing the weighted standard deviation across the focal category of interest (e.g., income groups or political orientations).

2.4.1 Summary Statistics of Trading Sentiment and Disagreement

In Table 1, we summarize the buy/sell/hold responses and disagreement across demographic categories, confirming that the examples in Figure 1, Panel B capture first-order variation in disagreement across groups.

Panel A reports the full distribution of LLM-derived buy/hold/sell scores across all headline–profile pairs (N = 1.188 billion). The mean score is 0.114 on a $-1, 0, 1$ scale, indicating that, on average, the simulated agents view firm-specific news as mildly positive with more BUY responses than SELL responses: 25.7% BUY versus 14.3% SELL.

When we partition responses by demographics, the mean sentiment by different demographic groups varies most for income, with the mean ranging from -0.007 to 0.235; mean sentiment also varies significantly by political orientation, ranging from a mean of 0.069 among independents to a mean of 0.166 for liberals. Despite this significant variation, we observe that the only persona characteristic associated with more sell than buy responses on average is low-income (less than $50k). However, most groups have a wide dispersion of buy versus sell responses. The cumulative distribution of average responses by investor persona in Figure 3 provides a non-parametric illustration of these demographic differences, showing systematic shifts in average sentiment by income (Panel A) and political orientation (Panel B), while differences by age are more muted (Panel C).

In Panel B of Table 1, the first row reports headline-level disagreement, defined as the weighted standard deviation of responses across all 216 personas. In subsequent lines, for each demographic dimension (e.g., income, politics, race), we compute dimension-specific disagreement by first averaging the responses over non-focal dimensions and then taking the weighted standard deviation of the average response across the categories of the focal dimension. We also construct "Ex-Pol" and "Ex-PolInc" disagreement measures by averaging out the politics dimension, and both politics and income dimensions, respectively, before taking the cross-persona weighted standard deviation. To obtain sample-level aggregates, we weigh headline's disagreement by the FINRA survey's empirical share of investor profiles so that the reported means mirror the demographic composition of U.S. retail investors. Consistent with earlier examples, income- and politics-based disagreement are largest, followed by age; education, gender, and race contribute more modestly. These statistics identify which demographic dimensions account for the bulk of persona disagreement.

3 Results

We now turn to our empirical findings. First, we show that drivers of disagreement and responses vary systematically across news types. Second, we relate LLM-derived disagreement to established measures of uncertainty and document how it evolves over time. Third, we validate that persona disagreement captures meaningful, human-like variation in interpretation by comparing LLM personas' moral evaluations of corporate actions to survey evidence from real investors. Finally, we document that LLM-derived disagreement and its demographic components strongly predict trading activity, with effects that are robust across Llama's training and post-training periods. These findings provide a deeper understanding of the drivers of investor disagreement, and establish that the LLM-based disagreement measure is relevant to real-world trading.

3.1 Drivers of Investor Disagreement

3.1.1 Disagreement Across Types of News

We begin by leveraging RavenPack's news type classifications to examine the types of news associated with varying levels of disagreement. Figure 4 shows the top 20 and bottom 20 RavenPack news types, illustrating that the distribution of investor-profile disagreement varies widely across news categories.[12] The bars plot the average disagreement by RavenPack news type; we color-code the news types to illustrate systematic patterns more easily. Red bars are news types related to insider compensation, blue bars are informational releases and fundamental events, and black bars are other events, mostly containing technical trading events.

At the high-disagreement end, in Panel A, are headlines related to insider transactions and air pollution, along with other governance- and socially oriented events. News about executive and workforce compensation also ranks highly, suggesting that topics tied to firm governance and social or distributional considerations are more likely to generate disagreement, leading to the preponderance of red bars.

At the low-disagreement end, in Panel B, earnings-related forecasts and other standardized firm announcements (e.g., EBIT expectations, conference calls) exhibit relatively low disagreement, reflecting broad consensus across investor personas. Even salient negative events, such as the deaths of executives or board members, are associated with low disagreement. Together, these patterns imply that news involving social, ethical, or governance issues tend to lead to more disagreement, whereas clearly quantifiable firm disclosures leave less room for differential interpretation. This heterogeneity highlights the value of our LLM-based approach: by capturing how different investor personas diverge in their interpretations of "soft" news, we shed light on the demographic drivers of disagreement, even when the average buy/hold/sell response is not strongly tilted in one direction.

Next, in Table 2, we regress standardized buy/hold/sell scores on persona demographic characteristics estimated separately for five event types: executive pay increases, workforce pay increases, share buybacks, fossil fuel discoveries, and employee layoffs. By estimating these

[12] Table A4 provides descriptions of each news type.

relationships within specific news types, we isolate demographic variation in interpretation rather than differences driven by the composition of news itself, providing a disciplined test of whether persona responses align with economically intuitive demographic heterogeneity. We find that higher-income personas evaluate executive pay increases, workforce pay increases, and share buybacks more favorably than lower-income personas. Liberals are consistently more negative than conservatives on executive pay increases while more positive on worker pay increases. Other demographics also matter. For example, White non-Hispanic and male personas tend to assign somewhat more positive responses to executive pay and fossil fuel discoveries than non-White and female personas. These patterns reinforce that, while income and political orientation explain the largest share of divergence, other demographic attributes also shape how personas interpret firm news.

Next, for the news analyzed in Table 2, we perform a topic analysis of the rationales that personas provide for their responses. These textual rationales provide context for the demographic differences in Table 2. Specifically, we construct five topics via a word-embedding (word2vec) approach and construct 100-word dictionaries for each theme: Employee Morale, Fairness, Environment, Profit Decline, and Efficiency. For each rationale, we compute a normalized topic loading (tf-idf weighted and divided by rationale length), and then regress these topic loadings on the persona demographic characteristics. In Table 3, we report the results from these regressions, which help contextualize the reasons for different trading responses across demographic groups. For example, liberal personas emphasize employee-morale, fairness, and environmental topics, whereas independents emphasize similar topics but not employee morale; higher-income personas tilt toward efficiency while downplaying fairness and environmental topics. White non-Hispanic and male personas reference profit-decline more, while de-emphasizing fairness and environmental topics. These results provide support for our finding of significant demographic response differences by shedding light on the language and thematic framing through which different personas interpret firm news.

A key concern with using LLMs to analyze historical data is forward-looking bias: the model may have memorized specific news events and outcomes during training, rather than genuinely interpreting headlines as an investor would in real time. To address this concern, we repeat the

analysis from Tables 2 and 3 in Tables A5 and A6 separately for periods before and after the Llama 3.1 8B model's training cutoff of December 31, 2023. Specifically, we run the analysis separately for 2010–2023, the model's training window, and for 2024–2025, the post-training window. The signs and qualitative patterns remain consistent across both windows, suggesting that the demographic differences we document are not artifacts of memorized training data. Rather, they appear to reflect stable differences in how the different LLM personas interpret headlines, differences that persist even when applied to news the model could not have seen during training.

Tables A7 and A8 extend this analysis along two dimensions. Table A7 examines additional high-disagreement event categories from Figure 4: insider buys, insider gifts, and executive incentives. The demographic patterns mirror those in our main analysis, with higher-income personas responding more positively and liberal personas more negatively to insider transactions, and these relationships remain stable across both the training and post-training windows. Table A8 examines additional rationale themes: uncertainty, confidence, and accountability. Here too, demographic patterns are intuitive and consistent across windows. For instance, higher-income personas express less uncertainty in their rationales, personas with less education emphasize confidence more heavily, and liberal personas place greater weight on accountability. The stability of these patterns across time periods and news categories reinforces our conclusion that the persona-based approach captures genuine belief heterogeneity rather than memorized associations.

Next, we analyze how disagreement across investor personas depends on the type of news. Using RavenPack's news categories, we consider two dimensions of news article characteristics: (i) whether the news is hard versus soft (i.e., hard news is about firm fundamental information such as earnings, revenues, or credit ratings, whereas soft news covers less quantifiable topics such as executive changes or corporate strategy) and (ii) whether the news is about social, environmental, or ethical & legal issues. We expect hard news to generate less disagreement because it provides concrete, verifiable information that leaves less room for differential interpretation across demographic groups. By contrast, soft news and news touching on social dimensions may

invite greater disagreement because investors must rely on their own values and worldviews to assess its implications. See Appendix Table A2 for group definitions.

Formally, we estimate article-level regressions of standardized disagreement on indicator variables for these article characteristics, controlling for stock and calendar-date fixed effects and clustering standard errors by stock and date. To consider hard versus soft news, we estimate:

$$\text{Disag}_i \;=\; \alpha \;+\; \beta_{\text{Hard}}\,\text{HardNews}_i \;+\; \delta_{\text{stock}(i)} \;+\; \gamma_{\text{date}(i)} \;+\; \varepsilon_i, \tag{1}$$

where Disag_i is the overall disagreement for headline i (normalized to unit variance) and $\text{HardNews}_i = 1$ if the RavenPack "event group" for headline i corresponds to hard news, meaning items we classify as clearly related to firm fundamentals.

To examine how disagreement varies by social, environmental or ethical legal news, we estimate the analogous specification:

$$\text{Disag}_i = \alpha + \beta_S\,\text{SocialNews}_i + \beta_E\,\text{EnviroNews}_i + \beta_L\,\text{EthicalNews}_i + \delta_{\text{stock}(i)} + \gamma_{\text{date}(i)} + \varepsilon_i, \tag{2}$$

where each indicator variable equals one if RavenPack's event "type" for headline i falls into the corresponding category—Social Issues, Environmental Issues, or Ethical & Legal Issues.

The results reported in column (1) of Table 4 show that Hard News articles elicit significantly lower disagreement than soft news, confirming the intuition that fundamental information leaves less room for differential interpretation. In column (2), the large and statistically significant coefficient on Social Issues indicates that, relative to other article categories, articles about social issues (e.g., discrimination) are associated with significantly more disagreement. By contrast, Environmental Issues and Ethical & Legal Issues exhibit coefficients closer to zero, indicating that these categories are no more polarizing than the baseline.

In column (3), we examine whether disagreement varies with the valence of news. We rely on RavenPack's news sentiment score to identify particularly optimistic and pessimistic news. Specifically, we define two indicator variables — *High Ngtv Sent* = 1 if the RavenPack event sentiment for headline i falls below the 5th percentile and *High Pstv Sent* = 1 if it exceeds the

95th percentile. Both very negative and very positive news are statistically associated with lower dispersion relative to moderate-sentiment events, though the effect is stronger for positive news.

In columns (4) through (9), we repeat the analysis separately for the LLM training window (2010–2023, columns 4-6) and the post-training window (2024–2025, columns 7-9). The results are broadly consistent across both periods: hard news is associated with lower disagreement, and social issues remain highly polarizing in the post-training period. Environmental and ethical issues continue to show comparatively small coefficient magnitudes. Both very negative and very positive news are associated with lower dispersion. These findings suggest that the patterns we document are not driven by the LLM having memorized associations between specific headlines and market outcomes during training.

In Figure 5 we provide a complementary, non-parametric view of this stable relationship. We plot average disagreement for each RavenPack news type in the training window (x-axis) against the corresponding value in the post-training window (y-axis). If the LLM were simply retrieving memorized patterns, we would expect the relationship to break down for post-training news. Instead, the plot shows that the news types that generate high disagreement in the training window also generate high disagreement in the post-training window, as the points cluster roughly around the 45 degree line. While the post-training estimates are somewhat noisier due to the smaller sample, the consistency of this pattern reinforces that the persona-based approach can generate meaningful measures of disagreement outside of the LLM's training window.

3.1.2 Persona Disagreement and Other Measures of Uncertainty

Next, we relate LLM-derived disagreement to standard uncertainty and disagreement proxies (e.g., volatility, analyst dispersion, geopolitical/political risk measures, and StockTwits disagreement) to validate the measure and gauge how much independent variation remains. We present the results in Table 5. In univariate regressions in columns (1-7), we find that LLM-derived persona disagreement is positively related to most of these uncertainty and disagreement characteristics, except for Hassan, Hollander, Van Lent, and Tahoun (2019)'s political risk exposure, which is negatively related to LLM disagreement. In the multiple regression (column 8), the proxies that retain a positive relation to LLM disagreement are volatility, StockTwits disagree-

ment and geopolitical risk exposure. These generally positive relations help to validate our measure. However, even when combined, their aggregate explanatory power is less than 10% of the variation in LLM-derived disagreement. This suggests that our LLM-derived measure captures meaningful variation in investor beliefs that is not reflected in existing measures of disagreement.

One main advantage of the persona-based LLM disagreement measure is that it yields a daily measure of disagreement about traditional news, which is useful for analyzing higher frequency movements (in comparison to analyst dispersion, which is slower moving). Of the measures in this validation exercise, StockTwits disagreement is also available at the firm-day level. The firm-day correlation with StockTwits disagreement is statistically significant but economically negligible ($r = 0.0011$), consistent with other evidence that traditional and social media reflect distinct information ecosystems (Cookson, Lu, Mullins, and Niessner, 2024).

3.1.3 Disagreement Trends over Time

In this section, we examine how the LLM-derived disagreement measure evolves over the sample period. A natural concern is that any trend in average disagreement could reflect changes in the composition of news sources over time, rather than genuine shifts in how investors interpret information. To address this, we construct a residual disagreement measure that accounts for changes in news composition. Specifically, we regress stock-day disagreement on the log of the number of news articles per RavenPack news group (there are 50 such groups), and then compute the average of the residuals plus the constant term for each day.

Figure 6 plots the yearly average and interquartile range (the shaded area) of this daily residual persona disagreement. The figure shows that disagreement remains relatively stable in the early 2010s, gradually declines through the mid-to-late 2010s, and then begins to rise again around 2022. The increase becomes more pronounced in the most recent years. Because we have removed variation attributable to changes in news composition, this pattern suggests a genuine increase in the dispersion of investor views, particularly following the Covid-19 pandemic and the heightened social and political polarization of that period.

3.2 Benchmarking LLM Responses to Human Surveys

From the variation in disagreement across news types, a plausible inference is that the differences of opinion reflect non-pecuniary motives like fairness or care for the environment, as well as behavioral biases. To provide support for these mechanisms, we evaluate Llama's ability to capture these dimensions using two public-domain human surveys as a benchmark.

First, we evaluate how the responses of our Llama personas align with how human survey respondents view corporate decisions when financial value is held constant. To address this, we replicate *in silico* the recent survey of Iliewa, Kempf, and Spalt (2024), who asked a representative sample of U.S. adults to rank ten corporate actions (listed in Figure 7) of a typical S&P 500 firm under the assumption that each action had the same certain, positive NPV. By holding financial value constant, their survey design isolates nonpecuniary motives for evaluating corporate actions. In the human survey, respondents consistently rated layoffs and increases in CEO pay as the most objectionable, while actions such as share buybacks, cost cutting, and outsourcing were viewed as less so. We administer the same survey to our LLM personas to test whether they exhibit similar moral judgments.

