When LLMs Go Abroad: Foreign Bias in AI Financial Predictions

Sean Cao
Robert H. Smith School of Business
University of Maryland

Charles C.Y. Wang
Harvard Business School and ECGI

Xiang Yi*
Hong Kong Polytechnic University

May 2026

Abstract

We document "foreign bias" in AI-generated financial analysis, reversing the classic home bias. U.S.-based ChatGPT is systematically more optimistic than China-based DeepSeek about Chinese firms in price predictions, earnings forecasts, and qualitative business descriptions; its quantitative forecasts are significantly less accurate. Evidence supports an *effective* information-availability mechanism: the bias tracks cross-border news-coverage gaps, attenuates for cross-listed firms, disappears when we inject Chinese news at inference, and is absent when both models analyze U.S. firms. Placebo and robustness tests further rule out function-centric alternatives—alignment and elicitation—as drivers. The bias is not inherently optimistic: its sign tracks whichever direction of news U.S. sources underreport (negative or positive), and thus cannot be signed in advance. Suggestive evidence indicates that U.S.-analyst optimism toward Chinese firms shifted with these coverage gaps around ChatGPT's release. LLMs trained in different information environments can create divergent, hard-to-sign signals, with implications for investors and policymakers as AI increasingly intermediates global markets.

Keywords: Artificial intelligence; Financial statement analysis; Large language models; ChatGPT; DeepSeek

JEL: D83, L86, G14, G15, G41, M41, O33

*Cao (scao824@umd.edu) is an associate professor and the director and co-founder of the AI Initiative for Capital Market Research at the Robert H. Smith School of Business. Wang (charles.cy.wang@hbs.edu) is the Tandon Family Professor of Business Administration at Harvard Business School and a Research Member of the European Corporate Governance Institute (ECGI). Yi (yi.xiang@polyu.edu.hk) is an assistant professor of accounting at Hong Kong Polytechnic University. Sean Cao acknowledges support from the AI Initiative for Capital Markets Research at the University of Maryland. We gratefully acknowledge helpful comments and suggestions from Bradford Levy, David Trainer, Eugene Soltes, seminar participants at the University of Alabama, University of Toronto, University of Alberta, National Taiwan University, Carlos III University of Madrid, University of Kentucky, University of Maryland College Park, Harvard Business School, McMaster University, Lancaster University, McGill University, Columbia University, University of Chicago, University of Rochester, Nanyang Technological University, and Norwegian school of Economics where earlier versions of this paper were presented. The authors used Claude.ai to assist with wording and phrasing. The intellectual content, analysis, and conclusions are solely those of the authors, who remain responsible for any errors. Comments are welcome.

1 Introduction

The rapid adoption of Large Language Models (LLMs) in financial analysis represents a fundamental shift in how investment information is produced and consumed (Engelberg, Manela, Mullins, and Vulicevic, 2025; He, Lv, Manela, and Wu, 2025). ChatGPT reached 100 million users within two months of its November 2022 launch, with surveys indicating that over 40% of institutional investors now use AI tools for market analysis (Bloomberg Intelligence, 2024); other market participants also appear to be integrating the tool into their workflows (Blankespoor, DeHaan, and Li, 2025; Cheng, Lin, and Zhao, 2025; Law and Shen, 2025). A nascent literature documents LLMs' ability to process or synthesize financial statements (Bertomeu, Lin, Liu, and Ni, 2025; Bradshaw, Ma, Yost, and Zou, 2025; de Kok, 2025; Shaffer and Wang, 2024), and to forecast earnings (Cook, Kazinnik, Hansen, and McAdam, 2023; Kim, Muhn, and Nikolaev, 2024; Kim and Nikolaev, 2025), corporate policies (Jha, Qian, Weber, and Yang, 2024), and stock returns (Lopez-Lira and Tang, 2023; Chen, Green, Gulen, and Zhou, 2024).

However, a growing body of work in finance reveals potential perils in applying LLMs to financial analysis. One stream focuses on *function-centric* limitations, documenting extrapolation errors, miscalibration (Chen et al., 2024), and behavioral anomalies that can distort predictions (Bini, Cong, Huang, and Lawrence, 2025). A second stream studies *information-set* distortions, most notably temporal distortions that lead to look-ahead bias and leakage (Levy, 2025; Sarkar and Vafa, 2024; He et al., 2025). We extend this second stream by studying a novel information-set distortion: *spatial* (geographic) gaps in training corpora arising from cross-country differences in news coverage. As investors increasingly rely on LLMs trained in different countries—and potentially shaped by very different information environments—understanding whether and how these spatial gaps affect financial predictions becomes essential.

The parallel development of frontier models in the US and China provides a unique opportunity to address this question. Cross-border model comparisons offer three key advantages. First, when two frontier models trained in different environments analyze identical firms, systematic prediction differences provide insights into potential sources of bias. Second, this design controls for firm characteristics and allows for attribution of prediction differences to model-level factors, such as how models process information or how their predictions are affected by asymmetries in their underlying training data, rather than to cross-sectional variation in firm fundamentals. Third, as the United States and China emerge as the two leading AI powers, understanding how their leading models differ has direct implications for both practitioners and academics. These differences can shape cross-border investment analysis, market efficiency, and the global diffusion of AI.

Our investigation draws parallels between human and algorithmic decision-making under information asymmetries. While humans exhibit home bias—overweighting domestic or local assets despite diversification benefits (French and Poterba, 1991; Cooper and Kaplanis, 1994; Tesar and Werner, 1995; Lewis, 1999; Coval and Moskowitz, 1999)—the mechanism remains debated: information advantages (Portes and Rey, 2005; Van Nieuwerburgh and Veldkamp, 2009) versus behavioral biases that lead investors to favor familiar assets (Heath and Tversky, 1991; Huberman, 2001). Whether LLMs exhibit similar patterns is an open and compelling question because they learn from human-generated text that could embody human biases (Blodgett, Barocas, Daumé, and Wallach, 2020; Bender, Gebru, McMillan-Major, and Shmitchell, 2021; Shumailov, Shumaylov, Zhao, Gal, Papernot, and Anderson, 2023), or could produce new biases due to how information is incorporated.

We prompt both ChatGPT 4.1 and DeepSeek R1, frontier models from the US and China, respectively, to analyze nearly 5,000 firms listed on China's Shanghai and Shenzhen exchanges as of June 30, 2024. For each firm, we elicit three outputs designed to capture the broad spectrum of quantitative and qualitative financial analyses that investors and analysts might conduct using LLMs: a six-month-ahead stock price prediction, a six-month-ahead earnings-per-share (EPS) prediction, and a short qualitative business description, whose tone we score using the Loughran and Mcdonald (2011) financial sentiment dictionary. Examining quantitative forecasts alongside qualitative descriptions lets us characterize how the two models compare across a range of quantitative and qualitative tasks. This design allows us to observe how two different models assess identical firms at the same point in time. Both models share a common reported knowledge cutoff of June 30, 2024, ensuring our December 31, 2024 target date represents a genuine six-month out-of-sample forecast for both.

We document a striking reversal of traditional home bias when LLMs analyze foreign firms. While humans favor familiar domestic assets, we find ChatGPT (US-based) is systematically more optimistic than DeepSeek (China-based) about Chinese firms, exhibiting what we term "foreign bias." ChatGPT's price predictions are 2.415 RMB higher than DeepSeek's, representing 12.53% of the average and 19.68% of the median actual stock price. The same relative optimism appears in the other two tasks: ChatGPT's EPS predictions exceed DeepSeek's by 0.399 RMB per share, representing 68% of the average and 111% of the median actual EPS. Moreover, the business descriptions ChatGPT generates are significantly more positive in tone. This foreign bias is not explained by firm characteristics—the differences persist after controlling for size, profitability, leverage, past returns, and other observables that might influence AI assessments. We further confirm that ChatGPT's quantitative forecasts reflect optimism rather than superior information: its price and EPS predictions both exhibit significantly larger absolute errors than DeepSeek's: 12.3% larger relative to DeepSeek's average price prediction errors and 67.8% larger relative to DeepSeek's average EPS prediction errors, measured against realized December 31, 2024 outcomes.

What explains ChatGPT's foreign bias? We consider two broad channels: information-centric and function-centric.[1] Under an information-centric explanation, relevant information about Chinese firms may be sparse or absent from ChatGPT's effective training information set: that is, it is either missing from the corpus altogether or present in the corpus but only weakly encoded in the model's parameters. Under function-centric explanations, the two models may convert a

[1] We view an LLM as implementing a conditional distribution $p_{\theta_{\mathcal{I}}}(\cdot \mid x)$ over outputs given a prompt x, where $\theta_{\mathcal{I}}$ denotes parameters learned from a training information set $\mathcal{I}$. Cross-model differences can therefore arise from (i) differences in the information effectively available to the model—either present in the training corpus or encoded in its parameters—or (ii) differences in how the model converts a given prompt into an output, holding information constant. We refer to (i) as *information-centric*; we split (ii) into *alignment* (training-time differences in the model's response policy) and *elicitation* (inference-time differences in how prompts are mapped to outputs).

given prompt into an output differently even when their underlying information is the same: their post-training alignment may shape their default response policies in different ways (*alignment*), or they may map prompts to outputs differently at inference time (*elicitation*). These channels yield different testable implications. If missing or under-encoded information is the primary driver, ChatGPT's relative optimism should be smallest where cross-market information gaps are smallest, and supplying Chinese news in the prompt should attenuate or eliminate the bias. If alignment differences are central, the bias should be triggered based on cues the model has been tuned to respond to, such as Chinese-sounding firm names, even where no real firm-specific information is available. If elicitation differences are central, the bias should vary with prompt language, format, retrieval scaffolding, or model variant.

Our evidence points to *effective* information availability as a likely driver of this foreign bias. Using comprehensive news data from RavenPack (US media) and the China Finance News Database (Chinese media), we construct firm-level measures of cross-market news coverage gaps. Chinese outlets account for the vast majority of news for many Chinese firms, and a large fraction of them receive little or no US coverage. Consistent with this mechanism, ChatGPT's relative optimism is most pronounced at firms where US media coverage, especially negative news coverage, is relatively scarce. Controlling for these firm-level news coverage gaps substantially attenuates the baseline foreign bias. These news gaps also reveal that ChatGPT's foreign bias is not inherently optimistic: when we separate negative- and positive-news coverage, ChatGPT's relative optimism rises with the negative-news gap but falls—and reverses sign—with the positive-news gap. ChatGPT is relatively pessimistic in stock price and EPS forecasts about firms whose favorable news is concentrated in Chinese sources that US coverage misses.

Cross-listed firms provide another natural test of this mechanism. Chinese firms listed on US exchanges receive substantially more English-language disclosure and attention (e.g., SEC filings, analyst reports, and US financial press), narrowing the cross-market information gap. Consistent with this mechanism, we find that ChatGPT's relative optimism is significantly attenuated for cross-listed firms: the foreign bias is reduced by roughly 30% for price predictions and 55% for EPS predictions.

Our most direct evidence comes from manipulating the LLMs' information sets available at inference time. We augment the prompt for both ChatGPT and DeepSeek with all Chinese-language firm news from June 2022 to June 2024.[2] This intervention distinguishes between the two information-centric sub-variants we outlined above: information missing from the training corpus altogether versus information present in the corpus but weakly encoded in the model's parameters. Prompting can add firm-specific information to the model's context but cannot change the underlying parameters learned during training. If the bias reflected Chinese information being present but systematically underweighted in the parameters, injected news would compete with existing priors and would be expected only to partially attenuate the foreign bias. Instead, we find that information provision alone eliminates it: the foreign-bias coefficient becomes statistically indistinguishable from zero for stock price predictions, EPS predictions, and business-description tone alike. Also consistent with missing information, we find news-augmentation does not impact DeepSeek's output yet dampens the average optimism of ChatGPT's. Overall, this evidence is consistent with ChatGPT's relative optimism stemming from missing corpus information rather than from weakly encoded parameters

Moreover, effective information availability predicts that foreign bias need not be symmetric. If English-language information about US firms is abundant and comparably available to both models, while Chinese-language information is more extensively covered in DeepSeek's training data, then the informational gap should be larger when the LLMs evaluate Chinese firms than US firms.[3] Consistent with this asymmetry, when both models analyze US firms, prediction differences

[2]Because providing the full text of all articles exceeds practical limits, we supply the abstracts for all available Chinese news articles in that window (rather than full text). However, in an alternative implementation, we instead provide five randomly sampled full-text Chinese news articles from the same window, and find very similar in the results.

[3]This asymmetry could be accentuated if DeepSeek has broad access to English-language information—potentially including via exposure to outputs from leading English models, as has been alleged in public reporting—while additionally drawing on richer Chinese-language sources. See Nellis et al. (Reuters, Jan. 29, 2025) and Weatherbed (The Verge, Jan. 29, 2025) for reporting on OpenAI/Microsoft concerns about possible "distillation" / unauthorized

are small and statistically insignificant—as expected when information is comparably accessible to both.

While these results are consistent with information-centric explanations, they could also reflect function-centric differences: either differences in how the two models have been aligned during post-training (*alignment*) or differences in how they map a given prompt to outputs at inference time (*elicitation*). We therefore conduct placebo and robustness tests targeted at each. One test speaks primarily to alignment. We ask both models to predict about synthetic (nonexistent) firms with Chinese-sounding names; by design, neither model has any real firm-specific information, so any systematic bias would be consistent with a built-in response policy that triggers on Chinese-name cues. We find no prediction differences in this placebo setting. Four tests speak primarily to elicitation. Each varies a different element of the inference-time prompting scaffolding—prompt language, retrieval, model variant, and forecast horizon—to test whether elicitation differences could account for the foreign bias. In each case, the bias persists, narrowing the scope for both alignment- and elicitation-based explanations. Taken together, the weight of evidence point to *effective* information availability/access as a first-order driver of the AI foreign-bias we document.

Finally, we provide suggestive evidence that this foreign bias may not remain confined to AI outputs. Examining sell-side analyst forecasts around the public release of ChatGPT, we find that US-based analysts became modestly more optimistic about Chinese firms relative to China-based analysts, and that this relative shift is concentrated among the firms with the largest cross-border news gaps—reversing in sign with the positive-news gap, just as ChatGPT's AI bias does. We interpret this pattern cautiously: our research design cannot establish that AI adoption caused the change in analyst behavior, only that human analysts' forecast biases correlate with the direction of ChatGPT's foreign bias.

This paper makes three main contributions to the literature. First, we document a novel LLM phenomenon—*foreign bias*—in which (US-based) ChatGPT is systematically more optimistic

use; the claim was also reported by the Financial Times (Jan. 28, 2025).

than (China-based) DeepSeek about Chinese-firm stocks and less accurate ex post. This pattern reverses the classic "home bias" documented for human investors, underscoring that behavioral finance intuitions do not automatically transfer to LLMs.

Second, we provide evidence that this foreign bias is most consistent with *spatial information gaps* across countries—exogenous differences in the informational environments from which models learn. Whereas much of the AI-bias literature emphasizes endogenous sources (e.g., developer choices, alignment, prompt design, or temporal leakage), our findings highlight that model biases can also arise from *exogenous* cross-market information frictions. Because these frictions reflect what a model's training corpus underrepresents, and corpus composition is generally unobservable, the resulting bias cannot be confidently signed a priori: the optimism bias we document for ChatGPT is the on-average realization that this particular information environment produced, not a general property of LLMs' foreign-firm analysis.

Third, we offer a cross-border perspective on LLMs as information intermediaries. By comparing ChatGPT and DeepSeek—frontier models developed in different jurisdictions—we show that the same firm can receive systematically different assessments depending on which jurisdiction's model is used. This extends the information asymmetry literature to a new intermediary: unlike analysts or media, who can invest in information acquisition to mitigate asymmetries (Brennan and Subrahmanyam, 1995; Bushee, Core, Guay, and Hamm, 2010), LLMs can inherit and amplify existing cross-border information gaps through their training and inference behavior.

The resulting divergence in LLM forecasts has implications for cross-border investment analysis, market efficiency, and for policy discussions around transparency in model training data and sourcing. An immediate practical implication is that when investors or analysts use LLMs for cross-border financial analysis and prediction tasks, certain models' outputs can be systematically skewed, potentially distorting investment inferences and risk assessments, and mitigating these distortions may call for explicit local-information supplementation (e.g., local-language sources) and validation against alternative models or benchmarks. More broadly, our findings suggest that unknown and exogenous information environment differences can result in unexpected biases in LLM predictions. As AI tools increasingly mediate cross-border capital allocation, understanding and monitoring these biases will be important for market efficiency and investor protection.

2 Sample Construction and Data

2.1 Sample Selection

Our sample construction begins with the universe of Chinese firms listed on the Shanghai and Shenzhen stock exchanges with available data in Compustat Global. We focus on Chinese firms as they represent a major cross-border investment opportunity where information asymmetries between US and Chinese investors are particularly pronounced. The parallel development of sophisticated LLMs in both countries—ChatGPT in the US and DeepSeek in China—provides an ideal setting to study how AI models trained in different linguistic and informational environments assess the same set of firms.

We start with all Chinese firms in Compustat Global as of December 31, 2023, requiring non-missing data for key financial variables used as controls in our analysis. Specifically, we require firms to have available information on total assets (to calculate $FIRMSIZE$), return on assets (ROA), and leverage (LEV). These screens ensure that both AI models have access to basic financial information when making predictions, allowing us to isolate the effect of information asymmetries in news and contextual information rather than differences in fundamental financial data availability. This process yields 4,978 unique Chinese firms.

2.2 AI Prediction Collection

For each firm in our sample, we collect a set of forward-looking predictions and a qualitative assessment from two frontier LLMs: ChatGPT 4.1 (developed by OpenAI) and DeepSeek R1 (developed by DeepSeek). We access both models through their respective APIs to ensure consistency and replicability. Both ChatGPT 4.1 and DeepSeek R1 have training sample periods that ended around June and July 2024, and all predictions are made for the same future date of December 31, 2024, creating an almost six-month out-of-sample prediction horizon.[4] This mitigates the influence of look-ahead bias on our results (Didisheim, Fraschini, and Somoza, 2025).

For each firm, we collect three outputs, chosen to span the tasks that LLMs can be deployed to perform in equity analysis. The first is a *stock price prediction*: each model forecasts the firm's stock price as of December 31, 2024 ($STOCK\_PRICE$). The second is an *earnings prediction*: each model forecasts the firm's earnings per share as of the same date (EPS), the central object of sell-side analyst research. The third is a *qualitative business description*: we ask each model to generate a 150–200-word description of the firm and measure the tone of that description using the Loughran and Mcdonald (2011) financial sentiment dictionary, yielding a normalized sentiment score ($BUSI\_SENTIMENT$) for which higher values indicate more positive language.

We use standardized prompts to ensure comparability across models, presenting each model with the same firm identifiers (name and ticker), instructing it to act as a professional financial analyst, and directing it not to use current online data. We randomize the order in which firms are presented to each model to avoid potential ordering effects, and we set the temperature to zero. To assess data quality, we re-query both models for all our sample firms one week later and find correlation coefficients exceeding 0.95 for both models' predictions. Exhibit A reports the prompts used across all analyses reported in this paper, including the baseline specification and each robustness test.

This process generates 9,956 firm–AI observations (4,978 firms $\times$ 2 models), with each observation containing a stock price prediction ($STOCK\_PRICE$), an EPS prediction (EPS), and a business-description sentiment score ($BUSI\_SENTIMENT$).

[4]The cutoff dates reflect the vendors' public statements regarding each model's training data, with June 30, 2024 used for ChatGPT-4.1 and July 31, 2024 for DeepSeek R1. While ChatGPT and DeepSeek differ by one month in their reported training cutoff dates, this discrepancy is unlikely to affect our findings. Our U.S. firm tests yield no systematic differences across models, suggesting that minor differences in training horizons do not materially influence the results.

2.3 News Coverage Data

To examine the mechanism driving AI prediction differences, we collect comprehensive data on media coverage from both US and Chinese sources. This allows us to infer the role of cross-border asymmetries in news coverage about Chinese firms in explaining the observed foreign bias. Our hypothesis is that ChatGPT's training prioritizes US sources, whereas local Chinese news plays a greater role in DeepSeek.

For US media coverage, we use RavenPack News Analytics, which processes news from major US financial publications including the Wall Street Journal, Financial Times, Bloomberg, Reuters, and Dow Jones Newswires (Drake, Guest, and Twedt, 2014; Guest, 2021; Lock, 2024). RavenPack provides sentiment scores for each news article mentioning our sample firms. We aggregate all news articles from June 30, 2022, to June 30, 2024 (a three-year window) to capture the information environment likely reflected in the LLMs' training data. Articles are classified as negative if they have a sentiment score below 0 on RavenPack's [-1, +1] scale, and positive if above 0.

For Chinese media coverage, we use the China Finance News Database (CFND), which aggregates content from major Chinese financial media outlets including Caixin, Securities Times, China Securities Journal, and Shanghai Securities News (Li, Li, Wang, and Thatcher, 2021; Gao, Zhang, and Yang, 2023; Qiao, 2023). Similar to our approach with RavenPack, we collect all articles mentioning our sample firms during the same three-year period and classify them by sentiment using CFND's proprietary sentiment analysis, which is calibrated to be comparable to RavenPack's methodology.

For both sources, we impose a minimum article length requirement of at least 10 sentences. This criterion ensures that we eliminate those articles with minimal firm-specific information, such as brief wire alerts and index-inclusion notices, and focus on more substantive articles. To the extent these media sources are used in the training of LLMs, such short items are unlikely to meaningfully shape an LLM's learned representations of a firm.

We then construct three firm-level measures of cross-border information gaps. *NEG_NEWS_GAP* is defined as

$$\frac{\text{CN negative} - \text{US negative}}{\text{CN negative} + \text{US negative}} \tag{1}$$

computed over the three-year window prior to the models' training cutoff (June 30, 2022 to June 30, 2024). Here, *CN negative* refers to the number of negative news articles about a given firm from Chinese media sources (CFND), while *US negative* refers to the number of negative news articles from U.S. media sources (RavenPack). This variable lies in $[-1, 1]$, with $+1$ indicating only Chinese negative coverage and -1 indicating only U.S. negative coverage.

POS_NEWS_GAP is defined analogously using positive news articles:

$$\frac{\text{CN positive} - \text{US positive}}{\text{CN positive} + \text{US positive}} \tag{2}$$

where *CN positive* (*US positive*) denotes the number of positive news articles about the firm from Chinese (U.S.) media sources. As with *NEG_NEWS_GAP*, the variable lies in $[-1, 1]$, with values near $+1$ indicating that positive coverage is concentrated in Chinese outlets and values near -1 indicating concentration in U.S. outlets.

NET_NEG_NEWS_GAP captures the net negativity gap across countries and is defined as

$$\frac{(\text{CN negative} - \text{CN positive}) - (\text{US negative} - \text{US positive})}{\text{CN negative} + \text{CN positive} + \text{US negative} + \text{US positive}}, \tag{3}$$

computed over the same window. Higher values indicate that, on net, Chinese coverage of a firm is more negative than U.S. coverage. For firms with no articles in either market, we set the gap to zero; results are robust to dropping such firms or adding a no-news indicator.

2.4 Control Variables

We obtain firm-level financial data from Compustat Global measured as of fiscal year-end 2023, ensuring all control variables precede our AI predictions. Our control variables include standard predictors of stock returns and firm value from the asset pricing literature. *FIRMSIZE* is the natural logarithm of total assets. *LEV* is total debt divided by total assets. *ROA* is income before extraordinary items divided by total assets. *LOSS* is an indicator variable equal to one if the firm reported negative net income. *BM* is the book-to-market ratio, calculated as book value of equity divided by market value of equity. *PRE_RETURN* is the stock return during 2023. *RETURN_VOLATILITY* is the standard deviation of daily returns during 2023. *IO* is the percentage of shares held by institutional investors, obtained from the China Stock Market and Accounting Research (CSMAR) database.

For robustness tests and mechanism analyses, we collect additional data. Foreign institutional ownership (*FOREIGN_IO*) comes from CSMAR's detailed ownership files. Actual stock prices on December 31, 2024, used to calculate prediction errors, are obtained from Compustat Global.

2.5 Sample Composition and Descriptive Statistics

Table 1, Panel A provides descriptive statistics for our key variables. The AI predictions reveal informative baseline patterns. *STOCK_PRICE* has a mean of 20.76 RMB with substantial variation (standard deviation 16.07); the distribution is right-skewed, with a median of 15.33 below the mean and an interquartile range spanning 8.32 to 28.46 RMB. *EPS* predictions average 0.97 RMB per share, with a median of 0.85, and *BUSI_SENTIMENT*—the normalized tone of the AI-generated business descriptions—averages 0.53, ranging from 0.20 at the 10th percentile to 0.82 at the 90th. On average, both models produce positive assessments of Chinese firms across all three tasks, providing a baseline against which we measure the differential optimism between them.

Table 1, Panel B, provides more detailed summary statistics on the news coverage for Chinese listed firms in our sample, comparing coverage from Chinese media sources (B.1) and US media sources (B.2). We report four measures: total news articles (*Total News*), articles with negative news (*Negative News*), articles with positive news (*Positive News*), and the difference between negative and positive articles (*Net Negative News*). The bottom row (B.3) reports the difference between US and Chinese source coverage, with significance stars indicating whether the gap is statistically different from zero. On average, U.S. media generate far fewer news articles than Chinese media: the average firm in our sample has 1.68 U.S. articles compared to 26.93 Chinese articles. While both sources produce more negative than positive articles in absolute terms, Chinese media coverage is net negative (2.24 more negative than positive articles per firm), whereas U.S. coverage is slightly net positive (-0.10), meaning marginally more positive than negative articles. All differences between US and Chinese coverage are statistically significant at the 1% level.

These coverage differences result in meaningful information gaps. Returning to Panel A, the mean *NEG_NEWS_GAP* of 0.78 indicates that negative news is heavily concentrated in Chinese media relative to US media; the mean *POS_NEWS_GAP* of 0.39 shows that positive news is similarly, though less severely, concentrated in Chinese sources; and the mean *NET_NEG_NEWS_GAP* of 0.35 indicates that, on net, the cross-border gap is tilted toward negative news. These asymmetries raise the possibility that a US-based LLM like ChatGPT incorporates systematically different firm-level news about Chinese companies than China-based DeepSeek. Because negative and positive coverage are unevenly missing, the net effect on relative optimism need not be uniform.

Table 1, Panel C reports the industry composition of our sample using the Fama–French 12-industry classification and, within each industry, the average cross-border news gap measures. Manufacturing is the largest sector (24.89%), followed by Business Equipment (20.37%), with sizable representation from Chemicals (8.86%) and Healthcare (8.44%). Consumer Nondurables (7.73%) and Consumer Durables (6.77%) also account for meaningful shares of the sample, while Financials represent 4.38%. Telecommunications is the smallest sector (0.52%).[5]

Panel C also shows substantial cross-industry heterogeneity in cross-border information gaps. For most industries, *NEG_NEWS_GAP* is strongly positive (roughly 0.67 to 0.85) and *POS_NEWS_GAP* is also positive but smaller (roughly 0.29 to 0.52), indicating that both negative and positive coverage are, on average, more concentrated in Chinese outlets than in U.S. outlets. Financials stand out with markedly lower gaps—a *NEG_NEWS_GAP* of 0.2618 and a *NET_NEG_NEWS_GAP* of

[5] Panel C counts firm–AI observations (9,956 total). Because each underlying firm appears twice (once per model), the industry shares are identical to those computed at the firm level.

0.0896—indicating relatively balanced cross-border coverage for this sector. This industry-level variation in the information environment motivates our use of industry fixed effects and industry-clustered inference in the ensuing empirical analyses.

3 Empirical Results

3.1 Foreign Bias in AI Predictions

We begin by establishing our central empirical finding: when evaluating Chinese firms, US-based ChatGPT is systematically more optimistic than China-based DeepSeek. This pattern—which we term "foreign bias"—reverses the traditional home bias observed among human investors.

To test for systematic differences in AI outputs, we estimate the following specification:

$$Y_{i,m} = \alpha + \beta \cdot US\_AI_m + \gamma' \mathbf{X}_i + \delta_j + \varepsilon_{i,m} \tag{4}$$

where $Y_{i,m}$ denotes the output for firm i by model m. In our main tests, $Y_{i,m}$ is one of three outputs: *STOCK_PRICE*, the predicted stock price for December 31, 2024 (approximately six months after the knowledge cutoff date); *EPS*, the predicted earnings per share as of the same date; or *BUSI_SENTIMENT*, the sentiment of the AI-generated business description. The key independent variable, US_AI_m, is an indicator equal to one for outputs generated by ChatGPT 4.1 and zero for those from DeepSeek R1. Thus, the coefficient β captures the systematic difference between the US-based and China-based models: a positive β indicates that ChatGPT is relatively more optimistic.