Figure 7 compares persona-based rankings to the human survey results. Because the survey post-dates the training cutoff, this comparison is naturally out of sample. Panel A plots the average wrongfulness rankings across all personas (x-axis) against the corresponding human rankings (y-axis), showing a strong positive relationship: corporate actions deemed most objectionable by personas are also judged most objectionable by humans, with layoffs and CEO pay increases consistently ranked as the most wrong by both groups. Panels B through D disaggregate the LLM persona responses by politics, income, and gender, plotting each subgroup's average rankings against the human rankings within each subgroup in the survey. The strong positive relationship persists within all subgroups, indicating that the alignment between LLM personas and human respondents is not driven by any particular demographic segment. The close correspondence between the LLM-elicited and human responses suggests that persona-based elicitation captures meaningful heterogeneity in attitudes toward corporate decisions, and that a significant portion of this heterogeneity stems from nonpecuniary motives.

Beyond non-pecuniary motives, psychological biases may also generate divergent interpretations of the same news. To evaluate Llama's ability to capture such behavioral biases, we replicate the large-scale "digital twins" study of Toubia, Gui, Peng, Merlau, Li, and Chen (2025) using Llama 3.1 8B, seeding the synthetic respondents only with the demographic information available in the digital twins database. The original study created digital twins by providing an LLM with both demographic information and a transcript of each respondent's prior survey answers, then showed that these twins responded similarly to human respondents on held-out questions designed to reveal classic behavioral biases. We take a stricter approach by seeding only demographics (no individual transcripts) and testing whether this is sufficient to reproduce held-out bias questions. Our replication largely shows that it is.

Specifically, we replicate Toubia, Gui, Peng, Merlau, Li, and Chen (2025) by focusing on single-response questions (excludes the "matrix" style surveys), since this format most closely resembles the format of our buy/hold/sell prompt. We obtain responses to each survey question using four different methods: (i) Llama 3.1 8B seeded with demographics only, (ii) GPT-4.1 mini seeded with demographics only, (iii) GPT-4.1 mini seeded with demographics and the full question and answer transcript from the human survey, and (iv) random guessing. We compare the answers from each of these conditions with the ground-truth human survey responses. The average accuracy rates are 64.87% (Llama with demographics only), 69.34% (GPT-4.1 mini with demographics only), 71.72% (GPT-4.1 mini with demographics and transcript), and 59.16% (random guessing). These results indicate that demographics alone provide sufficient context for LLMs to replicate well-known behavioral biases in human survey responses; adding the full human survey question and answer transcript only modestly increases accuracy. Moreover, the similar performance of Llama 3.1 8B and GPT-4.1 mini suggests that our main findings are unlikely to be driven by idiosyncrasies of a particular model. This finding supports our use of demographic-seeded personas to capture heterogeneity in how investors interpret information.

Having established that LLM personas capture meaningful variation in beliefs, we next relate this variation to trading volume in financial markets, which verifies that the LLM-derived measures are market-relevant.

3.3 Disagreement and Abnormal Trading Volume

To capture how LLM-based disagreement relates to trading, we estimate the following panel regression:

$$\text{AbnVol}_{i,t+h} = \alpha + \beta\,\text{AbnDisag}_{i,t} + \phi_1\,\text{AbnVol}_{i,t-1} + \delta' X_{i,t} + \eta_i + \lambda_t + \varepsilon_{i,t+h}, \tag{3}$$

where the dependent variable $\text{AbnVol}_{i,t+h}$ is the abnormal log trading volume for stock i on day $t+h$ (with $h = 0$ for contemporaneous volume and $h = 1, 2, ...5$ for volume over the next five days).[13] We construct this measure using total trading volume, as well as retail trading volume drawn from TAQ (Boehmer, Jones, Zhang, and Zhang, 2021). The key explanatory variable $\text{AbnDisag}_{i,t}$ is abnormal disagreement at time t, computed by subtracting the trailing 20-day average of disagreement for firm i. Most variables are standardized to facilitate interpretation and comparison of coefficient magnitudes.

The regression includes a standard suite of controls: lagged abnormal volume, $\text{AbnVol}_{i,t-1}$ to capture persistence in daily trading, recent return volatility (5- and 25-day windows), recent abnormal returns (5- and 25-day windows), and RavenPack's media sentiment score. We also control for alternative proxies for disagreement and media attention: the standard deviation of RavenPack media sentiment scores across articles, the log number of media articles, StockTwits disagreement, and average persona response. η_i and λ_t are stock and date fixed effects, and standard errors are clustered by stock and date.

Table 6 shows that LLM persona disagreement has a large and statistically significant positive association with both total trading volume (odd columns) and retail trading volume (even columns). In the full sample (columns 1-2), a one-standard-deviation increase in disagreement is associated with a 0.42% rise in same-day abnormal log volume and a 0.51% increase in same-day abnormal retail trading volume, holding constant other factors known to predict daily trading activity. In columns (3) through (6), we evaluate this relation inside versus outside the Llama 3.1 training window, which ended in 2023. The relation between disagreement and trading vol-

[13] We compute abnormal log trading volume on day t for firm i as current log trading volume minus the average log trading volume from the prior six months, excluding the most recent month – i.e., dates in $[t-140, t-21]$, following the construction in (Cookson and Niessner, 2020; Cookson, Engelberg, and Mullins, 2023).

ume, if anything, strengthens in the post-training sample (2024-2025). This larger coefficient both confirms the viability of our approach to elicit disagreement outside of the LLM training window, and reflects a strengthening of the disagreement-trading relation in the later part of our sample.[14]

Importantly, this strong connection persists after controlling for other well-known drivers of trading volume such as recent returns, volatility, media sentiment, StockTwits disagreement, and lagged abnormal volume, suggesting that the effect is not driven by mechanical volume persistence or overall news intensity. The specification also includes the log number of media articles and the dispersion of RavenPack sentiment across articles as an alternative proxy for media disagreement. Beyond showing that LLM Disagreement contains information incremental to these other predictors of daily trading, the coefficient magnitudes are comparable to those for recent volatility and recent returns while being smaller than those for social media disagreement.

To test the robustness of these findings to alternative weighting assumptions, we re-estimate the analysis using an equal-weighted version of persona disagreement. Table A10 shows that the association between disagreement and trading volume remains virtually unchanged when all personas receive equal weight instead of FINRA-based sample weights. This similarity confirms that our findings are not driven by the weighting of particular demographic groups and that the relationship between disagreement and trading is broadly representative across personas.

We also assess whether the results depend on specific demographic dimensions used to construct persona disagreement. Table A11 reports regressions that average out the politics dimension, and both politics and income dimensions, respectively. The coefficients remain highly significant, and if anything, they strengthen in magnitude. These results indicate that the positive relationship between disagreement and trading volume is not solely attributable to politically- or income-driven variation in responses, but rather reflects broader heterogeneity across demographic characteristics that shape how investor personas interpret firm news.

[14] In Table A9 we estimate the specification across various subsamples. The coefficient magnitudes are similar in the last two years of the training window (2022–2023) and the post-training window (2024–2025), and they remain stable when we estimate them in the years around the end of the training window. This finding indicates that the magnitude differences we see in the full sample reflect longer-term differences in the relation between disagreement and trading.

To examine how long disagreement-driven trading persists, we estimate our baseline specification at horizons $h = 0, 1, ...5$, and plot the coefficients in Figure 8. In Panel A, the dependent variable is abnormal log trading volume at date $t + h$, while Panel B focuses on abnormal log retail trading volume. In both cases, the coefficients are largest contemporaneously and decline out to day 5, but remain positive and statistically significant at day 1, indicating that LLM-based disagreement not only correlates with same-day trading activity but also *predicts* elevated volume on the following day. The estimated coefficients are larger for retail volume, consistent with personas being weighted to match the FINRA retail-investor distribution. Figure A2 replicates this analysis using equal-weighted persona disagreement and shows a nearly identical pattern, confirming that the dynamics of disagreement-driven trading do not depend on the FINRA-based weighting scheme.

3.3.1 LLM Disagreement and Trading Across Persona Types

Finally, we examine whether the source of disagreement (e.g., income, political outlook, race) matters for the relation between disagreement and trading. For example, disagreement stemming from income disparities may reflect conflicting views on a policy's distributive effects, whereas political or racial heterogeneity might capture ideological or social preferences that impact trading differently. It is an empirical question which dimensions of disagreement are more strongly related to trading. To explore this question, we re-estimate Eq. (3) using each within-dimension disagreement measure in turn.

In Table 7 Panel A, we find that total contemporaneous trading volume responds positively to within-dimension disagreement across all six demographic dimensions, though the magnitudes vary considerably. Education-based disagreement has the largest coefficient: a one-standard-deviation increase is associated with a 0.76% increase in same-day abnormal log volume. Age (0.42%) and income (0.38%) disagreement also show strong associations, while racial (0.25%), political (0.08%), and gender (0.06%) disagreement exhibit smaller, though still positive, effects. All coefficients except gender are statistically significant at the 1% level. Controls for lagged volume, volatility, returns, media sentiment, and fixed effects ensure these patterns are not driven by mechanical volume autocorrelation or overall news intensity. Panel B repeats the

analysis using retail trading volume as the dependent variable. The pattern is similar with slightly larger coefficient magnitudes throughout. Education-based disagreement remains the strongest predictor (0.90%), followed by age (0.52%) and income (0.44%). Political and gender disagreement again show the weakest associations.

To examine the predictive power of different demographic dimensions for future trading activity, we extend the analysis to next-day trading volume in Table A12. The broad pattern persists: age, education, and income disagreement remain the strongest predictors of both total and retail trading volume, while political and gender disagreement continue to show weaker associations. Although coefficient magnitudes generally attenuate from day t to $t+1$, the relative ranking of demographic dimensions remains largely stable, reinforcing that certain characteristics consistently capture investment-relevant interpretation differences.

These results suggest that not all sources of disagreement contribute equally to trading activity. Across both same-day and next-day specifications, disagreement rooted in education, age, and income translates most strongly into trading volume, while political and gender-based disagreement, despite generating meaningful differences in responses, appear less connected to actual trading behavior. This pattern is consistent with the idea that certain demographic dimensions capture differences in investment-relevant interpretations, whereas others may reflect opinion differences that do not translate as directly into portfolio actions.

4 Conclusion

This paper shows that open-source LLMs can serve as effective "mirrors" of human belief differences. By seeding the Llama 3.1 8B model with 216 investor profiles drawn from the FINRA survey, we generated high-frequency buy, hold and sell responses for a comprehensive sample of S&P 500 firm-specific news, comprising over 5.5 million unique headlines and 1.188 billion headline–profile response pairs. We find that these investor personas replicate well-known demographic splits and exhibit clear disagreement around socially charged and governance-related news such as executive pay, workforce compensation, and environmental issues. Disagreement is greatest for news linked to social or ethical dimensions of firms' actions, while "hard" finan-

cial news elicits the least dispersion. Income and political orientation emerge as particularly important drivers of disagreement.

Several findings support the validity of our approach. LLM persona responses align closely with human survey responses on the perceived wrongfulness of corporate actions, and the patterns we document persist outside the model's training window. Moreover, we show that higher disagreement predicts significantly greater trading volume, particularly among retail investors. These relationships are robust to controls for volatility, returns, media sentiment, social-media disagreement, and lagged volume, suggesting that LLM-derived disagreement captures genuine heterogeneity in investor beliefs rather than mechanical trading patterns.

Our persona-based approach offers a scalable, transparent, and high-frequency framework for capturing belief heterogeneity, particularly in settings where traditional surveys are infeasible. We anticipate future research will adopt similar approaches to study a wide range of financial market phenomena, especially in settings that require large-scale, high-frequency belief data. Overall, our results demonstrate that LLM personas provide a new lens into how demographics shape the interpretation of news and the behavior of investors in financial markets.

References

Aumann, R. J., 1976, "Agreeing to disagree," *The Annals of Statistics*, 4(6), 1236–1239.

Axelrod, R., and W. D. Hamilton, 1981, "The evolution of cooperation," *science*, 211(4489), 1390–1396.

Babina, T., A. Fedyk, A. He, and J. Hodson, 2024, "Artificial intelligence, firm growth, and product innovation," *Journal of Financial Economics*, 151, 103745.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel, 2018, "The economic effects of social networks: Evidence from the housing market," *Journal of Political Economy*, 126(6), 2224–2276.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong, 2018, "Social connectedness: Measurement, determinants, and effects," *Journal of Economic Perspectives*, 32(3), 259–280.

Baker, S. R., N. Bloom, and S. J. Davis, 2016, "Measuring economic policy uncertainty," *The Quarterly Journal of Economics*, 131(4), 1593–1636.

Barber, B. M., and T. Odean, 2001, "Boys will be boys: Gender, overconfidence, and common stock investment," *The quarterly journal of economics*, 116(1), 261–292.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang, 2021, "Tracking retail investor activity," *The Journal of Finance*, 76(5), 2249–2305.

Brynjolfsson, E., D. Li, and L. Raymond, 2025, "Generative AI at work," *The Quarterly Journal of Economics*, p. qjae044.

Bybee, J. L., 2023, "The ghost in the machine: Generating beliefs with large language models," *arXiv preprint arXiv:2305.02823*.

Caldara, D., and M. Iacoviello, 2022, "Measuring geopolitical risk," *American economic review*, 112(4), 1194–1225.

Chen, S., T. C. Green, H. Gulen, and D. Zhou, 2024, "What does chatgpt make of historical stock returns? extrapolation and miscalibration in llm stock return forecasts," *arXiv preprint arXiv:2409.11540*.

Cookson, J. A., C. Dim, and M. Niessner, 2024, "Disagreement on the Horizon," *Available at SSRN*.

Cookson, J. A., J. E. Engelberg, and W. Mullins, 2020, "Does partisanship shape investor beliefs? Evidence from the COVID-19 pandemic," *Review of Asset Pricing Studies*, 10(4), 863–893.

——— , 2023, "Echo chambers," *The Review of Financial Studies*, 36(2), 450–500.

Cookson, J. A., R. Lu, W. Mullins, and M. Niessner, 2024, "The social signal," *Journal of Financial Economics, Forthcoming*.

——— , 2025, "Market signals from social media," *Available at SSRN*.

Cookson, J. A., W. Mullins, and M. Niessner, 2024, "Social Media and Finance," *Available at SSRN 4806692*.

Cookson, J. A., and M. Niessner, 2020, "Why don't we agree? Evidence from a social network of investors," *The Journal of Finance*, 75(1), 173–228.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998, "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 53(6), 1839–1885.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons, 2012, "Journalists and the stock market," *The Review of Financial Studies*, 25(3), 639–679.