The vector $\mathbf{X}_i$ includes firm-level controls measured as of year-end 2023: *FIRMSIZE* (natural logarithm of total assets), *LEV* (total debt divided by total assets), *ROA* (income before extraordinary items divided by total assets), *LOSS* (indicator for negative net income), *BM* (book-to-market ratio), *PRE_RETURN* (stock return during 2023), *RETURN_VOLATILITY* (standard deviation of daily returns during 2023), and *IO* (percentage of shares held by institutional investors). We include industry fixed effects δ_j based on the Fama-French 12-industry classification and cluster standard errors at the industry level.

Table 2 presents the results. Column (1) reveals that ChatGPT predicts systematically higher stock prices than DeepSeek. The coefficient on *US_AI* is 2.4153 (significant at the 1% level), indicating that ChatGPT's price predictions exceed DeepSeek's by approximately 2.4 RMB per share. This difference is economically substantial, representing 12.53% of the mean actual stock price (19.28 RMB) and approximately 19.68% of the median actual stock price (12.27 RMB). To contextualize this magnitude, a 2.4 RMB difference for the median firm translates to a market capitalization difference of approximately 240 million RMB, assuming 100 million shares outstanding.

The control variables exhibit intuitive relationships with price predictions. More profitable firms (*ROA*) receive higher predictions, while firms with higher book-to-market ratios (*BM*) or prior returns (*PRE_RETURN*) receive lower predictions, consistent with AI models incorporating growth characteristics and mean reversion into their forecasts.

Columns (2) and (3) show that this relative optimism extends to ChatGPT's other two outputs. In Column (2), where the dependent variable is the predicted *EPS*, the coefficient on *US_AI* is 0.3993 (significant at the 1% level): ChatGPT's EPS forecasts exceed DeepSeek's by roughly 0.40 RMB per share, representing 68% of the average and 111% of the median actual EPS. In Column (3), where the dependent variable is *BUSI_SENTIMENT*, the coefficient on *US_AI* is 0.1115 (significant at the 1% level): the business descriptions ChatGPT generates are more positive in tone, by roughly one-fifth of the sample mean and about 0.4 standard deviations. That ChatGPT is more optimistic whether the task is a price target, an earnings figure, or a paragraph of narrative prose indicates that the foreign bias is a general feature of how the model assesses Chinese firms.

3.2 Prediction Accuracy

To ascertain whether ChatGPT's relatively higher forecasts reflect systematic error or superior information, we compare the LLMs' quantitative predictions against realized outcomes. If Chat-

GPT's higher predictions stem from superior information, we would expect them to result in lower forecast errors; if they reflect optimism bias, we would expect greater forecast errors.

We estimate Eq., (4) using measures of prediction accuracy as the dependent variables, one for each quantitative forecast. *STOCK_PRICE_ERROR* is the absolute difference between the predicted and realized stock price on December 31, 2024, and *EPS_ERROR* is the absolute difference between predicted and realized EPS. In both specifications, a positive coefficient β indicates that ChatGPT exhibits larger prediction errors than DeepSeek.

Table 3 presents the results. In Column (1), the coefficient on *US_AI* is 1.0863 (significant at the 1% level): ChatGPT's price predictions exhibit significantly larger absolute errors than DeepSeek's—approximately 12.3% larger relative to DeepSeek's average price prediction errors. In Column (2), the coefficient on *US_AI* is 1.8116 (significant at the 1% level), indicating that ChatGPT's EPS predictions are likewise substantially less accurate—approximately 67.8% larger relative to DeepSeek's average EPS prediction errors. On both quantitative tasks, ChatGPT's greater optimism is accompanied by larger, not smaller, forecast errors.

Taken together, Table 2 and Table 3 establish the paper's central stylized facts: ChatGPT is systematically more optimistic than DeepSeek when evaluating Chinese firms across price targets, earnings forecasts, and qualitative descriptions. And, on the two quantitative tasks where accuracy can be assessed, this optimism translates into lower predictive accuracy. The foreign bias therefore reflects genuine bias rather than superior information processing by the US-based model. We now turn to understanding what drives this pattern.

3.3 Conceptual Framework and Empirical Strategy

What explains ChatGPT's foreign bias? We conceptualize an LLM as implementing a conditional distribution $p_{\theta_{\mathcal{I}}}(\cdot \mid x)$ over outputs given a prompt x, where $\theta_{\mathcal{I}}$ denotes parameters learned from a training information set $\mathcal{I}$. Under this view, cross-model differences in predictions for the same firm can arise from two distinct sources: (i) differences in the information effectively available to each model—either present in the training corpus or encoded in its parameters—or (ii) differences in how the model converts a given prompt into an output, holding information constant.

3.3.1 Information-Centric Explanations

Information-centric explanations posit that prediction differences arise primarily from asymmetries in what information is effectively available to each model, leading to systematically different assessments. Two related possibilities fall under this umbrella: information about Chinese firms may be missing from ChatGPT's training corpus altogether, or it may be present in the corpus but only weakly encoded in the model's learned parameters (sometimes called *parameter encoding*). Both manifest as the model effectively having less information about Chinese firms than DeepSeek does.

Several features of the information environment support this possibility. First, Chinese-language news and financial media represent a vast corpus of firm-specific information that may be underrepresented in English-centric training data. Second, US media coverage of Chinese firms is often limited to a subset of larger or more internationally prominent companies, leaving smaller firms with sparse English-language coverage. Third, even when US media do cover Chinese firms, the coverage may emphasize different aspects or carry different sentiment than local Chinese coverage.

Critically, if Chinese-language news contains disproportionately more negative information about Chinese firms on average than US-language coverage—perhaps because local journalists have better access to firm-specific developments or because US coverage focuses on growth narratives—then a model trained primarily on English sources could systematically underweight negative signals, manifesting as relative optimism. The same logic, however, runs in both directions. If favorable news about a firm is instead concentrated in Chinese sources, a model reliant on English coverage would underweight *positive* signals, manifesting as relative pessimism. The direction of the resulting bias therefore depends on which type of news US coverage tends to miss.

The information-availability mechanism yields several testable predictions. First, ChatGPT's relative optimism should track the cross-market news coverage gap: it should be most pronounced for firms whose negative news is concentrated in Chinese rather than US media, and it should weaken—and potentially reverse—for firms whose positive news is substantially concentrated in China. Second, the bias should attenuate for firms where cross-market information asymmetries are naturally smaller, such as cross-listed firms that receive greater English-language disclosure and analyst attention. Third, providing both models with Chinese news at inference time should substantially reduce or eliminate the prediction gap. This intervention also helps distinguish between the two information-centric variants noted above: under parameter encoding, supplying information at inference should only partially close the gap, since the under-weighted internal prior would still compete with the new input. In a corpus-level missing-information story, providing information should fully close the gap. Fourth, because English-language information about American public firms is likely to be comparably accessible to both models, we expect the LLMs' prediction differences to be small or absent when they analyze US firms.

3.3.2 Function-Centric Explanations

Function-centric explanations focus on differences in how the two models convert a given prompt into an output, holding the underlying information set constant. Two such mechanisms operate at different stages of the model's life cycle: one at training time and one at inference time.

Under an *alignment* mechanism, the two models may carry different response policies, reflecting how each has been instruction-tuned during post-training. ChatGPT may have been aligned in ways that systematically favor optimistic, hedged, or stylistically distinctive responses to certain prompt types: for example, the bias could be triggered by cues the model has been tuned to respond to, such as a Chinese-sounding firm name in the prompt, even when no real firm-specific information is available. If so, the bias would not be expected to vanish when firm-level information is supplied at inference.

Under an *elicitation* mechanism, the two models may map a given prompt into outputs differently at inference time, driven by prompt language, prompt format, retrieval scaffolding, or other inference-time factors that depend on the prompt rather than on the model's underlying response policy or information. Under this mechanism, the bias could vary with prompt language, format, retrieval access, or model variant.

3.3.3 **Empirical Strategy**

The information-centric and function-centric explanations can produce observationally equivalent outcomes in many settings, and because we cannot directly observe training corpora or inspect how information is encoded in model parameters, LLMs remain, in important respects, black boxes. We therefore develop empirical strategies that provide indirect evidence on the likely sources of the foreign bias.

Our empirical strategy proceeds in two stages. First, the tests in Section 3.4 take up the information-centric predictions outlined above: cross-sectional variation in cross-market coverage gaps, cross-listed firms, news injection at inference time, and a US-firm symmetric-information benchmark. Second, the tests in Section 3.5 address function-centric alternatives. The synthetic-firms placebo speaks primarily to *alignment*: do the models respond differently to Chinese-named cues alone, even when no real firm-specific information is available? Then, we conduct a series of robustness tests that speak primarily to *elicitation*: each varies a different element of the inference-time prompting scaffolding—prompt language, retrieval access, the reasoning-oriented model variant used at inference, and forecast horizon—to test whether elicitation differences could account for the foreign bias.

3.4 Evidence for Information-Centric Mechanisms

We now examine whether the foreign bias is driven by information asymmetries as our conceptual framework predicts. We conduct four tests, exploiting both cross-sectional variation across firms and experimental manipulation of the information environment.

3.4.1 Cross-Border Media Coverage Gaps

If ChatGPT's foreign bias stems from limited access to Chinese-language news, its outputs should move with the cross-border coverage gap—relatively more optimistic where negative news is concentrated in Chinese rather than US media, and, by the same logic, relatively less optimistic where positive news is so concentrated. To test these predictions, we augment our baseline specification with interactions between US_AI and measures of the cross-border news coverage gap:

$$\begin{aligned} Y_{i,m} \quad = \quad & \alpha + \beta_1 \cdot US\_AI_m \times NEWS\_GAP\_Measure_i + \beta_2 \cdot NEWS\_GAP\_Measure_i + \beta_3 \cdot US\_AI_m \\ & + \gamma' \mathbf{X}_i + \delta_j + \varepsilon_{i,m} \end{aligned} \tag{5}$$

where $Y_{i,m}$ is one of our three AI outputs and $NEWS\_GAP\_Measure_i$ captures the extent to which news coverage of firm i is concentrated in Chinese versus US media. We employ the three measures defined in Section 2: $NET\_NEG\_NEWS\_GAP$ (the net negativity of news in China relative to the US), NEG_NEWS_GAP (the cross-border gap in negative coverage), and POS_NEWS_GAP (the gap in positive coverage). For each measure, values near $+1$ indicate that the relevant coverage is concentrated in Chinese outlets and values near -1 indicate concentration in US outlets. We report cluster-robust standard errors, clustering by industry to account for potential within-industry correlation in model outputs. In untabulated robustness checks, our inferences are unchanged when instead clustering at the firm level or when using heteroskedasticity-robust (White, 1980) standard errors.

Under the information-availability hypothesis, the interaction coefficient β_1 carries a sign that differs across measures. For NEG_NEWS_GAP and $NET\_NEG\_NEWS\_GAP$, β_1 should be positive: ChatGPT's relative optimism should rise as negative news becomes increasingly concentrated in Chinese sources it underweights. For POS_NEWS_GAP, β_1 should be negative: ChatGPT's relative optimism should fall as positive news becomes increasingly concentrated in those same sources.

Table 4, Panel A, reports results using the net negativity gap. The interaction US_AI $\times$

NET_NEG_NEWS_GAP is positive and significant for all three outputs: 5.5227 (significant at the 1% level) for stock price predictions, 0.1885 (significant at the 1% level) for EPS predictions, and 0.0172 (significant at the 1% level) for business-description sentiment. ChatGPT's relative optimism is thus concentrated where Chinese outlets are, on net, more negative than US outlets. Controlling for the gap, the direct *US_AI* effect attenuates and, for price predictions, becomes statistically insignificant (0.4824); for EPS and business-description sentiment it shrinks but remains significant, indicating that the net negativity gap accounts for much, though not all, of the cross-model difference in those two outcomes.

Panel B focuses on the asymmetry in negative- and positive-news coverage. The results show that the foreign bias is not inherently a tendency toward optimism. The interaction *US_AI* $\times$ *NEG_NEWS_GAP* is positive for price and EPS predictions (2.7121 and 0.1234, both significant at the 1% level): ChatGPT grows more optimistic, relative to DeepSeek, as negative news becomes more concentrated in Chinese outlets. The interaction *US_AI* $\times$ *POS_NEWS_GAP*, by contrast, carries the opposite sign and a comparable magnitude (-5.1233 and -0.1690, both significant at the 1% level): ChatGPT grows more *pessimistic* as positive news becomes more concentrated in Chinese outlets. A firm whose positive news is essentially confined to Chinese sources (*POS_NEWS_GAP* near one), the model-implied ChatGPT–DeepSeek price gap is approximately -0.7 RMB, meaning that ChatGPT is relatively pessimistic. On the other hand, for a firm whose negative news is essentially confined to Chinese sources (*NEG_NEWS_GAP* near one), the implied gap is roughly $+3$ RMB, meaning that ChatGPT is relatively optimistic. The two interactions carry the same opposing signs for business-description sentiment, though there they are not statistically significant.

Taken together, the evidence indicates that ChatGPT's foreign bias is associated with what US coverage underreports for a given Chinese firm: relative optimism when the missing news is unfavorable, relative pessimism when it is favorable. Crucially, because the composition of a model's effective information set is not observable ex ante, the sign of this bias cannot be anticipated; only its dependence on the cross-border coverage gap can be.

3.4.2 Cross-Listed Firms

Cross-listed firms provide another natural test of the information-asymmetry mechanism. Chinese firms listed on US exchanges are subject to broader disclosures, bilingual filings, and more extensive analyst and media coverage across markets, naturally narrowing the information gap between US and Chinese sources. If cross-border information frictions drive the foreign bias, the effect should weaken for these firms. We test this by estimating:

$$Y_{i,m} = \alpha + \beta_1 \cdot US\_AI_m + \beta_2 \cdot US\_AI_m \times CROSS\_LISTED_i + \beta_3 \cdot CROSS\_LISTED_i + \gamma' \mathbf{X}_i + \delta_j + \varepsilon_{i,m} \quad (6)$$

where $CROSS\_LISTED_i$ is an indicator equal to one if firm i is cross-listed on a US exchange. A negative coefficient β_2 would indicate that ChatGPT's relative optimism is attenuated for cross-listed firms, consistent with the information-availability mechanism.