Eisfeldt, A. L., G. Schubert, and M. B. Zhang, 2023, "Generative AI and firm values," working paper, National Bureau of Economic Research.

Engelberg, J., J. Guzman, R. Lu, and W. Mullins, 2022, "Partisan entrepreneurship," working paper, National Bureau of Economic Research.

Engelberg, J., A. Manela, W. Mullins, and L. Vulicevic, 2025, "Entity Neutering," *Available at SSRN*.

Fedyk, A., A. Kakhbod, P. Li, and U. Malmendier, 2024, "Chatgpt and perception biases in investments: An experimental study," *Available at SSRN 4787249*.

Garcia, D., 2013, "Sentiment during recessions," *The journal of finance*, 68(3), 1267–1300.

Geiecke, F., and X. Jaravel, 2024, "Conversations at scale: Robust AI-led interviews with a simple open-source platform," *Available at SSRN 4974382*.

Gerken, W. C., and M. O. Painter, 2023, "The value of differing points of view: Evidence from financial analysts' geographic diversity," *The Review of Financial Studies*, 36(2), 409–449.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus, 2021, "Five facts about beliefs and portfolios," *American Economic Review*, 111(5), 1481–1522.

Goetzmann, W. N., D. Kim, and R. J. Shiller, 2024, "Emotions and Subjective Crash Beliefs," working paper, National Bureau of Economic Research.

Goldman, E., N. Gupta, and R. Israelsen, 2024, "Political polarization in financial news," *Journal of Financial Economics*, 155, 103816.

Goldsmith-Pinkham, P., and K. Shue, 2023, "The gender gap in housing returns," *The Journal of Finance*, 78(2), 1097–1145.

Greenwood, R., and A. Shleifer, 2014, "Expectations of Returns and Expected Returns," *The Review of Financial Studies*, 27(3), 714–746.

Gupta, A., C. Hansman, and P. Mabille, 2024, "Financial constraints and the racial housing gap," *Journal of Financial Economics, Forthcoming*.

Hansen, A. L., J. J. Horton, S. Kazinnik, D. Puzzello, and A. Zarifhonarvar, 2024, "Simulating the survey of professional forecasters," *Available at SSRN*.

Hassan, T. A., S. Hollander, L. Van Lent, and A. Tahoun, 2019, "Firm-level political risk: Measurement and effects," *The Quarterly Journal of Economics*, 134(4), 2135–2202.

Hong, H., and J. C. Stein, 1999, "A unified theory of underreaction, momentum trading, and overreaction in asset markets," *Journal of Finance*, 54(6), 2143–2184.

——— , 2007, "Disagreement and the stock market," *Journal of Economic perspectives*, 21(2), 109–128.

Horton, J. J., 2023, "Large language models as simulated economic agents: What can we learn from homo silicus?," working paper, National Bureau of Economic Research.

Iliewa, Z., E. Kempf, and O. G. Spalt, 2024, "Corporate actions as moral issues," *Available at SSRN*.

Kandel, E., and N. D. Pearson, 1995, "Differential interpretation of public signals and trade in speculative markets," *Journal of Political Economy*, 103(4), 831–872.

Kapons, M., and P. Kelly, 2023, "Biased inference due to prior beliefs: Evidence from the field," *Journal of Political Economy: Microeconomics*.

Karpoff, J. M., 1986, "A Theory of Trading Volume," *The Journal of Finance*, 41(5), 1069–1087.

Kelly, B., S. Malamud, and K. Zhou, 2024, "The virtue of complexity in return prediction," *The Journal of Finance*, 79(1), 459–503.

Kempf, E., and M. Tsoutsoura, 2021, "Partisan professionals: Evidence from credit rating analysts," *The journal of finance*, 76(6), 2805–2856.

Kozlowski, A., and J. Evans, 2024, "Simulating subjects: The promise and peril of ai stand-ins for social agents and interactions," .

Lopez-Lira, A., and Y. Tang, 2023, "Can chatgpt forecast stock price movements? return predictability and large language models," *arXiv preprint arXiv:2304.07619*.

Lopez-Lira, A., Y. Tang, and M. Zhu, 2025, "The Memorization Problem: Can We Trust LLMs' Economic Forecasts?," *arXiv preprint arXiv:2504.14765*.

Lusardi, A., and O. S. Mitchell, 2008, "Planning and financial literacy: How do women fare?," *American economic review*, 98(2), 413–417.

Madrian, B. C., and D. F. Shea, 2001, "The power of suggestion: Inertia in 401 (k) participation and savings behavior," *The Quarterly journal of economics*, 116(4), 1149–1187.

Meeuwis, M., J. A. Parker, A. Schoar, and D. Simester, 2022, "Belief disagreement and portfolio choice," *The Journal of Finance*, 77(6), 3191–3247.

Miller, E. M., 1977, "Risk, Uncertainty, and Divergence of Opinion," *Journal of Finance*, 32(4), 1151–1168.

Noy, S., and W. Zhang, 2023, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 381(6654), 187–192.

Qin, X., M. Huang, and J. Ding, 2024, "AITurk: Using ChatGPT for social science research," *Available at SSRN 4922861*.

Schelling, T. C., 1971, "Dynamic models of segregation," *Journal of mathematical sociology*, 1(2), 143–186.

Tetlock, P. C., 2007, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 62(3), 1139–1168.

——— , 2011, "All the news that's fit to reprint: Do investors react to stale information?," *The Review of Financial Studies*, 24(5), 1481–1512.

Toubia, O., G. Z. Gui, T. Peng, D. J. Merlau, A. Li, and H. Chen, 2025, "Database report: Twin-2k-500: A data set for building digital twins of over 2,000 people based on their answers to over 500 questions," *Marketing Science*, 44(6), 1446–1455.

Varian, H. R., 1985, "Divergence of opinion in complete markets: A note," *The Journal of Finance*, 40(1), 309–317.

Figure 1: LLM Prompt Template and Example Output. This figure shows the prompt template used to endow the LLM with human profiles (persona) and elicit beliefs in Panel A, and two examples of news headlines and persona responses to the headline in Panel B.

A: Prompt Template

```
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
```

B: Sample Headlines and Persona Responses

News Headline: Wells Fargo CEO Stumpf total 2010 compensation valued at $19 million		
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	Wells Fargo's high CEO compensation suggests a strong company with potential for future growth and increased stock value.
Non-White Female, Age 55+, with some college or less of education, household income less than $50K, and politically independent.	SELL	I'm worried about the huge CEO compensation, it's unfair to regular folks who can barely make ends meet.
News Headline: Chevron confirms new oil discoveries off Congo		
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.

Figure 2: Illustration of Article-Level Disagreement. To provide an illustration of article-level disagreement, this figure shows the relative frequency of buy/hold/sell sentiment for the two news headlines featured in Figure 1.

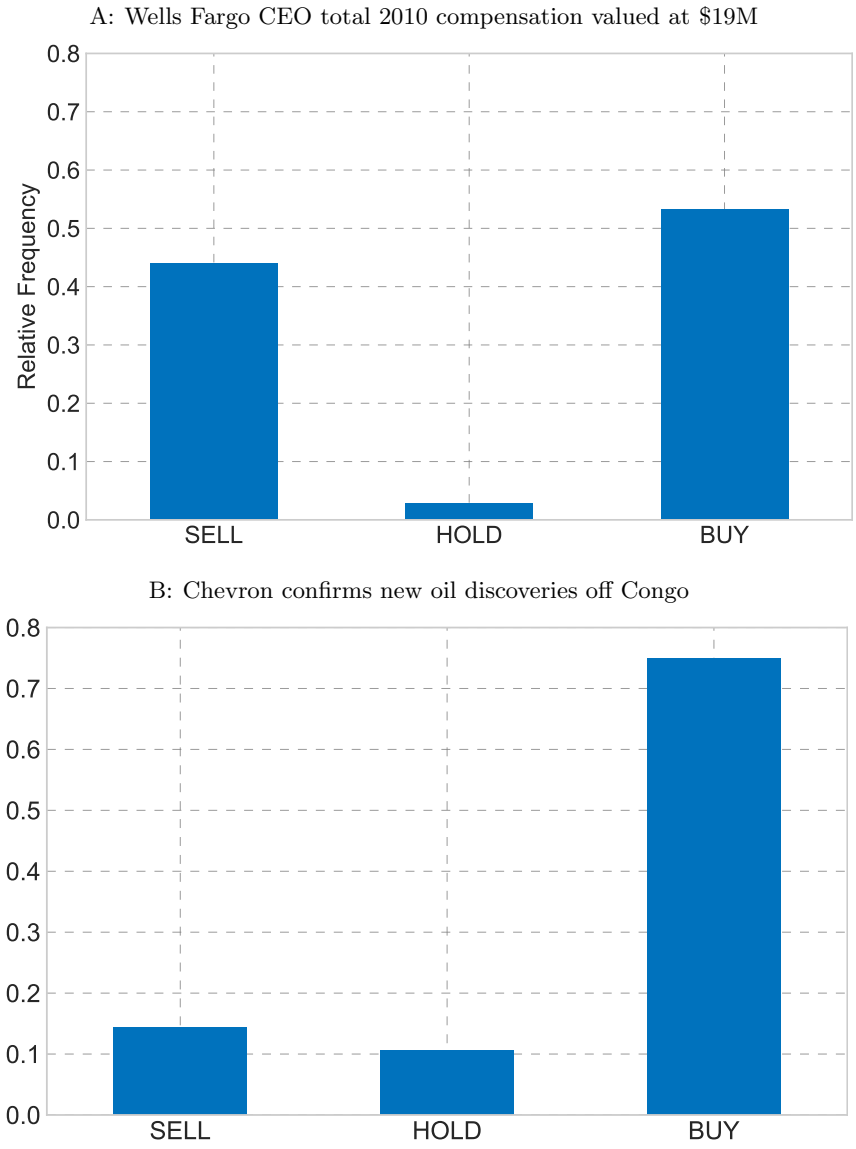


Figure 3: Sentiment Differences Across Investor Personas, Averaged Across All News Headlines. This figure shows the average sentiment per profile across all news plotted against its percentile ranks. The dots in the figures are colored by the income level (Panel A), political leaning (Panel B), and age (Panel C) of the profiles.

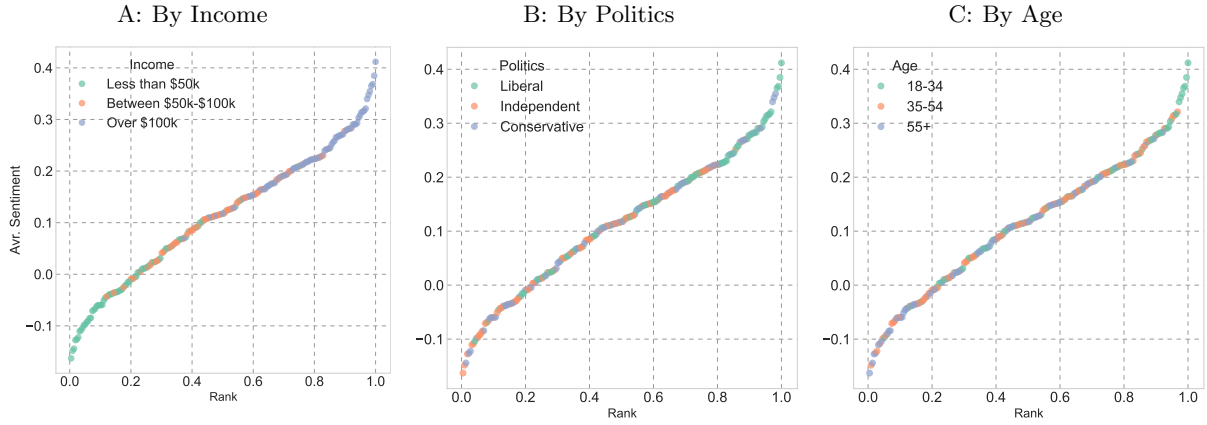


Figure 4: High versus Low Disagreement News Categories. This figure presents the top 20 (Panel A) and bottom 20 (Panel B) RavenPack news types sorted on average LLM Disagreement across headlines. The RavenPack type descriptions are presented in Appendix Table A4.

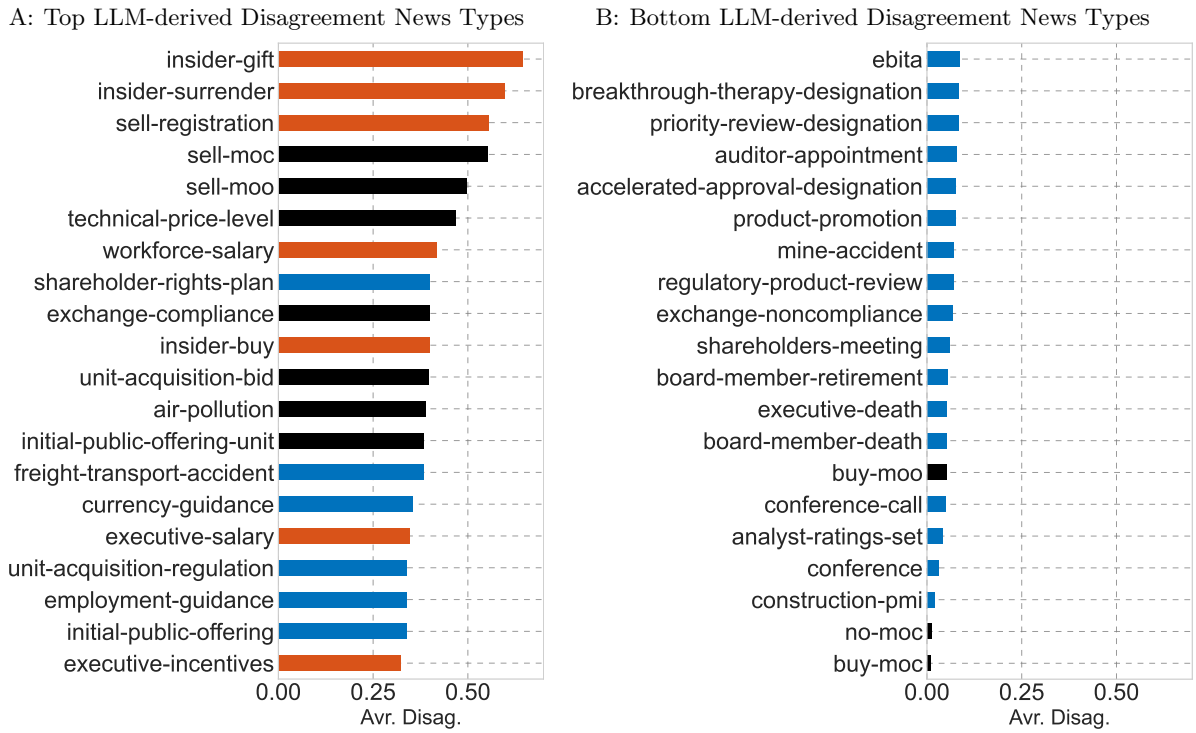


Figure 5: Persistence of Disagreement into Llama 3.1's Post-Training Window. This figure shows the scatter plot of average disagreement per RavenPack news type in the LLM training window (ending 2023) plotted against the corresponding values in the post-training window (2024 - 2025). We require at least five news articles per news type in both samples for the news type to be included in the plot.

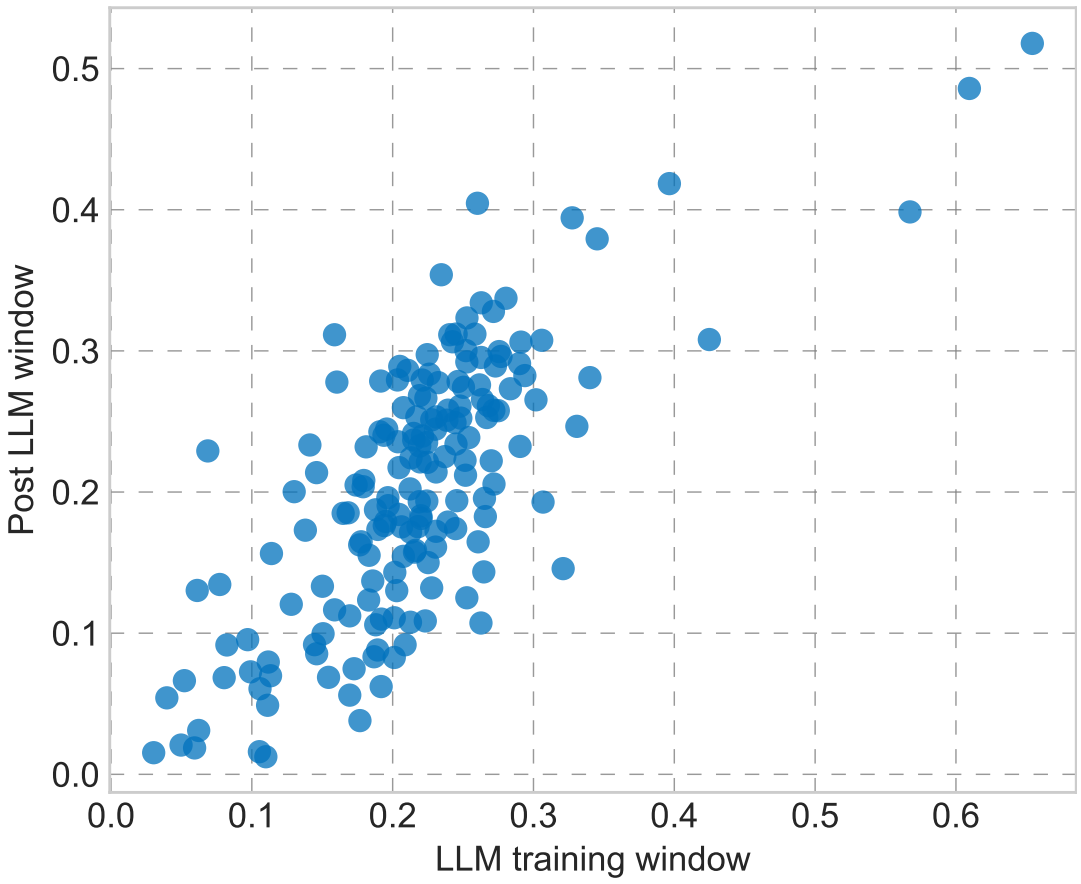


Figure 6: Time-series Evolution of the LLM-derived Disagreement. This figure plots the yearly average and inter-quartile range (shaded area) of daily residual disagreement. To account for news composition changes over time, daily residual disagreement is obtained from regressing stock-day disagreement on the log of the number of news articles per RavenPack news group per day. The sum of the regression residuals and the constant term is then averaged daily to obtain the daily time series.

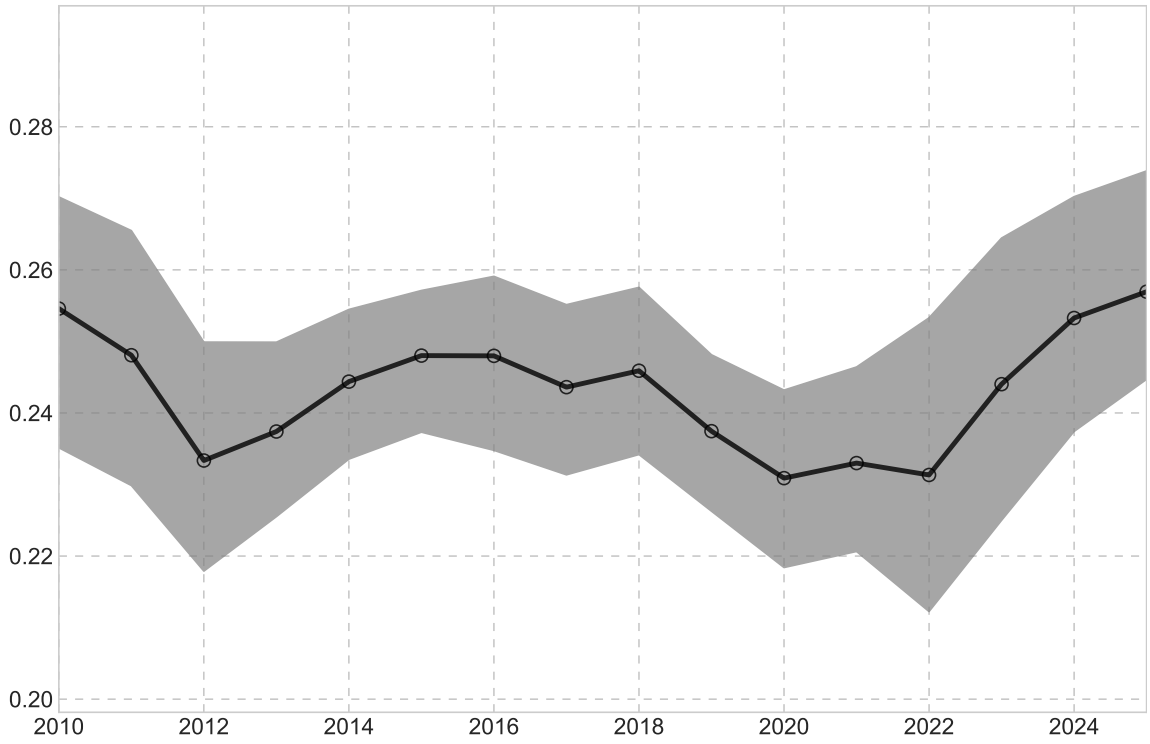


Figure 7: LLM Persona Responses Closely Match Human Responses Out of Sample. This figure reports a synthetic replication of human survey responses using Llama persona responses. Following Iliewa, Kempf, and Spalt (2024), which became public in the post-training window, we assign the LLM personas the exact same task of ranking 10 different corporate actions, all with positive NPV for a firm, on a scale of most wrong to least wrong. The figures plot the average human ranking from Iliewa, Kempf, and Spalt (2024) per corporate action against the LLM personas' average rankings. Higher values mean an action is perceived as more wrong. Panel A shows the plot based on the overall average per action, while Panels B - D show the average within specific demographic dimensions (Politics, Income, and Gender, respectively). Figure A1 depicts the prompt used to obtain the LLM persona responses on the survey task.

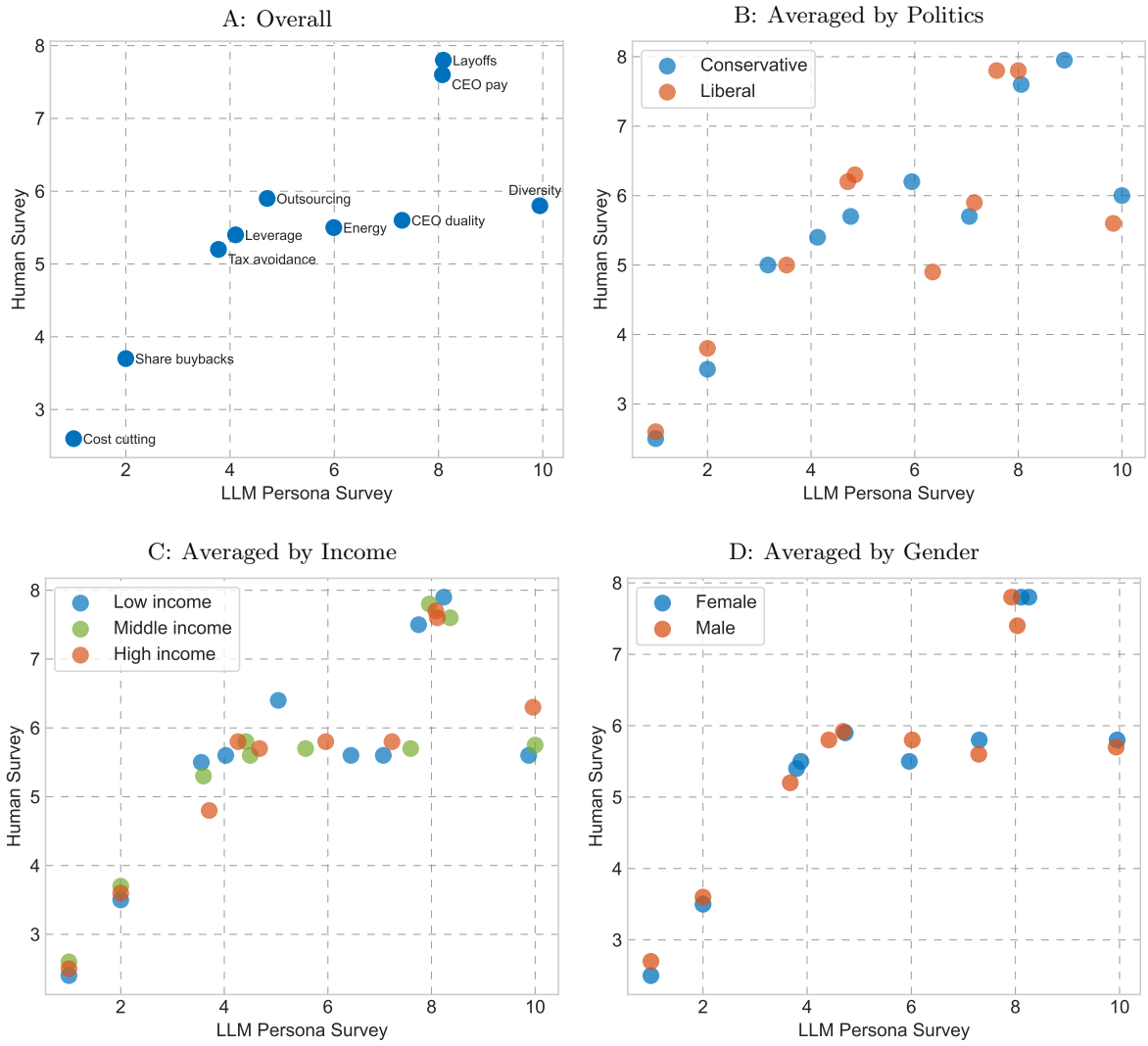


Figure 8: How Future Abnormal Volume Relates to Disagreement. This figure plots the coefficients for abnormal disagreement and its 95% confidence intervals from regressions of Abn. Log Volume$_{t+h} \times 100$ (Panel A) and Abn. Log Retail Volume$_{t+h} \times 100$ (Panel B) on Abn Disag$_t$ for h between 0 and 5 trading days. Abn Disag$_t$ is calculated as described under Table 6 and is standardized to zero mean and unit variance. All regressions include stock and date fixed effects, as well as the full set of controls listed in Table 6. Standard errors are clustered by stock and date.

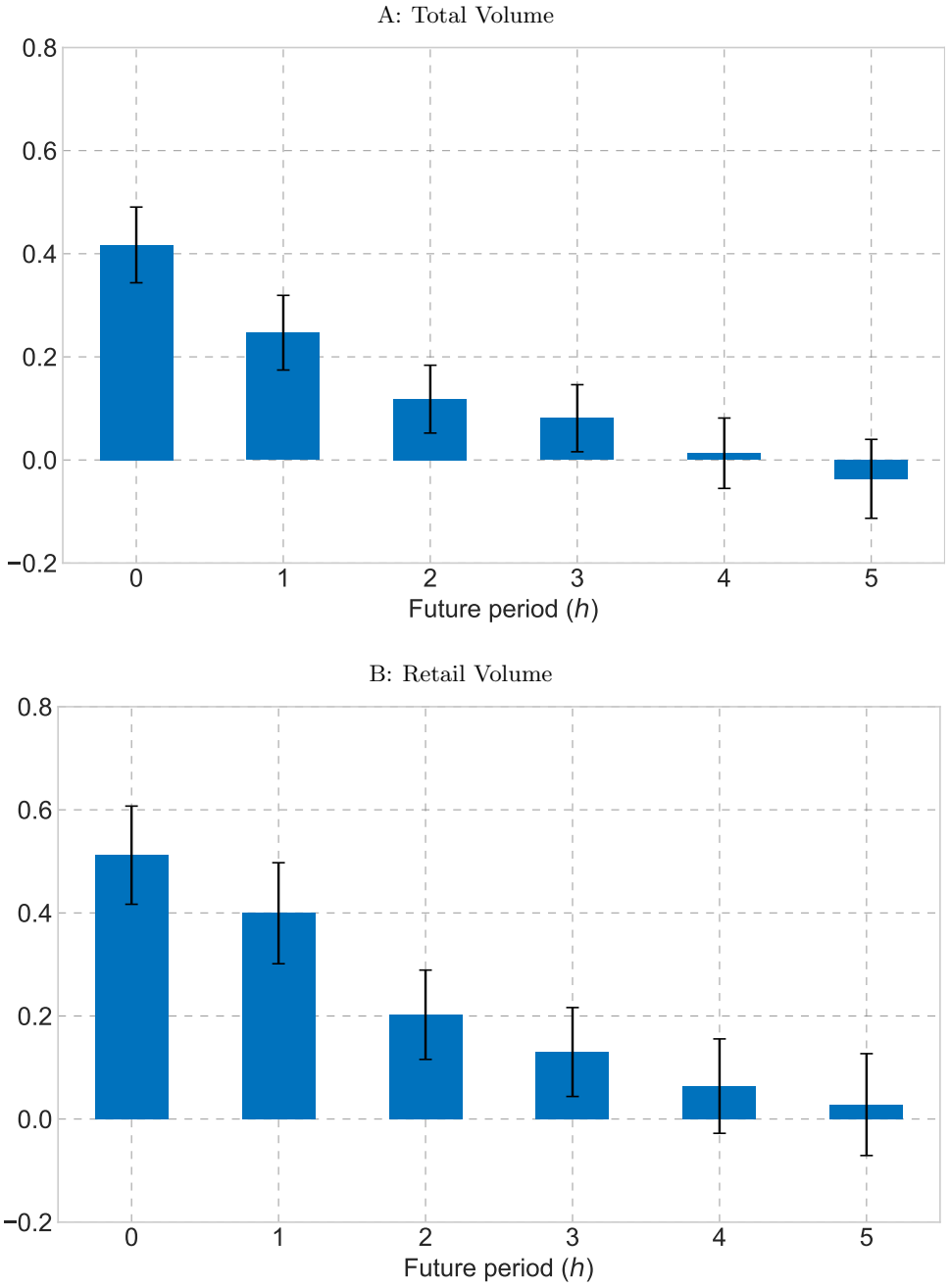


Table 1: Summary of Sentiment and Disagreement. This table reports summary statistics of the sentiment and disagreement measures inferred from the LLM. Panel A summarizes the LLM-derived sentiment for each news and individual profile pair (denoted "Sentiment") and the sentiment by each demographic dimension after averaging out the other demographic dimensions per news headline. The "BUY", "HOLD", and "SELL" sentiment labels are mapped to numerical values, 1, 0, and -1, respectively. These summary statistics are based on 1.188 billion prompts (216 × 5.5 million news headlines). Panel B summarizes the disagreement measures at the article level obtained from the LLM-derived sentiment for 5,501,751 news headlines. "Disagreement" is obtained as the weighted standard deviation of sentiment across individual profiles per news headline. For each demographic dimension, we calculate disagreement across that focal dimension (e.g., Income) by first averaging out the sentiment of the other dimensions per news headline, then obtain the standard deviation across the levels of that dimension per news article (e.g., Income Disagreement). "Ex-Pol Disag." denotes the disagreement measure obtained after averaging out the politics dimension from the sentiment. "Ex-PolInc Disag." denotes the disagreement measure obtained after averaging out both the politics and income dimensions from the sentiment.