Table 5 presents the results. Consistent with the prediction, the interaction term $US\_AI \times CROSS\_LISTED$ is negative for all three outputs. It is -0.7660 (significant at the 1% level) for price predictions and -0.2242 (significant at the 1% level) for EPS predictions; relative to the main US_AI effects (2.4467 and 0.4077), these imply that the foreign bias is roughly 30% smaller for price predictions and 55% smaller for EPS predictions among cross-listed firms. For business-description sentiment, the interaction is also negative (-0.0033) but small and only marginally significant (significant at the 10% level), so the cross-listing attenuation is clearest for the two quantitative forecasts. The main effect of US_AI remains positive and significant throughout, indicating that ChatGPT's optimism persists but is muted for cross-listed firms. The positive coefficient on $CROSS\_LISTED$ itself (0.5600 for prices and 0.2060 for EPS, both significant at the 1% level) indicates that cross-listed firms receive higher predictions overall from both models, consistent with their greater transparency and international visibility.

The attenuation of the foreign bias for cross-listed Chinese firms provides further support that the bias arises from differences in information environments rather than inherent model-specific optimism. When informational asymmetries between the US and Chinese markets are naturally smaller, so too is the magnitude of ChatGPT's foreign bias.

3.4.3 Closing the News Gap

Our most direct test of the information-availability mechanism involves experimentally manipulating the information available to both models at inference time. Table 6 reports results of an intervention in which we augment the prompt for both ChatGPT and DeepSeek with Chinese-language news about each firm drawn from CFND over June 2022 to June 2024. Because providing the full text of all articles exceeds practical limits, we instead supply the *abstracts* (i.e., the short summary fields) for *all* available Chinese news articles in that window.[6] This design directly targets the hypothesized information gap: it equalizes access to local Chinese news at prediction time while leaving the models' learned parameters unchanged.

Panel A of Table 6 shows that once both models are conditioned on the same Chinese news abstracts, the baseline foreign bias disappears across all three outputs. The coefficient on *US_AI* is statistically indistinguishable from zero for price predictions (−0.7980), EPS predictions (0.0241), and business-description sentiment (0.0109)—economically negligible next to the baseline effects of 2.4153, 0.3993, and 0.1115, respectively, from Table 2. Once both ChatGPT and DeepSeek have access to Chinese news in the inference context, ChatGPT's relative optimism is essentially eliminated.

To clarify which model drives this convergence, Panel B compares the baseline outputs with their news-augmented counterparts. We stack the two sets of outputs and estimate an interaction specification, using an indicator (*NEWS_INJECTED*) for the news-augmented observations and an interaction between *NEWS_INJECTED* and *US_AI*. This resembles a difference-in-differences design in which *NEWS_INJECTED* is the treatment and *US_AI* identifies the affected group.

[6] In an alternative implementation, we instead randomly sample five full-text Chinese news articles from the same June 2022–June 2024 window and provide the full text of those articles in the prompt. Untabulated results are very similar, indicating that our findings are not driven by using abstracts rather than full articles.

In Panel B, the positive and significant *US_AI* coefficient reproduces the baseline foreign bias of Table 2. The main effect of *NEWS_INJECTED* is essentially zero for all three outputs, implying that augmenting the prompt with Chinese news abstracts has no detectable average effect on DeepSeek's output. The interaction term *US_AI* × *NEWS_INJECTED*, by contrast, is negative and significant throughout: −3.2136 for price predictions and −0.3754 for EPS predictions (both significant at the 1% level), and −0.3958 (significant at the 5% level) for business-description sentiment.

Together, the *NEWS_INJECTED* main effect and the interaction indicate that news injection operates almost entirely on ChatGPT. For the two quantitative forecasts, the interaction roughly offsets the baseline *US_AI* effect—ChatGPT revises its price predictions down by about 3.2 RMB and its EPS predictions down by about 0.38 RMB—which is why the news-augmented gap in Panel A is close to zero. DeepSeek's outputs, already informed by Chinese news, are essentially unchanged.

The fact that the baseline foreign-bias collapses under news augmentation is helpful in distinguishing between the two information-centric sub-variants outlined earlier: corpus-missing vs. weakly encoded parameters. Prompting fundamentally cannot alter a model's pre-trained weights; it can only provide new information to process. If the bias reflected Chinese information being present but systematically underweighted in ChatGPT's parameters (the parameter-encoding sub-variant), injected news would compete with existing priors and only partially attenuate the foreign bias. Instead, we find that information provision alone eliminates it, consistent with the corpus-missing sub-variant. The fact that the prediction changes are concentrated in ChatGPT rather than DeepSeek further supports the effective information-availability interpretation.

3.4.4 US Firms: A Symmetric Information Benchmark

The information-availability mechanism predicts a foreign-bias asymmetry: such a bias should emerge when evaluating firms from the information-disadvantaged market but not when evaluating firms where both models have comparably rich information. For US firms, English-language coverage is abundant and accessible to both models. Moreover, as noted in footnote 3, if allegations that DeepSeek was trained in part on ChatGPT outputs are correct, information ChatGPT possesses about US firms may also be effectively embedded in DeepSeek's training data. Both channels point toward more symmetric information environments when analyzing US firms.

Table 7 tests this prediction by repeating the baseline analysis for a sample of 5,844 US firms. After collecting both models' outputs for these firms, we estimate the baseline regression (Eq., (4)). The coefficients on *US_AI* are small and statistically insignificant for all three outputs: −0.1539 for price predictions, −0.1046 for EPS predictions, and 0.0291 for business-description sentiment. These null results stand in stark contrast to the statistically strong and positive coefficients observed for Chinese firms, confirming that the foreign bias is specific to the cross-border setting where information asymmetries are severe.

3.5 Assessing Function-Centric Explanations

While the evidence presented thus far favors the effective information-availability explanation, two function-centric alternatives remain. Under *alignment*, the two models may carry different post-training response policies that systematically favor optimistic, hedged, or stylistically distinctive outputs in certain tasks, regardless of the firm-specific information they actually have. Under *elicitation*, the two models may map a given prompt to outputs differently at inference time—driven by prompt language, retrieval scaffolding, the model variant used at inference, or the forecast horizon, rather than by any training-time response policy or underlying information. We conduct a battery of placebo and robustness tests to assess the scope of each.

3.5.1 **Synthetic Firms Placebo**

We begin with an experiment designed to test whether the models respond differently to Chinese-sounding company names in the prompt, holding information differences constant. The logic is as follows: if ChatGPT's relative optimism reflects a built-in response policy that is triggered by Chinese-name cues (an *alignment* story), the bias should emerge even when neither model has any real information about the firm. If, instead, the bias requires genuine information asymmetries, it should disappear when both models are equally uninformed.

Table 8 implements this test using a placebo sample of 100 synthetic Chinese firm names. These are fictitious firms with Chinese-sounding names but no corresponding real-world entities or media coverage. By design, there is no underlying firm-specific information for either model to differentially access; any systematic bias in this setting would therefore be consistent with name-triggered alignment rather than information availability.

After collecting both models' outputs for these synthetic firms, we again estimate the baseline regression (Eq., (4)). Table 8 shows that the coefficient on *US_AI* is small and statistically insignificant for all three outputs—price predictions, EPS predictions, and business-description sentiment. This null result is inconsistent with a pure alignment explanation, which would predict that Chinese-sounding names alone could trigger differential responses between models. It also suggests that when firm-specific information is absent and both models operate under symmetric informational conditions, the excess optimism disappears, consistent with the effective information availability explanation.

3.5.2 Robustness Tests

While the synthetic-firms placebo rules out name-triggered alignment as a driver of the foreign bias, function-centric mechanisms may still operate through inference-time elicitation. We conduct four additional tests to assess these possibilities, reported in Table 9, each of which implements our baseline regression analysis (Eq., (4)) under a different inference-time scaffolding: the prompt language (Panel A), the retrieval access available at inference (Panel B), the model variant used at inference (Panel C), or the forecast horizon (Panel D), applied to the same 4,978 Chinese firms as in Table 2.

We first examine whether prompt language drives the results. If ChatGPT exhibits greater optimism because it responds differently to English-language prompts—perhaps due to language-specific fine-tuning—then delivering prompts in Mandarin should attenuate or eliminate the bias. Panel A shows that the foreign bias persists when prompts are delivered in Mandarin: the coefficient on *US_AI* remains positive and significant for price predictions (1.5173, significant at the 5% level), EPS predictions (0.2479, significant at the 1% level), and business-description sentiment (0.0785, significant at the 1% level). Relative to the baseline estimates in Table 2, the magnitudes attenuate by roughly a third for each output, but the bulk of the foreign bias clearly survives the switch to Mandarin prompts, indicating that it is not simply an artifact of English-language prompt framing.

We next examine whether inference-time scaffolding—either through expanded information access or enhanced reasoning capabilities—changes the magnitude of the bias. Panel B reports results when online search is enabled for both models, which expands the information available to the LLM at prediction time. If the foreign bias is purely elicitation-driven—reflecting how the models respond to prompts rather than what information they possess—then expanding information access may not systematically attenuate the bias. In contrast, if the bias reflects missing information, then allowing models to search the web should reduce the gap.

Table 9, Panel B shows that enabling online search attenuates, but does not eliminate, the bias. The coefficient on *US_AI* falls to 0.5338 for price predictions, 0.1019 for EPS predictions, and 0.0626 for business-description sentiment (a 78%. 74%, and 44% reduction relative to the baseline), each still statistically significant. We read this as more consistent with an information-availability explanation than a pure elicitation story: expanding the models' information access narrows the gap, as a missing-information account predicts, while the residual bias indicates that ad hoc retrieval does not fully close their effective information gap. The residual likely reflects a well-known limitation of automated web retrieval Shi, Chen, Misra, Scales, Dohan, Chi, Schärli, and Zhou (2023); Liu, Lin, Hewitt, Paranjape, Bevilacqua, Petroni, and Liang (2024): returned content varies in relevance, and so has a lower signal-to-noise ratio than a curated context window populated with directly relevant material, such as as in our news-injection experiment where we supply firm-specific news articles into the model's prompt.

Next, we test whether the bias attenuates with improved model capabilities. Recent advances in LLM development have produced reasoning-oriented model variants that are designed to improve multi-step reasoning performance. To the extent such variants reduce sensitivity to superficial prompting cues, a purely elicitation-driven mechanism could suggest the elimination of the AI foreign-bias when prompting such reasoning-oriented models.

Table 9, Panel C reports results using ChatGPT-5 nano, a smaller reasoning-oriented ChatGPT variant; we use the nano version for processing efficiency. Because ChatGPT-5 does not allow the temperature parameter to be set to zero (unlike DeepSeek R1), we construct each ChatGPT-5 output as the average across 50 repeated prompts.

Panel C shows that the coefficient on *US_AI* remains positive and significant across all three outputs: 1.6622 (significant at the 5% level) for price predictions, 0.1786 (significant at the 1% level) for EPS predictions, and 0.0535 (significant at the 1% level) for business-description sentiment.[7] The magnitudes are somewhat attenuated relative to the baseline in Table 2, but the bulk of the foreign bias survives the switch to a reasoning-oriented US model. To the extent that reasoning-oriented models reduce the reliance on superficial prompting cues, the persistence of bias in such a model suggests that it is unlikely to be driven by elicitation differences.

Finally, we test whether the bias is confined to short-horizon predictions. If elicitation differences cause ChatGPT to respond more optimistically to near-term forecasting tasks, extending the prediction horizon should attenuate the effect. Panel D extends the forecast horizon to 12 months, asking both models to predict stock prices and EPS with a target date of June 30, 2025. The coefficient on *US_AI* remains positive and significant: 1.8490 (significant at the 5% level) for price predictions and 0.3195 (significant at the 1% level) for EPS predictions. The persistence of the bias at a longer horizon indicates that it is not an artifact of short-term prediction framing.

[7] In untabulated results, taking the median rather than the mean of the 50 price predictions yields very similar results, with *US_AI* remaining positive (1.4373) and significant at the 5% level.

3.6 Discussion

A fundamental challenge in interpreting our results is that LLMs remain, in important respects, black boxes. We cannot directly observe the contents of ChatGPT's or DeepSeek's training corpora, the weighting schemes applied during training, or the specific objectives used in post-training fine-tuning. Any conclusion about why the foreign bias arises must therefore be inferred indirectly from observable patterns in model outputs.

With this caveat in mind, we summarize the weight of evidence, which is largely consistent across the three outputs we study—price predictions, EPS predictions, and business descriptions. The foreign bias: (i) tracks the cross-border coverage gap, with ChatGPT growing more optimistic where negative news is concentrated in Chinese media and less optimistic—even reversing sign—where positive news is; (ii) attenuates for cross-listed firms, which have greater salience in the US and coverage by US media; (iii) disappears when both models receive Chinese news at inference time; (iv) is absent when analyzing US firms, where information environments are likely to be more symmetric; (v) is absent for synthetic firms, where neither model possesses real information; and (vi) persists across prompt language, reasoning-oriented model variants, and forecast horizons.

The first four patterns are consistent with information-centric explanations; the fifth rules out name-triggered alignment; and the sixth narrows the scope for elicitation-based mechanisms. On balance, the evidence all points to *effective* information availability/access as a first-order driver of the AI foreign-bias we document. While we cannot rule out that some residual bias reflects alignment or elicitation differences, such mechanisms appear insufficient to explain the core phenomenon.

3.7 Spillovers to Human Analysts

The analysis so far concerns the predictions of the LLMs themselves. A natural further question is whether the cross-border information gaps that bias these models also bear on the human analysts who increasingly use them. We close the empirical analysis with a test that speaks to this question.

We emphasize at the outset that it provides only suggestive evidence: we do not observe which analysts use AI tools, and our research design cannot rule out contemporaneous factors unrelated to AI adoption.