Panel A: Sentiment Measures

	Mean	% BUY	% HOLD	% SELL
Sentiment	0.114	25.7	59.9	14.3
Sentiment by Gender:				
Female	0.088	22.9	63.0	14.1
Male	0.141	28.6	56.9	14.5
Sentiment by Age:				
18-34	0.173	31.0	55.4	13.7
35-54	0.101	24.9	60.3	14.8
55+	0.068	21.3	64.1	14.5
Sentiment by Race:				
Non-White	0.098	24.2	61.4	14.4
White non-Hispanic	0.130	27.3	58.4	14.3
Sentiment by Education:				
A college degree or more	0.146	26.7	61.3	12.1
Some college or less	0.082	24.8	58.6	16.6
Sentiment by Income:				
Over $100K	0.235	32.5	58.4	9.0
Between $50K-$100K	0.114	24.3	62.8	12.9
Less than $50K	-0.007	20.4	58.6	21.0
Sentiment by Politics:				
Conservative	0.107	26.7	57.3	16.0
Independent	0.069	20.6	65.7	13.7
Liberal	0.166	29.9	56.8	13.3

Panel B: Disagreement Measures

	Mean	SD
Disagreement	0.227	0.231
Income Disagreement	0.109	0.145
Political Disagreement	0.067	0.089
Age Disagreement	0.046	0.063
Education Disagreement	0.033	0.050
Gender Disagreement	0.035	0.047
Racial Disagreement	0.027	0.040
Ex-Pol Disagreement	0.172	0.187
Ex-PolInc Disagreement	0.092	0.099

Table 2: Illustrating Demographic Differences in Sentiment. The table presents results for the regressions of sentiment declared by the LLM personas on demographic characteristics. We illustrate sentiment differences by focusing on several illustrative news categories related to polarizing news categories connected to compensation and environment. The column headers indicate the news categories comprising the analysis sample for that specification. The dependent variable, sentiment (BUY = 1, HOLD = 0, and SELL = −1), is divided by its standard deviation in each specification. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	(1) Executive Pay Increase	(2) Workforce Pay Increase	(3) Share Buyback	(4) Fossil Fuel Discovery	(5) Employee Layoff
Income Over $100K	1.040***	0.303***	0.168***	0.245***	0.434***
	(0.041)	(0.035)	(0.008)	(0.053)	(0.015)
Income $50K - $100K	0.618***	0.124***	0.074***	0.159***	0.222***
	(0.028)	(0.023)	(0.004)	(0.038)	(0.010)
Independent	-0.107***	0.475***	-0.108***	-0.107***	0.008
	(0.022)	(0.089)	(0.010)	(0.027)	(0.006)
Liberal	-0.271***	1.029***	-0.001	-0.112***	0.048***
	(0.030)	(0.142)	(0.003)	(0.038)	(0.009)
White non-Hispanic	0.151***	-0.048***	0.057***	0.058***	0.023***
	(0.008)	(0.015)	(0.002)	(0.016)	(0.003)
Male	0.065***	0.043***	0.081***	0.081***	0.030***
	(0.008)	(0.013)	(0.006)	(0.023)	(0.002)
Age 35-54	-0.092***	-0.119***	-0.067***	-0.050***	-0.049***
	(0.006)	(0.009)	(0.003)	(0.012)	(0.002)
Age 55+	-0.128***	-0.183***	-0.122***	-0.096***	-0.063***
	(0.009)	(0.016)	(0.009)	(0.024)	(0.003)
Some college or less	-0.190***	-0.086***	-0.038***	-0.040***	-0.141***
	(0.009)	(0.009)	(0.002)	(0.013)	(0.005)
R-squared Adj.	0.577	0.532	0.500	0.754	0.677
Obs.	206,496	73,656	3,315,384	61,560	1,131,840

Table 3: Text-Based Rationale Differences Across Demographics. This table presents regressions of text-based rationale topics (derived from a word2vec procedure) on persona demographic characteristics. To focus the topic analysis, the sample is restricted to the news types analyzed in the previous Table 2 and in Table A7. We use word2vec embeddings of the textual rationales to construct a 100-word dictionary for five salient themes: Employee Morale, Fairness, Environment, Profit Decline, and Efficiency. Each rationale's loading on each theme is the sum of its *tf-idf* scores on the theme's dictionary words divided by the rationale's number of words. Table A3 lists the seed words for each theme and provides details on the dictionary construction. To aid interpretation of coefficient magnitudes, each dependent variable is divided by its standard deviation. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	(1) Employee Morale	(2) Fairness	(3) Environment	(4) Profit Decline	(5) Efficiency
Income Over $100K	0.041***	-0.018***	-0.050***	-0.007***	0.032***
	(0.004)	(0.003)	(0.004)	(0.001)	(0.003)
Income $50K - $100K	0.010***	-0.012***	-0.047***	-0.002**	0.018***
	(0.002)	(0.002)	(0.004)	(0.001)	(0.002)
Independent	-0.004**	0.002**	0.002***	0.005***	-0.032***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)
Liberal	0.052***	0.015***	0.074***	-0.003***	-0.016***
	(0.004)	(0.002)	(0.008)	(0.001)	(0.003)
White non-Hispanic	-0.012***	-0.003***	-0.031***	0.004***	-0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Male	0.009***	-0.002***	-0.012***	0.003***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 35-54	-0.023***	-0.001	0.008***	0.002***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.000)	(0.001)
Age 55+	-0.035***	-0.001	0.014***	-0.001	-0.010***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Some college or less	-0.004**	-0.005***	-0.018***	-0.002	-0.034***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared Adj.	0.033	0.043	0.010	0.033	0.020
Obs.	28,079,567	28,079,567	28,079,567	28,079,567	28,079,567

Table 4: News Article-Level Disagreement and News Characteristics. This table reports the results for regressions of the LLM disagreement measure on news event type indicators. We conduct the analysis in three samples: Full sample (columns 1 - 3), LLM training window (Columns 4 - 6), and post-LLM training window (Columns 7 - 9). The analysis is at the news headline level, and the disagreement measure is normalized to unit variance to facilitate exposition. In Columns 1, 4 and 7, we categorize news events as either hard news or soft news, and then define *Hard News* as a dummy variable that equals one if the news is clearly tied to firm fundamentals according to our own classification. The hard news groups are listed in Appendix Table A2. In column 2, 5 and 8, we define *Social Issues*, *Environmental Issues*, and *Ethical & Legal Issues* as indicator variables based on Ravenpack's event type variable. Appendix Table A2 lists the event types for each category. In Columns 3, 6 and 9 we classify unambiguously high versus low sentiment using Ravenpack's event sentiment score. *High Ngtv Sent* (*High Pstv Sent*) equals one if the Ravenpack event sentiment for the news headline is in the bottom (top) 5%. All regressions include stock and date fixed effects. Standard errors clustered by stock and date are reported in parentheses.

	Disagreement								
	Full sample			Training window			Post-training		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hard News	-0.167***			-0.157***			-0.291***		
	(0.008)			(0.008)			(0.015)		
Social Issues		1.405***			1.432***			1.089***	
		(0.024)			(0.024)			(0.034)	
Environmental Issues		-0.071*			-0.078**			0.151	
		(0.037)			(0.037)			(0.125)	
Ethical Legal Issues		0.022			-0.027*			0.243***	
		(0.015)			(0.015)			(0.027)	
High Ngtv Sentiment			-0.246***			-0.270***			-0.070***
			(0.013)			(0.013)			(0.027)
High Pstv Sentiment			-0.667***			-0.678***			-0.658***
			(0.010)			(0.010)			(0.020)
R-squared Adj.	0.051	0.083	0.052	0.049	0.084	0.052	0.090	0.095	0.081
Obs.	5,501,751	5,501,751	5,501,751	4,989,153	4,989,153	4,989,153	512,598	512,598	512,598

Table 5: LLM Disagreement versus Existing Measures of Disagreement, Volatility and Uncertainty. This table reports the results for regressions of the LLM disagreement, normalized to unit variance, on alternative uncertainty, volatility and disagreement measures. The specification in column 1 is at the stock-quarter frequency. For each quarterly earnings announcement day, it regresses the average disagreement measure for the trailing quarter on the dispersion of analysts' forecasts for the focal earnings release. The rest of the specifications (columns 2 - 8) are at the stock-month frequency, regressing the stock-month average disagreement on the existing measure plus year-month fixed effects. Policy Uncert. Expo is the absolute value of a firm's beta on the Economic Policy Uncertainty (EPU) Index of Baker, Bloom, and Davis (2016) based on a monthly time series regression of stock returns on monthly changes in the EPU Index using data over the past five years. Similarly, the Geopol. Risk Expo is the absolute value of a firm's beta on the Geopolitical Risk (GPR) Index of Caldara and Iacoviello (2022). Market Beta is a stock's CAPM beta based on a similar monthly time series regression. Political Risk Expo is the most recent firm-level political risk measure of Hassan, Hollander, Van Lent, and Tahoun (2019) based on earnings conference calls. Stocktwits Disag is the stock-month average of the Stocktwits disagreement measure from Cookson and Niessner (2020). Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. All the regressions include time fixed effects. Standard errors clustered by stock and time are reported in parentheses.

	Disagreement							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Analyst Forecast Disp. (z)	0.007							
	(0.011)							
Volatility (z)		0.078***						0.060***
		(0.014)						(0.018)
Stocktwits Disag. (z)			0.077***					0.049***
			(0.016)					(0.018)
Political Risk Expo (z)				-0.037***				-0.037***
				(0.008)				(0.008)
Policy Uncert. Expo (z)					0.032***			-0.006
					(0.009)			(0.010)
Geopol. Risk Expo (z)						0.046***		0.020**
						(0.010)		(0.010)
Market Beta (z)							0.038***	-0.005
							(0.011)	(0.015)
R-squared Adj.	0.088	0.076	0.079	0.077	0.071	0.072	0.071	0.084
Obs.	42,609	132,157	120,994	100,003	131,906	131,906	131,906	90,238

Table 6: Relating Stock-Day Trading Volume to LLM-derived Disagreement. This table reports the results from regressing day t stock abnormal log trading volume (odd-numbered columns) and abnormal log retail trading volume (even-numbered columns) on day t abnormal LLM disagreement, denoted Abn Disag. Abn Disag is calculated as day t's disagreement minus 20-day trailing average. Abnormal log trading volume measures are calculated following Cookson and Niessner (2020). Both abnormal log trading volume and abnormal log retail trading volume are multiplied by 100 for the regressions. We conduct the analysis in the full sample 2010-2025 (columns denoted "Full sample"), LLM training window, ending 2023 (columns labeled "Training window"), and post-LLM training window 2024-2025 (columns labeled "Post-training") Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects. Standard errors clustered by stock and date are reported in parentheses.

	Full sample		Training window		Post training	
	Total Volume	Retail Volume	Total Volume	Retail Volume	Total Volume	Retail Volume
	(1)	(2)	(3)	(4)	(5)	(6)
Abn Disag (t) (z)	0.417***	0.512***	0.309***	0.371***	1.475***	1.772***
	(0.037)	(0.049)	(0.038)	(0.050)	(0.089)	(0.111)
Volatility (t-5, t-1) (z)	-0.591***	1.034***	-0.555***	1.302***	-0.223	-0.203
	(0.154)	(0.226)	(0.158)	(0.241)	(0.459)	(0.697)
Volatility (t-25, t-6) (z)	1.395***	2.039***	1.471***	2.223***	1.519***	1.703**
	(0.165)	(0.235)	(0.195)	(0.292)	(0.569)	(0.691)
Abn Ret (t-5, t-1) (z)	-0.564***	-1.036***	-0.573***	-1.071***	-0.556***	-0.786***
	(0.088)	(0.111)	(0.089)	(0.115)	(0.197)	(0.234)
Abn Ret (t-25, t-6) (z)	-1.213***	-2.090***	-1.198***	-2.081***	-1.435***	-2.051***
	(0.097)	(0.141)	(0.097)	(0.147)	(0.261)	(0.355)
Ave Media Sentiment (t) (z)	-0.047	-0.053	-0.068	-0.086	0.143	0.209
	(0.053)	(0.066)	(0.055)	(0.069)	(0.139)	(0.162)
Std Media Sentiment (t) (z)	0.661***	0.825***	0.787***	0.948***	-0.790***	-0.652***
	(0.092)	(0.114)	(0.092)	(0.115)	(0.142)	(0.200)
Stocktwits Disag. (t) (z)	5.239***	7.111***	5.264***	7.292***	6.368***	7.717***
	(0.141)	(0.208)	(0.147)	(0.217)	(0.233)	(0.319)
Log # News (t) (z)	8.326***	10.788***	8.349***	10.925***	9.360***	10.681***
	(0.248)	(0.313)	(0.250)	(0.319)	(0.341)	(0.410)
Ave Persona Sentiment (t) (z)	-0.058	-0.173***	-0.064	-0.201***	-0.204	-0.172
	(0.050)	(0.067)	(0.052)	(0.070)	(0.129)	(0.163)
Abn. Log Volume (t-1) (%)	0.594***		0.593***		0.572***	
	(0.005)		(0.005)		(0.015)	
Abn Log Retail Volume (t-1) (%)		0.534***		0.525***		0.586***
		(0.008)		(0.008)		(0.027)
R-squared Adj.	0.534	0.426	0.529	0.415	0.585	0.544
Obs.	1,578,464	1,542,942	1,407,011	1,372,197	171,453	170,745

Table 7: Abnormal Trading and LLM Disagreement, Differences Across Demographic Dimensions. This table reports the results from regressing stock abnormal log trading volume (Panel A) and abnormal log retail trading volume (Panel B) on disagreement specific to each demographic characteristic. These within-dimension disagreement measures are constructed by first aggregating sentiment across the other dimensions (e.g., everything but 'education'), and then computing the weighted standard deviation for the focal dimension (e.g., 'education'). Abnormal disagreement is calculated as described in Table 6. Both abnormal trading volume and abnormal retail trading volume are multiplied by 100 for the regressions. Control variables included in the regressions are also the same as in Table 6. Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects. Standard errors clustered by stock and date and reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Abn Log Volume ×100						
Abn Education Disag (t) (z)	0.758***					
	(0.041)					
Abn Age Disag (t) (z)		0.417***				
		(0.037)				
Abn Income Disag (t) (z)			0.377***			
			(0.037)			
Abn Racial Disag (t) (z)				0.252***		
				(0.035)		
Abn Political Disag (t) (z)					0.077**	
					(0.036)	
Abn Gender Disag (t) (z)						0.055
						(0.035)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.534	0.534	0.534	0.534	0.534	0.534
Obs.	1,578,464	1,578,464	1,578,464	1,578,464	1,578,464	1,578,464
Panel B: Abn Log Retail Volume ×100						
Abn Education Disag (t) (z)	0.900***					
	(0.052)					
Abn Age Disag (t) (z)		0.523***				
		(0.049)				
Abn Income Disag (t) (z)			0.436***			
			(0.048)			
Abn Racial Disag (t) (z)				0.366***		
				(0.047)		
Abn Political Disag (t) (z)					0.087*	
					(0.046)	
Abn Gender Disag (t) (z)						0.075
						(0.048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.426	0.426	0.426	0.426	0.426	0.426
Obs.	1,542,942	1,542,942	1,542,942	1,542,942	1,542,942	1,542,942

Internet Appendix

"The Market's Mirror: Revealing Investor Disagreement with LLMs"

Figure A1: Prompt for Replication of Human Survey. Depicts the prompt used to survey LLM personas' perceived wrongfulness of corporate actions based on the human survey of Iliewa, Kempf, and Spalt (2024).

A: System Prompt

```
{PROFILE_TEXT}

Respond as this person would, reflecting their likely worldview and values.
```

B: User Prompt

```
XYZCorp is a multinational corporation headquartered in the United States.
Currently, it:

• Employs 40,000 workers domestically
• Pays an average tax rate of 21% on its profits
• Pays its CEO a total compensation of $16.7 million per year, which is 272
  times the pay of the average employee at the company
• Pays out 50% of its earnings to its shareholders in the form of share
  buybacks in a normal year (a share buyback is when a company uses
  corporate funds to buy back its own shares from its shareholders)
• Has debt outstanding that equals 20% of its total assets
• Has 13% of its energy usage coming from renewable energy sources
• Has a minorities share of 7% and a female share of 16% among top
  leadership positions
• Has recently acquired a competitor

Below, I provide a list of corporate actions that the management of XYZCorp
is considering for implementation in the near future. Please assume that, if
implemented, each of these corporate actions would have a positive financial
value to the firm's shareholders. This means that, for shareholders, all
current and future financial benefits of the action would outweigh all
current and future financial costs and risks. Assume that the financial value
to the firm's shareholders is certain and equal across all corporate actions
listed below. Also, assume that all actions by XYZCorp are legal. Here are
the actions labeled A to J:

• A: Implement a new cost cutting program at the recently acquired firm.
• B: Increase share buybacks (i.e., pay out a greater fraction of corporate
  funds to shareholders in the form of Share buybacks), thereby reducing
  corporate funds available for other purposes.
• C: Take out a loan in order to pay a dividend to its shareholders (thereby
  increasing the risk of bankruptcy and Reducing corporate tax payments).
• D: Lay off employees.
• E: Outsource parts of its operations in the United States to a foreign
  country with lower wages.
• F: Reduce the taxes it pays in the United States by having more of its
  profits taxed in low-tax countries.
• G: Increase the total compensation of the CEO.
• H: Discontinue existing personnel programs which increase the share of
  women and minorities in corporate Leadership roles.
• I: Increase the usage of fossil energy sources (e.g., oil, coal, and
  natural gas) in its operations.
• J: Appoint the current CEO also as the Chairman of the Board, giving the
  CEO more power inside the company.

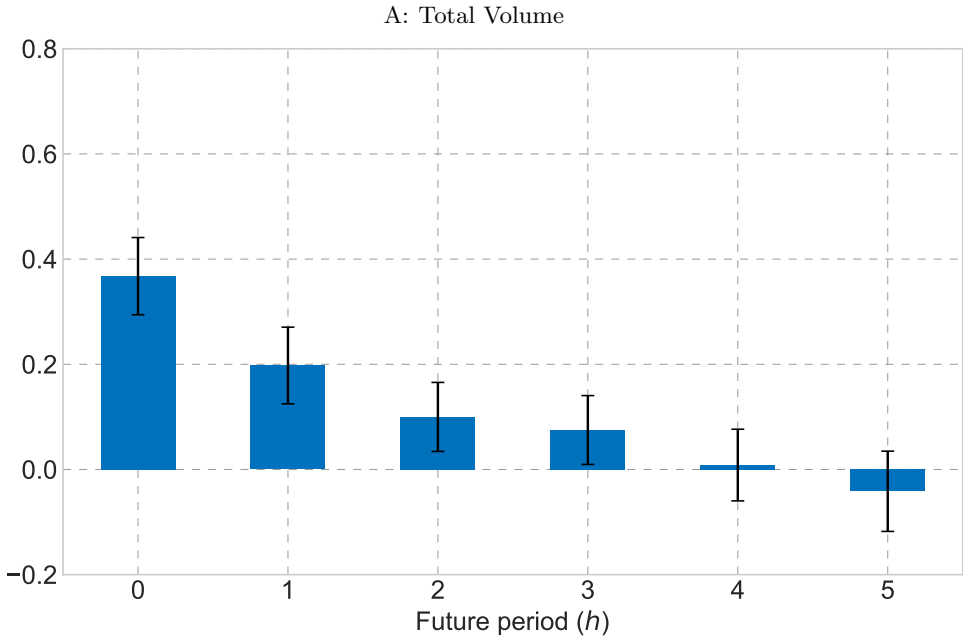
From the perspective of your persona, to what degree do you feel these
corporate actions are wrong? Please rank the actions by assigning numbers
from 1 (least wrong) to 10 (most wrong).

Assign 1 to the action you view as the least wrong, assign 10 to the action
you view as the most wrong, and use intermediary values to rank the remaining
actions. There is no correct or incorrect answer; we are interested in the
opinion of your persona.

Please respond ONLY with a list of the 10 ranked actions in increasing order
of wrongfulness. Use only the action label (A-J), excluding the full action
text, in your response as follows:

Rank: Label
```

Figure A2: Using Equal-weighted Disagreement in Trading Volume Regressions. This figure plots the coefficients for abnormal equal-weighted disagreement(Abn Equal-weighted Disag) and its 95% confidence intervals from regressions of Abn. Log Volume$_{t+h} \times 100$ (Panel A) and Abn. Log Retail Volume$_{t+h} \times 100$ (Panel B) on Abn Equal-weighted Disag$_t$ for h between 0 and 5 trading days. Equal-weighted Disag is obtained by assigning equal weights to the personas rather than using the FINRA sample weights when calculating the LLM persona disagreement. Abn Equal-weighted Disag is calculated as described under Table 6 and is standardized to zero mean and unit variance. All regressions include stock and date fixed effects, as well as the full set of controls listed in Table 6. Standard errors are clustered by stock and date.



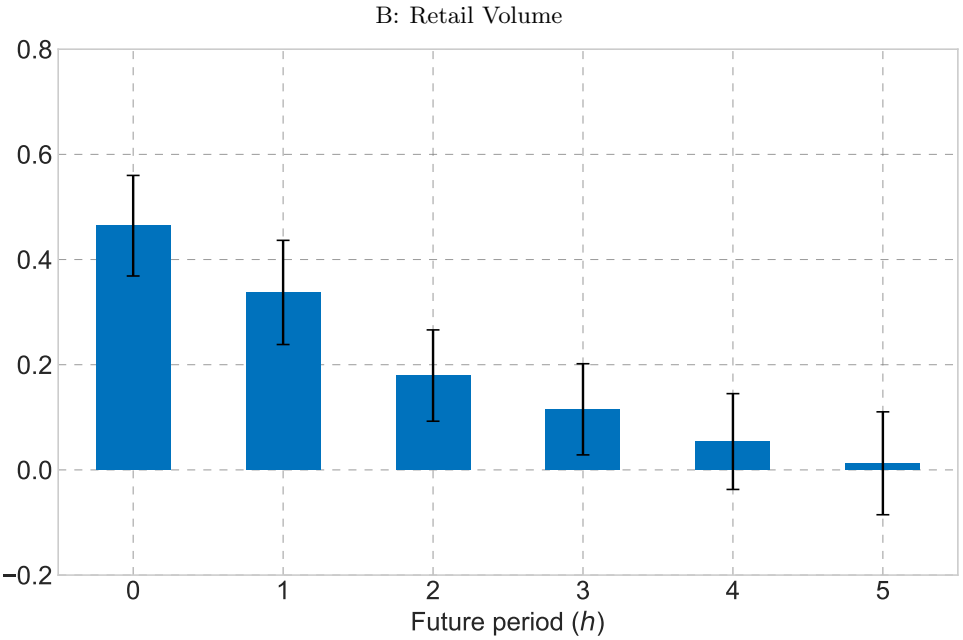


Table A1: Most Salient Bigrams by Demographics and News Type. This table shows the most salient bigrams used in the short sentiment rationales, computed separately for conservative and liberal profiles with high versus low income levels, indicated in the column labels. For each illustrative news type drawn from RavenPack, we select the top 10 news articles with the largest disagreement and then construct a document-term matrix based on bigrams and TF-IDF, using justifications elicited for all our profiles. Finally, we average the TF-IDF scores across personas' political leanings and income levels, and then select the top 10 bigrams. Panel A shows the top bigrams for news about executive compensation. Panel B shows the top bigrams for news about resource discovery. Panel C shows the top bigrams for news about workforce salary increase.

	Conservative		Liberal	
	Less than $50k	Over $100k	Less than $50k	Over $100k
Panel A: Executive Compensation				
1	increased costs	strong performance	excessive compensation	strong performance
2	salary increased	investment opportunity	corporate greed	compensation strong
3	concerned high	good investment	compensation outrageous	warranting investment
4	excessive compensation	compensation strong	worried high	investment opportunity
5	negatively impacting	strong commitment	increased costs	good investment
6	compensation package	salary strong	salary increased	performance warranting
7	decreased profitability	commitment growth	potential mismanagement	optimistic future
8	decreased profits	warranting investment	raises concerns	future growth
9	potential mismanagement	performance warranting	huge compensation	appointment substantial
10	worried high	strong financial	compensation package	substantial salary
Panel B: Resource Discovery				
1	growing offshore	growing offshore	environment impact	concerned environment
2	industry forecast	bodes well	concerned environment	liberal concerned
3	market industry	well offshore	impact offshore	growth potential
4	expected grow	drilling future	long term	socially conscious
5	market expected	future prospects	worried environment	environment impact
6	future offshore	market industry	term prospects	impact offshore
7	positive sign	industry forecast	drilling industry	industry growth
8	market increasing	projected bodes	drilling harm	market projected
9	increasing demand	market projected	liberal concerned	indicator offshore
10	services offshore	forecast bodes	environment concerns	drilling industry
Panel C: Workforce Salary Increase				
1	profit margins	operational costs	employee morale	commitment raising
2	costs hurt	increase operational	boost employee	employee morale
3	margins value	costs negatively	hour boost	morale productivity
4	hour increase	negatively impacting	morale productivity	long term
5	labor costs	profit margins	commitment raising	boost employee
6	hurt profit	costs hurt	positively impacting	socially responsible
7	operational costs	hour increase	impacting future	positively impacting
8	increase operational	labor costs	productivity benefiting	productivity positively
9	concerned raising	margins value	liberal values	demonstrates socially
10	negatively impacting	profitability value	aligns liberal	hour boost

Table A2: Keywords to Define Social, Environmental and Ethical Issues News Categories. This table shows the specific Ravenpack news groups categorized as hard news, and the news types categorized as reflecting Social, Environmental, and Legal & Ethical Issues in our analysis.

Panel A: Hard News Groups
acquisitions-mergers, analyst-ratings, assets, credit, credit-ratings, dividends, earnings, equity-actions, revenues, bankruptcy, partnerships, products-services
Panel B: Social Issues
discrimination, protest, workers-strike, union-pact, workforce-salary, executive-salary, layoffs, board-member-salary, executive-compensation, executive-incentives, insider-surrender, insider-gift, epidemic, suicide, hostage-situation, robbery, murder, vandalism, violence, shooting, bombing, suicide-bombing, terrorism, power-outage, transportation-disruption, airspace-closure, airspace-violation
Panel C: Environmental Issues
air-pollution, drilling, pipeline-accident, spill, mine-accident, refinery-accident, factory-accident, facility-accident, product-catastrophe, product-fault, product-recall, resource-discovery
Panel D: Ethical & Legal Issues
corruption, fraud, embezzlement, defamation, executive-scandal, insider-trading-lawsuit, regulatory-investigation, settlement, verdict, legal-issues, blackmail, tax-evasion, corporate-espionage, appeal, confidentiality-pact, cyber-attacks

Table A3: Seed Words for Constructing Rationale Theme Dictionaries. This table shows the seed words used to construct the respective 100-word theme dictionaries. To apply these seed words, we follow the steps for constructing dictionaries from a set of seed words and for scoring a text's loading on the dictionary outlined in Cookson, Dim, and Niessner (2024). To summarize, we first construct phrases (bigrams and bigrams connected with English connector words) from our rationales. Next, we fit a 300-dimensional *word2vec* embedding on the rationales, now composed of unigrams and the bigram phrases. For each theme, we feed the *word2vec* with the theme's seed words shown in the table and obtain the 100 words that are closest to the seed words in terms of their meaning and usage based on cosine similarity. These 100 words per theme comprise the theme dictionaries.