We use sell-side analyst EPS forecasts from I/B/E/S and compare the forecast optimism of US-based and China-based analysts covering Chinese firms, before and after the public release of ChatGPT in late 2022. $ANALYST\_FORECAST\_OPTIMISM$ is the analyst's EPS forecast minus realized EPS, scaled by stock price, so that higher values denote more optimistic forecasts; $POST\_CHATGPT$ indicates forecasts issued after ChatGPT's release; and $US\_ANALYSTS$ indicates analysts at US-based brokerages. The coefficient of interest is on the interaction $POST\_CHATGPT \times US\_ANALYSTS$, which captures whether US analysts became relatively more optimistic about Chinese firms once ChatGPT became available.

Table 10, Panel A, Column (1), reports a positive and significant interaction coefficient (0.0040, significant at the 1% level): relative to China-based analysts, US-based analysts' forecasts for Chinese firms became more optimistic following ChatGPT's release. Column (2) repeats the test for US firms—where we document no AI foreign bias—and finds no comparable effect (interaction coefficient 0.0026, statistically insignificant). As with the AI predictions, a differential shift in US-analyst optimism appears for Chinese firms but not for US firms.

Panel B examines whether this shift tracks the same cross-border news gaps that drive the AI bias. Across the three gap measures, the triple interaction $POST\_CHATGPT \times US\_ANALYSTS \times$ *News Gap* follows the pattern of the AI predictions: it is positive for the negative-news gap (0.024) and the net negativity gap (0.016), and negative for the positive-news gap (-0.035). US analysts became relatively more optimistic about Chinese firms whose negative news is concentrated in Chinese outlets, and relatively less optimistic about firms whose positive news is so concentrated. The post-ChatGPT change in US-analyst optimism thus moves with the news gaps—sign reversal included—in the same way as the AI foreign bias.

We interpret these results cautiously. The release of ChatGPT is a coarse proxy for AI adoption;

we do not observe individual analysts' use of LLMs, and other developments around the same period could plausibly affect US and Chinese analysts differently. The US-firm comparison in Panel A mitigates, but does not eliminate, this concern. We therefore read the evidence as consistent with—rather than dispositive of—the possibility that the AI foreign bias has begun to influence human forecasts. At a minimum, it indicates that the cross-border information gaps we document are associated with shifts in human analyst behavior along the same dimensions, a pattern we believe warrants further study.

4 Conclusion

This paper documents a novel *foreign bias* in AI-generated financial analysis: when evaluating Chinese listed firms, ChatGPT is systematically more optimistic than a China-developed frontier model (DeepSeek)—in stock price forecasts, earnings forecasts, and the tone of qualitative business descriptions alike—while being less accurate ex post on the quantitative tasks. The bias thus surfaces across the principal quantitative and qualitative tasks for which practitioners can apply LLMs in equity analysis. This pattern reverses the traditional home bias observed in human investors, suggesting that LLMs' biases need not mirror human psychology, even when trained on human-generated text.

Our evidence is most consistent with foreign bias arising from *spatial* information-set distortions. ChatGPT's excess optimism is concentrated precisely where cross-market information gaps are plausibly largest (firms with limited US coverage), attenuated where those gaps are smaller (cross-listed firms with greater English-language disclosure), and absent in settings where both models have comparably rich information (US firms). In addition, supplying firm-specific Chinese news at inference time eliminates ChatGPT's excess optimism. Because prompting can add information to the model's context but cannot alter learned parameters, this intervention strongly suggests that differences in effective information availability drive the foreign-bias.

A further implication concerns the *direction* of such biases. Our baseline finding—that Chat-

GPT is overly optimistic about Chinese firms—might suggest that foreign bias is synonymous with excess optimism. The evidence indicates otherwise. ChatGPT's relative optimism rises when US coverage disproportionately misses negative news about a firm, but it falls, and reverses, when US coverage disproportionately misses positive news. The foreign bias is therefore not a fixed, signable tendency; it is a reflection of whatever a model's training corpus happens to underrepresent. Because the composition of those corpora is generally unobservable—a known unknown—the direction of the resulting bias cannot be predicted in advance. What our two-model design makes visible is the existence and economic magnitude of such gaps; the optimistic sign we document is specific to this pairing of models and this information environment, and need not carry over to other models, markets, or tasks.

A natural question is why such biases may persist even for state-of-the-art systems. Expanding, balancing, and auditing global coverage can be costly and operationally difficult: high-quality financial information is fragmented across languages, paywalls, licensing regimes, and uneven disclosure standards, and continuous retraining or curation to close coverage gaps may compete with other engineering priorities. Consistent with our finding that enabling web browsing only partially attenuates the bias, ad hoc search may not reliably recover the most relevant local-language information or integrate it into forecasts without careful prompting. Together, these constraints imply that information-set distortions may be persistent features of frontier LLMs.

More broadly, our findings speak to AI development in an increasingly multi-polar landscape. As different jurisdictions build frontier models shaped by different informational environments, the same firm may receive systematically different assessments depending on which model is used. Rather than acting as neutral information technology, LLMs can introduce *model-specific* perspectives that interact with existing cross-border information frictions. This raises a practical concern: the diffusion of AI tools may not automatically democratize sophisticated analysis globally but could instead create new disparities based on which models and which information sources users can access and incorporate.

For practitioners, the implication is straightforward. LLM outputs in cross-border valuation and prediction tasks should be treated as potentially *systematically* skewed. Mitigating these distortions may require explicit local-information supplementation (e.g., by including local-language sources) and validation against alternative models or benchmarks. For policymakers and platform designers, our results highlight the importance of transparency around training data provenance, language coverage, and retrieval protocols, as well as the value of standardized audits that stress-test models across jurisdictions and information environments.

Finally, our analysis of analyst forecasts offers preliminary and admittedly suggestive evidence that these biases may not stay contained within AI outputs. Around the release of ChatGPT, US analysts' optimism toward Chinese firms shifted relative to that of Chinese analysts, and did so in step with the same cross-border news gaps—sign reversal included—that drive the model bias. We do not read this as proof of an AI-to-human channel, and establishing one is beyond the scope of our research design. But it raises the possibility that, as LLMs are absorbed into analyst workflows, the information-set distortions they carry may propagate into human forecasts and, ultimately, into prices. We believe this possibility merits further study.

In sum, the foreign bias we document illustrates how exogenous and often unobserved differences in informational environments can generate unexpected and economically meaningful distortions in LLM predictions. As AI tools increasingly mediate cross-border investment decisions, measuring, monitoring, and mitigating information-set distortions will be essential to maintain market efficiency and protect investors.

References

Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–23.

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni. 2025. The impact of generative AI on information processing: Evidence from the ban of ChatGPT in Italy. *Journal of Accounting and Economics* 80:101782–.

Bini, P., L. W. Cong, X. Huang, and J. Lawrence. 2025. Behavioral Economics of AI: LLM Biases and Corrections. *Working Paper* .

Blankespoor, E., E. DeHaan, and Q. Li. 2025. Generative AI in Financial Reporting. *Working Paper* 1–23.

Blodgett, S. L., S. Barocas, H. Daumé, and H. Wallach. 2020. Language (Technology) is power: A critical survey of "bias" in NLP. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 5454–76.

Bloomberg Intelligence. 2024. Generative AI 2024: Assessing Opportunities and Disruptions in an Evolving Trillion-Dollar Market. Working Paper.

Bradshaw, M. T., C. Ma, B. P. Yost, and Y. Zou. 2025. Generative AI Use by Capital Market Information Intermediaries : Evidence from Seeking Alpha. *Working Paper* .

Brennan, M. J., and A. Subrahmanyam. 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics* 38:361–81.

Bushee, B. J., J. E. Core, W. Guay, and S. J. Hamm. 2010. The role of the business press as an information intermediary. *Journal of Accounting Research* 48:1–19.

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. *Working Paper* .

Cheng, Q., P. Lin, and Y. Zhao. 2025. Does Generative AI Facilitate Investor Trading? Evidence from ChatGPT Outages Does Generative AI Facilitate Investor Trading? Evidence from ChatGPT Outages. *Singapore Management University School of Accountancy Research Paper Series* 13:101821–.

Cook, T., S. Kazinnik, A. Hansen, and P. McAdam. 2023. Evaluating Local Language Models: An Application to Bank Earnings Calls. *The Federal Reserve Bank of Kansas City Research Working Papers* .

Cooper, I., and E. Kaplanis. 1994. Home Bias in Equity Portfolios, Inflation Hedging, and International Capital Market Equilibrium. *Review of Financial Studies* 7:45–60.

Coval, J. D., and T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54:2045–73.

de Kok, T. 2025. ChatGPT for Textual Analysis? How to Use Generative LLMs in Accounting Research. *Management Science (Forthcoming)* .

Didisheim, A., M. Fraschini, and L. Somoza. 2025. Ai's predictable memory and its implications for finance. *Available at SSRN 5331177* .

Drake, M. S., N. M. Guest, and B. J. Twedt. 2014. The media and mispricing: The role of the business press in the pricing of accounting information. *Accounting Review* 89:1673–701.

Engelberg, J., A. Manela, W. Mullins, and L. Vulicevic. 2025. Entity Neutering. *Working Paper* 1–53.

French, K. R., and J. M. Poterba. 1991. Investor diversification and international equity markets. *American Economic Review* 81:222–6.

Gao, Y., M. Zhang, and H. Yang. 2023. Looking good in the eyes of stakeholders: corporate giving and corporate acquisitions. *Journal of Business Ethics* 185:375–96.

Guest, N. M. 2021. The information role of the media in earnings news. *Journal of Accounting Research* 59:1021–76.

He, S., L. Lv, A. Manela, and J. Wu. 2025. Chronologically Consistent Large Language Models. *Working Paper* .

Heath, C., and A. Tversky. 1991. Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty* 4:5–28.

Huberman, G. 2001. Familiarity breeds investment. *Review of Financial Studies* 14:659–80.

Jha, M., J. Qian, M. Weber, and B. Yang. 2024. ChatGPT and Corporate Policies. *SSRN Electronic Journal* .

Kim, A., M. Muhn, and V. Nikolaev. 2024. Financial Statement Analysis with Large Language Models .

Kim, A. G., and V. V. Nikolaev. 2025. Context-Based Interpretation of Financial Information. *Journal of Accounting Research (Forthcoming)* .

Law, K. K., and M. Shen. 2025. How Does Artificial Intelligence Shape Audit Firms? *Management Science* 71:3641–66.

Levy, B. 2025. Caution Ahead: Numerical Reasoning and Look-ahead Bias in AI Models. *Working Paper* .

Lewis, K. K. 1999. Trying to Explain Home Bias in Equities and Consumption. *Journal of Economic Literature* 37:571–608.

Li, J., M. Li, X. Wang, and J. B. Thatcher. 2021. Strategic directions for ai: The role of cios and boards of directors. *MIS quarterly* .

Liu, N. F., K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang. 2024. Lost in the Middle: How Language Models Use Long Contexts. *Transactions of the Association for Computational Linguistics* 12:157–73.

Lock, B. 2024. The impact of news media coverage on voluntary disclosure. *Contemporary Accounting Research* 41:2354–83.

Lopez-Lira, A., and Y. Tang. 2023. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *SSRN Electronic Journal* .

Loughran, T., and B. Mcdonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* 66:35–65.

Portes, R., and H. Rey. 2005. The determinants of cross-border equity flows. *Journal of International Economics* 65:269–96.

Qiao, F. 2023. Do analysts disseminate anomaly information in china? *Journal of Banking and Finance, Forthcoming* .

Sarkar, S., and K. Vafa. 2024. Lookahead Bias in Pretrained Language Models. *SSRN Electronic Journal* .

Shaffer, M., and C. C. Y. Wang. 2024. Scaling Core Earnings Measurement with Large Language Models. *Working Paper* .

Shi, F., X. Chen, K. Misra, N. Scales, D. Dohan, E. Chi, N. Schärli, and D. Zhou. 2023. Large Language Models Can Be Easily Distracted by Irrelevant Context. *Proceedings of Machine Learning Research* 202:31210–27.

Shumailov, I., Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, and R. Anderson. 2023. The Curse of Recursion: Training on Generated Data Makes Models Forget .

Tesar, L. L., and I. M. Werner. 1995. Home bias and high turnover. *Journal of International Money and Finance* 14:467–92.

Van Nieuwerburgh, S., and L. Veldkamp. 2009. Information immobility and the home bias puzzle. *Journal of Finance* 64:1187–215.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48:817–38.

Exhibit A. Prompts Used in the Analysis

This exhibit reports the exact prompts (temperature = 0) used to elicit stock price predictions and directional forecasts from large language models across the main analysis and robustness tests. Note: In the actual API calls, placeholders such as {`name`} and {`ticker`} were replaced with the specific company name and ticker symbol.

Panel A. Prompt Used for the Main Analysis (Table 2)

```
Here is a firm called {name} (ticker: {ticker}), listed in China. As a professional
financial analyst:

  1. Predict the firm's stock price on December 31, 2024 (three decimal places).
  2. Predict whether the stock price will increase from the end of June to December
     31, 2024 (1 = increase, 0 = not increase).

Do not use current online data or random numbers. Output only a JSON object, without
any additional text or explanation.
```

Panel B. Prompt Used for Closing the News Coverage Gap (Table 6)

```
Here is a firm called {name} (ticker: {ticker}), listed as an A-share in China.
Below is a collection of Chinese-language news articles related to this firm
published between June 2022 and June 2024.
[Chinese news abstracts inserted here]
Please act as a professional financial analyst and evaluate this firm based on the
information above:

  1. Predict the firm's stock price on December 31, 2024 (three decimal places).
  2. Predict whether the stock price will increase from the end of June to December
     31, 2024 (1 = increase, 0 = not increase).

Do not use current online data or random numbers. Output only a JSON object, without
any additional text or explanation.
```

Panel C. Prompt Used for U.S. Firms (Table 7)

Here is a firm called {name} (ticker: {ticker}), listed in the United States. Please act as a professional financial analyst and evaluate this firm based on the information above:

1. Predict the firm's stock price on December 31, 2024 (three decimal places).
2. Predict whether the stock price will increase from the end of June to December 31, 2024 (1 = increase, 0 = not increase).

Do not use current online data or random numbers. Output only a JSON object, without any additional text or explanation.