Employee Morale: morale, employee_morale, morale_and_productivity, employee_satisfaction
Fairness: priorities_and_fairness, greed_and_prioritization, prioritized_profits, people_needs
Environment: environmental_concerns
Profit Decline: lower_profit
Efficiency: efficiency
Uncertainty: uncertainty
Accountability: transparency_and_accountability, governance_and_transparency
Confidence: suggests_confidence, shows_confidence, loss_of_confidence, showing_faith, lost_faith

Table A4: RavenPack Descriptions for News Types in the Top and Bottom Average Disagreement. This table describes the RavenPack news types that comprise our top and bottom disagreement news types shown in Figure 4.

News Type	Description
insider-gift	News about an executive of the Company giving away corporate stock, usually as a charitable donation
insider-surrender	News about an executive surrendering shares back to the Company typically to pay taxes or cover the cost of an option exercise
sell-registration	News about an Executive of the Company registering the sell corporate stock
insider-buy	News about an executive of the Company buying corporate stock
executive-salary	News about wages or compensation paid to an executive
executive-incentives	News about the Company granting shares or options to an executive
workforce-salary	News about workforce salary
air-pollution	News about air pollution
buy-moo	News about exchange reporting an excess of Market On Open (MOO) buy orders for the Company's stock that cannot be matched to orders of the opposite type
buy-moc	News about exchange reporting an excess of Market On Close (MOC) buy orders for the Company's stock that cannot be matched to orders of the opposite type
no-moc	News about a market on close order imbalance for the Company's stock being resolved
sell-moc	News about an exchange reporting an excess of Market On Close (MOC) sell orders for the Company's stock that cannot be matched to orders of the opposite type
sell-moo	News about an exchange reporting an excess of Market On Open (MOO) sell orders for the Company's stock that cannot be matched to orders of the opposite type
technical-price-level	News about technical analysis indicating that the Entity's price has approached a resistance level
exchange-compliance	News about the Company's listing compliance plan or actions
exchange-noncompliance	News about exchange non-compliance
initial-public-offering	News about Initial Public Offering (IPO)
initial-public-offering-unit	News about the Company announcing the Initial Public Offering (IPO) of one of its units
shareholder-rights-plan	News about shareholder rights plan
shareholders-meeting	News about general meeting of the shareholders
unit-acquisition-bid	News about a bid to acquire the unit or division of the Company
unit-acquisition-regulation	News about regulatory action on the acquisition of the company's unit/division or the company's acquisition of another's unit/division
accelerated-approval-designation	News about the accelerated approval designation of the Company
analyst-ratings-set	News about analyst rating of the Company's stock
auditor-appointment	News about the Company's auditor appointment
conference-call	News about the Company conference call to report financial results or other corporate information
conference	News about a conference or business event to promote the Company's product or service
construction-pmi	News about the Construction Purchasing Managers Index (PMI)
employment-guidance	News about employment guidance figures or projections
currency-guidance	News about an entity's estimate or forecast on the value of the Currency
ebita	News about the company Earnings Before Interest, Taxes, and Amortization (EBITA)
product-promotion	News about the Company promoting a product or service through marketing activities including discount sales
regulatory-product-review	News about a regulatory body reviewing the Company's product or service
breakthrough-therapy-designation	News about breakthrough therapy designation status of the Company's product
priority-review-designation	News about priority review designation status of the Company's product
board-member-retirement	News about a board member of the company retiring
board-member-death	News about a board member dying
executive-death	News about an executive of the Company dying
freight-transport-accident	News about the Entity experiencing an accident in a freight transport service
mine-accident	News about the Entity experiencing an accident in an establishment for mining a commodity

Table A5: Demographic Differences in Sentiment in LLM Training & Post-training Windows. The table presents results for the regressions of sentiment declared by LLM personas on demographic characteristics in two subsamples: the LLM training window, 2010-2023 (Panel A) and the Post-LLM training window, 2024-2025 (Panel B). In each subsample, we regress LLM personas' sentiment declarations for a news category on persona demographic indicator variables. The column headers indicate the news categories comprising the analysis sample for that specification. The dependent variable (sentiment) is divided by its standard deviation in each specification. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	(1) Executive Pay Increase	(2) Workforce Pay Increase	(3) Share Buyback	(4) Fossil Fuel Discovery	(5) Employee Layoff
			Panel A: LLM training window		
Income Over $100K	1.042***	0.296***	0.165***	0.240***	0.416***
	(0.041)	(0.037)	(0.006)	(0.053)	(0.015)
Income $50K - $100K	0.620***	0.117***	0.074***	0.160***	0.211***
	(0.028)	(0.025)	(0.004)	(0.039)	(0.009)
Independent	-0.112***	0.480***	-0.099***	-0.101***	0.005
	(0.022)	(0.093)	(0.006)	(0.025)	(0.006)
Liberal	-0.277***	1.052***	-0.000	-0.112***	0.043***
	(0.030)	(0.149)	(0.003)	(0.039)	(0.009)
White non-Hispanic	0.151***	-0.050***	0.058***	0.058***	0.024***
	(0.008)	(0.016)	(0.002)	(0.016)	(0.003)
Male	0.065***	0.044***	0.077***	0.077***	0.031***
	(0.008)	(0.013)	(0.004)	(0.022)	(0.002)
Age 35-54	-0.091***	-0.119***	-0.065***	-0.045***	-0.048***
	(0.006)	(0.010)	(0.002)	(0.010)	(0.002)
Age 55+	-0.127***	-0.184***	-0.114***	-0.087***	-0.062***
	(0.009)	(0.016)	(0.005)	(0.022)	(0.004)
Some college or less	-0.190***	-0.083***	-0.037***	-0.039***	-0.133***
	(0.009)	(0.010)	(0.002)	(0.014)	(0.005)
R-squared Adj.	0.576	0.534	0.501	0.764	0.686
Obs.	200,448	68,688	3,033,504	59,616	1,035,720
			Panel B: Post-training window		
Income Over $100K	0.975***	0.450***	0.213***	0.432**	0.663***
	(0.143)	(0.137)	(0.041)	(0.182)	(0.055)
Income $50K - $100K	0.538***	0.254**	0.073***	0.132*	0.364***
	(0.110)	(0.108)	(0.019)	(0.075)	(0.040)
Independent	0.064	0.408***	-0.207***	-0.301*	0.049***
	(0.046)	(0.144)	(0.058)	(0.157)	(0.016)
Liberal	-0.049	0.677***	-0.008	-0.150	0.113***
	(0.089)	(0.164)	(0.009)	(0.139)	(0.021)
White non-Hispanic	0.135***	-0.022	0.059***	0.069	0.019*
	(0.037)	(0.028)	(0.007)	(0.055)	(0.010)
Male	0.067***	0.038**	0.130***	0.219**	0.016*
	(0.018)	(0.019)	(0.035)	(0.089)	(0.008)
Age 35-54	-0.125***	-0.124***	-0.094***	-0.232***	-0.063***
	(0.034)	(0.028)	(0.016)	(0.086)	(0.008)
Age 55+	-0.185***	-0.182***	-0.217***	-0.388***	-0.075***
	(0.049)	(0.041)	(0.053)	(0.150)	(0.012)
Some college or less	-0.199***	-0.159***	-0.048***	-0.073	-0.235***
	(0.042)	(0.038)	(0.012)	(0.058)	(0.020)
R-squared Adj.	0.606	0.532	0.648	0.416	0.656
Obs.	6,048	4,968	281,880	1,944	96,120

Table A6: Text-Based Rationale Differences Across Demographics in LLM Training & Post-training Windows. The table presents regressions of text-based rationale topics (derived from a word2vec procedure) on persona demographic characteristics. To focus the topic analysis, the sample is restricted to news types analyzed in the Table 2 and in Table A7. The analysis is conducted in two subsamples: the LLM training window, 2010-2023 (Panel A) and the Post-LLM training window, 2024-2025 (Panel B). Table 3 provides details of the construction of the rationale theme loadings and the set of news categories used for the analysis. Each dependent variable is divided by its standard deviation. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	(1) Employee Morale	(2) Faireness	(3) Environment	(4) Profit Decline	(5) Efficiency
Panel A: LLM training window					
Income Over $100K	0.043***	-0.018***	-0.049***	-0.008***	0.030***
	(0.005)	(0.003)	(0.004)	(0.002)	(0.002)
Income $50K - $100K	0.011***	-0.012***	-0.046***	-0.003**	0.017***
	(0.002)	(0.002)	(0.004)	(0.001)	(0.002)
Independent	-0.005**	0.002**	0.002***	0.004***	-0.026***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)
Liberal	0.053***	0.015***	0.074***	-0.004***	-0.013***
	(0.004)	(0.002)	(0.008)	(0.001)	(0.002)
White non-Hispanic	-0.012***	-0.003***	-0.031***	0.003***	-0.004***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Male	0.010***	-0.002***	-0.012***	0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 35-54	-0.023***	-0.001	0.008***	0.002***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.000)	(0.001)
Age 55+	-0.036***	-0.001	0.014***	-0.001	-0.010***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Some college or less	-0.005**	-0.006***	-0.018***	-0.000	-0.031***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R-squared Adj.	0.034	0.044	0.010	0.034	0.022
Obs.	25,551,935	25,551,935	25,551,935	25,551,935	25,551,935
Panel B: Post-training window					
Income Over $100K	0.011***	-0.020***	-0.053***	-0.003	0.054***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.006)
Income $50K - $100K	-0.007***	-0.016***	-0.052***	-0.000	0.027***
	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)
Independent	-0.003	0.000	0.001	0.011**	-0.074***
	(0.004)	(0.002)	(0.001)	(0.005)	(0.009)
Liberal	0.040***	0.016***	0.072***	0.000	-0.038***
	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
White non-Hispanic	-0.021***	-0.007***	-0.033***	0.005*	0.005*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Male	-0.004	-0.006*	-0.014***	0.003	-0.021***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)
Age 35-54	-0.014***	-0.001	0.009***	0.000	-0.013***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Age 55+	-0.018***	0.000	0.020***	-0.004	-0.009**
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Some college or less	0.002	-0.001	-0.014***	-0.018***	-0.059***
	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)
R-squared Adj.	0.040	0.008	0.009	0.084	0.023
Obs.	2,527,632	2,527,632	2,527,632	2,527,632	2,527,632

Table A7: Illustrating Sentiment Differences across Demographics for Additional Event Categories. The table presents the extension of Tables 2 and A5 using additional news categories. Columns (1) - (3) use the full sample (2010-2025) as in Tables 2. Columns (4) - (6) use the LLM training window (2010-2023) as in Tables A5 Panel A. Columns (7) - (9) use the post-LLM training window (2024-2025) as in Tables A5 Panel B. For each subsample and news category (indicated in the column headers), we regress the LLM personas' sentiment declarations for articles in that category on dummy variables for the personas' demographics. The dependent variable (sentiment) is divided by its standard deviation in each specification. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	Full sample			Training window			Post Training		
	(1) Insider Buy	(2) Insider Gift	(3) Executive Incentives	(4) Insider Buy	(5) Insider Gift	(6) Executive Incentives	(7) Insider Buy	(8) Insider Gift	(9) Executive Incentives
Income Over $100K	0.820***	1.185***	0.606***	0.817***	1.190***	0.649***	0.852***	1.170***	0.423***
	(0.016)	(0.016)	(0.070)	(0.015)	(0.016)	(0.077)	(0.031)	(0.035)	(0.139)
Income $50K - $100K	0.460***	0.619***	0.272***	0.469***	0.621***	0.293***	0.378***	0.638***	0.168***
	(0.009)	(0.010)	(0.047)	(0.009)	(0.010)	(0.052)	(0.014)	(0.020)	(0.049)
Independent	-0.379***	-0.120***	-0.115***	-0.358***	-0.131***	-0.104***	-0.580***	0.078***	-0.332**
	(0.008)	(0.009)	(0.031)	(0.008)	(0.009)	(0.031)	(0.021)	(0.016)	(0.149)
Liberal	-0.196***	0.170***	-0.085**	-0.194***	0.162***	-0.086**	-0.214***	0.325***	-0.124
	(0.009)	(0.012)	(0.036)	(0.009)	(0.012)	(0.039)	(0.016)	(0.021)	(0.081)
White non-Hispanic	0.320***	0.150***	0.115***	0.323***	0.153***	0.117***	0.293***	0.109***	0.158***
	(0.006)	(0.004)	(0.021)	(0.006)	(0.004)	(0.023)	(0.010)	(0.008)	(0.055)
Male	0.307***	0.216***	0.088***	0.303***	0.220***	0.082***	0.348***	0.146***	0.225***
	(0.005)	(0.004)	(0.016)	(0.005)	(0.004)	(0.015)	(0.011)	(0.009)	(0.085)
Age 35-54	-0.167***	-0.261***	-0.086***	-0.163***	-0.260***	-0.090***	-0.216***	-0.285***	-0.087
	(0.003)	(0.003)	(0.012)	(0.003)	(0.003)	(0.013)	(0.008)	(0.010)	(0.054)
Age 55+	-0.344***	-0.348***	-0.163***	-0.332***	-0.349***	-0.168***	-0.462***	-0.366***	-0.198***
	(0.006)	(0.005)	(0.018)	(0.006)	(0.005)	(0.020)	(0.015)	(0.015)	(0.077)
Some college or less	-0.163***	-0.256***	-0.101***	-0.167***	-0.256***	-0.104***	-0.130***	-0.277***	-0.122***
	(0.004)	(0.005)	(0.016)	(0.004)	(0.005)	(0.018)	(0.006)	(0.009)	(0.040)
R-squared Adj.	0.331	0.480	0.730	0.327	0.483	0.710	0.420	0.506	0.686
Obs.	19,988,424	3,275,424	26,784	18,071,856	3,057,696	24,408	1,916,568	217,728	2,376

Table A8: Text-Based Rationales across Demographics – Additional Themes. The table presents the extension of Tables 3 and A6 using additional rationale themes. Columns (1) - (3) use the full sample (2010-2025) as in Tables 3. Columns (4) - (6) use the LLM training window (2010-2023) as in Tables A6 Panel A. Columns (7) - (9) use the post-LLM training window (2024-2025) as in Tables A6 Panel B. For each subsample, we regress the rationales' theme loadings (indicated in the column headers) on dummy variables for the personas' demographics. Tables 2 and Tables 3 provide details of the construction of the rationale theme loadings and the set of news categories used for the analysis. Each dependent variable is divided by its standard deviation. All specifications include stock and date fixed effects. Standard errors shown in parentheses are clustered by stock and date.

	Full sample			Training window			Post Training		
	(1) Uncertainty	(2) Confidence	(3) Accountability	(4) Uncertainty	(5) Confidence	(6) Accountability	(7) Uncertainty	(8) Confidence	(9) Accountability
Income Over $100K	-0.037***	-0.141***	0.010***	-0.039***	-0.142***	0.010***	-0.018*	-0.128***	0.007*
	(0.003)	(0.005)	(0.002)	(0.003)	(0.006)	(0.002)	(0.010)	(0.010)	(0.004)
Income $50K - $100K	-0.028***	-0.013***	0.003**	-0.030***	-0.013***	0.003**	-0.015*	-0.009**	-0.003
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.008)	(0.005)	(0.003)
Independent	-0.014***	-0.086***	-0.005***	-0.011***	-0.082***	-0.005***	-0.037***	-0.136***	-0.008***
	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.010)	(0.010)	(0.002)
Liberal	-0.001	-0.040***	0.017***	0.001	-0.037***	0.017***	-0.011	-0.067***	0.016***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.008)	(0.004)	(0.004)
White non-Hispanic	-0.037***	0.029***	-0.006***	-0.038***	0.027***	-0.006***	-0.037***	0.055***	-0.002
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.004)	(0.004)	(0.002)
Male	-0.014***	0.049***	-0.002*	-0.014***	0.047***	-0.002*	-0.009***	0.076***	0.002
	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.001)	(0.003)	(0.007)	(0.002)
Age 35-54	0.014***	0.028***	0.000	0.014***	0.028***	0.001	0.011***	0.027***	-0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)	(0.002)
Age 55+	0.021***	0.051***	0.003**	0.021***	0.053***	0.003**	0.021***	0.035***	-0.001
	(0.002)	(0.003)	(0.001)	(0.001)	(0.003)	(0.001)	(0.006)	(0.007)	(0.003)
Some college or less	-0.029***	0.273***	-0.021***	-0.032***	0.274***	-0.021***	-0.004	0.268***	-0.013***
	(0.002)	(0.006)	(0.002)	(0.002)	(0.007)	(0.002)	(0.007)	(0.010)	(0.003)
R-squared Adj.	0.025	0.061	0.014	0.025	0.062	0.015	0.040	0.080	0.010
Obs.	28,079,567	28,079,567	28,079,567	25,551,935	25,551,935	25,551,935	2,527,632	2,527,632	2,527,632

Table A9: Abnormal Trading Volume and Disagreement Around End of LLM Training Window. This table reports the results for the contemporaneous regressions of stock abnormal trading volume (Panel A) and abnormal retail trading volume (Panel B) on the abnormal LLM persona disagreement over bi-annual (Columns (1) - (2)) and annual (Columns (3) - (6)) subsamples indicated in the column headers. Abn Disag is calculated as described under Table 6. The intention is to illustrate changes in the loading before versus after the end of Llama 3.1's training window in December 2023. Both abnormal trading volume and abnormal retail trading are multiplied by 100 for the regressions. Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects and all the control variables listed in Table 6. Standard errors clustered by stock and date are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
Sample period:	[2022, 2023]	[2024, 2025]	2022	2023	2024	2025
			Panel A: Abn Log Volume (%)			
Abn Disag (t) (z)	1.457***	1.475***	1.410***	1.536***	1.515***	1.474***
	(0.073)	(0.089)	(0.094)	(0.103)	(0.102)	(0.159)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.572	0.585	0.608	0.543	0.572	0.577
Obs.	260,547	171,453	131,068	129,479	127,850	43,603
			Panel B: Abn Log Retail Volume (%)			
Abn Disag (t) (z)	1.575***	1.772***	1.500***	1.711***	1.926***	1.493***
	(0.085)	(0.111)	(0.115)	(0.117)	(0.123)	(0.181)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.476	0.544	0.478	0.486	0.534	0.547
Obs.	259,409	170,745	130,430	128,979	127,346	43,399

Table A10: Equal-weighted Disagreement and Trading Volume. This table reports the results from regressing stock abnormal trading volume (odd columns) and abnormal log retail trading volume (even columns) on *equal weighted* abnormal LLM disagreement. The specification is the same as in Table 6, but using equal weights across personas instead of the FINRA weights to calculate disagreement. Both abnormal trading volume and abnormal retail trading volume are multiplied by 100 for the regressions. We conduct the analysis in the full sample 2010-2025 (columns denoted "Full sample"), LLM training window, ending 2023 (columns labeled "Training window"), and post-LLM training window 2024-2025 (columns labeled "Post-training"). Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects. Standard errors clustered by stock and date are reported in parentheses.

	Full sample		Training window		Post training	
	Total Volume	Retail Volume	Total Volume	Retail Volume	Total Volume	Retail Volume
	(1)	(2)	(3)	(4)	(5)	(6)
Abn Equal-weighted Disag (t) (z)	0.368***	0.464***	0.270***	0.337***	1.343***	1.642***
	(0.037)	(0.049)	(0.038)	(0.050)	(0.090)	(0.114)
Volatility (t-5, t-1) (z)	-0.591***	1.034***	-0.555***	1.302***	-0.226	-0.206
	(0.154)	(0.226)	(0.158)	(0.241)	(0.459)	(0.697)
Volatility (t-25, t-6) (z)	1.394***	2.039***	1.470***	2.223***	1.520***	1.705**
	(0.165)	(0.235)	(0.195)	(0.292)	(0.569)	(0.691)
Abn Ret (t-5, t-1) (z)	-0.564***	-1.036***	-0.573***	-1.071***	-0.558***	-0.788***
	(0.088)	(0.111)	(0.089)	(0.115)	(0.197)	(0.234)
Abn Ret (t-25, t-6) (z)	-1.213***	-2.090***	-1.198***	-2.081***	-1.436***	-2.053***
	(0.097)	(0.141)	(0.097)	(0.147)	(0.262)	(0.356)
Ave Media Sentiment (t) (z)	-0.046	-0.050	-0.067	-0.085	0.143	0.211
	(0.053)	(0.066)	(0.055)	(0.069)	(0.139)	(0.162)
Std Media Sentiment (t) (z)	0.658***	0.822***	0.785***	0.946***	-0.813***	-0.679***
	(0.092)	(0.114)	(0.092)	(0.115)	(0.142)	(0.200)
Stocktwits Disag. (t) (z)	5.239***	7.111***	5.264***	7.292***	6.373***	7.722***
	(0.141)	(0.208)	(0.147)	(0.217)	(0.233)	(0.319)
Log # News (t) (z)	8.325***	10.787***	8.348***	10.925***	9.329***	10.647***
	(0.248)	(0.313)	(0.250)	(0.319)	(0.341)	(0.409)
Ave Persona Sentiment (t) (z)	-0.074	-0.192***	-0.076	-0.215***	-0.224*	-0.197
	(0.050)	(0.067)	(0.052)	(0.070)	(0.129)	(0.164)
Abn. Log Volume (t-1) (%)	0.594***		0.593***		0.572***	
	(0.005)		(0.005)		(0.015)	
Abn Log Retail Volume (t-1) (%)		0.534***		0.525***		0.586***
		(0.008)		(0.008)		(0.027)
R-squared Adj.	0.534	0.426	0.529	0.415	0.585	0.543
Obs.	1,578,464	1,542,942	1,407,011	1,372,197	171,453	170,745

Table A11: Excluding Politics and Income: LLM Disagreement and Trading Volume. This table reports the results from regressing stock abnormal trading volume and abnormal log retail trading volume on abnormal LLM disagreement excluding the politics and / or the income dimension (by averaging out these dimensions before computing the disagreement measure). The specification is the same as in Table 6, but using equal weights across personas instead of the FINRA weights to calculate disagreement. Both abnormal trading volume and abnormal retail trading volume are multiplied by 100 for the regressions. We conduct the analysis in the full sample 2010-2025 (columns denoted "Full sample"), LLM training window, ending 2023 (columns labeled "Training window"), and post-LLM training window 2024-2025 (columns labeled "Post-training") Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects. Standard errors clustered by stock and date are reported in parentheses.

	Full sample				Training window				Post training			
	Total Volume		Retail Volume		Total Volume		Retail Volume		Total Volume		Retail Volume	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Abn Ex-Politics Disag (t) (z)	0.514***		0.623***		0.416***		0.496***		1.476***		1.774***	
	(0.038)		(0.049)		(0.039)		(0.051)		(0.089)		(0.115)	
Abn Ex-PoliInc Disag (t) (z)		0.593***		0.735***		0.465***		0.576***		1.790***		2.078***
		(0.037)		(0.049)		(0.038)		(0.051)		(0.088)		(0.111)
Volatility (t-5, t-1) (z)	-0.591***	-0.590***	1.034***	1.035***	-0.555***	-0.554***	1.302***	1.304***	-0.222	-0.219	-0.201	-0.199
	(0.154)	(0.154)	(0.226)	(0.226)	(0.158)	(0.158)	(0.241)	(0.241)	(0.460)	(0.458)	(0.697)	(0.696)
Volatility (t-25, t-6) (z)	1.396***	1.394***	2.040***	2.038***	1.471***	1.470***	2.224***	2.222***	1.522***	1.509***	1.708**	1.691**
	(0.165)	(0.165)	(0.236)	(0.236)	(0.195)	(0.195)	(0.292)	(0.292)	(0.569)	(0.568)	(0.691)	(0.690)
Abn Ret (t-5, t-1) (z)	-0.566***	-0.566***	-1.037***	-1.038***	-0.574***	-0.576***	-1.072***	-1.074***	-0.569***	-0.520***	-0.801***	-0.743***
	(0.088)	(0.088)	(0.111)	(0.111)	(0.089)	(0.089)	(0.115)	(0.115)	(0.197)	(0.196)	(0.235)	(0.234)
Abn Ret (t-25, t-6) (z)	-1.215***	-1.208***	-2.092***	-2.083***	-1.199***	-1.195***	-2.082***	-2.077***	-1.451***	-1.392***	-2.070***	-2.001***
	(0.097)	(0.097)	(0.141)	(0.141)	(0.097)	(0.097)	(0.147)	(0.147)	(0.262)	(0.261)	(0.356)	(0.355)
Ave Media Sentiment (t) (z)	-0.035	-0.049	-0.038	-0.055	-0.057	-0.067	-0.073	-0.085	0.157	0.133	0.226	0.190
	(0.053)	(0.054)	(0.066)	(0.066)	(0.055)	(0.055)	(0.069)	(0.069)	(0.139)	(0.139)	(0.162)	(0.162)
Std Media Sentiment (t) (z)	0.659***	0.656***	0.824***	0.820***	0.785***	0.783***	0.947***	0.944***	-0.791***	-0.812***	-0.652***	-0.678***
	(0.092)	(0.092)	(0.114)	(0.114)	(0.092)	(0.092)	(0.115)	(0.115)	(0.142)	(0.142)	(0.200)	(0.201)
Stocktwits Disag. (t) (z)	5.239***	5.236***	7.110***	7.107***	5.264***	5.261***	7.292***	7.289***	6.368***	6.353***	7.716***	7.701***
	(0.141)	(0.141)	(0.208)	(0.207)	(0.147)	(0.147)	(0.217)	(0.217)	(0.232)	(0.232)	(0.319)	(0.319)
Log # News (t) (z)	8.335***	8.333***	10.798***	10.797***	8.357***	8.352***	10.934***	10.930***	9.370***	9.445***	10.693***	10.771***
	(0.248)	(0.248)	(0.313)	(0.313)	(0.250)	(0.250)	(0.320)	(0.319)	(0.341)	(0.344)	(0.410)	(0.413)
Ave Persona Sentiment (t) (z)	-0.034	-0.106**	-0.144**	-0.233***	-0.043	-0.097*	-0.177**	-0.242***	-0.074	-0.561***	-0.016	-0.584***
	(0.050)	(0.050)	(0.067)	(0.067)	(0.052)	(0.051)	(0.070)	(0.070)	(0.131)	(0.130)	(0.164)	(0.167)
Abn. Log Volume (t-1) (%)	0.594***	0.594***			0.593***	0.593***			0.572***	0.571***		
	(0.005)	(0.005)			(0.005)	(0.005)			(0.015)	(0.015)		
Abn Log Retail Volume (t-1) (%)			0.534***	0.534***			0.525***	0.525***			0.586***	0.585***
			(0.008)	(0.008)			(0.008)	(0.008)			(0.027)	(0.027)
R-squared Adj.	0.534	0.534	0.426	0.426	0.530	0.530	0.415	0.415	0.585	0.585	0.543	0.544
Obs.	1,578,464	1,578,464	1,542,942	1,542,942	1,407,011	1,407,011	1,372,197	1,372,197	171,453	171,453	170,745	170,745

Table A12: Day-Ahead Abnormal Trading and LLM Disagreement, Across Demographic Dimensions. This table reports the results from regressing day $t+1$ stock abnormal log trading volume (Panel A) and day $t+1$ abnormal log retail trading volume (Panel B) on disagreement specific to each demographic characteristic. These within-dimension disagreement measures are constructed by first aggregating sentiment across the other dimensions (e.g., everything but 'education'), and then computing the weighted standard deviation for the focal dimension (e.g., 'education'). Abnormal disagreement is calculated as described in Table 6. Both abnormal trading volume and abnormal retail trading volume are multiplied by 100 for the regressions. Control variables included in the regressions are the same as in Table 6. Variable names that include "(z)" mean that they are standardized to zero mean and unit variance to facilitate comparability. The regressions include stock and date fixed effects. Standard errors clustered by stock and date and reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Abn Log Volume $(t+1) \times 100$						
Abn Education Disag (t) (z)	0.289***					
	(0.040)					
Abn Age Disag (t) (z)		0.351***				
		(0.036)				
Abn Income Disag (t) (z)			0.171***			
			(0.036)			
Abn Racial Disag (t) (z)				0.257***		
				(0.036)		
Abn Political Disag (t) (z)					0.060	
					(0.037)	
Abn Gender Disag (t) (z)						0.107***
						(0.036)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.426	0.426	0.426	0.426	0.426	0.426
Obs.	1,577,937	1,577,937	1,577,937	1,577,937	1,577,937	1,577,937
Panel B: Abn Log Retail Volume $(t+1) \times 100$						
Abn Education Disag (t) (z)	0.447***					
	(0.054)					
Abn Age Disag (t) (z)		0.515***				
		(0.048)				
Abn Income Disag (t) (z)			0.285***			
			(0.049)			
Abn Racial Disag (t) (z)				0.414***		
				(0.047)		
Abn Political Disag (t) (z)					0.149***	
					(0.050)	
Abn Gender Disag (t) (z)						0.184***
						(0.048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.329	0.329	0.329	0.329	0.329	0.329
Obs.	1,542,426	1,542,426	1,542,426	1,542,426	1,542,426	1,542,426