Panel D. Prompt Used for Mandarin-Language Elicitation Test (Table 9, Panel A)

以下是一家名为{name}（股票代码：{ticker}）的公司，在中国A 股市场上市。请你作为一名专业的金融分析师，基于你已有的知识对该公司进行评估：

1. 预测该公司在2024 年12 月31 日的股票价格（保留三位小数）。
2. 判断该公司股票价格是否会从2024 年6 月底上涨至2024 年12 月31 日（1 = 上涨，0 = 不上涨）。

请勿使用当前的在线数据或编造数据。仅输出一个JSON 对象，不要包含任何额外的文字或解释。

Panel E. Prompt Used for Alternative Prediction-Horizon Test (Table 9, Panel D)

Here is a firm called {name} (ticker: {ticker}), listed as an A-share in China. Please act as a professional financial analyst and evaluate this firm based on your existing knowledge:

1. Predict the firm's stock price on June 30, 2025 (three decimal places).
2. Indicate whether the stock price will increase from the end of June 2024 to June 30, 2025 (1 = increase, 0 = not increase).

Do not use current online data or perform any external search. Return only a JSON object, without any additional text or explanation.

Appendix A Description of Variables

This appendix defines the variables used in our analyses. Variable definitions are based on data obtained from ChatGPT 4.1, DeepSeek R1, RavenPack News Analytics, the China Finance News Database (CFND), Compustat, CSMAR, and I/B/E/S, as described in Section 2.

Variable	Description	Source
Table 1 and Baseline Variables		
STOCK_PRICE	Six-month-ahead stock price prediction generated by an AI model for a Chinese firm as of December 31, 2024.	ChatGPT 4.1, DeepSeek R1
EPS	Six-month-ahead EPS prediction generated by an AI model for a Chinese firm as of December 31, 2024.	ChatGPT 4.1, DeepSeek R1
BUSI_SENTIMENT	Normalized sentiment score of the AI-generated business description for a firm. For each firm, the AI model is prompted to generate a 150–200-word business description, and sentiment is measured using the Loughran and McDonald financial dictionary. Higher values indicate more positive sentiment.	ChatGPT 4.1, DeepSeek R1; Loughran and McDonald dictionary
US_AI	Indicator equal to one for observations generated by ChatGPT and zero for observations generated by DeepSeek.	ChatGPT 4.1, DeepSeek R1
NEG_NEWS_GAP	Difference between negative Chinese-media coverage and negative U.S.-media coverage, scaled by their sum, measured from June 30, 2022 to June 30, 2024. It is set to zero when the measure is missing. Higher values indicate that negative firm-level news is more concentrated in Chinese media than in U.S. media.	RavenPack News Analytics, CFND
POS_NEWS_GAP	Difference between positive Chinese-media coverage and positive U.S.-media coverage, scaled by their sum, measured from June 30, 2022 to June 30, 2024. It is set to zero when the measure is missing. Higher values indicate that positive firm-level news is more concentrated in Chinese media than in U.S. media.	RavenPack News Analytics, CFND
NET_NEG_NEWS_GAP	Net negativity gap between Chinese and U.S. media, defined as the negative news gap minus the positive news gap. Higher values indicate that unfavorable news is relatively more concentrated in Chinese media than in U.S. media.	RavenPack News Analytics, CFND
FIRM_SIZE	Natural logarithm of total assets measured before the AI prediction date.	Compustat
ROA	Return on assets, calculated as income before extraordinary items divided by total assets.	Compustat
LEVERAGE	Total debt divided by total assets.	Compustat
LOSS	Indicator equal to one if the firm reports negative net income, and zero otherwise.	Compustat
BM	Book-to-market ratio, calculated as book value of equity divided by market value of equity.	Compustat
PRE_RETURN	Stock return prior to the AI prediction period.	Compustat / CSMAR
RETURN_VOLATILITY	Standard deviation of daily stock returns prior to the AI prediction period.	Compustat / CSMAR
IO	Percentage of shares held by institutional investors.	CSMAR
Table 3: Prediction Error Variables		

Continued on next page

Appendix A continued

Variable	Description	Source
STOCK_PRICE_ERROR	Absolute stock price prediction error, calculated as the absolute difference between the AI-predicted stock price and the realized stock price on December 31, 2024.	ChatGPT 4.1, DeepSeek R1, Compustat
EPS_ERROR	Absolute EPS prediction error, calculated using the AI-predicted EPS and realized EPS.	ChatGPT 4.1, DeepSeek R1, Compustat
Table 5: Cross-Listing Variable		
CROSS_LISTED	Indicator equal to one if the Chinese firm is cross-listed on a U.S. exchange, and zero otherwise.	Compustat / CSMAR
Table 6: News-Intervention Variables		
STOCK_PRICE_INJECTED	Six-month-ahead stock price prediction generated after the AI model is provided with Chinese-language firm news in the prompt.	ChatGPT 4.1, DeepSeek R1, CFND
EPS_INJECTED	Six-month-ahead EPS prediction generated after the AI model is provided with Chinese-language firm news in the prompt.	ChatGPT 4.1, DeepSeek R1, CFND
BUSI_SENTIMENT_INJECTED	Sentiment score of the AI-generated business description after the AI model is provided with Chinese-language firm news in the prompt.	ChatGPT 4.1, DeepSeek R1, CFND
NEWS_INJECTED	Indicator equal to one if the prediction is generated from a news-augmented prompt that provides available Chinese-language news summaries for the firm from June 30, 2022 to June 30, 2024, and zero for baseline-prompt observations.	ChatGPT 4.1, DeepSeek R1, CFND
STOCK_PRICE_STACKED	Stock price prediction in the stacked sample combining baseline and news-augmented predictions.	ChatGPT 4.1, DeepSeek R1, CFND
EPS_STACKED	EPS prediction in the stacked sample combining baseline and news-augmented predictions.	ChatGPT 4.1, DeepSeek R1, CFND
BUSI_SENTIMENT_STACKED	Business-description sentiment score in the stacked sample combining baseline and news-augmented predictions.	ChatGPT 4.1, DeepSeek R1, CFND
Table 7: U.S.-Firm Placebo Variables		
STOCK_PRICE_US	Six-month-ahead stock price prediction generated by an AI model for a U.S. firm as of December 31, 2024.	ChatGPT 4.1, DeepSeek R1
EPS_US	Six-month-ahead EPS prediction generated by an AI model for a U.S. firm as of December 31, 2024.	ChatGPT 4.1, DeepSeek R1
BUSI_SENTIMENT_US	Sentiment score of the AI-generated business description for a U.S. firm.	ChatGPT 4.1, DeepSeek R1
Table 8: Synthetic-Firm Placebo Variables		
STOCK_PRICE_SYNTHETIC	Stock price prediction generated for a synthetic Chinese-sounding firm name with no corresponding real-world entity.	ChatGPT 4.1, DeepSeek R1
EPS_SYNTHETIC	EPS prediction generated for a synthetic Chinese-sounding firm name with no corresponding real-world entity.	ChatGPT 4.1, DeepSeek R1
BUSI_SENTIMENT_SYNTHETIC	Sentiment score of the AI-generated business description for a synthetic Chinese-sounding firm name with no corresponding real-world entity.	ChatGPT 4.1, DeepSeek R1
Table 9: Robustness-Test Variables		
STOCK_PRICE_Mandarin	Six-month-ahead stock price prediction for a Chinese firm generated using a Mandarin prompt.	ChatGPT 4.1, DeepSeek R1
EPS_Mandarin	Six-month-ahead EPS prediction for a Chinese firm generated using a Mandarin prompt.	ChatGPT 4.1, DeepSeek R1

Continued on next page

Appendix A continued

Variable	Description	Source
BUSI_SENTIMENT_Mandarin	Sentiment score of the AI-generated business description for a Chinese firm generated using a Mandarin prompt.	ChatGPT 4.1, DeepSeek R1
STOCK_PRICE_Web	Six-month-ahead stock price prediction for a Chinese firm generated when online search is enabled.	ChatGPT 4.1, DeepSeek R1
EPS_Web	Six-month-ahead EPS prediction for a Chinese firm generated when online search is enabled.	ChatGPT 4.1, DeepSeek R1
BUSI_SENTIMENT_Web	Sentiment score of the AI-generated business description for a Chinese firm generated when online search is enabled.	ChatGPT 4.1, DeepSeek R1
STOCK_PRICE_ChatGPT5	Stock price prediction for a Chinese firm using ChatGPT-5-based predictions for the U.S.-AI observations. Because ChatGPT-5 does not allow the temperature parameter to be set to zero, the continuous outcome is constructed from repeated prompts following the paper's ChatGPT-5 procedure.	ChatGPT-5, DeepSeek R1
EPS_ChatGPT5	EPS prediction for a Chinese firm using ChatGPT-5-based predictions for the U.S.-AI observations.	ChatGPT-5, DeepSeek R1
BUSI_SENTIMENT_ChatGPT5	Business-description sentiment score using ChatGPT-5-based predictions for the U.S.-AI observations.	ChatGPT-5, DeepSeek R1
STOCK_PRICE_ALT	Stock price prediction for a Chinese firm using the alternative prediction horizon.	ChatGPT 4.1, DeepSeek R1
EPS_ALT	EPS prediction for a Chinese firm using the alternative prediction horizon.	ChatGPT 4.1, DeepSeek R1
Table 10: Analyst-Level Variables		
ANALYST_FORECAST_OPTIMISM	Analyst forecast optimism, defined as the analyst's EPS forecast minus actual EPS, scaled by the stock price as of February 7, 2023. Higher values indicate more optimistic analyst forecasts.	I/B/E/S, Compustat
POST_CHATGPT	Indicator equal to one for analyst forecasts issued after the release of ChatGPT, and zero for forecasts issued before the release.	I/B/E/S
US_ANALYSTS	Indicator equal to one for analysts working at brokerage houses located in the United States, and zero for analysts working at brokerage houses located in China.	I/B/E/S
News Gap Proxy	Generic notation used in Table 10 for one of the three news gap measures: *NEG_NEWS_GAP*, *POS_NEWS_GAP*, or *NET_NEG_NEWS_GAP*.	RavenPack News Analytics, CFND
FIRM_EXPERIENCE	Analyst's firm-specific experience, measured based on the analyst's prior forecasting history for the same firm.	I/B/E/S
GENERAL_EXPERIENCE	Analyst's general forecasting experience, measured based on the analyst's prior forecasting history across firms.	I/B/E/S
ANALYST_BUSY	Analyst busyness, measured based on the analyst's forecasting workload.	I/B/E/S

Table 1. Summary Statistics

This table reports descriptive statistics for the main variables in our sample of Chinese firms used in the baseline analysis. Panel A provides the distribution of the primary dependent and independent variables across 9,956 firm–AI observations, including AI stock price predictions, EPS predictions, business-description sentiment, news gap measures, and key firm characteristics. Panel B reports means and standard errors of news coverage for Chinese listed firms in the sample, comparing coverage from Chinese media sources (B.1) and US media sources (B.2). We report four measures: total news articles (*Total News*), articles with negative news (*Negative News*), articles with positive news (*Positive News*), and the difference between negative and positive articles (*Net Negative News*). Panel B.3 reports the difference between US and Chinese source coverage, with significance stars indicating whether the gap is statistically different from zero. Panel C reports the industry distribution of firms in our sample by Fama-French 12 industry classification and the average news gap measures. All variables are defined in Appendix A and all continuous variables are winsorized at the 1st and 99th percentiles.

Panel A: Distribution of Main Variables

Variable	*N*	*Mean*	*SD*	*p10*	*p25*	*p50*	*p75*	*p90*
STOCK_PRICE	9,956	20.76	16.07	5.33	8.32	15.33	28.46	45.33
EPS	9,956	0.97	1.42	0.15	0.42	0.85	1.25	1.85
BUSI_SENTIMENT	9,956	0.53	0.26	0.20	0.33	0.56	0.71	0.82
US_AI	9,956	0.50	0.50	0.00	0.00	0.50	1.00	1.00
NEG_NEWS_GAP	9,956	0.78	0.45	0.00	1.00	1.00	1.00	1.00
POS_NEWS_GAP	9,956	0.39	0.58	0.00	0.00	0.00	1.00	1.00
NET_NEG_NEWS_GAP	9,956	0.35	0.59	-0.38	-0.02	0.27	1.00	1.00
FIRM_SIZE	9,956	8.60	1.40	7.13	7.62	8.29	9.30	10.50
ROA	9,956	0.02	0.06	-0.05	0.00	0.03	0.05	0.09
LEVERAGE	9,956	0.42	0.22	0.13	0.24	0.40	0.57	0.71
LOSS	9,956	0.21	0.41	0.00	0.00	0.00	0.00	1.00
BM	9,956	0.57	0.46	0.18	0.29	0.45	0.70	1.02
PRE_RETURN	9,956	0.03	0.32	-0.32	-0.17	0.00	0.17	0.39
RETURN_VOLATILITY	9,956	2.36	0.88	1.36	1.71	2.19	2.88	3.59
IO	9,956	41.11	25.12	7.28	19.63	40.43	61.78	75.79

Panel B: News Coverage by Media Source

	Total News	*Negative News*	*Positive News*	*Net Negative News*
B.1: Chinese Sources				
Mean	26.9339	14.5882	12.3457	2.2425
	(0.6284)	(0.3348)	(0.4082)	(0.4032)
B.2: US Sources				
Mean	1.6826	0.7921	0.8905	-0.0984
	(0.0748)	(0.0415)	(0.0406)	(0.0339)
B.3: Difference (US − Chinese)				
Difference	-25.2513***	-13.7961***	-11.4552***	-2.3409***
	(0.6216)	(0.3334)	(0.4055)	(0.4059)

Table 1. Summary Statistics [Continued]

Panel C: Average News Gap by Fama-French 12-industry classification

Industry	*Number of Firms*	*%*	*NEG_NEWS_GAP*	*POS_NEWS_GAP*	*NET_NEG_NEWS_GAP*
Consumer Nondurables	385	7.73	0.8506	0.5178	0.3077
Consumer Durables	337	6.77	0.7820	0.4029	0.3611
Manufacturing	1,239	24.89	0.7900	0.3924	0.3628
Energy	62	1.25	0.6762	0.5165	0.1602
Chemicals	441	8.86	0.8122	0.4283	0.3473
Business Equipment	1,014	20.37	0.8035	0.2914	0.4459
Telecommunications	26	0.52	0.8385	0.5087	0.3449
Utilities	88	1.77	0.8353	0.4755	0.3321
Wholesale and Retail	185	3.72	0.7921	0.4676	0.3355
Healthcare	420	8.44	0.8294	0.4654	0.3134
Financials	218	4.38	0.2618	0.2053	0.0896
Others	563	11.31	0.7747	0.4232	0.3304

Table 2. ChatGPT vs. DeepSeek: EPS, Price Predictions, and Business Description Sentiment

This table presents baseline regressions comparing ChatGPT and DeepSeek in six-month-ahead stock price predictions, six-month-ahead EPS predictions, and business description sentiment for Chinese firms. The dependent variables are *STOCK_PRICE* in Column (1), *EPS* in Column (2), and *BUSI_SENTIMENT* in Column (3). The key independent variable, *US_AI*, equals one for observations generated by ChatGPT and zero for those generated by DeepSeek. *BUSI_SENTIMENT* is the normalized sentiment score of the AI-generated business description for a firm, with higher values indicating more positive sentiment. Control variables include firm size, profitability, leverage, loss indicator, book-to-market ratio, prior returns, return volatility, and institutional ownership. Industry fixed effects are included in all columns. The coefficients on *US_AI* capture whether ChatGPT produces systematically different outputs relative to DeepSeek. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)
Dep Var =	*STOCK_PRICE*	*EPS*	*BUSI_SENTIMENT*
US_AI	2.4153***	0.3993***	0.1115***
	(0.722)	(0.026)	(0.009)
FIRM_SIZE	0.7579	0.2435*	0.0119**
	(0.788)	(0.124)	(0.005)
ROA	33.7108***	4.6429***	0.1790**
	(7.558)	(1.040)	(0.078)
LEVERAGE	-15.9897***	-0.8200**	-0.1179***
	(3.649)	(0.360)	(0.015)
LOSS	0.6487	0.2576	-0.0229***
	(1.013)	(0.162)	(0.007)
BM	-4.1948***	-0.3554	-0.0376**
	(1.216)	(0.209)	(0.016)
PRE_RETURN	-8.6622***	-0.4008***	-0.0300**
	(0.951)	(0.050)	(0.013)
RETURN_VOLATILITY	4.9367***	0.1786***	0.0134
	(0.778)	(0.043)	(0.008)
IO	0.0702*	0.0010	0.0006***
	(0.033)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.3071	0.1199	0.1385

Table 3. Prediction Errors

This table compares the accuracy of ChatGPT and DeepSeek predictions against realized outcomes. Column (1) reports absolute stock price prediction errors, and Column (2) reports absolute EPS prediction errors. The key independent variable, *US_AI*, equals one for observations generated by ChatGPT and zero for those generated by DeepSeek. Positive coefficients on *US_AI* indicate that ChatGPT produces larger prediction errors than DeepSeek. Control variables include firm size, profitability, leverage, loss indicator, book-to-market ratio, prior returns, return volatility, and institutional ownership. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

	(1)	(2)
Dep Var =	*STOCK_PRICE_ERROR*	*EPS_ERROR*
US_AI	1.0863***	1.8116***
	(0.318)	(0.208)
FIRM_SIZE	0.3676	-0.3067
	(0.454)	(0.195)
ROA	24.6574***	-17.3425***
	(5.694)	(3.024)
LEVERAGE	-6.2347***	-2.5732*
	(1.926)	(1.432)
LOSS	1.8352**	2.2679*
	(0.657)	(1.130)
BM	-2.8381**	-0.9450***
	(0.946)	(0.170)
PRE_RETURN	-1.3305	2.3865**
	(0.815)	(0.861)
RETURN_VOLATILITY	2.6255***	-0.8394***
	(0.566)	(0.133)
IO	0.0462*	-0.0050
	(0.022)	(0.006)
Industry FE	Yes	Yes
Observations	9,956	9,956
Adj R^2	0.1487	0.0239

Table 4. Role of Media Coverage in AI Prediction: Gap in News Coverage

This table examines whether cross-border media coverage asymmetries explain the differences between ChatGPT and DeepSeek in AI-generated stock price predictions, EPS predictions, and business-description sentiment for Chinese firms. The dependent variables are *STOCK_PRICE*, *EPS*, and *BUSI_SENTIMENT*. *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for observations generated by DeepSeek. Panel A examines *NET_NEG_NEWS_GAP*, defined as the negative news gap minus the positive news gap. Panel B separately examines negative and positive news gaps. *NEG_NEWS_GAP* captures the extent to which negative firm-level news is more concentrated in Chinese media than in U.S. media, while *POS_NEWS_GAP* is constructed analogously using positive news coverage. Positive coefficients on *US_AI* × *NEG_NEWS_GAP* and *US_AI* × *NET_NEG_NEWS_GAP* indicate that ChatGPT's relative optimism is stronger when negative news is relatively more concentrated in Chinese media than in U.S. media. Negative coefficients on *US_AI* × *POS_NEWS_GAP* indicate that ChatGPT's relative optimism is weaker when positive news is relatively more concentrated in Chinese media than in U.S. media. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

Panel A: Negative News minus Positive News Gap

	(1) *STOCK_PRICE*	(2) *EPS*	(3) *BUSI_SENTIMENT*
US_AI×NET_NEG_NEWS_GAP	5.5227***	0.1885***	0.0172***
	(0.971)	(0.019)	(0.005)
NET_NEG_NEWS_GAP	-0.2494	0.0367**	0.0069
	(0.391)	(0.013)	(0.005)
US_AI	0.4824	0.3334***	0.1054***
	(0.343)	(0.024)	(0.010)
FIRM_SIZE	1.0361	0.2580*	0.0136**
	(0.801)	(0.125)	(0.005)
ROA	33.6656***	4.6405***	0.1787**
	(7.523)	(1.042)	(0.077)
LEVERAGE	-15.7076***	-0.8053**	-0.1162***
	(3.650)	(0.357)	(0.015)
LOSS	0.5157	0.2507	-0.0238***
	(0.984)	(0.160)	(0.007)
BM	-4.2135***	-0.3563	-0.0377**
	(1.184)	(0.208)	(0.016)
PRE_RETURN	-8.5130***	-0.3930***	-0.0291**
	(0.915)	(0.048)	(0.013)
RETURN_VOLATILITY	4.8234***	0.1726***	0.0127
	(0.762)	(0.042)	(0.008)
IO	0.0686*	0.0009	0.0006***
	(0.032)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.3248	0.1239	0.1398

Table 4. Role of Media Coverage in AI Prediction: Gap in News Coverage – Continued

Panel B: Negative News Gap and Positive News Gap

	(1)	(2)	(3)	(4)	(5)	(6)
	STOCK_PRICE		*EPS*		*BUSI_SENTIMENT*	
US_AI × *NEG_NEWS_GAP*	2.7121***		0.1234***		0.0199	
	(0.670)		(0.035)		(0.017)	
US_AI×*POS_NEWS_GAP*		-5.1233***		-0.1690***		-0.0086
		(1.004)		(0.025)		(0.008)
NEG_NEWS_GAP	-0.5303		-0.0162		0.0064	
	(0.632)		(0.048)		(0.013)	
POS_NEWS_GAP		-0.4152		-0.0648*		-0.0070
		(0.544)		(0.034)		(0.009)
US_AI	0.3079	4.4303***	0.3035***	0.4658***	0.0960***	0.1149***
	(0.225)	(0.977)	(0.039)	(0.031)	(0.021)	(0.011)
FIRM_SIZE	0.7959	0.8276	0.2456*	0.2470*	0.0127**	0.0122**
	(0.797)	(0.766)	(0.126)	(0.123)	(0.005)	(0.005)
ROA	33.1979***	36.4117***	4.6146***	4.7783***	0.1688*	0.1893**
	(7.573)	(6.940)	(1.024)	(1.026)	(0.078)	(0.077)
LEVERAGE	-16.0780***	-15.1213***	-0.8249**	-0.7764*	-0.1197***	-0.1146***
	(3.703)	(3.477)	(0.363)	(0.364)	(0.015)	(0.015)
LOSS	0.6011	0.8260	0.2550	0.2665	-0.0239***	-0.0223***
	(0.988)	(0.938)	(0.160)	(0.159)	(0.007)	(0.007)
BM	-4.2290***	-3.9274***	-0.3572	-0.3420	-0.0383**	-0.0366**
	(1.239)	(1.193)	(0.210)	(0.211)	(0.016)	(0.016)
PRE_RETURN	-8.6093***	-8.6719***	-0.3979***	-0.4013***	-0.0289*	-0.0300**
	(0.933)	(0.876)	(0.051)	(0.045)	(0.013)	(0.013)
RETURN_VOLATILITY	4.9214***	4.8237***	0.1777***	0.1729***	0.0131	0.0129
	(0.780)	(0.754)	(0.043)	(0.041)	(0.008)	(0.008)
IO	0.0690*	0.0731**	0.0009	0.0011	0.0006**	0.0006***
	(0.033)	(0.032)	(0.001)	(0.001)	(0.000)	(0.000)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,956	9,956	9,956	9,956	9,956	9,956
Adj R^2	0.3089	0.3266	0.1203	0.1244	0.1394	0.1390

Table 5. Cross-Listing Firms

This table examines whether the differences between ChatGPT and DeepSeek are less pronounced for Chinese firms that are cross-listed on U.S. exchanges. The dependent variables are *STOCK_PRICE*, *EPS*, and *BUSI_SENTIMENT*. *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for those generated by DeepSeek. *CROSS_LISTED* is an indicator equal to one if the firm is cross-listed on a U.S. exchange, and zero otherwise. The coefficient on *US_AI* × *CROSS_LISTED* captures whether ChatGPT's relative optimism, compared with DeepSeek, is attenuated for cross-listed firms. A negative coefficient on this interaction term indicates that the gap between ChatGPT and DeepSeek is smaller for firms with greater exposure to the U.S. information environment. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

	(1) *STOCK_PRICE*	(2) *EPS*	(3) *BUSI_SENTIMENT*
US_AI	2.4467***	0.4077***	0.1116***
	(0.716)	(0.024)	(0.009)
US_AI×*CROSS_LISTED*	-0.7660***	-0.2242***	-0.0033*
	(0.046)	(0.002)	(0.002)
CROSS_LISTED	0.5600***	0.2060***	0.0012
	(0.046)	(0.002)	(0.001)
FIRM_SIZE	0.7471	0.2401*	0.0119**
	(0.784)	(0.124)	(0.005)
ROA	33.6858***	4.5656***	0.1810**
	(7.636)	(1.057)	(0.079)
LEVERAGE	-15.9575***	-0.8119**	-0.1177***
	(3.644)	(0.355)	(0.015)
LOSS	0.6376	0.2472	-0.0228**
	(1.017)	(0.163)	(0.007)
BM	-4.1986***	-0.3650	-0.0374**
	(1.201)	(0.207)	(0.016)
PRE_RETURN	-8.6770***	-0.4045***	-0.0301**
	(0.949)	(0.049)	(0.013)
RETURN_VOLATILITY	4.9266***	0.1739***	0.0134
	(0.777)	(0.042)	(0.008)
IO	0.0704*	0.0010	0.0006***
	(0.033)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,954	9,956
Adj R^2	0.3085	0.1413	0.1385

Table 6. Closing the News Gap

This table reports results from an information intervention that augments the prompts with Chinese-language firm news. In Panel A, both ChatGPT and DeepSeek are provided with Chinese-language news summaries for each firm before generating predictions. The dependent variables are *STOCK_PRICE_INJECTED* in Column (1), *EPS_INJECTED* in Column (2), and *BUSI_SENTIMENT_INJECTED* in Column (3). *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for observations generated by DeepSeek. Panel B compares the baseline predictions with the news-augmented predictions using a stacked sample. The dependent variables are *STOCK_PRICE_STACKED* in Column (1), *EPS_STACKED* in Column (2), and *BUSI_SENTIMENT_STACKED* in Column (3). *NEWS_INJECTED* is an indicator equal to one for observations generated from news-augmented prompts and zero for baseline-prompt observations. The interaction term *US_AI* × *NEWS_INJECTED* captures whether augmenting prompts with Chinese-language news changes ChatGPT's outputs relative to DeepSeek's outputs. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

Panel A: News-Augmented Predictions

	(1) *STOCK_PRICE_INJECTED*	(2) *EPS_INJECTED*	(3) *BUSI_SENTIMENT_INJECTED*
US_AI	-0.7980	0.0241	0.0109
	(0.770)	(0.023)	(0.010)
FIRM_SIZE	0.6187	0.2443*	0.0128**
	(0.808)	(0.126)	(0.005)
ROA	33.5798***	4.7130***	0.1896**
	(8.068)	(1.020)	(0.081)
LEVERAGE	-16.5883***	-0.8060**	-0.1191***
	(3.731)	(0.362)	(0.014)
LOSS	0.4735	0.2633	-0.0242***
	(1.037)	(0.164)	(0.007)
BM	-4.0468***	-0.3644	-0.0391**
	(1.211)	(0.211)	(0.016)
PRE_RETURN	-9.0368***	-0.4042***	-0.0311**
	(0.986)	(0.053)	(0.014)
RETURN_VOLATILITY	5.1877***	0.1770***	0.0143
	(0.818)	(0.042)	(0.008)
IO	0.0752*	0.0009	0.0005**
	(0.035)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.3060	0.1012	0.0976

Table 6 [Continued]

Panel B: Comparing Baseline and News-Augmented Predictions

	(1) *STOCK_PRICE _STACKED*	(2) *EPS _STACKED*	(3) *BUSI_SENTIMENT _STACKED*
US_AI×NEWS_INJECTED	-3.2136***	-0.3754***	-0.3958**
	(0.062)	(0.004)	(0.160)
US_AI	2.4151***	0.3994***	0.1113***
	(0.722)	(0.026)	(0.009)
NEWS_INJECTED	0.0023	0.0001	-0.0002
	(0.002)	(0.000)	(0.000)
FIRM_SIZE	0.6884	0.2439*	0.0119**
	(0.797)	(0.125)	(0.005)
ROA	33.6489***	4.6780***	0.1794**
	(7.811)	(1.029)	(0.078)
LEVERAGE	-16.2887***	-0.8130**	-0.1179***
	(3.688)	(0.360)	(0.015)
LOSS	0.5618	0.2605	-0.0229***
	(1.025)	(0.163)	(0.007)
BM	-4.1208***	-0.3599	-0.0376**
	(1.213)	(0.210)	(0.016)
PRE_RETURN	-8.8492***	-0.4025***	-0.0300**
	(0.968)	(0.051)	(0.013)
RETURN_VOLATILITY	5.0622***	0.1778***	0.0134
	(0.797)	(0.042)	(0.008)
IO	0.0727*	0.0009	0.0006***
	(0.034)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	19,912	19,912	19,912
Adj. R^2	0.3087	0.1152	0.2864

Table 7. AI Predictions of U.S. Firms

This table reports a placebo test using U.S. firms. The dependent variables are *STOCK_PRICE_US* in Column (1), *EPS_US* in Column (2), and *BUSI_SENTIMENT_US* in Column (3). *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for observations generated by DeepSeek. This analysis examines whether the differences between ChatGPT and DeepSeek documented for Chinese firms also appear in a setting where firm-specific information is more readily available in the U.S. information environment. An insignificant coefficient on *US_AI* indicates that ChatGPT does not produce systematically different stock price predictions, EPS predictions, or business-description sentiment than DeepSeek for U.S. firms. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)
	STOCK_PRICE_US	*EPS_US*	*BUSI_SENTIMENT_US*
US_AI	-0.1539	-0.1046	0.0291
	(0.361)	(0.178)	(0.027)
FIRM_SIZE	10.8241***	0.6538***	0.0172***
	(2.254)	(0.173)	(0.002)
ROA	-3.2749***	-0.2532***	-0.0050
	(0.674)	(0.038)	(0.003)
LEVERAGE	-0.0487	0.0092	-0.0039
	(0.183)	(0.016)	(0.003)
LOSS	-31.9823***	-3.5062***	-0.0718***
	(5.347)	(0.751)	(0.016)
BM	-4.6968***	-0.0926	-0.0073**
	(0.931)	(0.060)	(0.003)
PRE_RETURN	0.7273***	0.0424	0.0004
	(0.184)	(0.027)	(0.002)
RETURN_VOLATILITY	-0.0857	-0.0090	-0.0032**
	(0.087)	(0.033)	(0.001)
IO	12.9281	-0.5755	0.0270***
	(11.715)	(0.587)	(0.007)
Industry FE	Yes	Yes	Yes
Observations	11,688	11,688	11,688
Adj R^2	0.2740	0.0375	0.0980

Table 8. Synthetic Firms

This table reports placebo tests using synthetic Chinese firm names. The dependent variables are *STOCK_PRICE_SYNTHETIC* in Column (1), *EPS_SYNTHETIC* in Column (2), and *BUSI_SENTIMENT_SYNTHETIC* in Column (3). *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for those generated by DeepSeek. The sample consists of fictitious firms with Chinese-sounding names but no corresponding real-world entities or firm-specific information. This design allows us to examine whether the differences between ChatGPT and DeepSeek are driven by elicitation effects triggered solely by Chinese-sounding firm names, rather than by underlying information asymmetries. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

	(1) *STOCK_PRICE _SYNTHETIC*	(2) *EPS _SYNTHETIC*	(3) *BUSI_SENTIMENT _SYNTHETIC*
US_AI	-0.2260	0.0816	0.0016
	(0.275)	(0.080)	(0.022)
FIRM_SIZE	-0.5009	0.2829	0.0091
	(0.801)	(0.195)	(0.030)
ROA	-47.7204*	11.3975	0.1309
	(22.845)	(7.894)	(0.523)
LEVERAGE	0.3245	-1.4124*	0.0006
	(6.373)	(0.671)	(0.105)
LOSS	-9.1091**	0.3875	0.0042
	(3.396)	(0.380)	(0.028)
BM	0.1858	-0.1184	-0.0504
	(1.098)	(0.345)	(0.048)
PRE_RETURN	-4.4750	-0.3182	-0.0486
	(3.371)	(0.282)	(0.076)
RETURN_VOLATILITY	2.5929**	0.1123	0.0076
	(0.863)	(0.115)	(0.025)
IO	0.0567	-0.0144	0.0021**
	(0.036)	(0.016)	(0.001)
Industry FE	Yes	Yes	Yes
Observations	200	200	200
Adj R^2	0.1773	0.1742	0.1246

Table 9. Testing the Elicitation Channel

This table presents a series of robustness analyses to examine whether the documented foreign bias is driven by differences in how ChatGPT and DeepSeek respond to prompts. Panel A repeats the baseline analysis using prompts written in Mandarin. Panel B allows online search for both models before generating predictions. Panel C repeats the baseline analysis using ChatGPT-5. Panel D extends the prediction horizon to an alternative horizon. The dependent variables are *STOCK_PRICE*, *EPS*, and *BUSI_SENTIMENT* in Panels A–C, and *STOCK_PRICE_ALT* and *EPS_ALT* in Panel D. *US_AI* is an indicator equal to one for observations generated by ChatGPT and zero for those generated by DeepSeek. In Panel C, because ChatGPT-5 does not allow the temperature parameter to be set to zero, the ChatGPT-5-based continuous outcomes are constructed from 50 repeated prompts following the paper's ChatGPT-5 procedure. Industry fixed effects are included in all columns. All variables are defined in Appendix A, and all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors, clustered at the Fama-French 12-industry level, are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

Panel A: AI Prediction with Prompt in Mandarin

	(1) *STOCK_PRICE _Mandarin*	(2) *EPS _Mandarin*	(3) *BUSI_SENTIMENT _Mandarin*
US_AI	1.5173**	0.2479***	0.0785***
	(0.644)	(0.026)	(0.008)
FIRM_SIZE	3.8150	0.2423*	0.0129*
	(2.549)	(0.124)	(0.007)
ROA	60.9911**	4.5934***	0.1255
	(23.239)	(1.033)	(0.093)
LEVERAGE	-22.6838**	-0.8198**	-0.1163***
	(7.758)	(0.358)	(0.019)
LOSS	5.5226	0.2522	-0.0282**
	(3.226)	(0.161)	(0.010)
BM	-7.3221	-0.3548	-0.0402**
	(4.620)	(0.208)	(0.017)
PRE_RETURN	-7.4429***	-0.4049***	-0.0295*
	(1.305)	(0.051)	(0.015)
RETURN_VOLATILITY	4.0061***	0.1781***	0.0079
	(0.928)	(0.043)	(0.009)
IO	0.0553	0.0010	0.0006***
	(0.039)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.0763	0.1073	0.0825

Table 9 [Continued]

Panel B: AI Prediction Allowing Online Search

	(1) STOCK_PRICE _Web	(2) EPS _Web	(3) BUSI_SENTIMENT _Web
US_AI	0.5338**	0.1019***	0.0626***
	(0.228)	(0.026)	(0.009)
FIRM_SIZE	0.9237***	0.2443*	0.0125**
	(0.151)	(0.125)	(0.005)
ROA	2.0662	4.7132***	0.1797**
	(2.513)	(1.032)	(0.080)
LEVERAGE	-0.4815	-0.8282**	-0.1309***
	(0.455)	(0.359)	(0.014)
LOSS	0.9130***	0.2672	-0.0224**
	(0.249)	(0.162)	(0.007)
BM	-1.3104***	-0.3483	-0.0346*
	(0.321)	(0.212)	(0.017)
PRE_RETURN	0.2101	-0.4009***	-0.0247
	(0.357)	(0.048)	(0.016)
RETURN_VOLATILITY	-0.5043**	0.1819***	0.0126
	(0.164)	(0.044)	(0.009)
IO	-0.0077*	0.0009	0.0007***
	(0.004)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.0186	0.1029	0.0922

Table 9 [Continued]

Panel C: AI Prediction Using ChatGPT-5

	(1) *STOCK_PRICE _ChatGPT5*	(2) *EPS _ChatGPT5*	(3) *BUSI_SENTIMENT _ChatGPT5*
US_AI	1.6622**	0.1786***	0.0535***
	(0.704)	(0.027)	(0.009)
FIRM_SIZE	3.5656	0.2429*	0.0124**
	(2.135)	(0.125)	(0.005)
ROA	53.6364**	4.6523***	0.1557*
	(19.005)	(1.033)	(0.081)
LEVERAGE	-21.8180***	-0.8259**	-0.1204***
	(6.630)	(0.358)	(0.015)
LOSS	3.9178	0.2578	-0.0263***
	(2.679)	(0.162)	(0.008)
BM	-8.6852**	-0.3539	-0.0379**
	(3.801)	(0.209)	(0.016)
PRE_RETURN	-7.4982***	-0.4048***	-0.0305**
	(0.958)	(0.052)	(0.013)
RETURN_VOLATILITY	4.1449***	0.1789***	0.0138
	(0.820)	(0.043)	(0.008)
IO	0.0502	0.0010	0.0006***
	(0.030)	(0.001)	(0.000)
Industry FE	Yes	Yes	Yes
Observations	9,956	9,956	9,956
Adj R^2	0.0823	0.1058	0.1072

Table 9 [Continued]

Panel D: Alternative Prediction Horizon

	(1) *STOCK_PRICE_ALT*	(2) *EPS_ALT*
US_AI	1.8490**	0.3195***
	(0.818)	(0.025)
FIRM_SIZE	4.2758	0.2462*
	(2.668)	(0.124)
ROA	71.6423**	4.6921***
	(24.110)	(1.028)
LEVERAGE	-26.2111***	-0.8343**
	(8.427)	(0.372)
LOSS	6.1597	0.2679
	(3.553)	(0.164)
BM	-10.1748*	-0.3639
	(4.901)	(0.207)
PRE_RETURN	-9.3972***	-0.4009***
	(1.374)	(0.051)
RETURN_VOLATILITY	4.8821***	0.1761***
	(1.037)	(0.041)
IO	0.0624	0.0010
	(0.039)	(0.001)
Industry FE	Yes	Yes
Observations	9,956	9,956
Adj R^2	0.1037	0.1121

Table 10. Consequences of GenAI Bias

This table examines whether U.S. analysts' forecast optimism for Chinese firms increased relative to Chinese analysts' forecast optimism after the release of ChatGPT, and whether this pattern is related to cross-border news gaps. The sample is constructed from I/B/E/S analyst forecasts. U.S. analysts are analysts working at brokerage houses located in the United States, and Chinese analysts are analysts working at brokerage houses located in China. The dependent variable is *ANALYST_FORECAST_OPTIMISM*, defined as the analyst's EPS forecast minus actual EPS, scaled by the stock price as of February 7, 2023. Higher values indicate more optimistic analyst forecasts. Panel A compares U.S. and Chinese analysts' forecasts for Chinese firms and U.S. firms around the release of ChatGPT. *POST_CHATGPT* is an indicator equal to one for forecasts issued after the release of ChatGPT, and zero for forecasts issued before the release. *US_ANALYSTS* is an indicator equal to one for U.S. analysts and zero for Chinese analysts. The coefficient on *POST_CHATGPT* × *US_ANALYSTS* captures whether U.S. analysts became relatively more optimistic than Chinese analysts after the release of ChatGPT. Panel B examines whether this relative increase in U.S. analysts' forecast optimism for Chinese firms is stronger when cross-border news gaps are larger. The news gap proxies are *NEG_NEWS_GAP*, *POS_NEWS_GAP*, and *NET_NEG_NEWS_GAP*. The coefficient on *POST_CHATGPT* × *US_ANALYSTS* × *News Gap Proxy* captures whether U.S. analysts became relatively more optimistic after ChatGPT's release for Chinese firms with larger news gaps. Controls and industry fixed effects are included where indicated. Standard errors are reported in parentheses. Significance levels are indicated by *, **, and *** for 10%, 5%, and 1%, respectively.

Panel A: U.S. and Chinese Analyst Forecasts for Chinese Firms and U.S. Firms

	(1)	(2)
Forecasted Firms =	*CN Firms*	*US Firms*
Dep Var =	*ANALYST_FORECAST_OPTIMISM*	
POST_CHATGPT × *US_ANALYSTS*	0.0040***	0.0026
	(0.001)	(0.013)
POST_CHATGPT	-0.0127***	-0.0055
	(0.001)	(0.014)
US_ANALYSTS	0.0009	-0.0263*
	(0.001)	(0.013)
FIRM_SIZE	-0.0045**	0.0097*
	(0.002)	(0.005)
ROA	-0.2181***	-0.0516***
	(0.046)	(0.011)
LEVERAGE	0.0037	0.0448**
	(0.015)	(0.016)
LOSS	0.0030	0.0414*
	(0.006)	(0.023)
BM	-0.0014	0.0061
	(0.002)	(0.015)
PRE_RETURN	-0.0032	0.0071
	(0.007)	(0.013)
RETURN_VOLATILITY	-0.0023	0.0010
	(0.002)	(0.005)
IO	0.0000	-0.0610**
	(0.000)	(0.024)
FIRM_EXPERIENCE	0.0005	0.0008
	(0.000)	(0.001)
GENERAL_EXPERIENCE	0.0000	-0.0002
	(0.000)	(0.000)
ANALYST_BUSY	0.0000	-0.0000
	(0.000)	(0.000)
Industry FE	Yes	Yes
Observations	9,588	7,262
Adj. R^2	0.0784	0.0458

Table 10 [Continued]

Panel B: The Effect of News Gaps on U.S. and Chinese Analyst Forecasts for Chinese Firms

Dep Var = *ANALYST_FORECAST_OPTIMISM*	(1)	(2)	(3)
News Gap Proxy =	*NEG_NEWS_GAP*	*POS_NEWS_GAP*	*NET_NEG_NEWS_GAP*
POST_CHATGPT × *US_ANALYSTS* × *News Gap Proxy*	0.024*	-0.035**	0.016**
	(0.012)	(0.014)	(0.007)
POST_CHATGPT × *US_ANALYSTS*	-0.018	0.035**	-0.002
	(0.011)	(0.012)	(0.001)
POST_CHATGPT × *News Gap Proxy*	-0.012	0.003	-0.022***
	(0.009)	(0.006)	(0.003)
US_ANALYSTS × *News Gap Proxy*	-0.002	0.028***	-0.024***
	(0.006)	(0.006)	(0.004)
POST_CHATGPT	-0.002	-0.015***	-0.007***
	(0.008)	(0.004)	(0.001)
US_ANALYSTS	0.003	-0.024***	0.007***
	(0.006)	(0.005)	(0.001)
News Gap Proxy	0.007	0.003	0.026***
	(0.006)	(0.004)	(0.003)
Observations	9,588	9,588	9,588
Controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Adjusted R^2	0.079	0.081	0.087